Evolutionary Dynamics : ID 4201

The Trailer

What is Life ?

NASA's definition:

Life is a self-sustained chemical system capable of undergoing Darwinian evolution.

----- Gerald Joyce

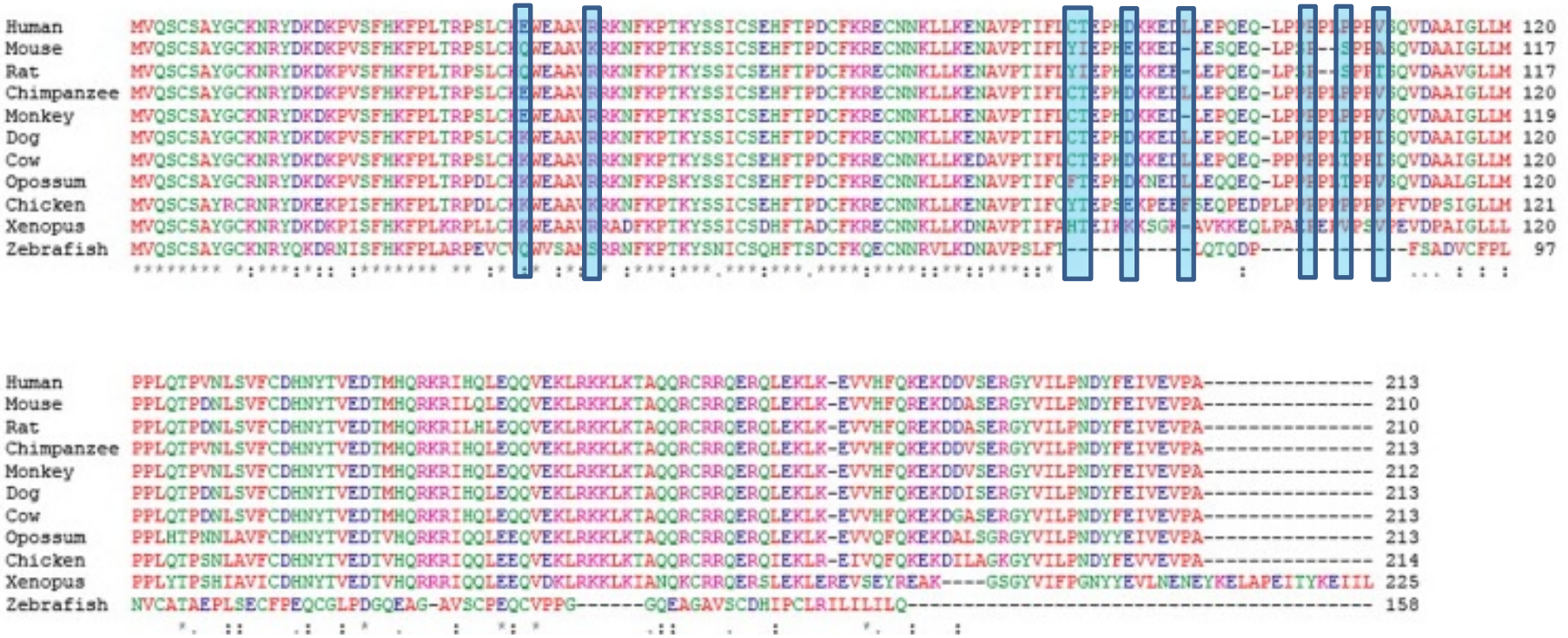
Life as we know it is much more constrained.

- Enclosed within a protective cell wall.
- Carbon-based, uses DNA for genetic information storage and synthesizes proteins to perform crucial tasks.
- Protein synthesis uses a genetic code which encodes only 20 amino acids.

Nothing in Biology makes sense except in the light of evolution.

----- Theodosius Dobzhansky

Multiple Sequence Alignment of a segment of a gene



Transparent blue strips indicate locations where the amino acids in the ***human*** and **mouse** gene segments differ
There is no difference between the ***human*** and **chimp** for this gene segment

Humans diverged from **chimps** approx. *5 million* years ago

Darwinian Evolution

- Mutations occur at ***random***

  Mutations can be advantageous, deleterious or neutral.

- A *random* mutation can confer a slight *selective advantage* to an individual which enhances its ability to survive and reproduce in the given environment with respect to its competitors.

- This slight advantage leads to its progeny surviving and propagating with greater probability.

- Fixation of the mutant - Over time this can lead to the mutant individual taking over the entire population.

The Peppered Moth Story (Pictures taken by Olaf Leillinger)

Pre-industrial revolution

Post-industrial revolution

Post pollution control

Topics

Introduction to Evolution

Concepts of sequence Spaces and Fitness Landscapes

Evolutionary Games

Effect of Finite Populations

Evolutionary Games in Finite Populations

Introduction to Evolutionary Graph Theory

Spatial Games: Cooperation and Conflict between interacting agents

Evolutionary Games on Networks

Evolutionary Dynamics of Cancer

Objective: To develop a quantitative understanding of evolutionary processes

Tools: Basic Calculus and basic statistical concepts, simple computer programs

Books : Evolutionary Dynamics by Martin Nowak (Primary)

Evolution and the Theory of Games by John Maynard Smith (Secondary)

Mathematical Models of Social Evolution: A Guide for the Perplexed by McElreath & Boyd

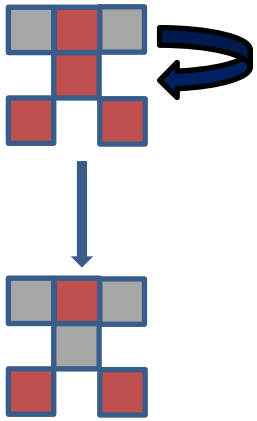
Marks based on assignments, mid-term exams, final exams, paper/proposal presentation

Big Questions

- How do populations evolve ?
- How does the ***mutation*** rates of agents in the population affect long-term population structure ?
- What is the effect of ***selection*** on population structure when there are mutants with varying fitness in the population ? *(Survival of the fittest ?)*
- Under what conditions can a ***neutral*** (no selective advantage or disadvantage relative to the wild type) mutant take over the entire population by chance ?
- How did ***cooperation*** emerge in a world where individuals try to maximize their benefits and minimize their costs ?
- Role of underlying network structure on the evolutionary dynamics of the population
- What is the effect of finite population size on evolutionary dynamics ?
- How can we explain the sudden proliferation of certain infected cell types which lead to cancer ?

Two Examples

1. ***Random Selection:*** Evolution of population structure of two types of individuals having the same fitness.



- Two types represented by two different colours, **red** and **grey.**
- The two colours are initially randomly distributed across the grid.
- Evolution is mimicked by each patch randomly picking another patch to update its own colour
- Over time, one colour will gain a slight dominance over the other.
- The dominant colour will spread as it is more likely to be picked to update the colour of a patch.
- The random nature of the process can lead to shifts in the dominant colour
- Eventually one colour takes over the entire grid.
- Since there is no advantage of one colour over the other, each colour is equally likely to take over the entire grid.

2. ***Fitness-dependent selection :*** Evolution of a population of cooperative/altruistic and selfish agents


Altruism.nlogo

- Evolution of the population depends upon fitness determined by two phenotypic traits.
- Cooperative/altruistic behaviour (phenotype) comes with a cost as well as a benefit to the agent.
- There is no cost for selfish behaviour.
- Due to the cost associated with altruism/cooperation, the fitness of altruistic agents are somewhat **less** than the fitness of selfish agents.
- Population evolution is mimicked by changes in the colour of each agent over time.
- Colour changes are determined by the fitness and number of altruistic and selfish agents surrounding each agent.
- Population evolves from an initial state of roughly equal numbers of altruists and selfish agents to a final stage in which either selfish or altruistic agents dominate.
- Gives insight into how cooperation/altruistic behaviour can be sustained in spite of the cost associated with such a phenotype.

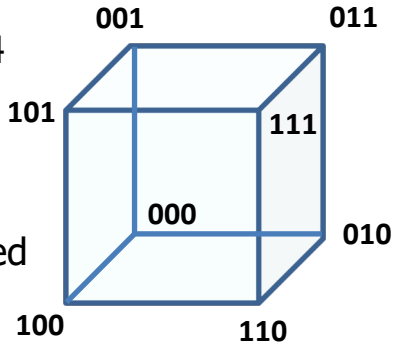
Sequence Spaces and Fitness Landscapes

Sequence Space : A sequence of length L is a point in **L-dimensional** space where each dimension has **4** discrete values corresponding to the 4 nucleotides.

There are $\mathbf{4^L}$ discrete points in the sequence space and each point corresponds to a distinct sequence of length **L**.

When only purines (A and G) and Pyrimidines (T and C) are distinguished instead of all 4 nucleotides, each of the L dimensions in sequence space has 2 discrete points ➔ the sequence space for a sequence of length L has $\mathbf{2^L}$ discrete points.

For L=3, the sequence space for sequences made up of purines (represented by 0) and pyrimidines (represented by 1) has a total of $\mathbf{2^3 = 8}$ discrete points and can be represented by the vertices of a cube.



Movement in sequence space occurs in discrete steps and ***Evolution is a trajectory through sequence space***

Fitness Landscape: Fitness determined by the phenotype i.e. characteristics like behaviour, morphology, structure, shape etc.

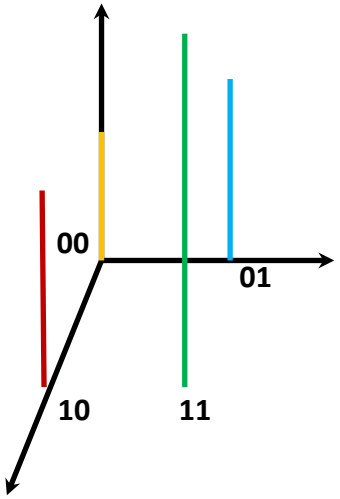
Genotype ➔ Phenotype mapping determines how changes in the genome affect the phenotype – the most challenging problem in Biology

Fitness landscape : Direct mapping between ***genotype and fitness.***

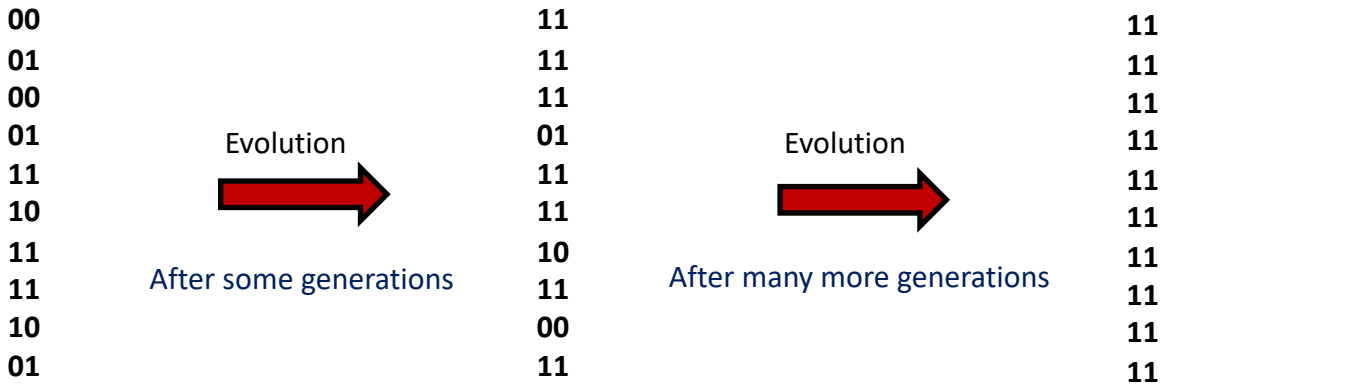
Fitness Landscape: (L+1)dimensional space where the first L dimensions describes the sequence space and the (L+1) th dimension represents the fitness associated with each of the 4^L (or 2^L) sequences

Evolution of sequences amounts to movement in the fitness landscape and ***Adaptation is an attempt to attain peaks in the fitness landscape.***

Single Peak Fitness Landscape



Adaptation of a population of 10 sequences of length L=2



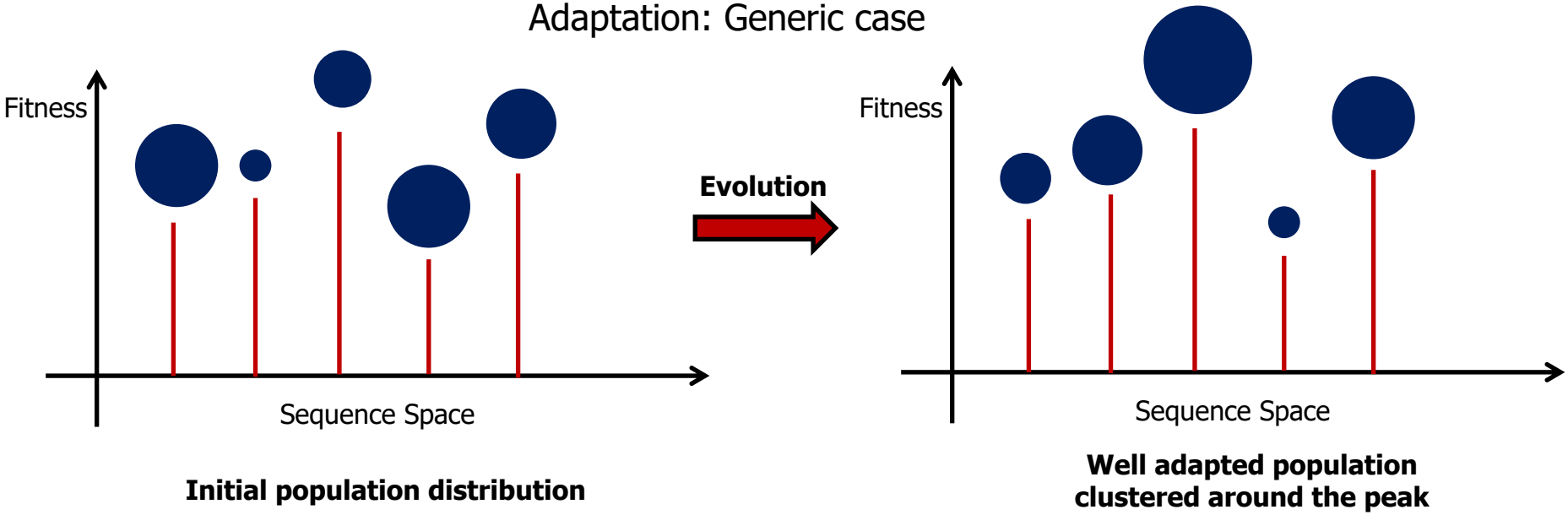
Type	Frequency
00	0.2
01	0.3
10	0.2
11	0.3

Type	Frequency
00	0.1
01	0.1
10	0.1
11	0.7

Type	Frequency
00	0
01	0
10	0
11	1

A population is well adapted if most (or all) of the members cluster around a fitness peak in the fitness landscape

Adaptation: Generic case

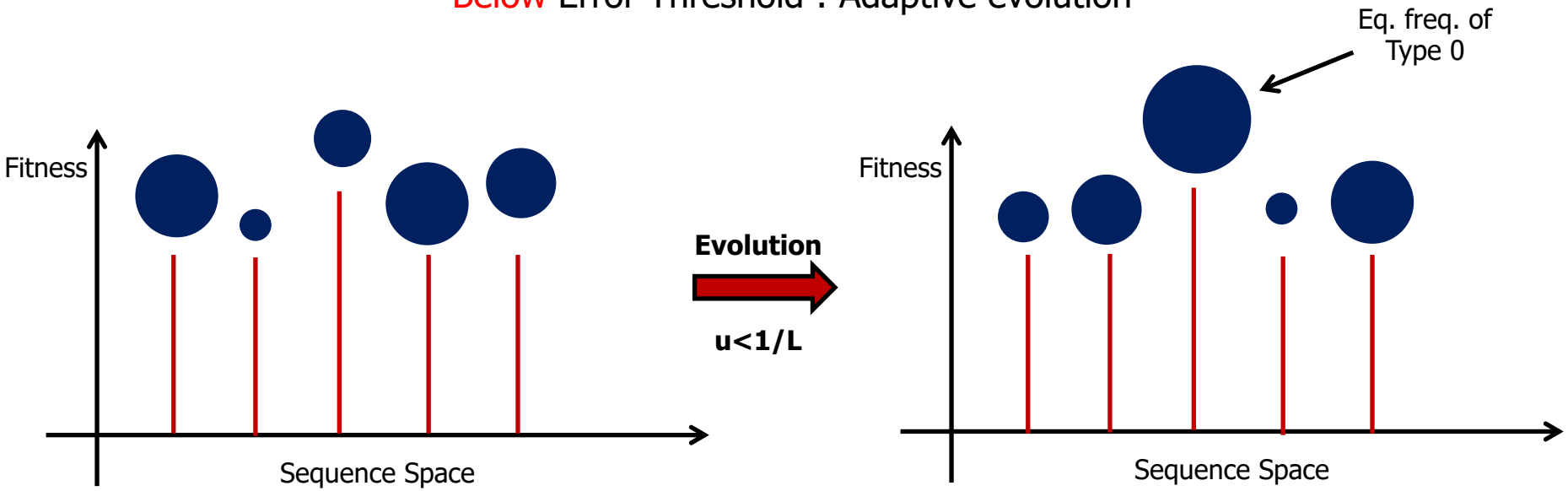


Initial population distribution

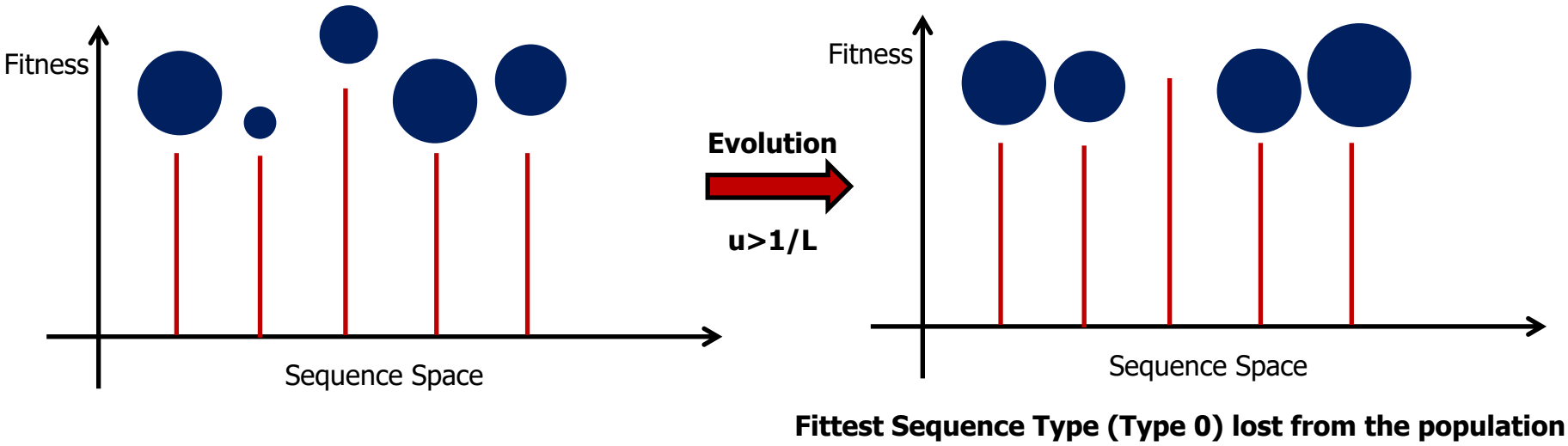
Well adapted population clustered around the peak

Size of the circles represent the frequency of the corresponding sequence in the population

Below Error Threshold : Adaptive evolution



Above Error Threshold : Non-adaptive evolution



Fittest Sequence Type (Type 0) lost from the population

Eigen's Paradox

- For a replicating molecule to be viable without error-correcting enzymes, its length should be small.
- For a replicating molecule to encode error-correcting enzymes its length should be substantially larger than several thousand bases.

RNA virus Error Catastrophe

Anti-viral effect manifest by enhanced mutagenesis of the Polio virus genome

High mutation rates ➔ loss of viability of the Polio virus genome

Table 2. The antiviral effects of ribavirin can be directly attributed to lethal mutagenesis

	Normal	100 μM ribavirin	400 μM ribavirin	1,000 μM ribavirin
RNA-specific infectivity loss	—	3.3	18	140
Loss of total viral RNA	—	—	6	16
Total predicted titer reduction	1	3.3	100	2,200
Actual titer reduction*	1	3.2	71	2,000

*Untreated ("normal") poliovirus titer in this experiment was 1.2×10^{10} PFU per plate of HeLa cells (6×10^6 cells). Data are the average of three experiments.

Polio viruses reside near the edge of the Error Threshold

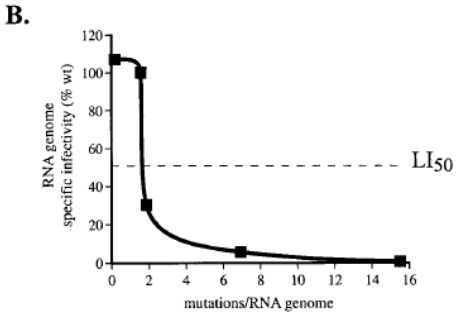
Modest (less than 2-fold) increase in mutation rate ➔ 50% of the viral population becomes unviable

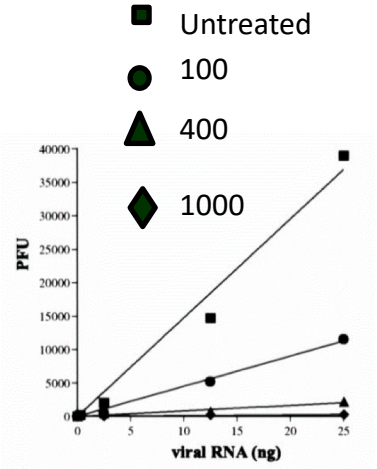
4-fold increase in mutation rate ➔ 95% of the viral population becomes unviable.

Table 3. Mutation frequency in ribavirin-treated RNA virus populations

	G→A	C→T	Total mutation frequency*
Normal population	0.5	1.2	2.1
100 μM ribavirin	—	1.3	2.5
400 μM ribavirin	4.4	5.0	9.3
1,000 μM ribavirin	6.8	12.0	20.8

*Mutations per 10,000 nt sequenced.





High mutation rate produces Polio-virus mutants having low infectivity

The amount of infectious virus genomes in the population is reduced several fold as concentration of Ribavarin increases.

Ref.: Crotty, Cameron, Andino; PNAS 98, 6895-6900 (2001)

PFU: plaque forming unit

How does Mutation Rate per site (u) across organisms compare with the Error Threshold ($1/L$) ?

Organism	Genome length in bases	Mutation rate per base	Mutation rate per genome
RNA viruses			
Lytic viruses			
Qβ	4.2×10^3	1.5×10^{-3}	6.5
Polio	7.4×10^3	1.1×10^{-4}	0.84
VSV	1.1×10^4	3.2×10^{-4}	3.5
Flu A	1.4×10^4	7.3×10^{-6}	0.99
Retroviruses			
SNV	7.8×10^3	2.0×10^{-5}	0.16
MuLV	8.3×10^3	3.5×10^{-6}	0.029
RSV	9.3×10^3	4.6×10^{-5}	0.43
Bacteriophages			
M13	6.4×10^3	7.2×10^{-7}	0.0046
λ	4.9×10^4	7.7×10^{-8}	0.0038
T2 and T4	1.7×10^5	2.4×10^{-8}	0.0040
E. coli	4.6×10^6	5.4×10^{-10}	0.0025
Yeast *(S. cerevisiae)*	1.2×10^7	2.2×10^{-10}	0.0027
Drosophila	1.7×10^8	3.4×10^{-10}	0.058
Mouse	2.7×10^9	1.8×10^{-10}	0.49
Human *(H. sapiens)*	3.5×10^9	5.0×10^{-11}	0.16

Sources: Drake (1991, 1993) and Drake et al. (1998).

Note: Most organisms have a mutation rate per genome which is less than one, as predicted by the error threshold theory. Why Qβ and VSV have such a high mutation rate is at present unexplained.

Note: Organisms remain viable only if uL<1)

Pictorial representation of evolution by mutation without selection

Time

#A

#B

Time=t

nA

nB

Pick Random members

Check type **A** or **B**

Check if RN < u1, if true mutate A to B. If False, leave A unchanged

Repeat process for other members of population

Time=t+1

nA-1

nB+1

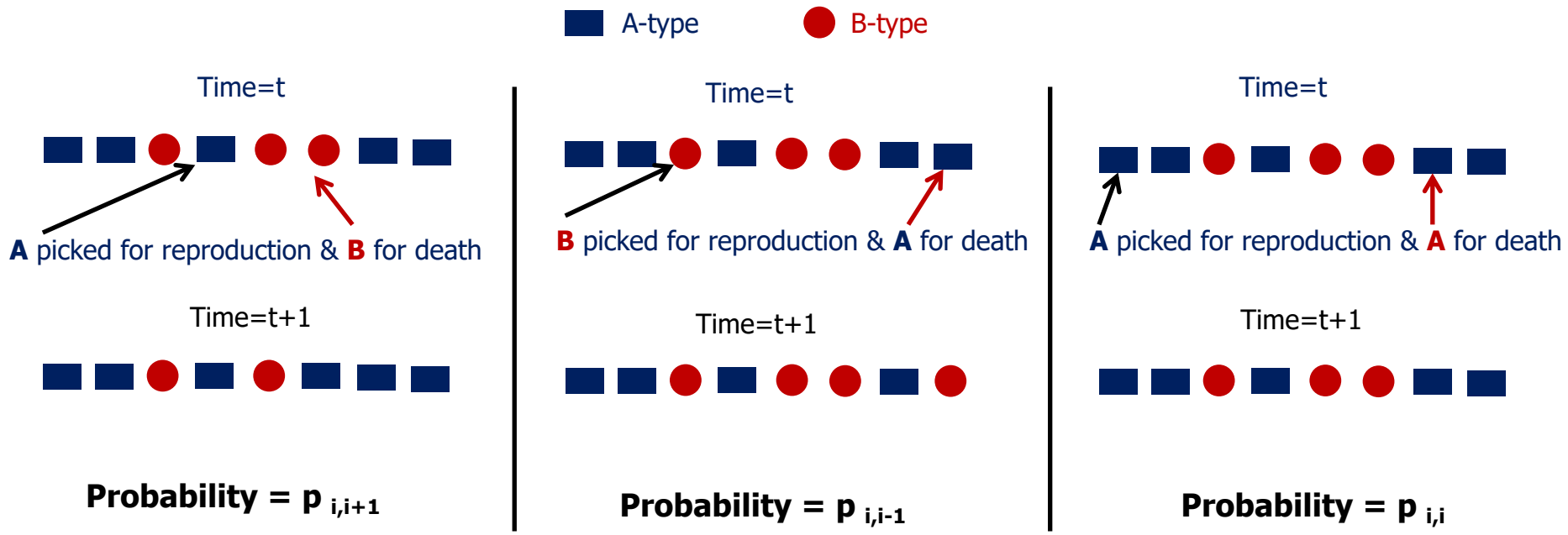
Moran Process

Assume two types of individuals A and B in the population of N individuals, both types having the *same* fitness.

In every generation, one individual is picked at random for reproduction and another individual is picked at random for death.

Question: What is the ***fixation probability*** of A ?

Fixation Probability of A: Probability that the frequency of A increases from an initial value of i/N to a final value of 1 i.e. the final population consists entirely of A



$$p_{i,i-1} + p_{i,i} + p_{i,i-1} = 1$$

Difference between Invasion and Fixation

Invasion is a special case of Fixation

In *deterministic* simulations involving the replicator equation,

Invasion by type A ➔ frequency of A increases from a very small fraction $x=\varepsilon << 1$ ➔ $x=1$

Fixation of A ➔ frequency of A increases from any x ➔ $x=1$

In *stochastic* simulations like evolution by Moran process

Invasion by type A ➔ frequency of A increases from $x=1/N$ ➔ $x=1$

Fixation of A ➔ frequency of A increases from any x (such that $0 < x=\#A/N < 1$) ➔ $x=1$

NOTE

The inability of A to *invade* a population of B does not mean that A cannot get *fixed* in the population if its initial frequency is sufficiently large.

Evolution with selection via Moran Process

Reproduction: Occurs with probability proportional to the fitness of the agent

Death: Occurs at random, independent of the fitness of the agent

$$p_{i,i+1} = \left(\frac{ri}{ri + 1(N-i)}\right)\left(\frac{N-i}{N}\right)$$

Prob. of picking A for reproduction

Prob. of picking B for death

$$p_{i,i-1} = \left(\frac{(N-i)1}{ri + 1(N-i)}\right)\left(\frac{i}{N}\right)$$

Prob. of picking B for reproduction

Prob. of picking A for death

$$\gamma_i = \frac{p_{i,i-1}}{p_{i,i+1}} = \frac{1}{r}$$

$$x_i = \frac{(1-\frac{1}{r^i})}{(1-\frac{1}{r^N})}$$

For r> 1; as N➔ ∞ $x_1 \equiv \rho_A \quad 1-\frac{1}{r}$

Invasion Probabilities for N=100

r	$x_1 \equiv \rho_A$
2	0.5
1.01	0.016
1	0.01

r	$x_1 \equiv \rho_A$
0.99	0.0058
0.9	0.000003

➔ Invasion by a selectively advantageous mutant is not guaranteed even for large population sizes

Standard Game Theory

Two players A and B play a game

***Neither* player knows the opponent's strategy**

Primary Question: What strategy must each player adopt to maximize its payoff ?

Brings into play the concept of rational agents

If both A and B act rationally, each will try to maximize his payoff.

However, there is no guarantee that a player will act rationally.

Frequency-dependent Fitness

Individuals have ***fixed strategies*** that are known to other individuals in the population.

Random interactions occur with other individuals (including those belonging to the same type)

In the Biological Context

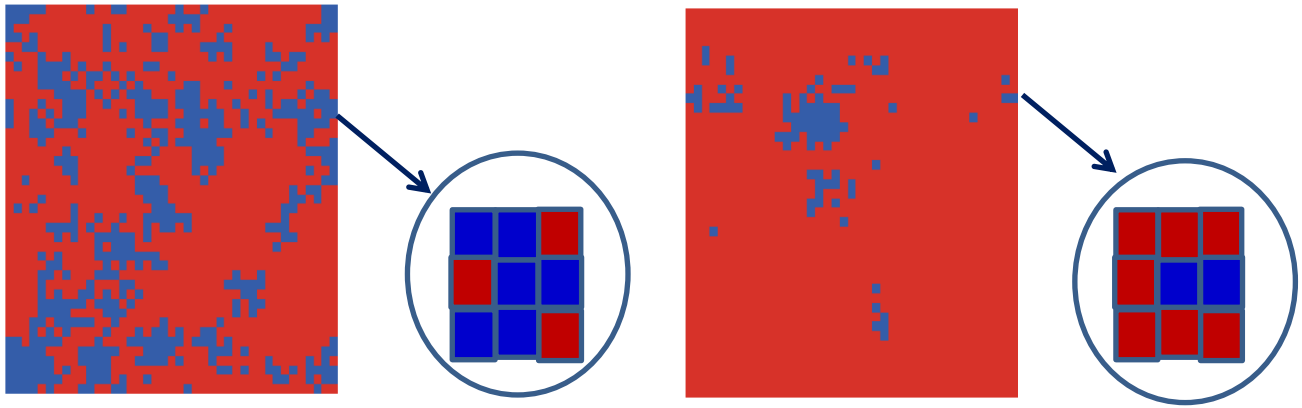
strategy ⟷ ***phenotype*** and **payoff** ⟷ ***fitness***

Fitness is a measure of reproductive success

A component of ***Fitness*** of an individual is determined by the ***cumulative*** payoff to that individual resulting from the encounter with other individuals of the same type as well as different types.

The population is *updated* every generation when individuals reproduce with

Probability ***proportional to*** fitness



PD N-Person Iterated-1C-10D.nlogo

PD N-Person Iterated-10C-1D.nlogo

PD N-Person Iterated-10C-10D.nlogo

George Price

Evolutionary Games

John Maynard-Smith

Competition between different types of individuals with frequency-dependent fitness can be thought of as a game in which each type employs a distinct strategy and gets a certain payoff in an encounter with another individual.

Fitness is a measure of reproductive success and **strategies** that yield higher cumulative **payoff**, reproduce at a faster rate.

Consider a population with two types of individuals **A** and **B** whose fitness are $\mathbf{f_A}$ and $\mathbf{f_B}$ respectively.

Assuming ***linear*** dependence of fitness on frequencies $\mathbf{x_A}$,$\mathbf{x_B}$;

$\mathbf{f_A} = a\,\mathbf{x_A} + b\,\mathbf{x_B}$

$\mathbf{f_B} = c\,\mathbf{x_A} + d\,\mathbf{x_B}$

Can be written in matrix notation as $\mathbf{f} = \mathbf{Mx}$, $\mathbf{f}=[\mathbf{f_A}, \mathbf{f_B}]$; $\mathbf{x}=[\mathbf{x_A}, \mathbf{x_B}]$

where **M** is the payoff matrix.

$$\mathbf{M}= \begin{array}{c} \\ \mathbf{A} \\ \mathbf{B} \end{array} \begin{array}{c} \begin{array}{cc} \mathbf{A} & \mathbf{B} \end{array} \\ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \end{array}$$

E(A,A)=a : Payoff to **A** when it interacts with another A

E(A,B)=b : Payoff to **A** when it interacts with **B**

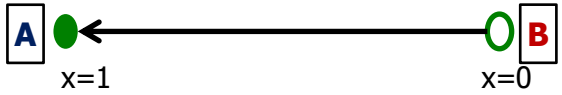
E(B,A)=c : Payoff to **B** when it interacts with **A**

E(B,B)=d : Payoff to **B** when it interacts with another **B**

Replicator (Deterministic) Dynamics

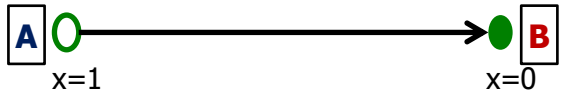
Case 1: a>c; b >d ➔ Payoff to **A** > Payoff to **B** : **A** dominated **B**

Only one stable equilibrium solution exists.



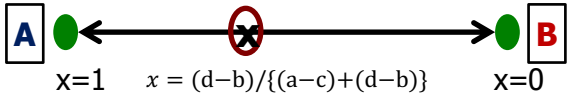
Case 2: a<c; b<d ➔ Payoff to **B** > Payoff to **A** : **B** dominates **A**

Only one stable equilibrium solution exists.



Case 3: a>c; b<d ➔ $x = 0, 1, (\mathrm{d}-\mathrm{b})/\{(\mathrm{a}-\mathrm{c})+(\mathrm{d}-\mathrm{b})\}$

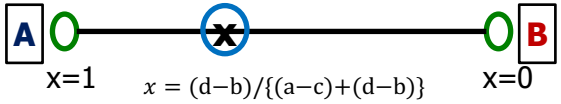
3 equilibrium solution exists; **A** and **B** are ***bistable***, mixed-state solution is ***unstable***.



Case 4: a<c; b>d ➔ $x = 0, 1, (\mathrm{d}-\mathrm{b})/\{(\mathrm{a}-\mathrm{c})+(\mathrm{d}-\mathrm{b})\}$

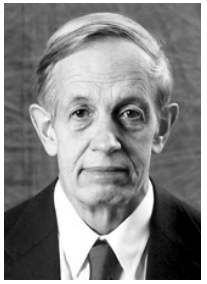
Only one stable equilibrium solution exists. ***A*** & ***B*** stably co-exist.

x=0, 1 are ***unstable*** equilibrium solutions.



Nash Equilibrium

A strategy is said to be a *Nash Equilibrium* if the person adopting the strategy cannot increase his payoff by changing to a different strategy.



A is a ***strict*** Nash Equilibrium if a > c

A is a Nash Equilibrium if a ≥ c

B is a strict Nash Equilibrium if d > b

B is a Nash Equilibrium if d ≥ b

For a strategy to be a ***strict*** Nash Equilibrium, the payoff of the person adopting that strategy must *decrease* if he changes the strategy.

Evolutionarily Stable Strategies (ESS)

Consider a large population of individuals employing strategy A. If a mutant employing strategy B is introduced into the population, **can the mutant invade the population consisting primarily of A-type players?**

In the *infinitely large population size limit*, let the number of B mutants (invaders) be infinitesimally small with frequency given by $\mathbf{x_B = \varepsilon}$. Frequency of **A's**: $\mathbf{x_A = 1 - \varepsilon}$

$\mathbf{f_A}$ = a **(1- ε)** + b **ε ;** $\mathbf{f_B}$ = c **(1- ε)** + d **ε**

B cannot invade A only if $\mathbf{f_A > f_B}$ **i.e.** a **(1- ε)** + b **ε >** c **(1- ε)** + d **ε**

Since ε is very small, neglecting terms of order ε gives a>c ➔ **E(A,A) > E(B,A)**

If however, a=c, $\mathbf{f_A > f_B}$ gives b>d ➔ **E(A,B) > E(B,B)**

Definition: ***Invasion Probability of B*** (in **finite** populations): Probability that a **single B** mutant in a population of (N-1) A-type individuals eventually gets fixed in the population i.e. the final population consists entirely of B.

General Condition for Nash Equilibrium and ESS in games with more than two strategies

Strategy $\mathbf{S_k}$ is a ***strict*** Nash equilibrium if $\mathbf{E(S_k , S_k) > E(S_i , S_k); i \neq k}$

Strategy $\mathbf{S_k}$ is a Nash equilibrium if $\mathbf{E(S_k , S_k) \geq E(S_i , S_k); \text{ for all } i \neq k}$

Strategy $\mathbf{S_k}$ is an **ESS** if for all $\mathbf{i \neq k}$

either $\mathbf{E(S_k , S_k) > E(S_i , S_k)}$

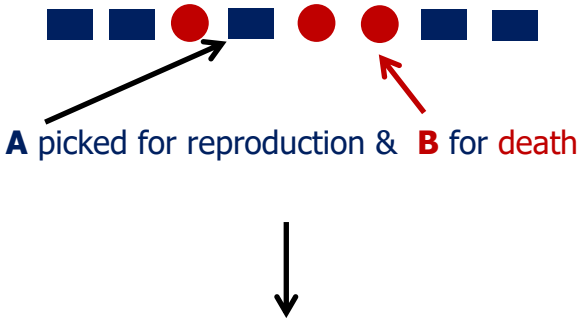
OR

$\mathbf{E(S_k , S_k) = E(S_i , S_k) \ \& \ E(S_k , S_i) > E(S_i , S_i)}$

If a strategy satisfies the conditions for either ***strict*** **Nash equilibrium or ESS,** it implies that the strategy **cannot** be invaded by a mutant strategy.

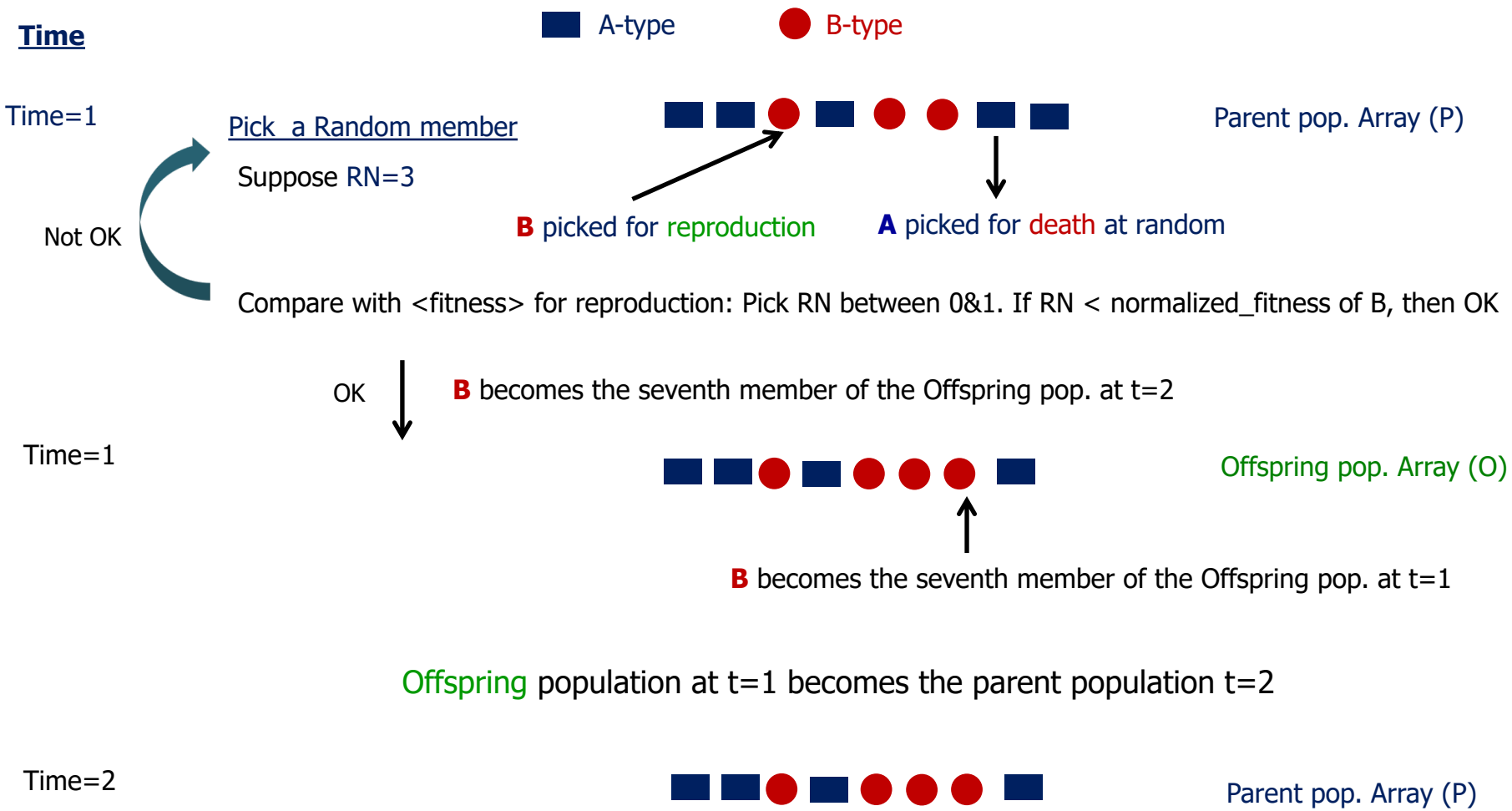
Pictorial representation of the Simulation of Moran Process: Neutral Evolution Case

Time		**#A**	**#B**
Time=t	Pick Random members Suppose RN1=6 ; RN2=4	**nA**	**nB**
Time=t+1		**nA+1**	**nB-1**



Pictorial representation of selection without mutation: Fixation of an ad/disadvantageous mutant

Evolution by Moran Process



Note: If fitness of A is r1 and fitness of B is r2,

normalized_fitness_of_**A**=r1/(r1+r2); normalized_fitness_of_**B**=r2/(r1+r2)

Pictorial representation of evolution via mutation and selection

Evolution by updating whole population every generation

A-type B-type

Time=1 Copy parent pop. to offspring pop.

Time=1 Offspring pop. after **mutations** ➔

mutations

Time=1 Pick a Random member

Suppose RN=3

Offspring pop. Array (O)

B picked for reproduction

Not OK

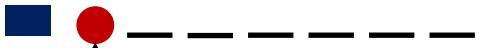
Compare with fitness for reproduction: Pick RN between 0&1. If RN< normalized_fitness_of_B, then OK

normalized_fitness_of_B = f1/(f0+f1)

OK

B becomes the second member of the Offspring pop. at t=2

Time=2



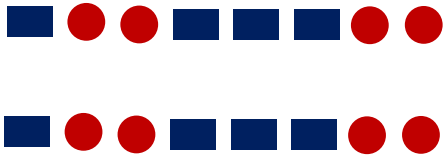
Parent pop. Array (P)

B becomes the second member of the Offspring pop. at t=2

Repeat above steps to fill up the parent pop. at t=2

Copy offspring array at t=2 to parent array such that it becomes the parent pop. For choosing offspring at t=3

Copy



Offspring pop. Array (O) at t=2

Parent pop. Array (P) at t=3

Hawk-Dove Game

Hawk (H) strategy escalates the fight at the cost of injury

Dove (D) strategy initially threatens but eventually backs off, avoiding injury, but getting a lower payoff

$$M = \begin{array}{c} \\ H \\ D \end{array} \begin{array}{c} \begin{array}{cc} H & D \end{array} \\ \begin{bmatrix} (b-c)/2 & b \\ 0 & b/2 \end{bmatrix} \end{array}$$

When two Hawks interact, each has an equal probability of winning the resource (b) but also an equal likelihood of loosing and getting injured (c)

➔ Expected payoff to each Hawk : E(H,H) = (b-c)/2

When two Doves interact, each has an equal probability of winning the resource (b) but also an equal likelihood of loosing and not getting anything.

➔ Expected payoff to each Dove : E(D,D) = b/2

Assumptions

There is no fitness difference within the Hawk population ➔ all Hawks are equally capable. The same is true for Doves.

Both players arrive at the resource simultaneously and there is no time-lag in the behavioral response of the two players.

Hawk-Dove-Retaliator Game

Retaliator plays Hawk against Hawk, but plays Dove against Dove and other Retaliators

Key Questions

Given an initial frequency of H,D,R, what is the final equilibrium state of the system ?

Under what conditions can one strategy invade the others ?

Sub-Population Dynamics

Are Retaliators stable against invasion by Hawks ?

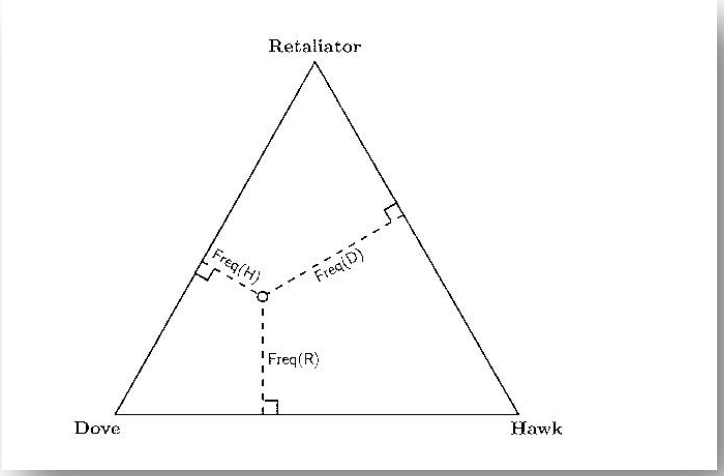
Are Doves stable against invasion by Retaliators ?

Can a small frequency of Hawks invade a mixed population of Doves & Retaliators ?

Can a small frequency of Retaliators invade an *equilibrated mixture* of Doves & Hawks ?

The frequency of all three strategies can be represented by a point in the Simplex S_3.

The *length of the perpendicular* drawn from the point to one face of the simplex gives the frequency of the strategy that is occupying the vertex *opposite* to that face.



Hawk-Dove-Retaliator Game

Retaliator plays Hawk against Hawk, but plays Dove against Dove and other Retaliators

Key Questions

Given an initial frequency of H,D,R, what is the final equilibrium state of the system ?

Under what conditions can one strategy invade the others ?

Sub-Population Dynamics

Are Retaliators stable against invasion by Hawks ? Yes!

Are Doves stable against invasion by Retaliators ? Yes!

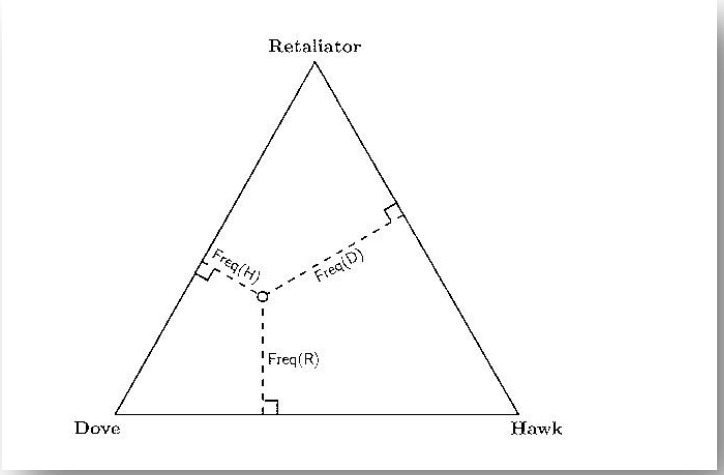
Can a small frequency of Hawks invade a mixed population of Doves & Retaliators ?

Yes! iff $x_R < b/(b+c)$; when $b < c$

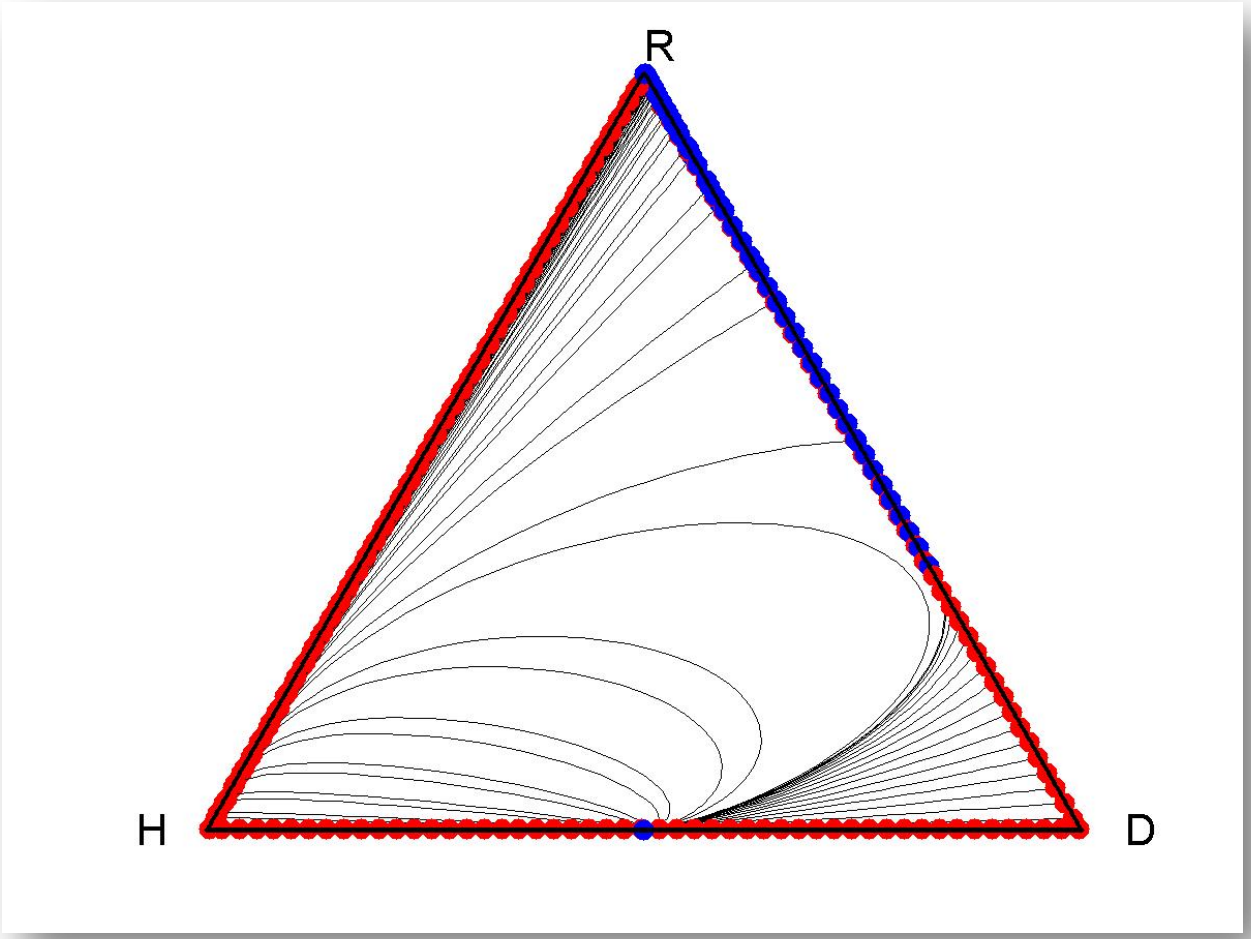
Can a small frequency of Retaliators invade an *equilibrated mixture* of Doves & Hawks ? No!

The frequency of all three strategies can be represented by a point in the Simplex S_3.

The *length of the perpendicular* drawn from the point to one face of the simplex gives the frequency of the strategy that is occupying the vertex *opposite* to that face.



Mapping of Frequencies onto a Simplex



- Represents initial frequencies in the population
- Represents final equilibrium frequencies in the population

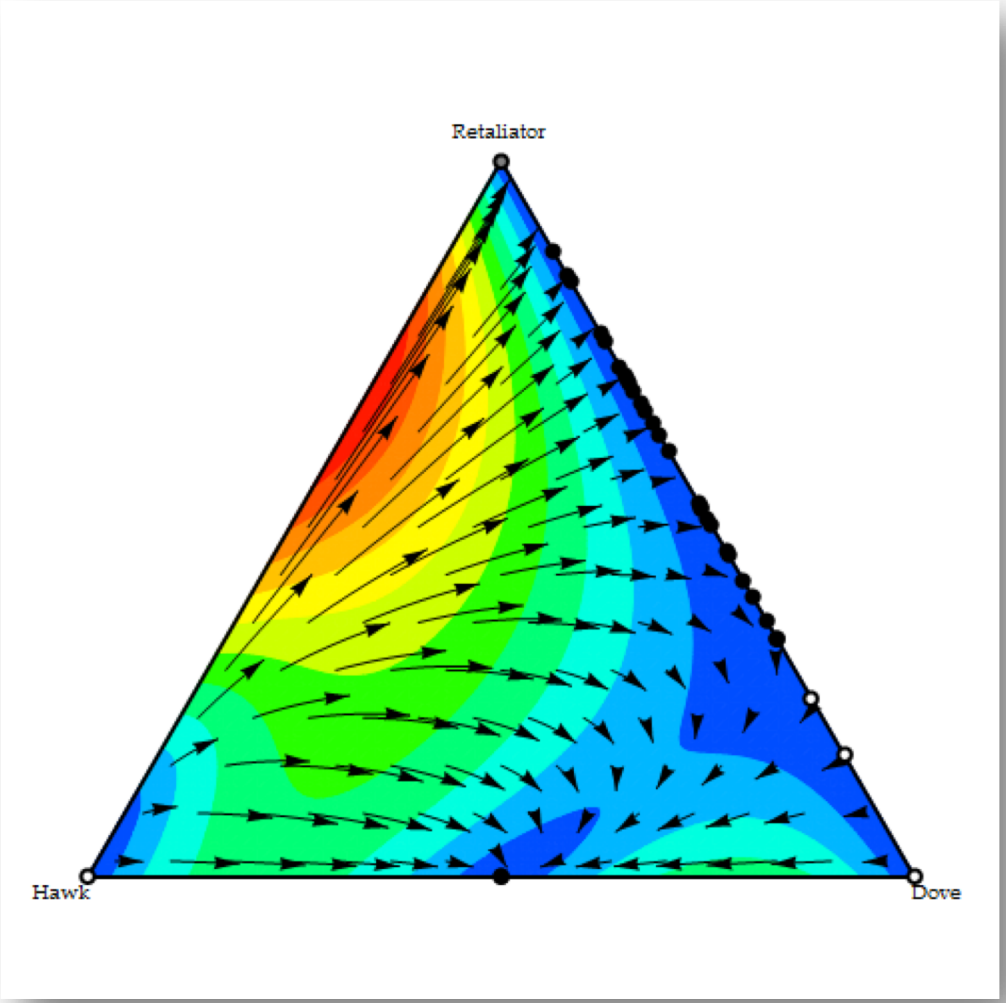
Trajectories show how the frequencies change over time & attain the final equilibrium state for different initial conditions.

Mapping of Frequencies onto a Simplex

- ○ Represents ***unstable*** equilibrium state
- ● Represents ***stable*** equilibrium state

red-fastest rate of change

blue-slowest rate of change



Generated by Dynamo3S

Arrows show how the frequencies change over time & attain the final equilibrium state

Coloured contours indicate how fast the frequencies are changing in the region.

red-fastest, blue-slowest

Hawk-Dove-Bourgeois Game

Asymmetry in claiming the resource: Relaxing the assumption of *simultaneous* claim to the resource

Bourgeois (B) plays Hawk when it arrives first and claims ownership of the resource but plays Dove when it arrives later to claim the resource

$$M = \begin{array}{c} \\ H \\ D \\ B \end{array} \begin{array}{c} \begin{array}{ccc} H & D & B \end{array} \\ \begin{bmatrix} (b-c)/2 & b & b/2+(b-c)/4 \\ 0 & b/2 & b/4 \\ (b-c)/4 & b/2 + b/4 & b/2 \end{bmatrix} \end{array}$$

Key Question

Can a small frequency of Bourgeois invade an *equilibrated* mixture of D & H ? Yes! iff $b < c$

Conclusions

Which strategy is an ESS depends not just on the payoff it receives when it interacts with a different strategy but also on the outcome of its interaction with itself and on the frequency

Pictorial representation of evolution via mutation and selection (Error Threshold Simulation)

Evolution by updating whole population every generation

A-type B-type

Time=1 Copy parent pop. to offspring pop.

Time=1 Offspring pop. after **mutations** ➔

Time=1 Pick a Random member

Suppose RN=3

B picked for reproduction

Offspring pop. Array (O)

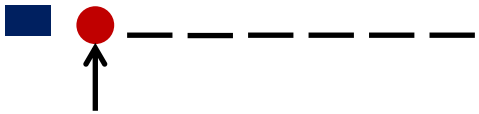
Not OK

Compare with fitness for reproduction: Pick RN between 0&1. If RN< normalized_fitness_of_B, then OK

normalized_fitness_of_B = f1/(f0+f1)

OK

Time=2 **B** becomes the second member of the Offspring pop. at t=2



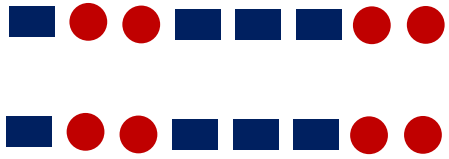
Parent pop. Array (P)

B becomes the second member of the Offspring pop. at t=2

Repeat above steps to fill up the parent pop. at t=2

Copy offspring array at t=2 to parent array such that it becomes the parent pop. For choosing offspring at t=3

Copy



Offspring pop. Array (O) at t=2

Parent pop. Array (P) at t=3

Evolution of Cooperation

What Don't We Know?

At *Science*, we tend to get excited about new discoveries that lift the veil a little on how things work, from cells to the universe. That puts our focus firmly on what has been added to our stock of knowledge. For this anniversary issue, we decided to shift our frame of reference, to look instead at what we *don't* know: the scientific puzzles that are driving basic scientific research.

We began by asking *Science*'s Senior Editorial Board, our Board of Reviewing Editors, and our own editors and writers to suggest questions that point to critical knowledge gaps. The ground rules: Scientists should have a good shot at answering the questions over the next 25 years, or they should at least know how to go about answering them. We intended simply to choose 25 of these suggestions and turn them into a survey of the big questions facing science. But when a group of editors and writers sat down to select those big questions, we quickly realized that 25 simply wouldn't convey the grand sweep of cutting-edge research that lies behind the responses we received. So we have ended up with 125 questions, a fitting number for *Science*'s 125th anniversary.

First, a note on what this special issue is not: It is not a survey of the big societal challenges that science can help solve, nor is it a forecast of what science might achieve. Think of it instead as a survey of our scientific ignorance, a broad swath of questions that scientists themselves are asking. As Tom Siegfried puts it in his introductory essay, they are "opportunities to be exploited."

We selected 25 of the 125 questions to highlight based on several criteria: how fundamental they are, how broad-ranging, and whether their solutions will impact other scientific disciplines. Some have few immediate practical implications—the composition of the universe, for example. Others we chose because the answers will have enormous societal impact—whether an effective HIV vaccine is feasible, or how much the carbon dioxide we are pumping into the atmosphere will warm our planet, for example. Some, such as the nature of dark energy, have come to prominence only recently; others, such as the mechanism behind limb regeneration in amphibians, have intrigued scientists for more than a century. We listed the 25 highlighted questions in no special order, but we did group the 100 additional questions roughly by discipline.

Our sister online publications are also devoting special issues to *Science*'s 125th anniversary. The Science of Aging Knowledge Environment, SAGE KE (www.sageke.org), is surveying several big questions confronting researchers on aging. The Signal Transduction Knowledge Environment, STKE (www.stke.org), has selected classic *Science* articles that have had a high impact in the field of cell signaling and is highlighting them in an editorial guide. And *Science*'s Next Wave (www.nextwave.org) is looking at the careers of scientists grappling with some of the questions *Science* has identified.

We are acutely aware that even 125 unknowns encompass only a partial answer to the question that heads this special section: What Don't We Know? So we invite you to participate in a special forum on *Science*'s Web site (www.sciencemag.org/sciext/eletters/125th), in which you can comment on our 125 questions or nominate topics we missed—and we apologize if they are the very questions you are working on.

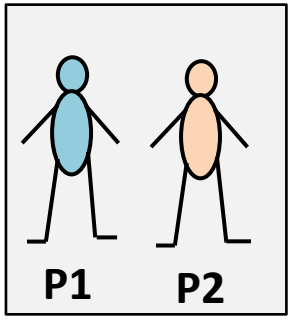
–Donald Kennedy and Colin Norman

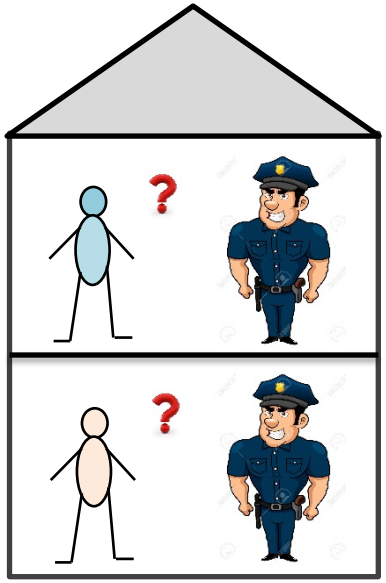
Contents >> News

76 In Praise of Hard Questions
78 What Is the Universe Made Of?
79 What Is the Biological Basis of Consciousness?
80 Why Do Humans Have So Few Genes?
81 To What Extent Are Genetic Variation and Personal Health Linked?
82 Can the Laws of Physics Be Unified?
83 How Much Can Human Life Span Be Extended?
84 What Controls Organ Regeneration?
85 How Can a Skin Cell Become a Nerve Cell?
86 How Does a Single Somatic Cell Become a Whole Plant?
87 How Does Earth's Interior Work?
88 Are We Alone in the Universe?
89 How and Where Did Life on Earth Arise?
90 What Determines Species Diversity?
91 What Genetic Changes Made Us Uniquely Human?
92 How Are Memories Stored and Retrieved?
93 How Did Cooperative Behavior Evolve?
94 How Will Big Pictures Emerge From a Sea of Biological Data?
95 How Far Can We Push Chemical Self-Assembly?
96 What Are the Limits of Conventional Computing?
97 Can We Selectively Shut Off Immune Responses?
98 Do Deeper Principles Underlie Quantum Uncertainty and Nonlocality?
99 Is an Effective HIV Vaccine Feasible?
100 How Hot Will the Greenhouse World Be?
101 What Can Replace Cheap Oil—and When?
102 Will Malthus Continue to Be Wrong?
78 So Much More to Know …

See also Editorial on p. 19 and www.sciencemag.org/sciext/125th

Prisoner's Dilemma





- Both prisoner's can *cooperate* (C) with each other by remaining silent
- P1 can *cooperate* by remaining silent but P2 can *defect* (D) by confessing
- P1 can *defect* (D) by confessing but P2 can *cooperate* (C) by remaining silent
- *Both* prisoner's *defect* (D) by confessing

$$M = \begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \end{array}$$

a=-1, b=-5, c=0, d=-3

What should your advise be ?

Prisoner's Dilemma

Which is the best strategy for a prisoner to adopt that would minimize his jail-term ?

Cooperation is an emergent phenomenon

Cooperation/Altruistic behavior comes with a cost as well as a benefit !

AIM

To understand how cooperation can be sustained in an environment where individuals are always trying to maximize their benefits and minimize their costs ?

Prisoner's Dilemma

Consider the Payoff Matrix **M=**

	C	D
C	a	b
D	c	d

a<c, d>b & a>d

a>d➔ better for both to cooperate than defect

Note: D is a strict Nash equilibrium as well as an ESS.

A mixed population consisting primarily of cooperators and a very small fraction of mutant defectors will eventually be invaded by the Defectors.

Direct Reciprocity : The game is not just played once but repeated several time between the same two players. Each player can adopt many distinct strategies specified by the sequence of cooperate (**C**) or defect (**D**) moves. **The sequence of C and D moves of a player can be informed by the corresponding set of moves by his opponent.** The payoff for each encounter is noted and the cumulative payoff is calculated at the end of the game by adding the payoffs for each encounter.

What is the strategy (i.e. sequence of C,D moves) that will maximize the payoff ?

Can a cooperative strategy be stable against invasion by selfish agents (defectors) ?

Suppose the game is played for **m** rounds. a=3, b=0, c=5, d=1

You **- GRIM: CDDD…..D**

Me **- ALLD: DDDD….D**

What is the payoff that you and I get after m rounds of the game ?

A critical number of rounds $\mathbf{m_c}$=2 has to be played before the more cooperative strategy **GRIM** becomes stable against invasion by more selfish strategy **ALLD**. If m > $\mathbf{m_c}$ and both you and me play **GRIM**, then neither of us can increase our payoff by changing to **ALLD.**

ALLD is also an ESS since md > b + (m-1)d ➔ d>b

➔ **ALLD** is also stable against invasion by **GRIM**

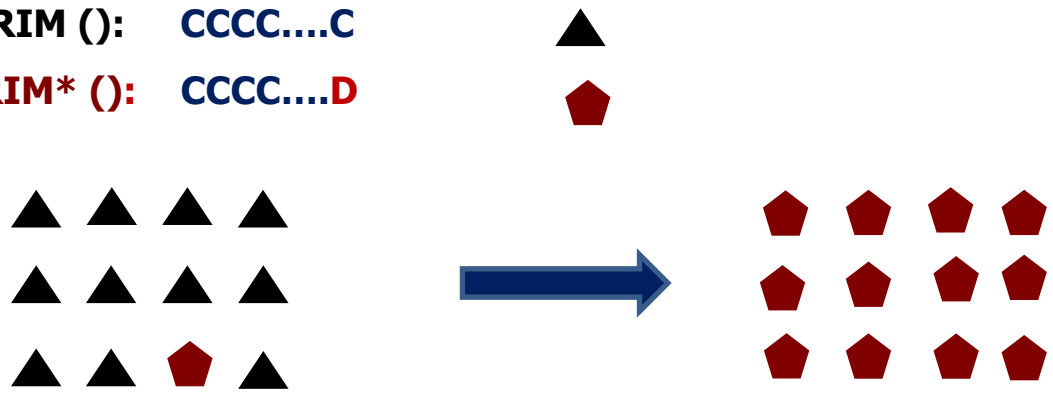
Problem with *fixed* #rounds: **No incentive to cooperate in the last round of the game.**

When a more selfish strategy than GRIM emerges, it can invade a population of GRIM players.

Consider the strategy **GRIM*:** Cooperate as long as your opponent cooperates but defect in the last round of the game

You – **GRIM ():** **CCCC….C**

Me – **GRIM* ():** **CCCC….D**



However, **GRIM**** (a more selfish strategy than **GRIM***) which cooperates in m-2 rounds but defects in the last two rounds of the game can invade **GRIM***

Always possible to find a more selfish strategy which defects one round before the opposing strategy which can invade the opposing strategy i.e. GRIM ➔ **GRIM*** ➔ **GRIM**** ➔ ……**ALLD**

A cooperative strategy will always be eliminated from the population ☹

The number of rounds is not always known. Let w be the probability that another round is played after one round is completed.

The probability that the game is over after one round is 1-w

Expected number of rounds played : $\overline{m} = \frac{1}{1-w}$

Key Questions in the Evolution of Cooperation

What is the best strategy in a repeated Prisoner's Dilemma game and is it possible to find such a strategy from a space of many different strategies ?

Can a cooperative strategy emerge and invade a population of selfish agents ?

Does the outcome of the competition depend on the presence of *other* strategies in the population ?

Strategy Space

Two possible classes of strategies exist:

Deterministic Strategies: Given the sequence of **C** and **D** moves that have been played in the previous rounds of the game, a deterministic strategy specifies which move to play in the *current* round of the game.

Stochastic Strategies: Given the sequence of **C** and **D** moves that have been played in the previous rounds of the game, a stochastic strategy specifies the probability of a C or D move in the *current* round of the game.

For m=1, the strategy space is 4-D and all 16 strategies can be represented by the vertices of a 4-D hypercube.

For m>1, the strategy space is $\mathbf{4^m}$ dimensional with each deterministic strategy corresponding to the vertex of a $\mathbf{4^m}$ dimensional hypercube. There are $\mathbf{2^{4\hat{}m}}$ such strategies.

For **deterministic strategies**, the strategy space is discrete and changing from one strategy to another is equivalent to moving from one vertex of the $\mathbf{4^m}$ dimensional hypercube to another.

For **stochastic strategies**, the strategy space is continuous since the probability of a **C** move can vary continuously from 0 to 1. The vertices of the strategy hypercube then correspond to cooperating with probability 1 or 0.

Evolutionary Game of Thrones : Axelrod's 1979 Tournament

People were invited to send in strategies in the form of computer program that would decide whether to play C or D in PD game with another strategy

Each strategy was made to interact with itself as well as every other strategy in the population.

The average payoff for each strategy was calculated every generation as a result of these interactions.

The population of strategies in the next generation was updated by selecting strategies from the current generation with a probability proportional to the average payoff (fitness) for each strategy.

The best strategy was the strategy that took over the entire population.

Result

15 strategies (including ALLC and ALLD) were submitted in the first tournament and 63 in the second tournament.

In both tournaments the winner was TFT

Drawbacks of TFT

Susceptibility to Mistakes: Two players playing **TFT** can end up with a very low payoff when one of them accidentally changes her strategy from C to D.

You: C C C C D C **D** C D C D

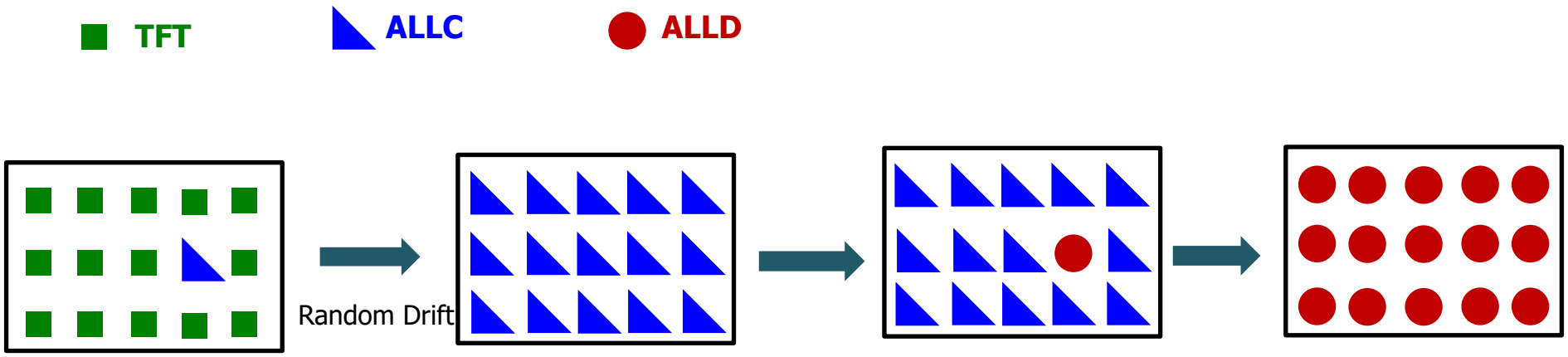
Me: C C C C C D **C** D C D C

A mistake on your part resulting in changing your move from C to D in the fifth round of the game will change the state from mutual cooperation to alternating rounds of cooperation and defection and lead to a very low net payoff for both players.

Payoff for a **TFT** player in the presence of a small amount of behavioral noise is

E(TFT,TFT) = (a+b+c+d)/4 < a since a> (b+c)/2 and a>d

TFT is susceptible to invasion by ALLC by random drift which is susceptible to invasion by ALLD
A population of TFT players can eventually be invaded by ALLD players.



Average Payoff to *each* strategy in an iterated PD game between two strategies

Use the adjacent NetLogo program to answer the following questions

PD Two Person Iterated.nlogo

Questions to Consider

- How does the average payoff to TFT and ALLD change as the number of rounds increases ?
- Is there any difference in the average payoff behaviour in TFT vs TFT compared to TFT vs TFTT, TFT vs ALLC, ALLC vs ALLC ?
- Is there any difference in the average payoff to ALLD in TFT vs ALLD compared to TFTT vs ALLD ? How does the average payoff to ALLD change as the number of rounds increases ?
- Compare the average payoff to ALLD & RANDOM in ALLD vs RANDOM with the average payoff to TFT & RANDOM in TFT vs RANDOM after several rounds of the game ?

  Which strategy TFT or ALLD gets a higher average payoff against RANDOM ?

  In which game does RANDOM get a higher average payoff ?

  Compare the average payoff to ALLD in ALLD vs RANDOM and ALLD vs TFT games.

shows how

The average payoff to each strategy depends on other strategies present in the population

Average Payoff when different strategies interact in a *finite* population

PD N-Person Iterated.nlogo

- There exists a finite population of distinct strategies
- Each strategy interacts at random with itself or another strategy
- The cumulative and average payoffs are calculated after each interaction

Shows how the average payoff to ***each*** strategy changes as a result of interaction with different strategies in the population.

NOTE

There is no *evolution* of population structure since the total number of players playing a given strategy remains *fixed*.

Can a strategy be found that is more cooperative than TFT and

(i) Resists invasion by a more selfish strategy

(ii) Does not suffer from the drawbacks of TFT

Reactive Strategies

A **Reactive strategy** is a **Stochastic** strategy that takes into account the opponent's last move to determine its move (C or D) in the current round.

If the opponent **cooperated** in the last round, it chooses to **cooperate** with probability **p** and **defect** with a probability **(1-p)**

If the opponent defected in the last round, it chooses to **cooperate** with a probability q and **defect** with a probability **(1-q)**

The **Reactive Strategy** **S(p,q)** can be represented by a point in an unit square. The vertices of the unit square correspond to the deterministic strategies **ALLD** (**S(0,0)**); **TFT** (**S(1,0)**); **ALLC** (**S(1,1)**) and **reverse of TFT** (**S(0,1)**)

The repeated Prisoner's Dilemma between two reactive strategies is a ***Markov Chain*** on the state space (CC, CD,DC, DD).

A ***Markov chain*** is a system that undergoes transitions from one state to another among a set of finite and discrete states. Probability of finding the system in the next state depends only on the current state and not on past memory.

	Last Round		Current Round
	My move	**Opponent's Move**	**My move**
State 1:	**C**	**C**	**?**
State 2:	**C**	**D**	**?**
State 3:	**D**	**C**	**?**
State 4:	**D**	**D**	**?**

For each of these states visited in the ***last*** round of the game, what is the probability that the ***current*** round will be found in the states **CC, CD, DC, DD** ?

Algorithm for finding the Reactive Strategy with the highest cumulative payoff

1. Generate ***n*** distinct Reactive Strategies that are distinguished by ***n*** distinct set of values of **(p,q).** Start with an initial configuration in which all the reactive strategies have the same frequency ***1/n***

2. Calculate the ***n*** x ***n*** payoff matrix E(S1*, S2*)

3. Use the replicator equation to study the evolution of population structure to determine which of the ***n*** strategies eventually remain in the population.

$$\frac{dx_i}{dt} = x_i(f_i - \phi)$$

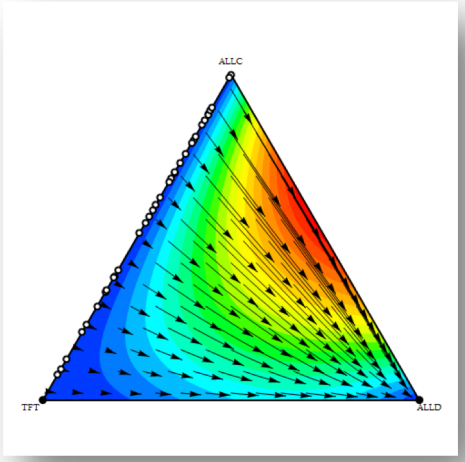
$$\varphi = \sum_i f_i x_i$$

$$f_i = E(S_i^*, S_j^*)x_j$$

4. Is there any cooperative strategy that is better than TFT in surviving against a selfish strategy like ALLD ?

Competition between ALLD, TFT and ALLC

$\bar{m} = 2$

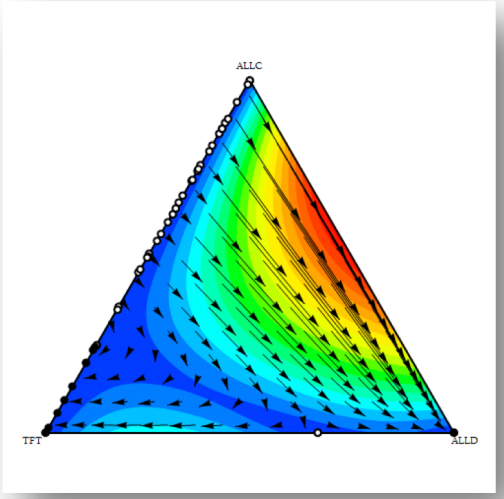


ALLD is the only ESS and x_ALLD ➔ 1 for *any* non-zero *initial* values of x0_ALLD, x0_TFT, x0_ALLC

ALLD vs TFT game: x^*_{TFT} *decreases* as $\bar{m}$ *increases*

$$x^*_{TFT} = \frac{d-b}{\bar{m}(a-d)+2d-b-c} = \frac{1}{2\bar{m}-1}$$

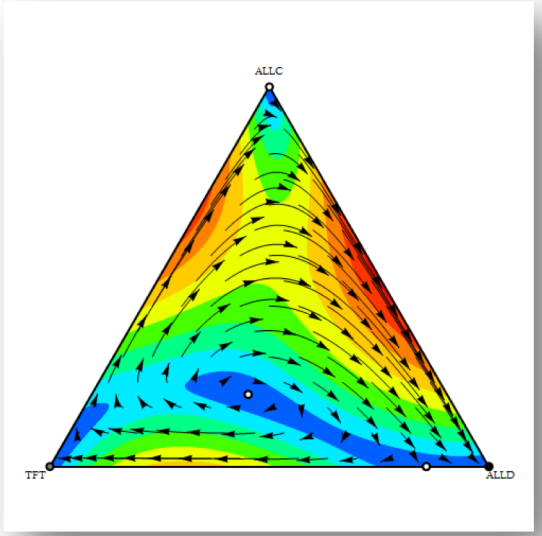
$\bar{m} = 3$



Both TFT and ALLD are an ESS

$$x^T_{ALLC} = \frac{\bar{m}(a-d)+(d-c)}{\bar{m}(c-d)+(d-c)} = \frac{\bar{m}-2}{2(\bar{m}-1)}$$

$\bar{m} = 10$



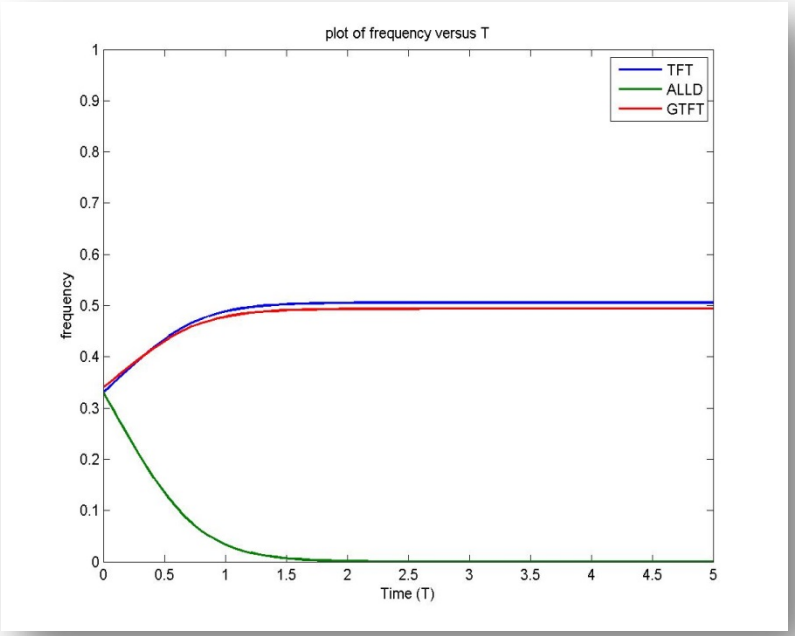
Generated using Dynamo3S by Prateek Verma

A small frequency of ALLD can invade TFT and ALLC if the frequency of x0_ALLC > xT_ALLC

x^T_{ALLC} is the *critical threshold* value

x^T_{ALLC} *increases* as $\bar{m}$ *increases*

Competition between ALLD, TFT and GTFT (without noise)



Obtained by solving the coupled set of replicator equations giving the time evolution of the frequencies of ALLD, TFT and GTFT

There is no behavioral noise in the system

GTFT cannot replace TFT

ALLD cannot replace TFT unless x_TFT is *very* small

ALLD cannot replace GTFT unless x_GTFT is small

Generated using Dynamo3S by Prateek Verma

○ Represents ***unstable*** equilibrium state

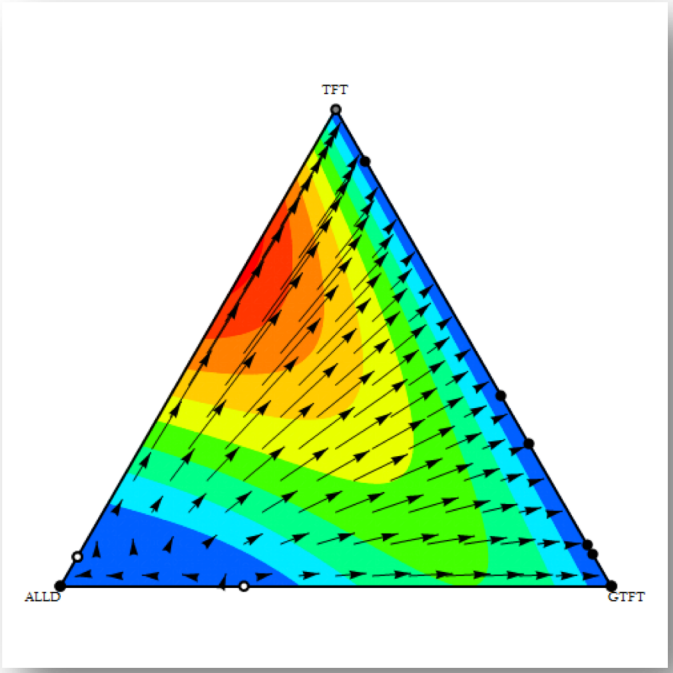
● Represents ***stable*** equilibrium state

Arrows show how the frequencies change over time & attain the final equilibrium state

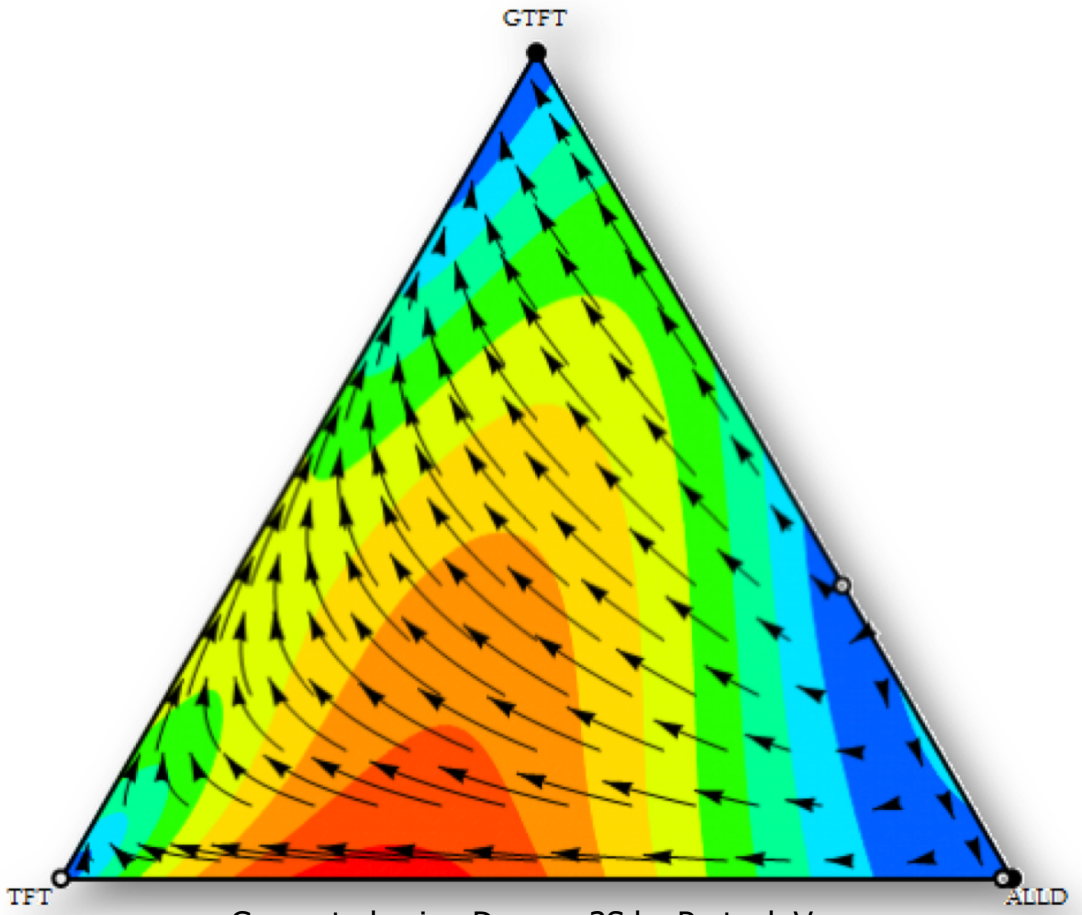
Coloured contours indicate how fast the frequencies are changing in the region.

red-fastest rate of change

blue-slowest rate of change



Competition between ALLD, TFT and GTFT (with noise)



Generated using Dynamo3S by Prateek Verma

GTFT can ***invade*** only in the presence of *both* TFT and ALLD provided x0_TFT is not very small

In a population consisting of only GTFT and ALLD, both strategies are an ESS

ALLD cannot replace GTFT unless x0_GTFT is small

Strategies with last round memory

These are **Stochastic** strategies which decide to cooperate or defect based on both player's move in the last round of the game.

Each strategy is characterized by the probabilities (p1,p2,p3,p4) for cooperating in the current round depending on whether the state in the last round was CC, or CD or DC, or DD.

Competition between two such strategies **S1(p1,p2,p3,p4)** and **S2(p1',p2',p3',p4')** in a repeated Prisoner's Dilemma game is a ***Markov Chain*** on the state space CC,CD,DC,DD.

TFT: p1=p3=1, p2=p4=0. **GTFT: p1=p3=1, p2=p4=1/3**

Is GTFT still the *best* strategy (i.e. strategy with the highest cumulative payoff) when strategies with last round memory are included ?

Probability definitions for Strategy S1 **my move**

p1: Probability of **me** cooperating in the current round when the state of the game in the *last* round was CC

p2: Probability of **me** cooperating when the opponent *defected* in the *last* round i.e. the state was CD

p3: Probability of **me** cooperating when the opponent *cooperated* in the *last* round i.e. the state was DC

p4: Probability of **me** cooperating in the current round when the state of the game in the *last* round was DD

Probability definitions for Strategy S2 **opp move**

p1': Probability of **opp** cooperating in the current round when the state of the game in the *last* round was CC

p2': Probability of **opp** cooperating when the opponent *defected* in the *last* round i.e. the state was DC

p3': Probability of **opp** cooperating when the opponent *cooperated* in the *last* round i.e. the state was CD

p4': Probability of **opp** cooperating in the current round when the state of the game in the *last* round was DD

my move

Strategies with last round memory

Algorithm for finding the *Stochastic Strategy with last round memory* having the highest cumulative payoff

1. Start with a homogeneous population consisting of a stochastic strategy defined by p1=p2=p3=p4=1/2
2. Every 100 generations introduce a new strategy that is chosen from a random distribution of strategies.
3. Solve the eigenvalue equation with the new strategy to determine the probability of finding the game in one of the 4 possible Markov states at equilibrium.

$$x = xT$$

$$E(S_1, S_2) = ax_1 + bx_2 + cx_3 + dx_4$$

4. Using the replicator equation, evolve the system to check whether the new strategy becomes extinct or coexists with other strategies or gets fixed in the population by eliminating all other strategies.

5. Is there any cooperative strategy that is better than GTFT in surviving against ALLD-like strategies ?

Win-Stay, Lose-Shift (WSLS) Strategy

WSLS : p1=1,p2=p3=0,p4=1

Strategy : Cooperate when both cooperates or both defects in the last round; otherwise defect.

WSLS strategy decides to keep playing the same move if it is winning i.e. getting a payoff of **a** or **c** but changes its move in the *current* round if it is loosing i.e. getting a smaller payoff of **b** or **d** in the *last* round.

WIN

Last Round		Current Round	
My move	C	My move	C
Opponent's Move	C		
My Payoff	a		

Last Round		Current Round	
My move	D	My move	D
Opponent's Move	C		
My Payoff	c		

LOSE

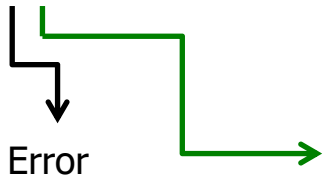
Last Round		Current Round	
My move	C	My move	D
Opponent's Move	D		
My Payoff	b		

Last Round		Current Round	
My move	D	My move	C
Opponent's Move	D		
My Payoff	d		

Advantages of WSLS Strategy

1. WSLS has error correcting ability. Cooperate when both cooperates or both defects in the last round; otherwise defect.

WSLS : **CCCC D DCCC**
WSLS : **CCCC C DCCC**

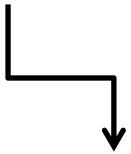


Error

Error corrected because strategy is changed to increase payoff

2. WSLS dominates ALLC in the presence of behavioral noise and therefore wont change to ALLC via random drift.

WSLS : **CCCC D DDDD**
ALLC : **CCCC C CCCC**



Error in WSLS increases its payoff and is maintained to exploit ALLC

3. Competition between **WSLS** and **ALLD**

WSLS : **C D C D C D C D**
ALLD : **D D D D D D D D**

WSLS cooperates with **ALLD** in every other round

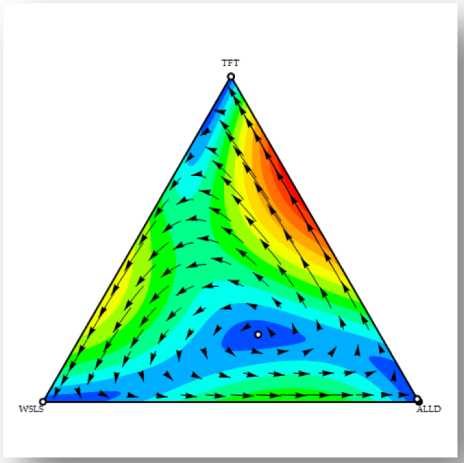
$$M = \begin{array}{c} \\ \text{WSLS} \\ \text{ALLD} \end{array} \begin{array}{c} \begin{array}{cc} \text{WSLS} & \text{ALLD} \end{array} \\ \begin{bmatrix} ma & (m/2)(b+d) \\ (m/2)(c+d) & md \end{bmatrix} \end{array}$$

WSLS is stable against invasion by ALLD iff E(WSLS,WSLS)>E(ALLD,WSLS) ➔ **a > (c+d)/2**

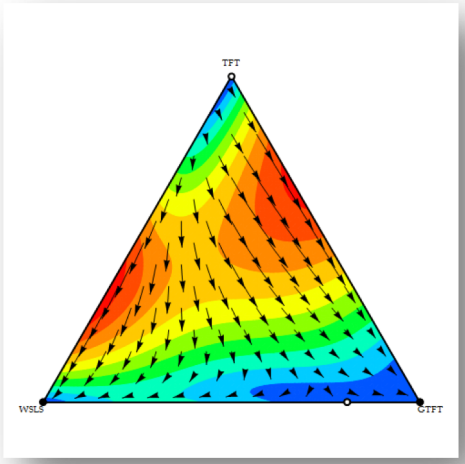
Note: ALLD is also an ESS since E(ALLD,ALLD)>E(WSLS,ALLD) ➔ d>b which is always true

Competition between WSLS and any two of ALLD, TFT, GTFT (with noise)

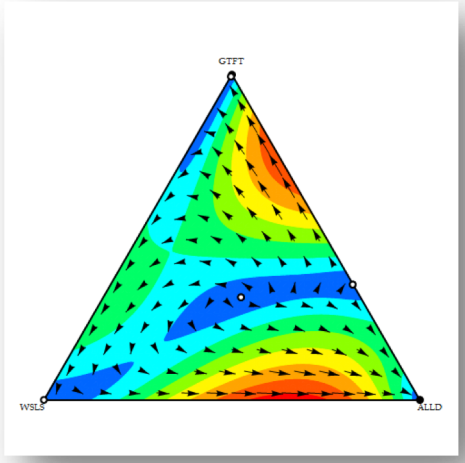
WSLS vs ALLD vs TFT



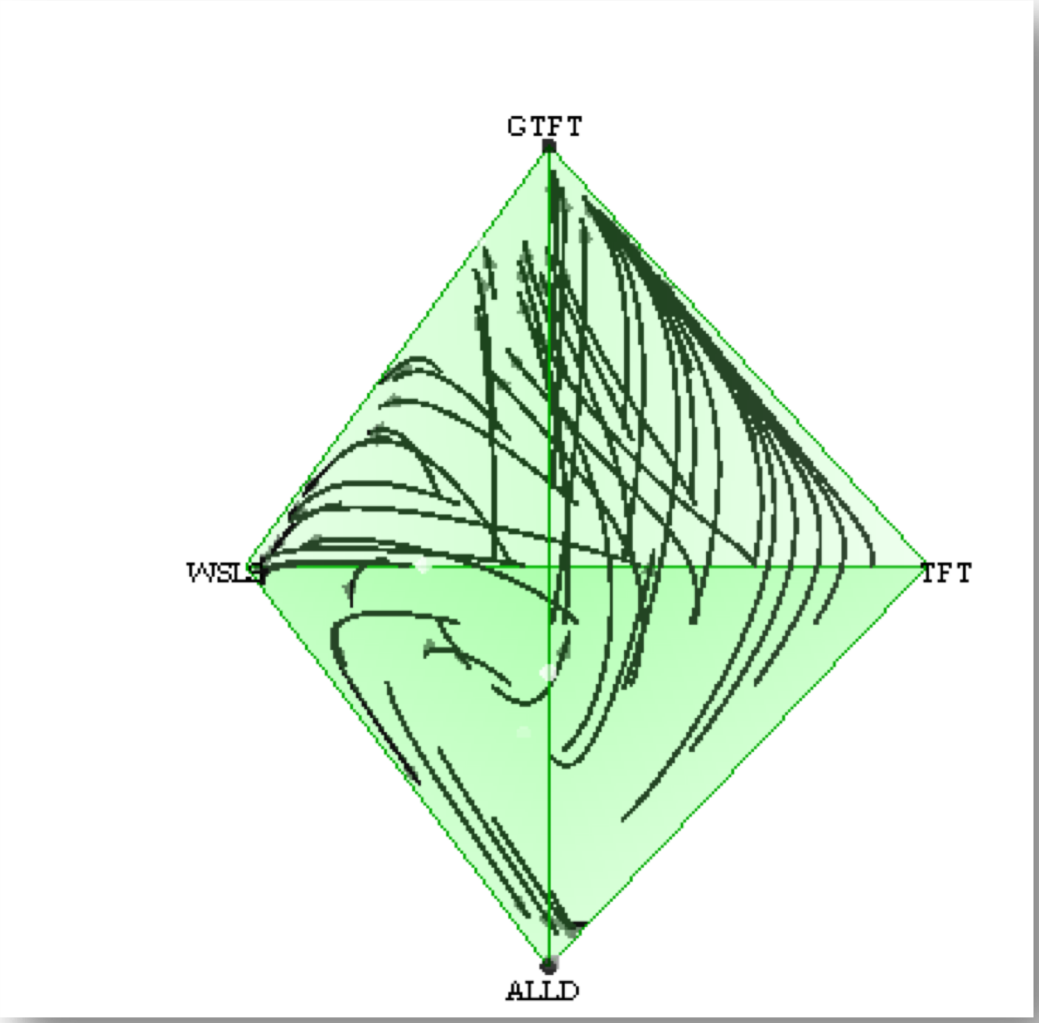
WSLS vs TFT vs GTFT



WSLS vs ALLD vs GTFT



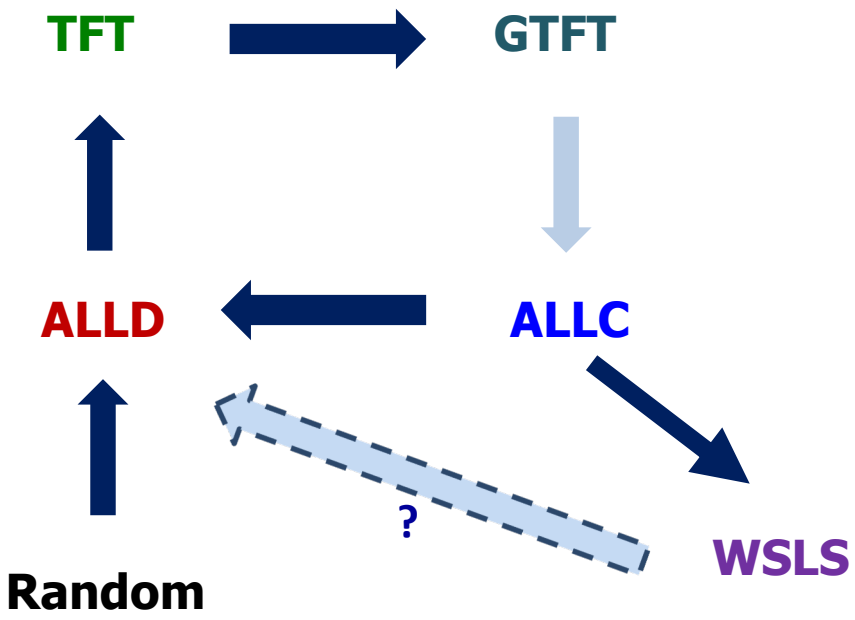
Competition between ALLD, TFT, GTFT and WSLS (with noise)



Generated using Dynamo4S by Prateek Verma

WSLS dominates ALLC and resists invasion by ALLD

Dynamics in Strategy Space



Games in Finite Populations

For the Payoff Matrix **M=**

	C	**D**
C	a	b
D	c	d

b<c and a>d

No. of C players = **i**

No. of D players = **N-i**

Prob. that C interacts with another C = (i-1)/(N-1)

Prob. that C interacts with another D = (N-i)/(N-1)

Prob. that D interacts with another C = i/(N-1)

Prob. that D interacts with another D = (N-i-1)/(N-1)

Expected payoff to C when it interacts with C = a(i-1)/(N-1)

Expected payoff to C when it interacts with D = b(N-i)/(N-1)

Total expected payoff for C : $\mathbf{F_i = (a(i-1)+b(N-i))/(N-1)}$

Total expected payoff for D : $\mathbf{G_i = (ci+d(N-i-1))/(N-1)}$

Define fitness of **C** as : $f_i = 1- w + w F_i$; fitness of **D** as : $g_i = 1- w + w G_i$

w=1 ➔ strong selection; fitness completely determined by interactions

w=0 ➔ no selection between C & D

w<<1 ➔ weak selection

$f_1>g_1$ ➔ $F_1>G_1$ ➔ $b(N-1) > c + d(N-2)$

Moran Process in Games in Finite Populations

C's and D's are picked for reproduction with a probability proportional to their mean fitness and for death randomly.

Probability of picking **C** for reproduction and **D** for death : $p_{i,i+1} \equiv \alpha_i = \left(\frac{{}^{i}f_i}{if_i + (N-i)g_i}\right)\left(\frac{N-i}{N}\right)$

Probability of picking **D** for reproduction and **C** for death : $p_{i,i-1} \equiv \beta_i = \left(\frac{(N-i)g_i}{if_i + (N-i)g_i}\right)\left(\frac{i}{N}\right)$

$$\gamma_i = \frac{\beta_i}{\alpha_i} = \frac{g_i}{f_i}$$

$$\rho_C = \frac{1}{1 + \sum_{k=1}^{N-1}\prod_{i=1}^{k}\frac{g_i}{f_i}} \qquad \rho_D = \frac{\prod_{i=1}^{N-1}\frac{g_i}{f_i}}{1 + \sum_{k=1}^{N-1}\prod_{i=1}^{k}\frac{g_i}{f_i}} \qquad \frac{\rho_C}{\rho_D} = \prod_{i=1}^{N-1}\frac{f_i}{g_i}$$

In the limit w➔0, $\rho_C > \frac{1}{N}$ leads to the inequality

a(N-2) + b(2N-1) > c(N+1) + d(2N-4) which in the limit N>>1,reduces to a + 2b > c + 2d

For fixed, a,b,c,d, the above inequality gives a lower bound on the population size N

N > Nc $\qquad N_C = \frac{2a + b + c - 4d}{a + 2b - c - 2d}$

Nc is the *minimum* size of the population necessary for selection to favour fixation of cooperators

Evolutionary Stability in Finite Populations

If a>c and b>d, **C** is a strict Nash equilibrium as well as an ESS and selection will always favour fixation of **C** and oppose fixation of **D** in a finite population of any size.

If a>c and b<d, both **C** and **D** is an ESS. According to the infinite population analysis, a small fraction of **C** mutants cannot invade a population consisting predominantly of **D** players.

What happens for finite populations ?

Fixation of **C** will be favoured by selection only if $\rho_C > \frac{1}{N}$ even if $F_1 > G_1$ ➔ $N > N_C$

Fixation of **D** can still be favoured even if $F_1 > G_1$ provided $\rho_D > \frac{1}{N}$ ➔ $N < N_C$

Condition for a strategy to be an ESS has to be modified for *finite* populations.

In a *finite* population of size N, a strategy **C** is an ***ESSN*** if

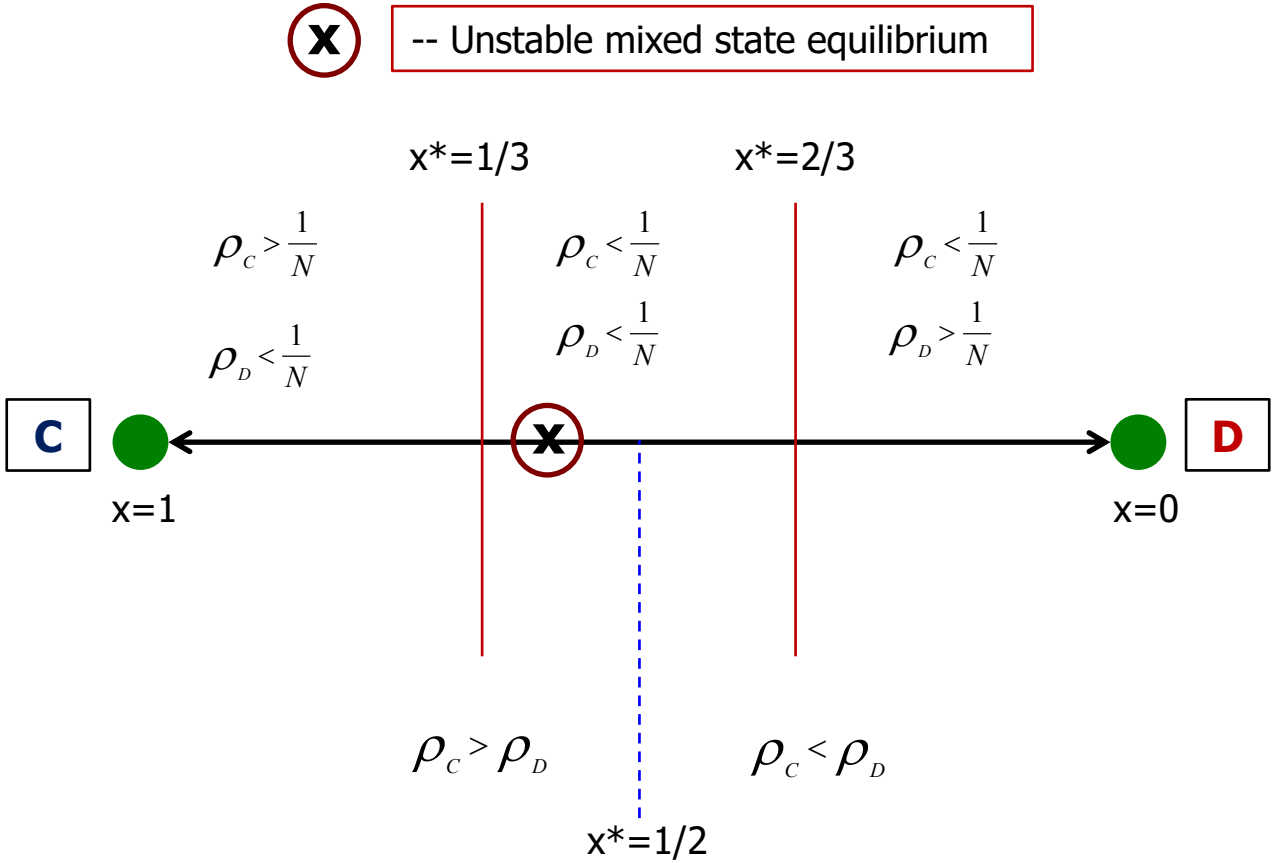
(i) A single mutant of any other strategy has lower fitness than **C**

(ii) The fixation probability of *every other strategy* must be smaller than the fixation probability of a neutral strategy and the fixation probability of **C** must be larger than the fixation probability of a neutral strategy

C is an ***ESSN*** if $F_1 > G_1$ and $\rho_C > \frac{1}{N}$ and $\rho_D < \frac{1}{N}$

D is an ***ESSN*** if $G_1 > F_1$ and $\rho_D > \frac{1}{N}$ and $\rho_C < \frac{1}{N}$

Fixation Probabilities and the 1/3 Law for w<<1 and N>>1



Risk Dominance: If both **C** and **D** is a strict Nash Equilibrium in the conventional sense i.e. if a>c and d>b then which strategy has a higher fixation probability ?

C is Risk Dominant if $\rho_C > \rho_D$ ➔ **a + b > c + d** when w<<1 and N>>1

D is Risk Dominant if $\rho_D > \rho_C$

A strategy is Risk Dominant if the total payoff for that strategy is larger than the total payoff for every other strategy.

The Risk Dominant strategy has a greater fixation probability in the limit w<<1 and N>>1

TFT can Invade ALLD in a Finite Population

For the Payoff Matrix **M =**

	TFT	**ALLD**
TFT	ma	b+(m-1)d
ALLD	C+(m-1)d	md

No. of TFT players = **i**

No. of ALLD players = **N-i**

c>a>d>b

According to the *infinite population* analysis, for m > (c-d)/(a-d), both TFT and ALLD are an ESS and each strategy is stable against invasion by either strategy.

In finite populations, **TFT** can get fixed in the population even if $F_{TFT} < G_{ALLD}$ provided $\rho_{TFT} > \frac{1}{N}$

If F_i and G_i is the fitness of **i TFT** and (**N-i) ALLD** players,

$$F_i = \frac{ma(i-1) + (b+(m-1)d)(N-i)}{N-1} \qquad G_i = \frac{(c+(m-1)d)i + md(N-i-1)}{N-1}$$

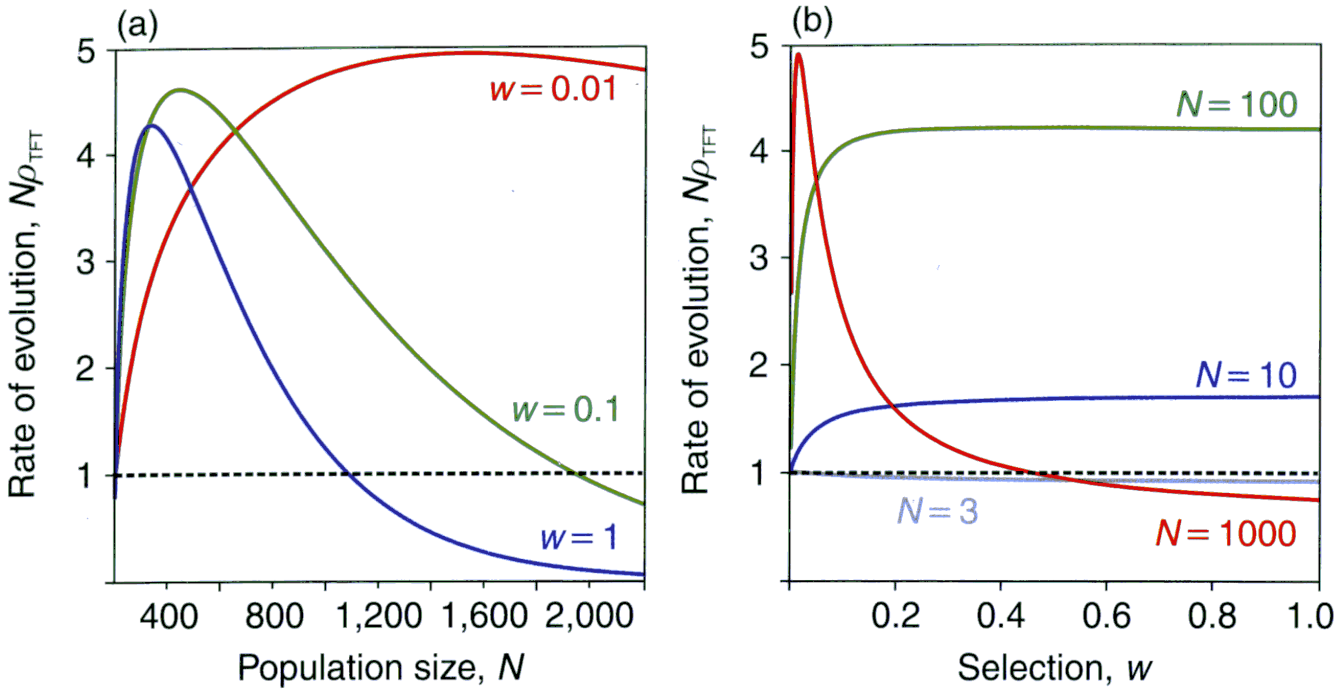
$F_1 = F_{TFT} = b+(m-1)d$ and $G_1 = G_{ALLD} = (c+(m-1)d + md(N-2))/(N-1)$

For w➔0 and fixed N, $\rho_{TFT} > \frac{1}{N}$ gives a lower bound on m: $m > \frac{c(N+1) + d(N-2) - b(2N-1)}{(a-d)(N-2)}$

When N=2, m>∞, When N=3: m>10 When N=4: m>6

For N>>1, lower bound on m: $m > \frac{c+d-2b}{(a-d)} \Rightarrow m > 3$ when a=3, b=0, c=5, d=1

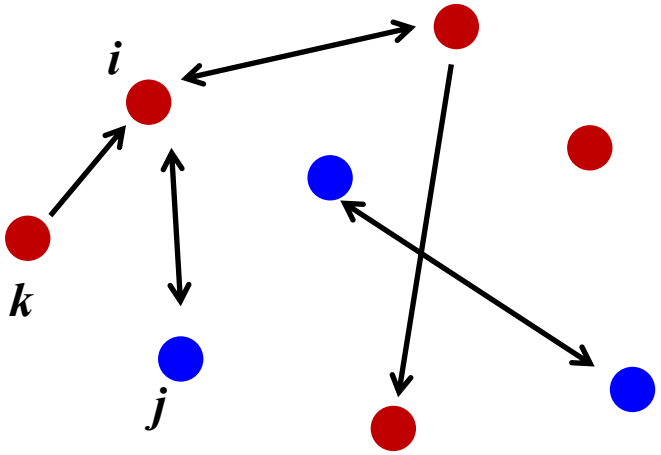
For fixed m, $\rho_{TFT} > \frac{1}{N}$ gives a lower bound on N: $N > \frac{2ma + b + c - 2d(m+1)}{ma + 2b - c - d(m+1)}$



Evolutionary Graph Theory

Questions

- How does the fixation probability of a mutant change when the population is structured i.e. only certain members of the population can replace others during the course of evolution.
- If a structured population is represented by a graph, with vertices representing members and edges representing interaction between corresponding members, is it possible to characterize ***all*** graphs that have the ***same*** evolutionary dynamics.
- Can certain structured populations increase the fixation probabilities of advantageous mutants ?
- Can certain structured populations eliminate the effect of selection ?



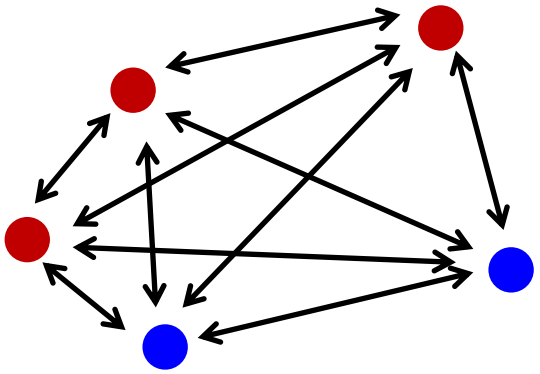
Structured Population

Not all vertices are connected by an edge

The edges can have different weights

i can replace j and j can replace i

k can replace i but i cannot replace k



Unstructured Population

There is an edge between ***any*** two vertices

All edges have the ***same*** weight

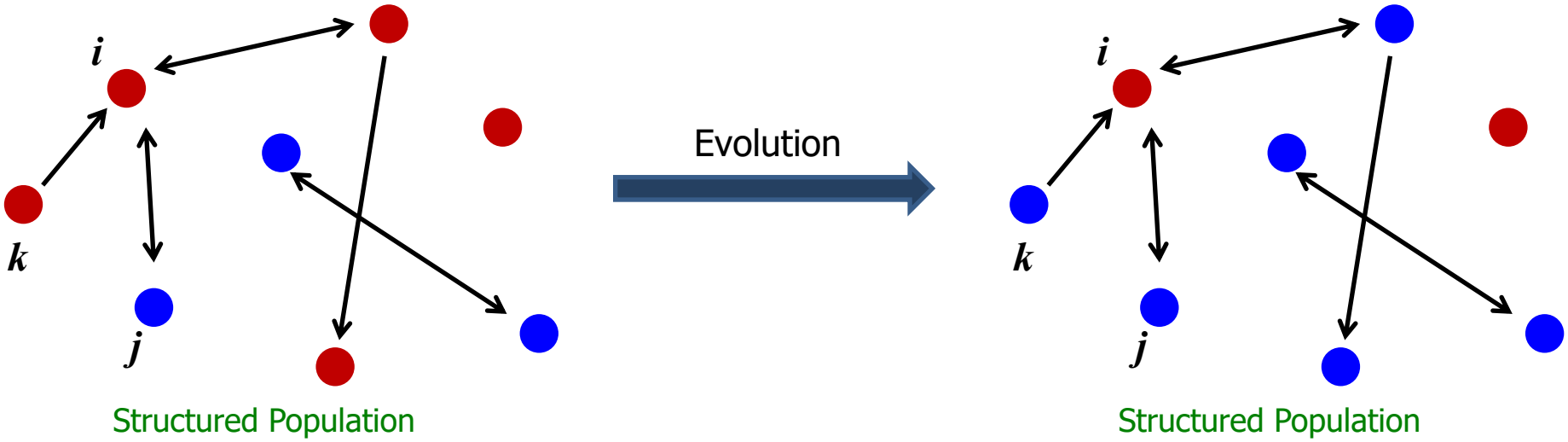
Formulating Evolution on Networks

A graph (network) can be completely specified by a stochastic matrix $W=[w_{ij}]$

$W=[w_{ij}]$ is an N x N stochastic matrix that determines the probability of replacing the j'th member of the population by the i'th member.

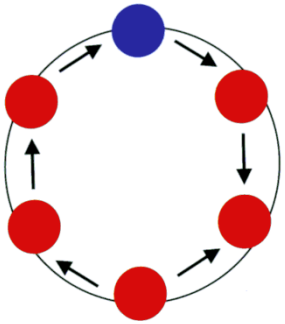
w_{ij} =0 if there is no directed edge *from* **i** *to* **j** ➔ offspring of i ***cannot*** replace j

$\sum_{j=1}^{N} w_{ij} = 1$ since the i'th member picked for reproduction has to replace someone

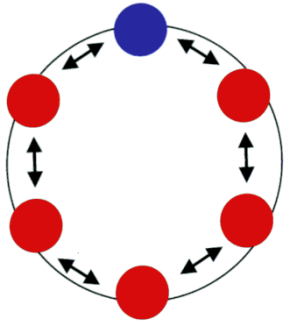


Population composition changes as the population evolves but the rules for replacement via the Moran process remain the same.

Fixation Probability of a mutant that arises in a structured population



Directed Cycle



Cycle

The i'th member can only be replaced by the member preceding it i.e the (i-1)th member.

Fitness of B (blue) = r
Fitness of A (red) = 1

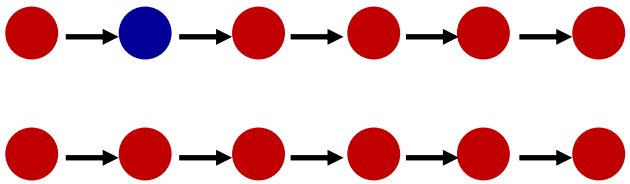
Due to the nature of the structured population (only nearest neighbour replacements are allowed), there can be only one cluster of B's. Fragmentation of clusters into two or more sub-clusters is not possible.

$$\rho_B = \frac{1}{1+\sum_{k=1}^{N-1}\prod_{i=1}^{k}\gamma_i} = \frac{1-1/r}{1-1/r^N}$$

Fixation probability of B on a directed cycle is identical to the fixation probability of B in the Moran process (unstructured population)

Fixation probability of a mutant randomly placed on a "Line" graph

Line



Rules of Replacement: Every member is replaced only by the member preceding it. The last member replaces itself.

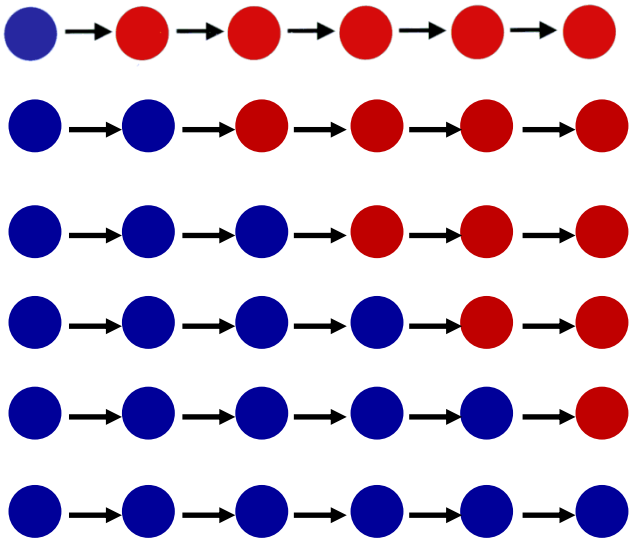
If the mutant B arises at any position other than the first position in the line, it will be replaced by A and become extinct.

Probability that B arises in any positions from i=2…N is (N-1)/N since there are N-1 such positions.

Probability that B appears in position 1: 1/N

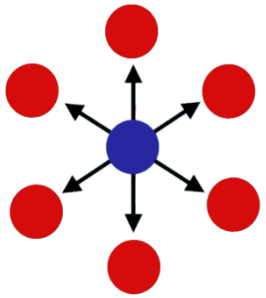
The mutant B will definitely be fixed if it arises in position 1

Fixation probability of B: $\rho_B = \frac{1}{N}$



Fixation probability differs from the Moran process and is independent of the fitness of members.

Invasion probability of a mutant randomly placed on a "Burst" graph



Burst

Rules of Replacement: Every member is replaced only by the member at the centre with equal probability. The central member *cannot* be replaced by any other peripheral member or itself.

If the mutant B arises at any position other than the ***central*** position in the star, it will be replaced by the A at the centre and become extinct.

Probability that B arises in any positions from i=2…N is (N-1)/N since there are N-1 such positions.

Probability that B appears in the central position: 1/N

The mutant B will definitely be fixed if it appears at the centre of the star

Invasion probability of B: $\rho_B = \frac{1}{N}$

Invasion is *independent* of the fitness of members and equivalent to that of a neutral mutant in the Moran process.

Both the "Line" and "Burst" graphs are suppressors of selection

Graphs which are suppressors or amplifiers of selection

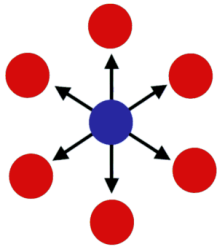
$$\rho_B = \frac{1-1/r}{1-1/r^N}$$ Invasion probability of a single mutant with a relative fitness ***r*** in a Moran process

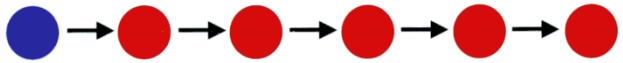
If the fixation probability of a single mutant with a relative fitness r on the *structured graph* G is ρ_G

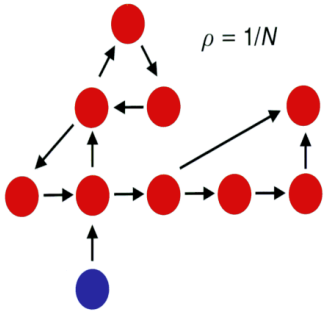
If $\rho_G > \rho_B$ when r>1 ➔ G is an amplifier of selection. G favours selection over drift

If $\rho_G < \rho_B$ when r>1 ➔ G is an suppressor of selection. G favours drift over selection

If $\rho_G = \frac{1}{N}$ when r>1 ➔ G is the strongest possible suppressor of selection.







All Graphs with a ***single root*** have the same invasion probability: $\rho_G = \frac{1}{N}$

The Isothermal Theorem

$T_j = \sum_{i=1}^{N} w_{ij}$ T_j : Sum of all weights that lead to the vertex.
A vertex with high temperature is replaced more frequently than a vertex with low temperature

Isothermal Graph: All vertices have the *same* temperature ➔ $T_j = \sum_{i=1}^{N} w_{ij} = const = 1$

Isothermal Theorem: The fixation probability of a single mutant on a graph G is equivalent to the fixation probability on an unstructured graph (Moran process) iff G is an isothermal graph.

$v_i = 1$ ➔ vertex i is occupied by B, $v_i = 0$ ➔ vertex i is occupied by A

$m = \sum_{i=1}^{m} v_i$ ➔ No. of B's in the population

$$p_{m,m+1} = \frac{r\sum_i \sum_j w_{ij} v_i (1-v_j)}{rm + N - m} \qquad p_{m,m-1} = \frac{\sum_i \sum_j w_{ij} v_j (1-v_i)}{rm + N - m}$$

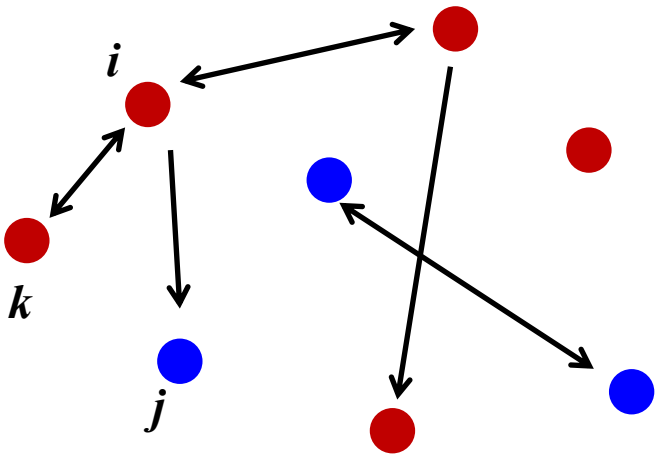
Fixation probability is the same as the Moran process if $\frac{p_{m,m-1}}{p_{m,m+1}} = \frac{1}{r}$ ➔

$$\sum_i \sum_j w_{ij} v_j (1-v_i) = \sum_i \sum_j w_{ij} v_i (1-v_j) \quad \rightarrow \quad \sum_i \sum_j w_{ji} v_i (1-v_j) = \sum_i \sum_j w_{ij} v_i (1-v_j)$$

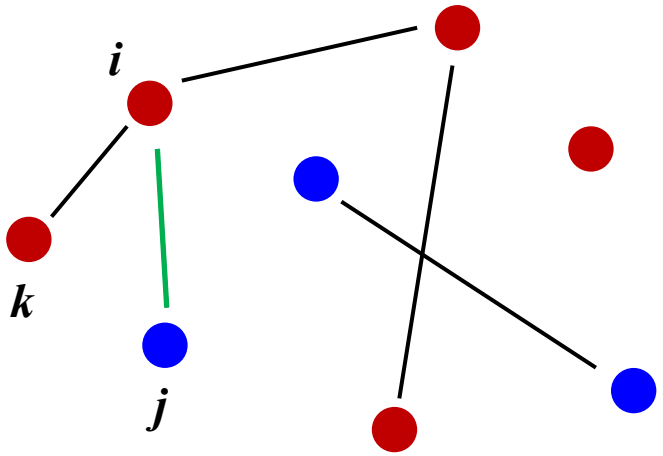
The above equation is valid for all i. Specifically for a particular i=k such that $v_k = 1, v_i = 0$ for all $i \neq k$

➔ $\sum_j w_{jk} = \sum_j w_{kj}$ Since $\sum_{j=1}^{N} w_{kj} = 1$ ➔ $T_k \equiv \sum_{j=1}^{N} w_{jk} = 1$

Alternative way to represent evolutionary dynamics on Graphs



Structured Population



Structured Population

Instead of picking a vertex for reproduction and another vertex (which it can replace) for death

Pick an Edge (ij) with probability proportional to (w_{ij} x f_i)

f_i - fitness of the member at vertex i

Arrows are no longer necessary: weight w_{ij} contains information about which member replaces which

W=[w_{ij}] need not be a stochastic matrix. w_{ij} can be any *non-negative* number $\sum_{j=1}^{N} w_{ij} \neq 1$

Fixation probability is same as that of a Moran Process in unstructured populations if the graph is a circulation i.e.

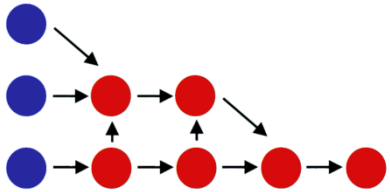
$$\sum_j w_{jk} = \sum_j w_{kj}$$

Every *Isothermal* graph is a *Circulation* but not every Circulation is *Isothermal*

A *single* mutant B appearing in a graph with multiple roots can never get fixed since it cannot replace A located in one or more of the multiple roots.

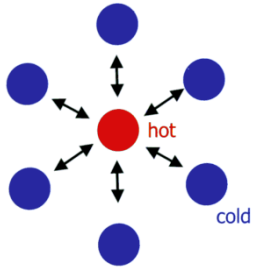
If a mutant appears at one of the roots, the lineage generated can never become extinct. Allows coexistence of both A and B in the population.

A root is a vertex with no edge pointing to it



Graph with multiple roots

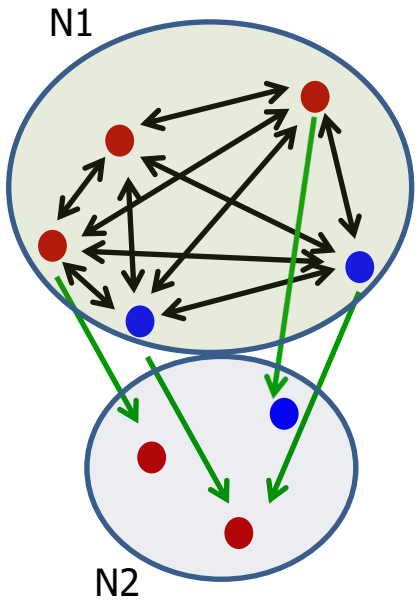
Star Graph is an *amplifier* of selection: For r>1, a single mutant with relative fitness r has a higher fixation probability that is equivalent to the fixation probability with relative fitness r^2 in a Moran process (unstructured graph) in the limit of ***large*** N



Star

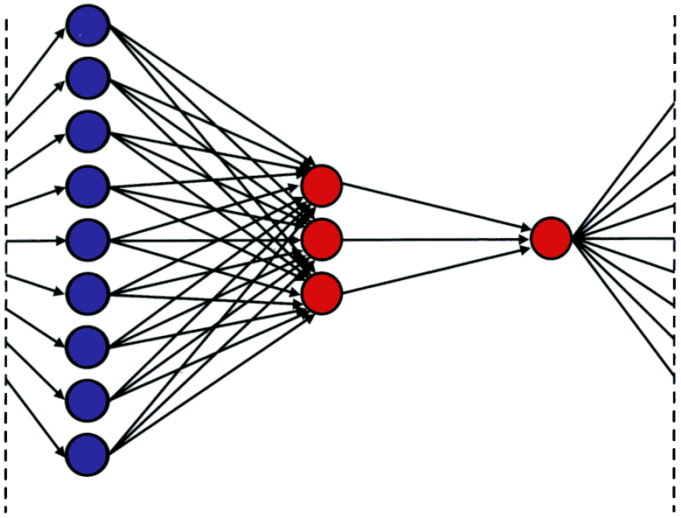
$$\rho_{B,star} = \frac{1-1/r^2}{1-1/r^{2N}}$$

A compartmentalized graph which has a complete sub-graph is a *suppressor* of selection: For r>1, a single mutant with relative fitness r has a lower fixation probability than the corresponding fixation probability in a Moran process (unstructured graph).
$N=N_1+N_2$; where N1 is the size of the complete (unstructured) sub-graph and N2 is the size of the second component.



$$\rho_B = \frac{1}{1+\sum_{j=1}^{N_1-1}\prod_{k=1}^{j}\gamma_k} = \frac{1-1/r}{1-1/r^{N_1}}$$

The Funnel graph is a strong amplifier of selection



The Funnel graph has k+1 layers labelled j=0,1,…k

The zeroth layer has just one vertex and the layer j has m^j vertices

Edges originating from vertices in layer j lead to vertices in layer j-1

Edges originating from the single vertex in the layer j=0 lead back to vertices in the layer j=k

In the limit of ***large*** k and ***large*** m,

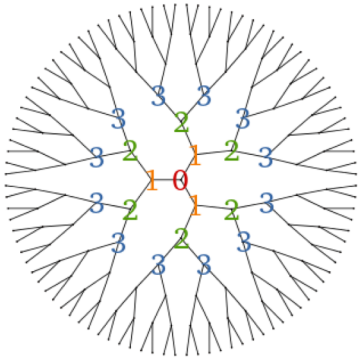
$\rho_G \to 1$ **for r>1**

$\rho_G \to 0$ **for r<1**

Reference: Lieberman, Hauert, Nowak; Evolutionary Dynamics on Graphs, Nature 2005

Constraints on the Theoretical Formulation

Image Source: Wikipedia



Caley Tree/Bethe Lattice with k=3

- Caley Tree/Bethe Lattice
  - *Regular* graph with each node having **k** neighbours
  - Graph does not have any loops

- Theoretical analysis valid for
  - N>>k
  - Weak selection limit holds i.e. w<<1 when separation of time-scales is possible
  - Uses the pair approximation which is valid only for Bethe lattices i.e. graphs without any loops.

Pair Approximation ➔ frequencies of larger clusters obtained from pair frequencies

Relations between stochastic variables for evolution of cooperation on graphs

$$p_A + p_B = 1 \Rightarrow p_B = 1 - p_A$$

$$q_{A|B} + q_{B|B} = 1; q_{B|A} + q_{A|A} = 1$$

$$p_{AA} = q_{A|A} p_A$$

$$p_{AB} = q_{A|B} p_B = q_{A|B}(1 - p_A)$$

$$p_{BB} = q_{B|B} p_B = (1 - q_{A|B})(1 - p_A)$$

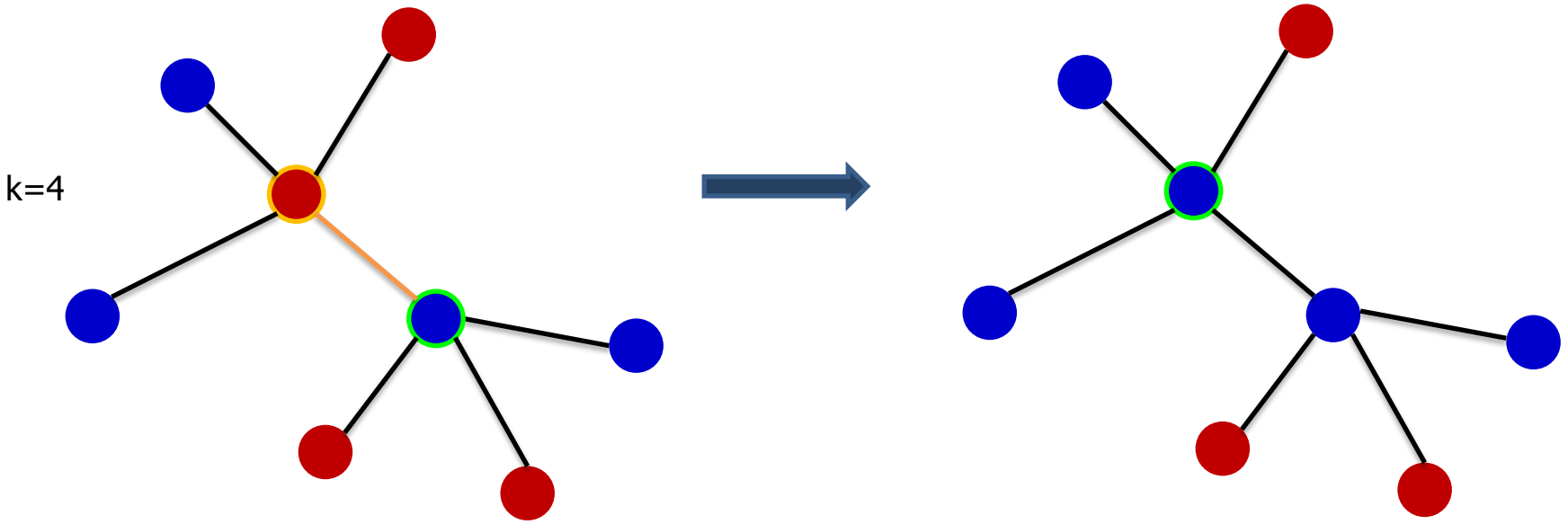
$$p_{BA} = p_{AB} \Rightarrow q_{B|A} p_A = q_{A|B} p_B \Rightarrow q_{A|B} = (1 - q_{A|A})\left(\frac{p_A}{1 - p_A}\right)$$

Only 2 of the 6 stochastic variables are *independent*: $p_A; p_{AA}$

AIM: Obtain dynamical equations in terms of these variables and solve them under certain approximations to obtain the condition for the fixation probability of A (ρ_A) > 1/N

Key approx.: *Pair frequencies equilibrate in a much faster time-scale than individual frequencies* in the weak selection limit (w<<1)

Illustration of the Death-Birth Process : Updating a B-player



Focal player selected for death is B

A-neighbour of focal B selected to replace B has payoff: $P_A = b + 3q_{A|A}a + 3q_{B|A}b$

B-neighbour of focal B *not* selected to replace B has payoff: $P_B = d + 3q_{B|B}d + 3q_{A|B}c$

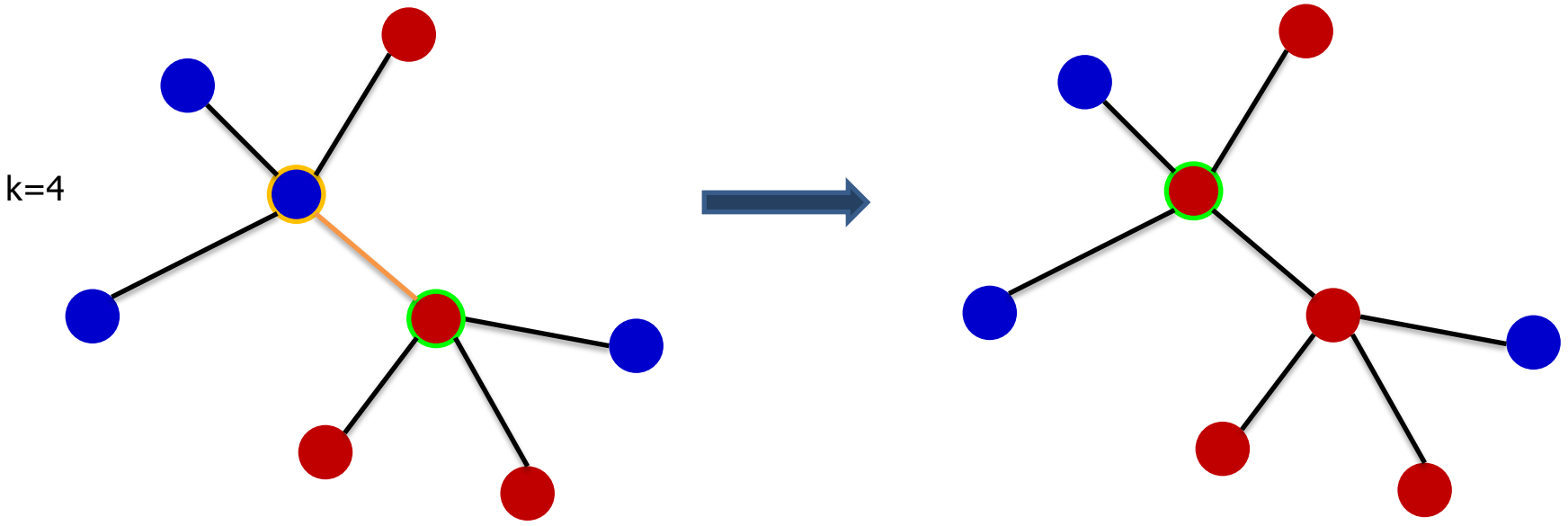
$f_A = 1 - w + wP_A$: Fitness of A-neighbour of focal B

$f_B = 1 - w + wP_B$: Fitness of B-neighbour of focal B

Probability that replaces the focal :

$$\frac{k_A f_A}{k_A f_A + k_B f_B}$$

Illustration of the Death-Birth Process : Updating an A-player



Focal player selected for death is A

A-neighbour of focal A *not* selected to replace A has payoff: $P_A = a + 3q_{A|A}a + 3q_{B|A}b$

B-neighbour of focal A selected to replace A has payoff: $P_B = c + 3q_{B|B}d + 3q_{A|B}c$

$g_A = 1 - w + wP_A$: Fitness of A-neighbour of focal A $\quad g_B = 1 - w + wP_B$: Fitness of B-neighbour of focal A

Probability that replaces the focal :

$$\frac{k_B f_B}{k_A f_A + k_B f_B} \; ; k_A + k_B = 4$$

Condition for Spread of Cooperation on Networks

Ohtsuki *et al.* A simple rule for evolution of cooperation on graphs and social networks; Nature 441 (2006) 502

Death-Birth (DB) updating: $\rho_C > \frac{1}{N} > \rho_D$ ➔ $\frac{b}{c} > k$

Imitation (IM) updating: Payoff of focal individual being updated also matters.

Focal individual (F) compares her fitness with her neighbours.

F retains her strategy if f_F > f_neighbour (neighbour is randomly selected)

F imitates neighbour with probability proportional to neighbour's fitness if f_F<f_neighbour

$f_F = 1-w+w(k_Ac+k_Bd)$ where f_F: fitness of the Focal B-player with k_A A-neighbours and k_B B-neighbours

Probability that a focal B-player adopts the strategy of an A-neighbour: $\frac{k_A f_A}{k_A f_A + k_B f_B + f_F}$

Compare with

Probability that a focal B is replaced by an A-neighbour in Death-Birth updating: $\frac{k_A f_A}{k_A f_A + k_B f_B}$

$g_F = 1-w+w(k_Aa+k_Bb)$ where g_F: fitness of the Focal A-player with k_A A-neighbours and k_B B-neighbours

Probability that a focal A-player adopts the strategy of a B-neighbour: $\frac{k_B g_B}{k_A g_A + k_B g_B + g_F}$

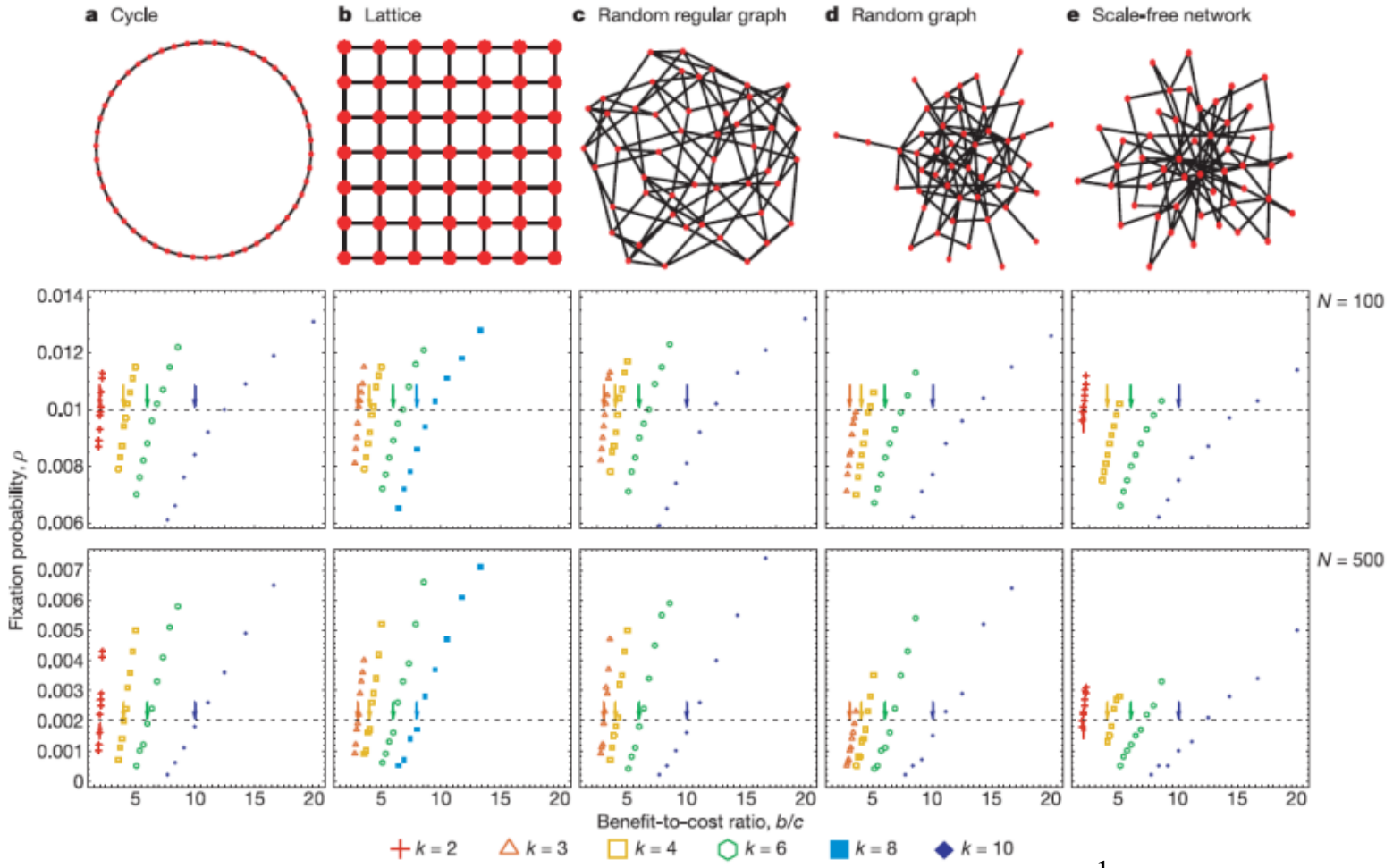
$$\rho_C > \frac{1}{N} > \rho_D \quad \rightarrow \quad \frac{b}{c} > k+2$$

Ratio of Benefit to Cost of cooperation determines whether selection favours spread of cooperation and fixation of cooperators

Condition for Spread of Cooperation on Networks

Ohtsuki *et al.* A simple rule for evolution of cooperation on graphs and social networks; Nature 441 (2006) 502

Birth-Death (BD) updating: $\rho_D > \frac{1}{N} > \rho_C$ ⟹ Selection never favours fixation of cooperators

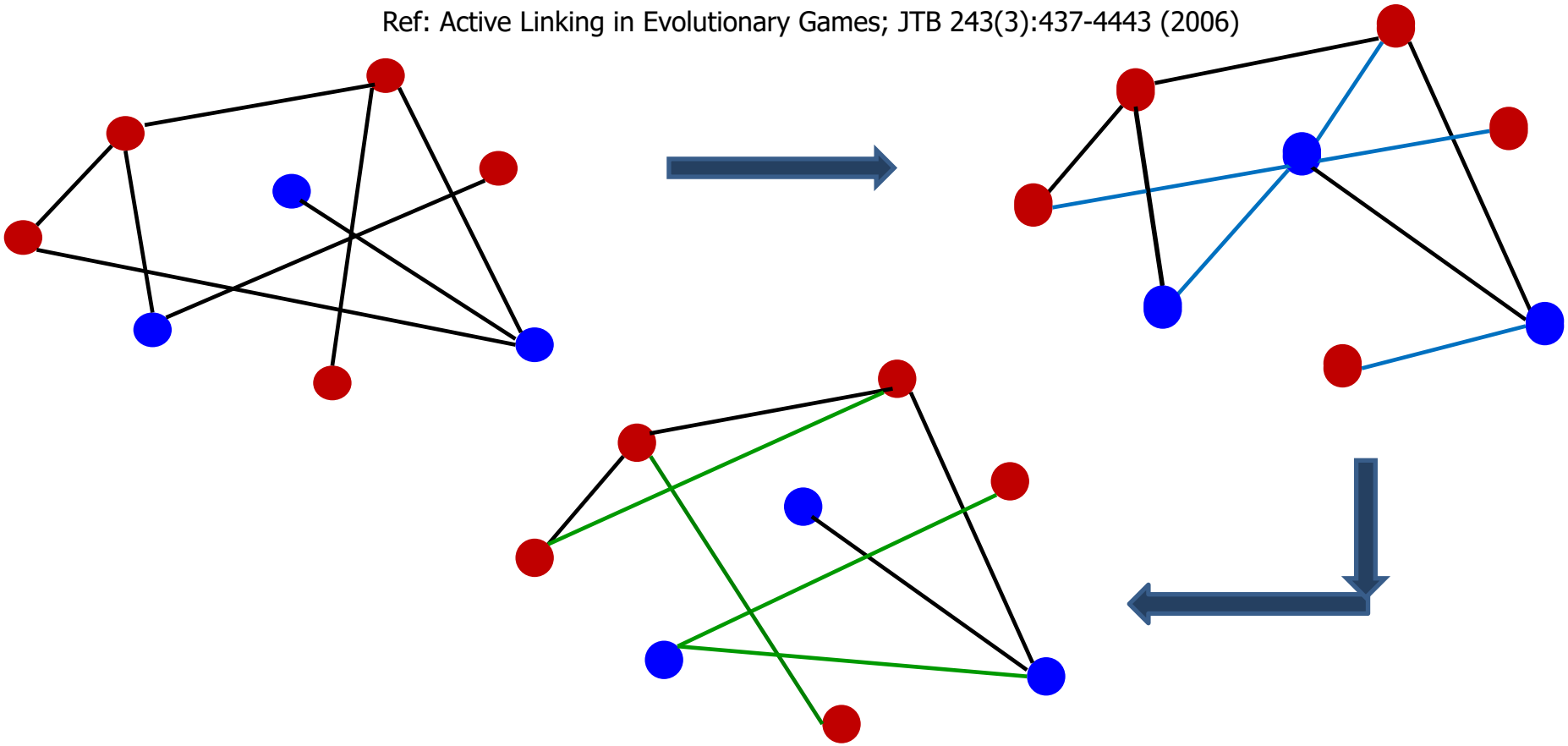


Arrow indicates b/c=k. Theoretical Prediction : b/c > k ➔ $\rho_C > \frac{1}{N} > \rho_D$

- Discrepancy with theoretical prediction observed for *non-regular* graphs
- Discrepancy increases with increasing k but decreases with increasing N

Evolutionary Games on Networks with Dynamic Linking

Ref: Active Linking in Evolutionary Games; JTB 243(3):437-4443 (2006)



Two Relevant Time-Scales

Time-Scale of Linking Dynamics: τ_a

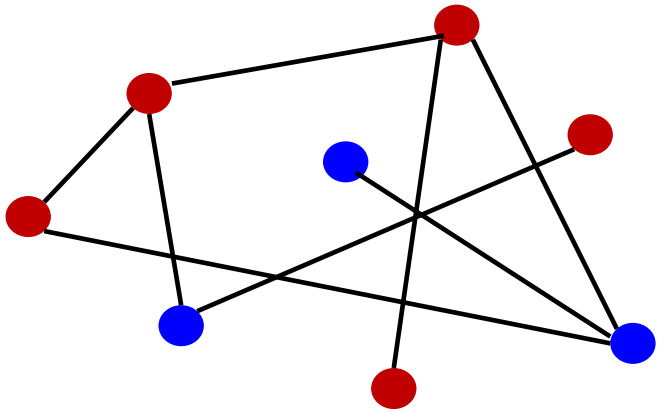
$\tau_a >> \tau_e$

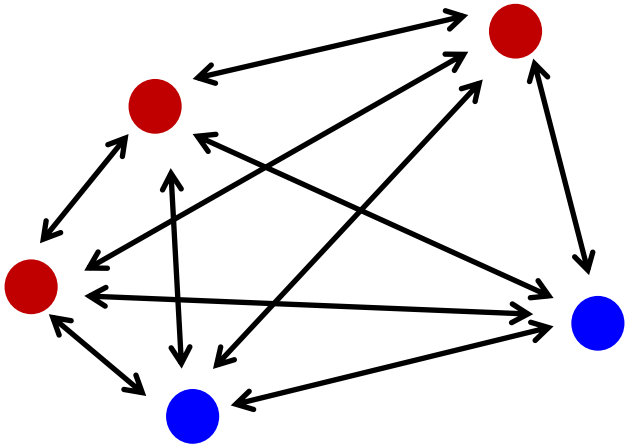
Effectively Static Network

Time-Scale of Strategy update Dynamics: τ_e

$\tau_a << \tau_e$

Links equilibrate before strategy updates occur





Spatial Games

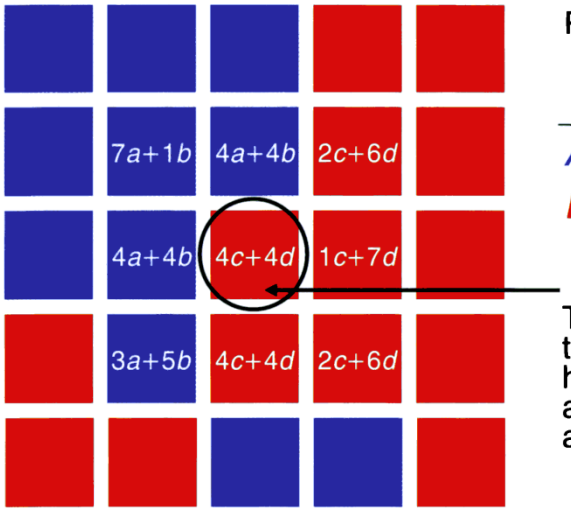
Rules for *Deterministic* Spatial Games

1. The payoff to each player is given by the total payoff obtained by playing each of its eight neighbours.

2. Rules for updating a cell are deterministic: The focal (central) cell is replaced either by itself or one of the eight neighbouring cells (Moore neighbourhood) depending on which has the highest payoff.

3. All cells are updated *simultaneously* (synchronous updating)

4. Periodic boundary condition is used to ensure all cells are treated in the same way and there are no boundary effects.



Payoff matrix:

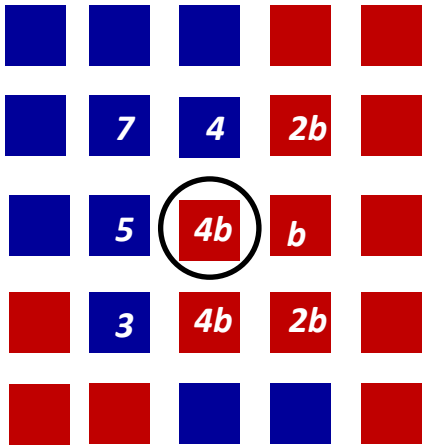
	A	B
A	a	b
B	c	d

The focal cell will be taken over by whoever has the highest payoff among the 8 neighbors and the cell itself

Cooperator Defector a=1,b=0,c=b,d=ε

b-Measure of benefit gained from *exploiting* an *altruistic* partner relative to the benefit gained from *cooperating* with an *altruistic* partner

The survival of a cell depends on its own strategy, the strategy of its eight neighbours as well as the strategies of their neighbours ➔ 25 cells in all

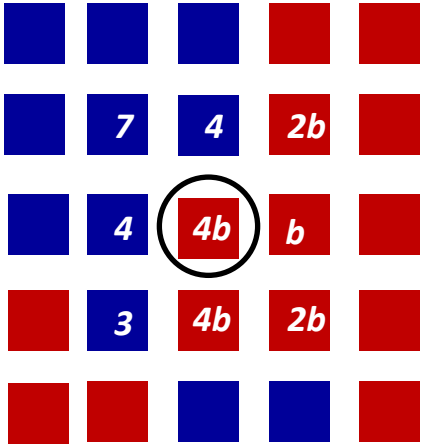


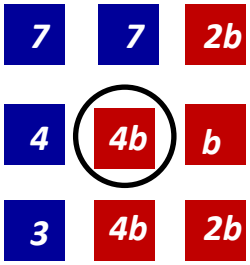
As ε➔0, the focal cell (D) has a total payoff = 4b since it is surrounded by 4 C's and 4 D's.

If **4b>7**, central cell remains a Defector in the next generation

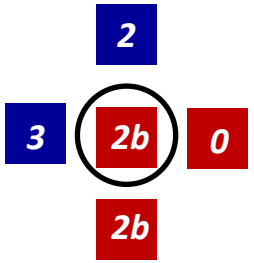
If **4b<7**, central cell transforms from Defector to Cooperator in the next generation

The Moore and Von-Neumann Neighbourhood





9-cell Moore neighbourhood



5-cell Von-Neumann neighbourhood

Rules for updating a cell: The focal (central) cell is replaced either by itself or one of the ***eight*** neighbouring cells (Moore neighbourhood) depending on which has the highest payoff.
The cells diagonally adjacent to the focal cell are ***also*** part of the neighbourhood.

Rules for updating a cell: The focal (central) cell is replaced either by itself or one of the ***four*** neighbouring cells (Von-Neumann neighbourhood) depending on which has the highest payoff.
The cells diagonally adjacent to the focal cell are ***not*** part of the neighbourhood.

Initial Condition: Half the cells are randomly chosen to be cooperators and remaining half as defectors.

Colour Code:

Blue : C that was C in the previous generation.

Green: C that was D in the previous generation.

Red: D that was D in the previous generation.

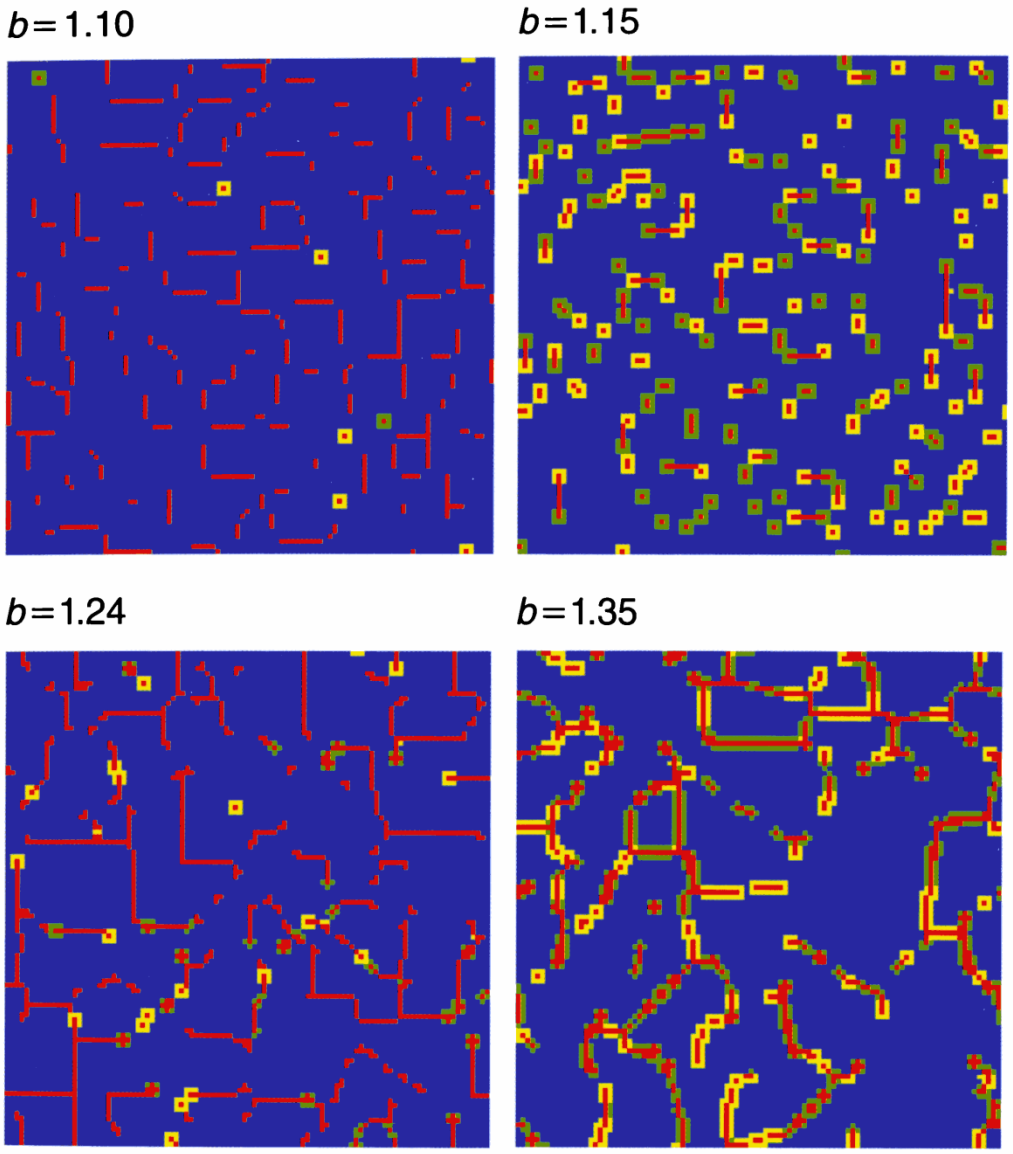
Yellow: D that was C in the previous generation.

b=1.10: Oscillation from isolated single defectors to squares of 9 defector and then back to single defector.

b=1.24: Larger lines of still mostly disconnected defectors are observed.

b=1.35: Lines of defectors now form a network with oscillating islands (yellow and green regions) around lines of defectors.

b the only parameter determining the evolution of the spatial distribution of C's and D's



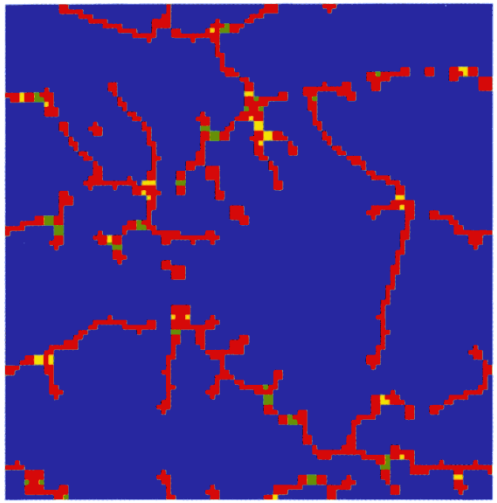
100x100 Lattice

b=1.55: Mostly static network of defectors in large islands of cooperators.

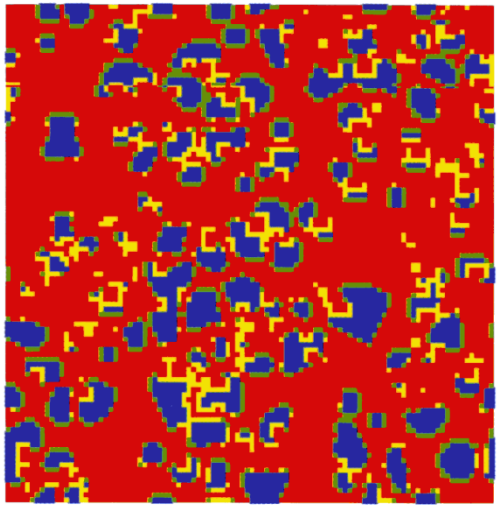
b=1.65: Defectors have attained majority by replacing most of the cooperators. Configuration with dynamic clusters of cooperators seen.

b=1.70: Static pattern showing a few clusters of cooperators in a sea of defectors. Lines of defectors now form a network with oscillating islands (yellow and green regions) around lines of defectors.

$b = 1.55$

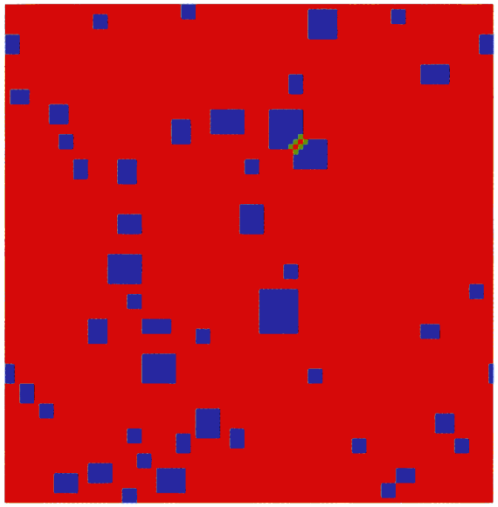


$b = 1.65$



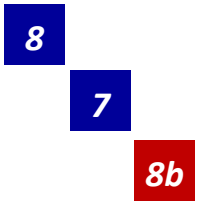
Altruism.nlogo

$b = 1.70$



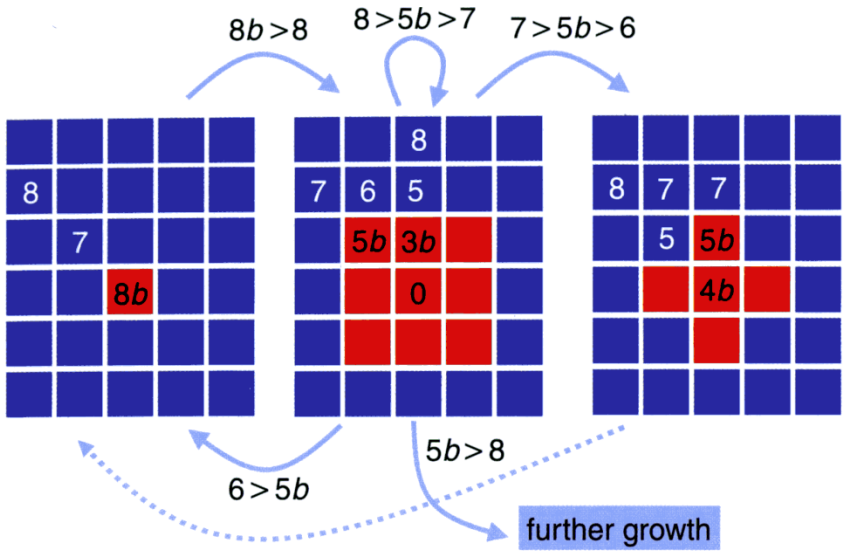
Defectors invading Cooperators

What is the likelihood that a single Defector mutant will take over a population consisting of Cooperators ?



If 7<8b<8, cell in the middle will remain C

If 8b>8, cell in the middle will become D, single D in the centre expands to give a cluster of 9 D's.



9D

5b<6: Red cell with payoff 5b is replaced by blue cell with payoff 6:

9D➔1D

7>5b>6: Red cell at the corners with payoff 5b replaced by blue cell with payoff 7:

9D➔5D

8>5b>7: 9D configuration remains unchanged

9D➔9D

5b>8: Blue cell with payoff 7 replaced by red cell with payoff 5b:

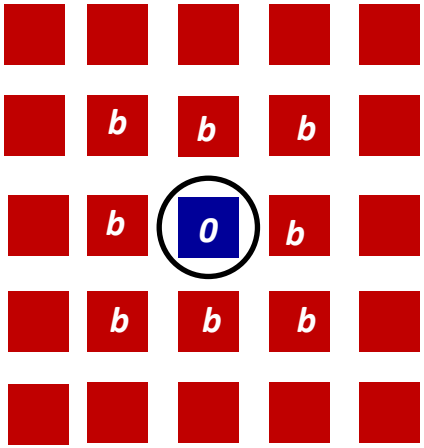
9D➔25D

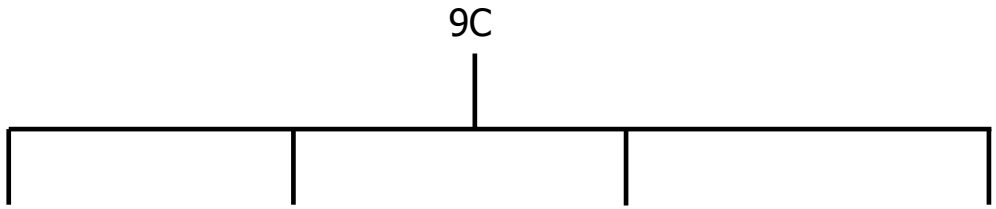
Cooperators invading Defectors

A single cooperator can never survive in a population of defectors and is immediately eliminated.

Cooperators can only survive in clusters.

Sometimes the cluster of cooperators can grow in size.





2b<3: Red cells with payoff 2b,3b are replaced by blue cells with payoff 3,5:

9C➔25C

3/2<b<5/3: Red cell with payoff 3b replaced by blue cell with payoff 5: **C** cluster expands along central axis but not along diagonals

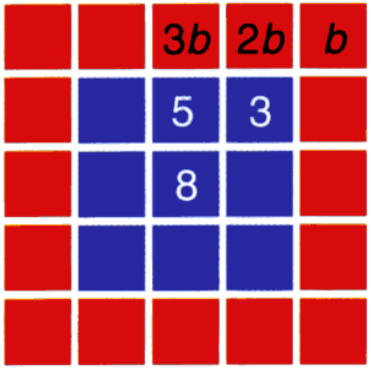
9C➔21C

5/3<b<8/3: 9C configuration remains unchanged

9C➔9C

b>8/3: 9C cluster disappears

9C➔0C



b-dependence on Spatial Evolution of Cooperators and Defectors

(i) b<8/5: Only C clusters keep growing; Cooperators dominate

(ii) b>5/3: Only D clusters keep growing; Defectors dominate

(iii) 8/5<b< 5/3: Both C and D clusters keep growing; *Co-existence* between cooperators and defectors

In (i) and (ii), final frequency of cooperators and defectors depends on the initial configuration.

In (iii) final frequency is independent of initial configuration. The pattern is dynamic but the final frequency remains nearly constant at 30% cooperators ➔ Dynamic Equilibrium

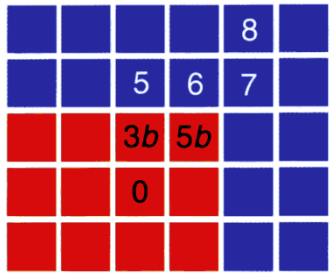
Dynamical Fractals and Evolutionary Kaleidoscopes

For **8/5<b<5/3**, configuration starting from a single defector shows repetitive (fractal-like) patterns.

A single D grows to form a 3x3 square.

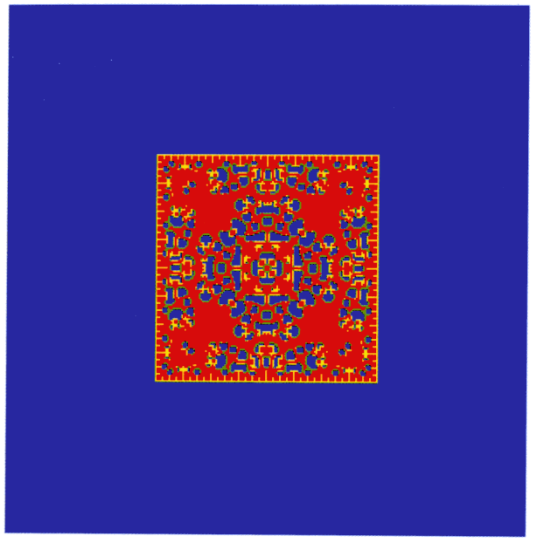
If **5b>8**, D wins at the corners and red cell with payoff 5b replaces the blue corner cell with payoff 7

If **3b<5**, C wins along lines and red cell with payoff 5b is replaced by blue cell with payoff 5

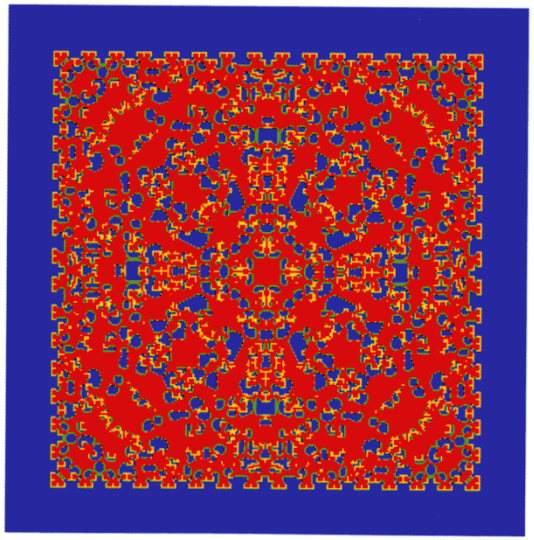


If **8/5<b<5/3 ➔ 1.6<b<1.67,** C wins along lines but loose along irregular boundaries.

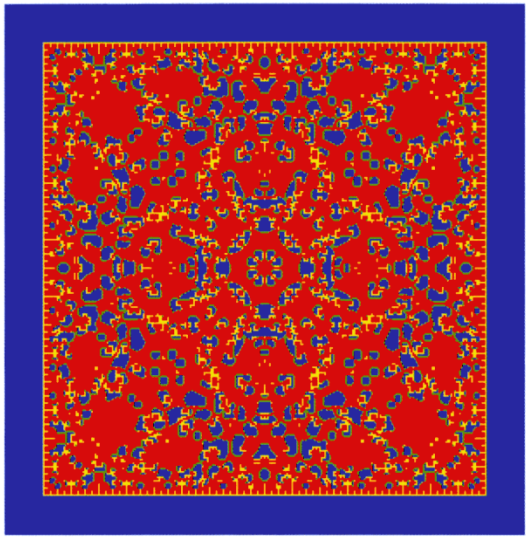
$t = 64$



$t = 124$



$t = 128$

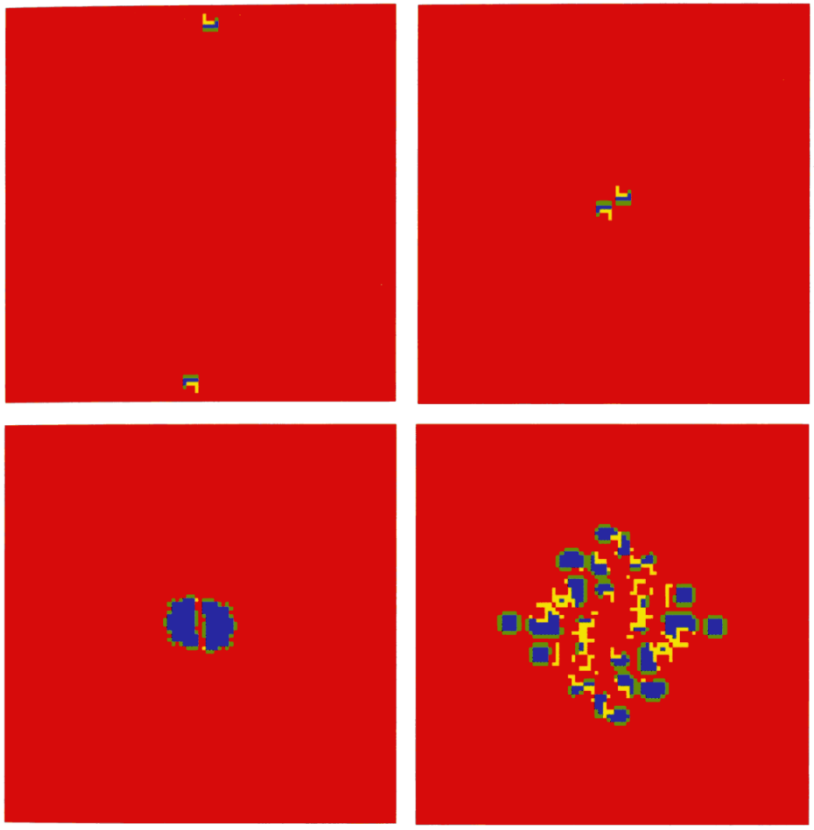


Invasion by Cooperators

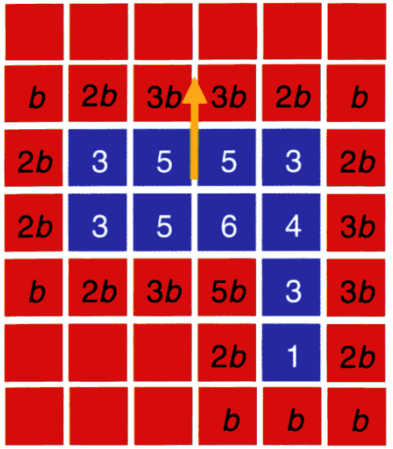
Two walkers can collide and generate two large clusters of cooperators.

A walker is a cluster of 10 cooperators moving in a direction shown by the arrow.

A collision of two "walkers" can lead to a "big bang" of cooperation



A "walker"



$8/5 < b < 5/3$

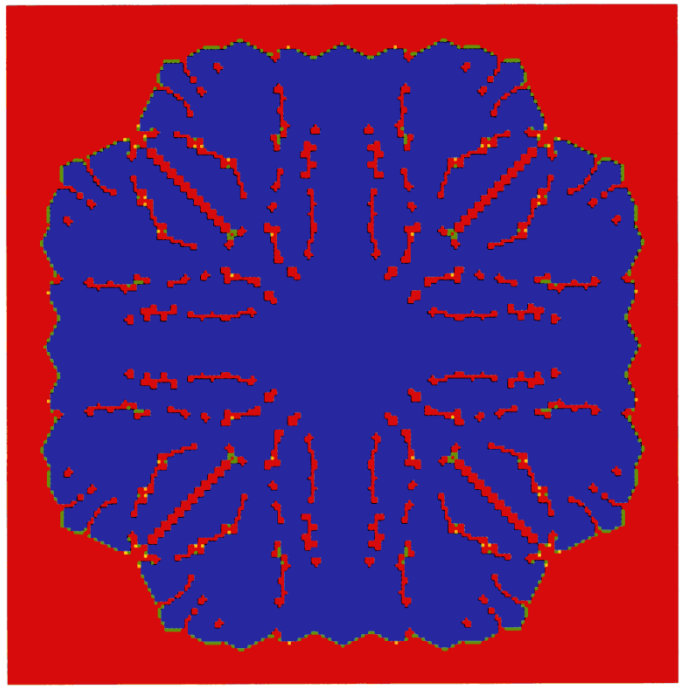
3b<5 since b<5/3

➔ 2b<5

2b>3 since b>1.6 implies **C**[payoff=1]➔**D**

Parameter region: **3/2<b<8/5**

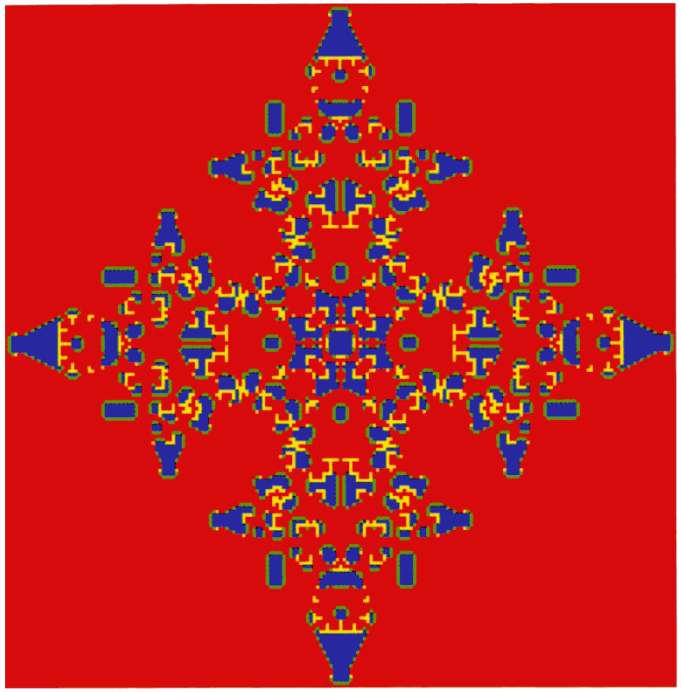
Initial Configuration: Single cluster of 3x3 C's



Increasing ***b***

Parameter region: **8/5<b<5/3**

Initial Configuration: Single cluster of 3x3 C's



Summary

A structured population can facilitate survival and sustenance of cooperators even under conditions that are unfavourable for their survival in a mixed-population scenario

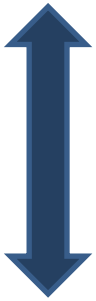
Cooperators survive by forming clusters (➔ strength in numbers) and the growth or shrinkage of the clusters depend on interactions at the boundary of the cluster

The extent of survival and spread of cooperators depend on the relative advantage that a selfish agent has over a cooperator/altruist

The greater the advantage, the lesser is the likelihood of survival of cooperators even in structured populations

Connection between ABS models and Stochastic Reaction-Diffusion Models

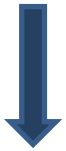
Fully Stochastic Agent-Based Simulation Models



Sometimes an *exact* *correspondence* exists

Stochastic Reaction-Diffusion (SRDE) Models

Exact correspondence



Results of an SRDE model are completely consistent with that of the ABS model for moderate to large population size (N)

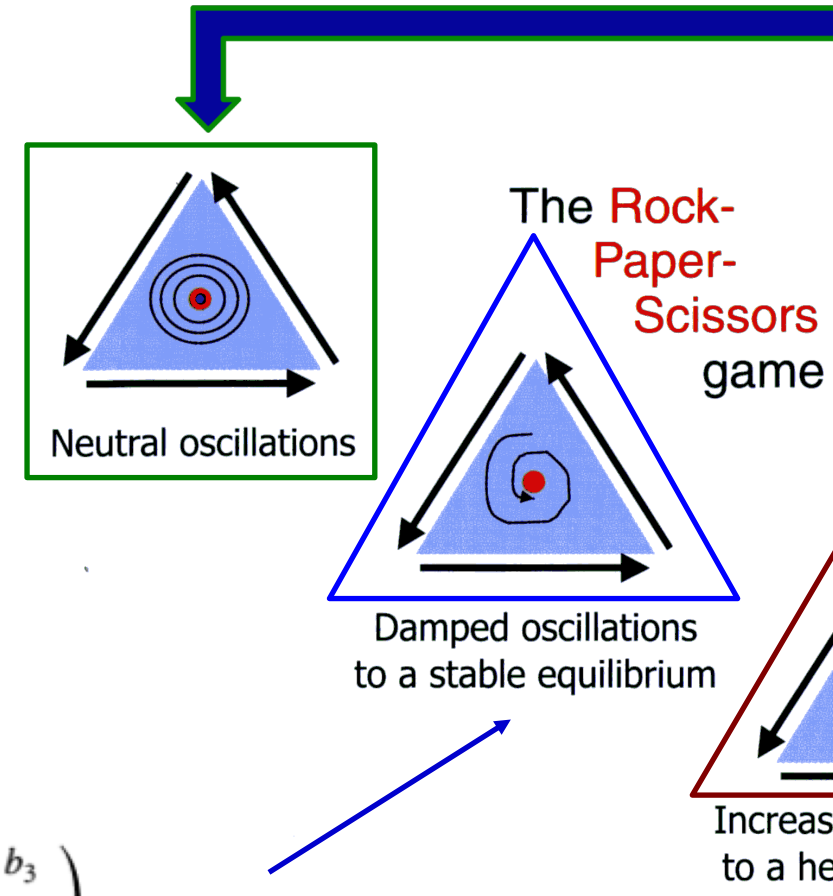
Correspondence can be established for a Rock-Paper-Scissors Game

A useful framework for understanding the conditions for multi-species coexistence in an ecosystem

Rock Paper Scissors: 3x3 Game with Cyclic Domination

(Example of an Evolutionary Game with non-transitive interactions)

$$A = \begin{array}{c} \\ R \\ S \\ P \end{array} \begin{array}{c} \begin{array}{ccc} R & S & P \end{array} \\ \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \end{array}$$



Det[A]<0 ➔ $a_1a_2a_3 > b_1b_2b_3$

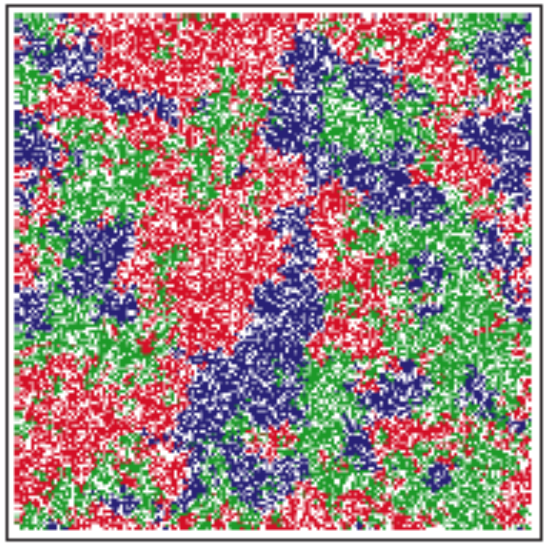
$$A = \begin{pmatrix} 0 & -a_2 & b_3 \\ b_1 & 0 & -a_3 \\ -a_1 & b_2 & 0 \end{pmatrix}$$

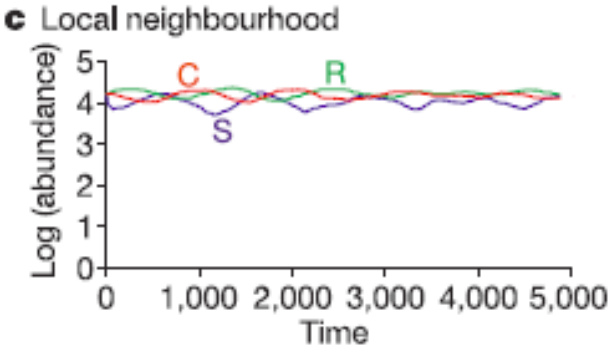
Det[A]>0 ➔ $a_1a_2a_3 < b_1b_2b_3$

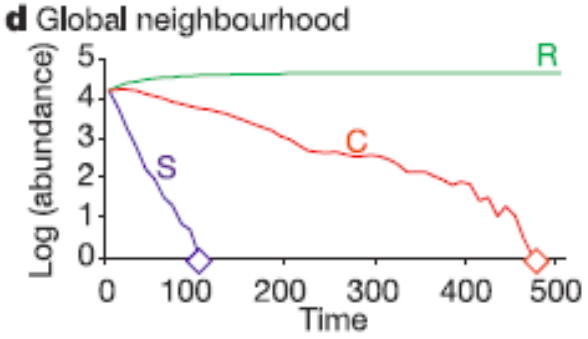
Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors

Benjamin Kerr*, Margaret A. Riley†, Marcus W. Feldman* & Brendan J. M. Bohannan*

A bactericidal model







Local vs **Global** Update Dynamics

C - colicin producing *E.coli*

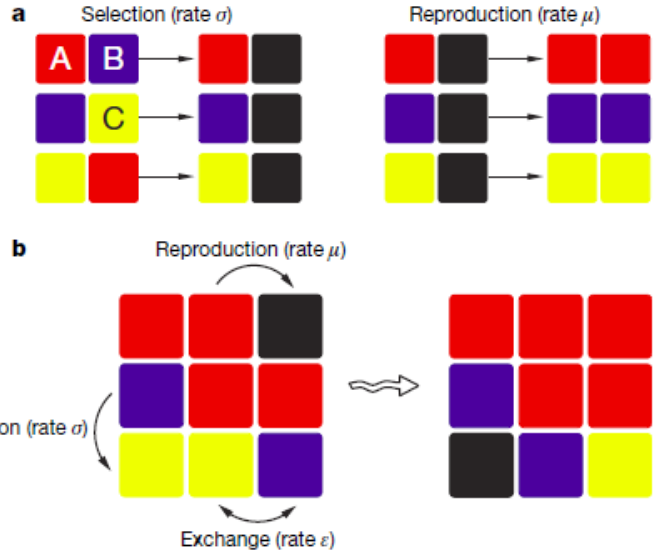
S - colicin sensitive *E.coli*

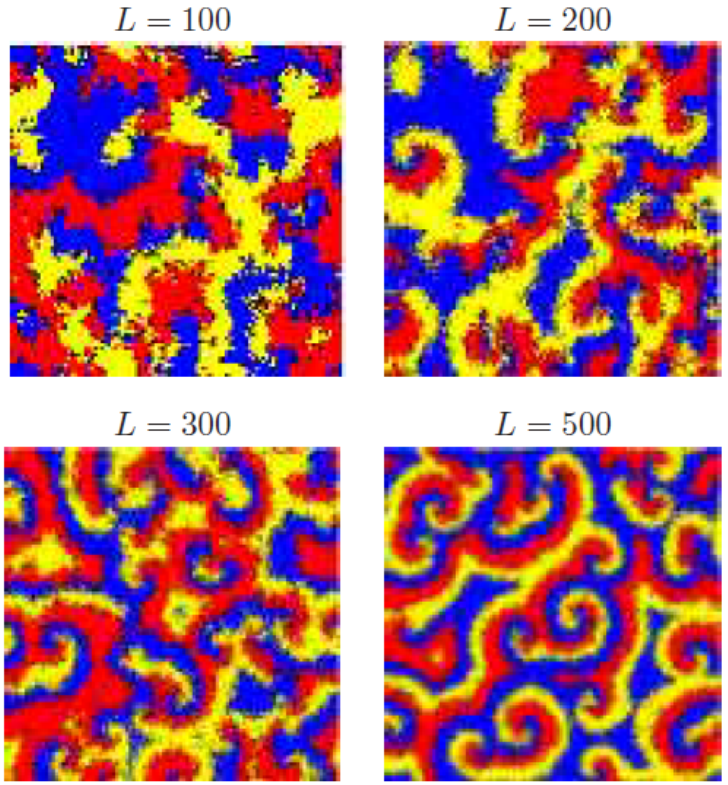
R - colicin resistant *E. coli*

1. Focal lattice site (F) chosen at random
2. If F is empty, it is filled with a cell of type i=C or S or R from its ***local* neighbourhood** (8 nearest neighbours surrounding F) or ***global* neighbourhood** (anywhere in the lattice except the focal site) with a probability proportional to fraction (f_i) of cell-type (i) present in the neighbourhood.
3. If focal lattice site F is occupied by cell-type i, it is killed with probability Δ_i
4. $\Delta_C = 0.33 \quad \Delta_R = 0.3125 \quad \Delta_S = \Delta_{S0} + \tau f_C ; \Delta_{S0} = 0.25$

$\tau =$ Measure of toxicity of colicin

Rock Paper Scissors on a Spatial 2D Lattice with Diffusion





Update Rules:

Choose one of 3 reactions with the following probabilities

Reproduction

$$\frac{\mu}{\mu+\sigma+\varepsilon}$$

Selection

$$\frac{\sigma}{\mu+\sigma+\varepsilon}$$

Exchange

$$\frac{\varepsilon}{\mu+\sigma+\varepsilon}$$

May-Leonard Model

$$AB \xrightarrow{\sigma} A\oslash, \quad BC \xrightarrow{\sigma} B\oslash, \quad CA \xrightarrow{\sigma} C\oslash,$$

Selection

$$A\oslash \xrightarrow{\mu} AA, \quad B\oslash \xrightarrow{\mu} BB, \quad C\oslash \xrightarrow{\mu} CC$$

Reproduction

Effect of Lattice Size on Pattern Formation
Shows important role of noise in perturbing the patterns for small values of L

Alternative Model

$$AB \xrightarrow{p} A\emptyset, \quad BC \xrightarrow{p} B\emptyset, \quad CA \xrightarrow{p} C\emptyset,$$
$$AB \xrightarrow{z} AA, \quad BC \xrightarrow{z} BB, \quad CA \xrightarrow{z} CC,$$
$$A\emptyset \xrightarrow{q} AA, \quad B\emptyset \xrightarrow{q} BB, \quad C\emptyset \xrightarrow{q} CC,$$

$$A \xrightarrow{\mu} \begin{cases} B \\ C \end{cases}, \quad B \xrightarrow{\mu} \begin{cases} A \\ C \end{cases}, \quad C \xrightarrow{\mu} \begin{cases} A \\ B \end{cases}$$

Relation between Deterministic and Stochastic Spatial Models of Rock-Paper-Scissors Game

Deterministic

Noise incorporated through a Gaussian noise term in the evolution equations

Partially Stochastic represented by Stochastic PDE's

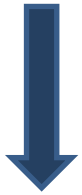
Fully Stochastic Agent-Based Simulations (ABS)

Relation between the Master Equation and SPDE's

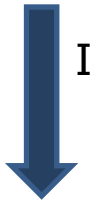
Master Equation

Kramers-Moyal expansion of Master Equation to second order in (δs) i.e. O(1/N)

Fokker-Planck (FP) Equation



Stochastic Differential Equations SDE's



Incorporate diffusion through exchange process to leading order in (δs≡1/N) i.e. $(1/N)^0$

Partially Stochastic Model represented by Stochastic PDE's

➔ Noise due to number fluctuations incorporated to $O(\frac{1}{\sqrt{N}})$

Source of Noise in SPDE's

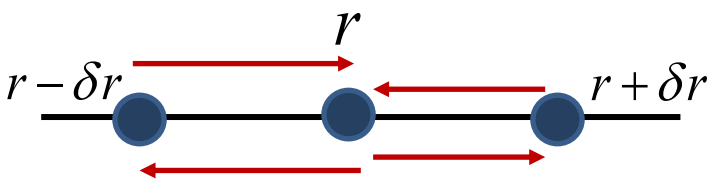
❖ Noise due to number fluctuations arising from birth-death processes

Appears in SDE at $\mathrm{O}(\frac{1}{\sqrt{N}})$

❖ Noise due to number fluctuations arising from exchange processes

The Diffusion term to leading order i.e. in the absence of fluctuations i.e. $O\left(1/N\right)^0$

$$T_1 = \frac{\varepsilon}{d}\sum_{i=1}^{d}\Big[\{a(r+\delta r)-a(r)\}+\{a(r-\delta r)-a(r)\}\Big]$$

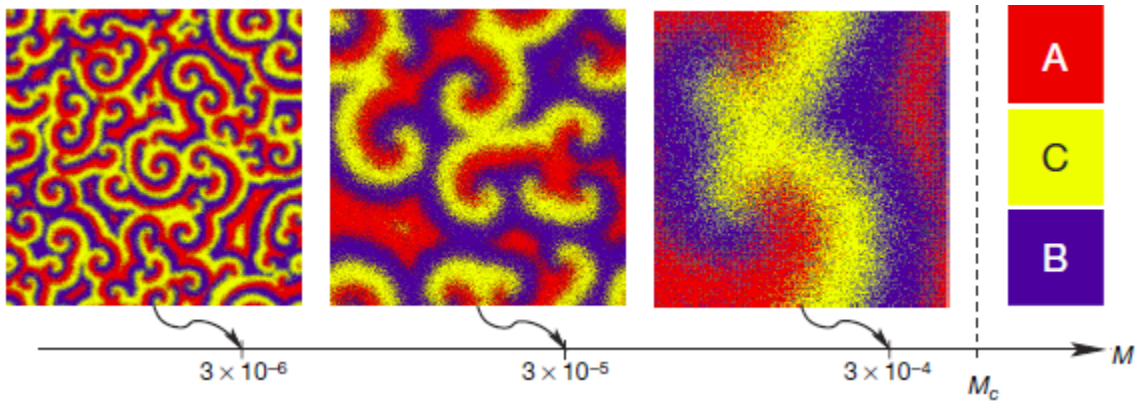


Noise due to number fluctuations arising from exchange processes comes from the term

$$B(r,r') = \frac{D}{N^2}\partial_r\partial_{r'}\left[\delta(r-r')a(r)(1-a(r))\right]$$

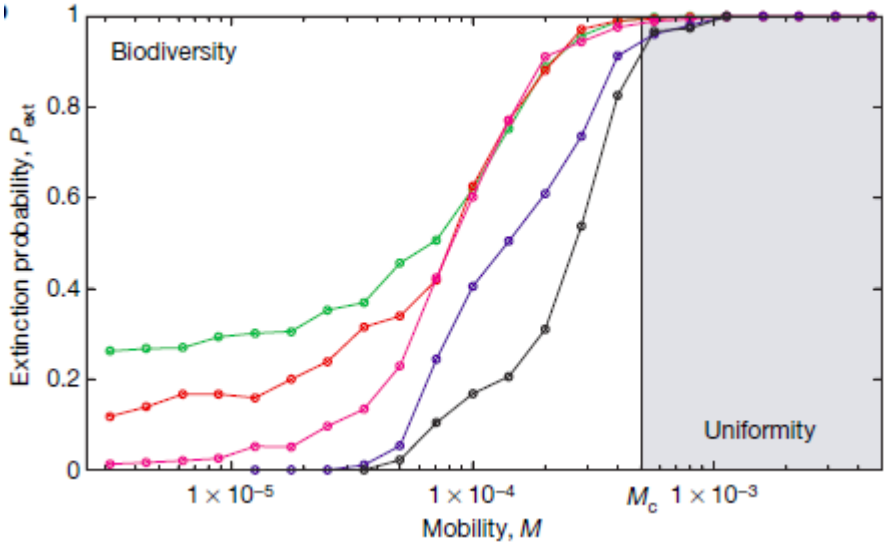
Since noise is proportional to $\sqrt{B(r,r')}$ it appears at $\mathrm{O}(\frac{1}{N})$ and can be ignored

Effect of Increasing Diffusion Coefficient: Loss of Diversity



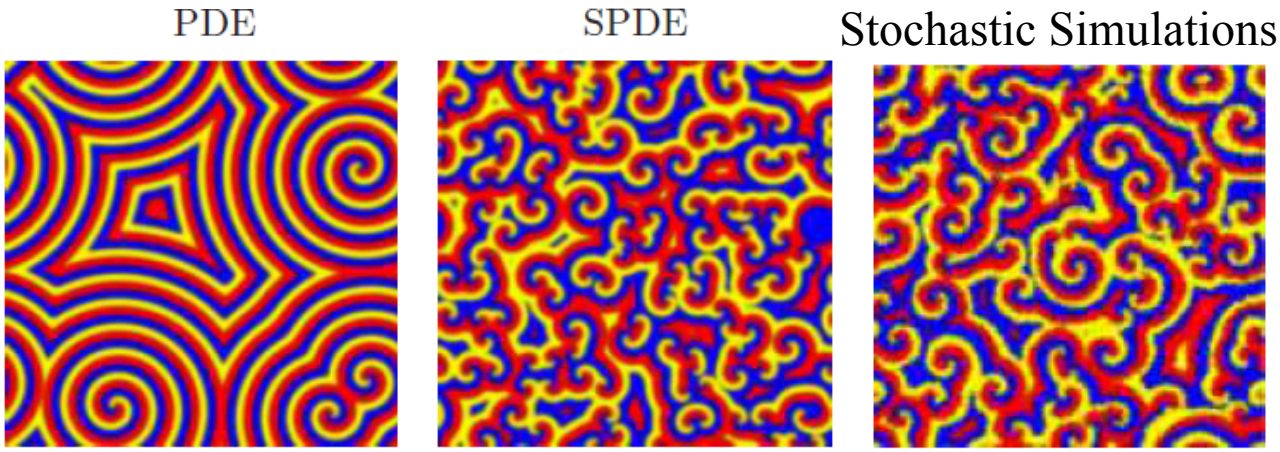
$M = 2 \in N^{-1} = 4D$: Area covered by a random walker in unit time in 2-dim

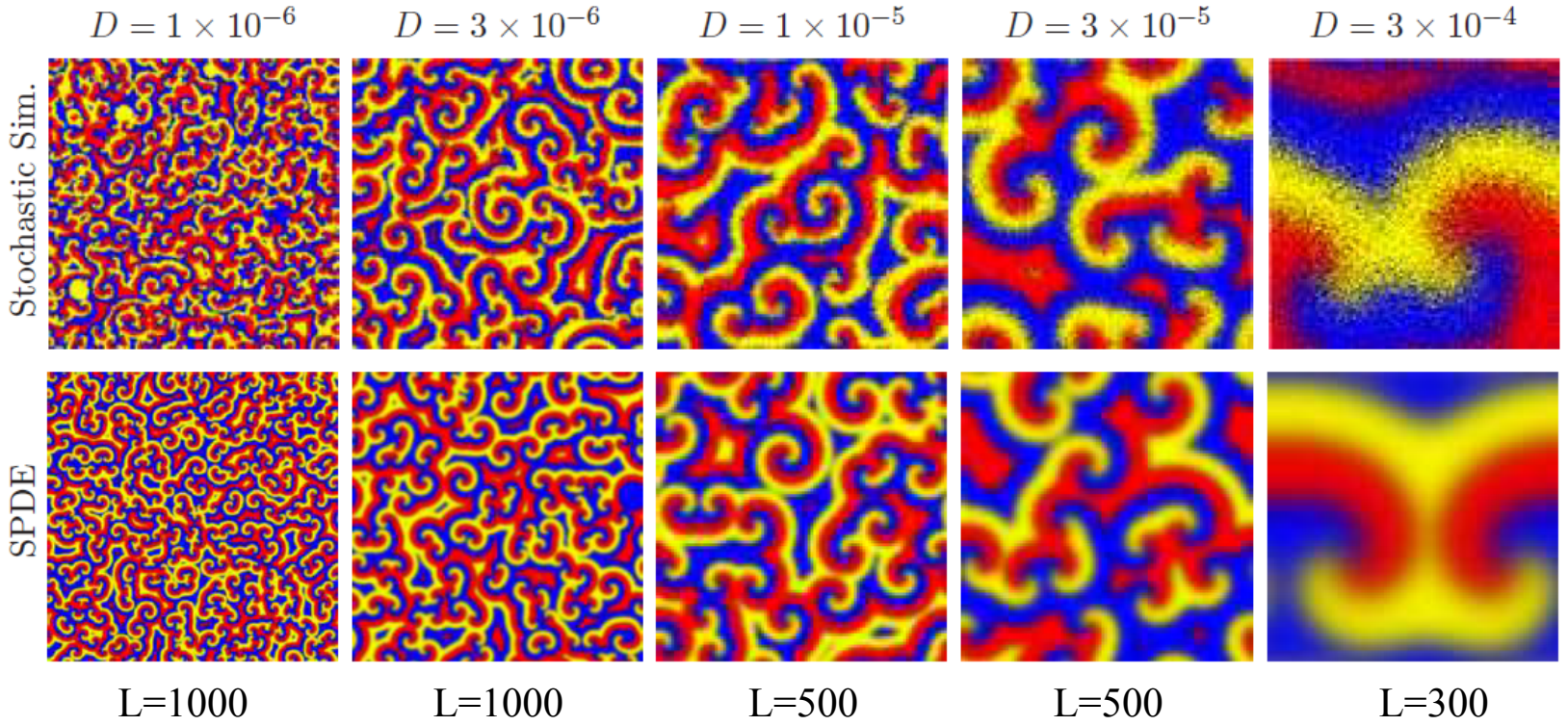
D : Diffusion Coefficient



Reichenbach, Mobilia, Frey; Nature 2007, Journal of Theoretical Biology 2008

Pattern Formation *without* and *with* Noise

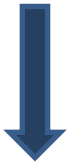




Ways in which Cooperation can evolve

1. Repeated interactions leading to evolution of cooperation due to behavioral book-keeping
2. Direct Reciprocity as manifest in the success of TFT-like strategies
3. Indirect Reciprocity: Reputation (as a cooperator or defector) determines the likelihood of being helped by others
4. Structured populations can facilitate survival and spread of altruistic behaviour

Altruists interacting more often with each-other than with non-altruists can acquire larger payoffs



Positive assortment

Can be facilitated by Kin-recognition

Evolution of Cooperation: The effect of interactions with kin

McElreath & Boyd (Chapter 3)

A- Altruists (Cooperators), N – Non-altruists (Defectors)

Pr(A|A) – probability that an altruist interacts with another altruist

Pr(A|N) – probability that a non-altruist interacts with an altruist

	A	N
A	b-c	-c
N	b	0

b – benefit to the recipient

c – cost to the donor

Altruists will increase in frequency if

$$\{Pr(A|A) - Pr(A|N)\}b > c$$

Altruism can evolve when **A**'s are more likely (than random) to interact with other **A**'s than with **N**'s

If Pr(A|A)=Pr(A|N), altruism *cannot* evolve regardless of the benefit (**b**) it confers since the cost **c>0**

Kin Recognition: One mechanism by which animals recognize and preferentially interact with kin

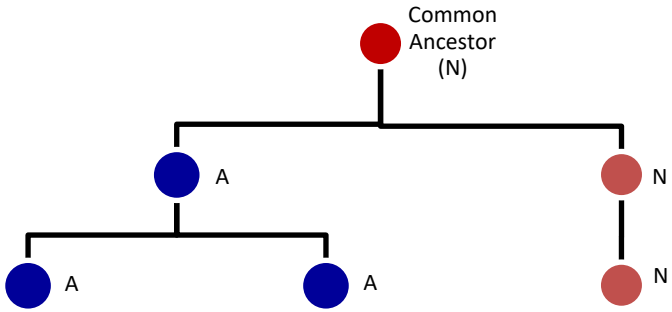
➔ A is more likely to interact with A and N is more likely to interact with N

Kin recognition requires that the relation between two individuals in the population must be incorporated while calculating the probability of interaction between them.

Inclusive Fitness

Altruistic behaviour (phenotype) encoded in genes (genotype).

Kins can share the altruism gene if it is passed onto each individual by a common ancestor in which the altruistic gene arose.



An allele coding for altruism can selectively help other copies of itself found in kin if altruistic behaviour is preferentially directed towards kin

Helps in explaining the evolution of social insect societies where the workers are infertile and hence has zero reproductive fitness.

Requires redefinition of fitness.

Inclusive Fitness: $f_i = F_i + \sum_{\rho} F_\rho r_\rho$

- F_i: Individual Fitness
- r_ρ: **Coefficient of relatedness** between i & ρ
- F_ρ: Fitness of ρ'th relative

The inclusive fitness can be ***non-zero*** even when the individual (reproductive) fitness is zero

Hamilton's Rule

Assume individuals interact in a Prisoner's Dilemma Game

Probability that two individuals possess the same (A or N) allele depends not just on the frequency (p) of allele in the population but also on whether the two individuals are ***related***.

Probability that the two individuals are related by common descent = ***r*** – **coefficient of relatedness**

$$Pr(A|A) = r(1) + (1-r)p$$

If the two individuals are related, both possess the altruism allele with probability **1**

If they are not related, both can still possess the altruism allele with probability **p**

$$Pr(N|A) = r(0) + (1-r)(1-p)$$

$$Pr(A|N) = r(0) + (1-r)p$$

$$Pr(N|N) = r(1) + (1-r)(1-p)$$

Haploid kin-selection model

Altruism phenotype specified by a *single* allele

$$Pr(A|A) + Pr(N|A) = 1$$

$$Pr(N|N) + Pr(A|N) = 1$$

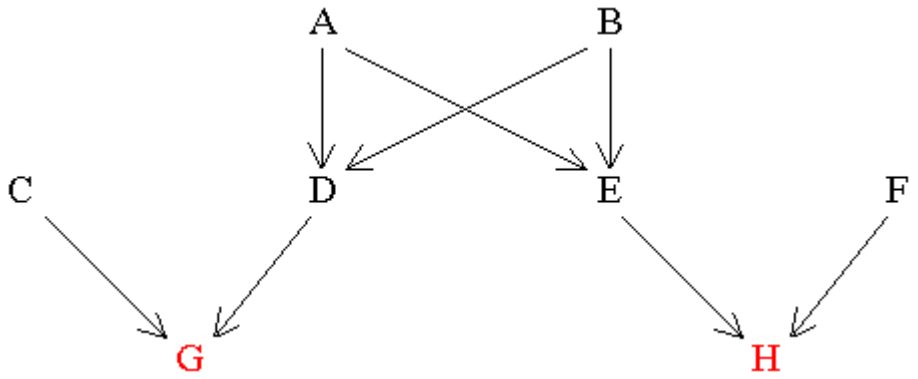
When **r=0**, Pr(A|A)=p; Pr(N|A)=1-p; Pr(N|N)=1-p; Pr(A|N)=p ➔ Random interaction

Whether A interacts with another A depends only on the frequency (p) of A

$\{Pr(A|A) - Pr(A|N)\}b > c$ ⟹ **$rb > c$**

Hamilton's Rule for altruism to evolve

Genealogy and "Coefficient of Relationship" between Kin



D and E are siblings : What is r_DE ?

Count number of steps required in going from D to E via the common ancestor

Multiply each step by (1/2) and sum over all possible independent paths between D & E to get r_DE

Two independent paths : D➔A➔E or D➔B➔E, for each path two steps are reqd. to go from D to E

r_DE = (1/2)^2 + (1/2)^2 = 1/2

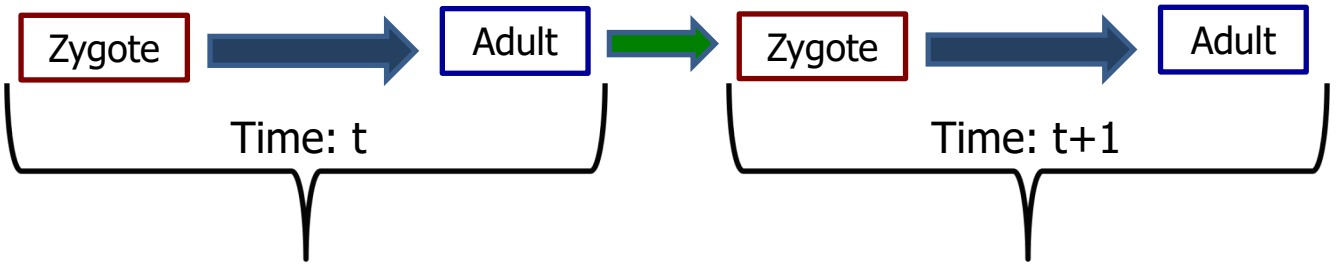
G and H are cousins : What is r_GH ?

Two independent paths: G➔D➔A➔E➔H or G➔D➔B➔E➔H; for each path 4 steps are required

r_GH = (1/2)^4 + (1/2)^4 =1/8

Valid for **diploid** organisms where each of the 2 alleles at a locus has an *equal* probability of being inherited. This explains multiplication by (1/2) for every step.

Heuristic derivation of Price Equation



A: possess the altruisim allele; B: does not possess the altruism allele (cheater)

Probability that a zygote A survives to adulthood α to the mean fitness of A (F_A) in the population

A-adults that survive at time t: NpF_A

Frequency of A-adults at time t: $$p' = \frac{F_A Np}{F_A Np + F_B N(1-p)} \equiv \frac{pF_A}{\bar{w}}$$

Assuming asexual reproduction and each adult produces z-zygotes at time t+1

Frequency of A-zygotes at time t+1: $$\frac{(\#A\_adults)z}{(\#A\_adults)z + (\#B\_adults)z} \equiv p' = \frac{pF_A}{pF_A + (1-p)F_B}$$

$$\Delta p \equiv p' - p = \frac{pF_A - p\bar{w}}{\bar{w}} \quad \Rightarrow \quad \bar{w}\Delta p = pF_A - p\bar{w} = p(1-p)(F_A - F_B)$$

$$F_A = \frac{\sum_i w_i p_i}{\sum_i p_i}$$

$\sum_i w_i p_i$ → Total fitness of Altruism alleles in the population

$\sum_i p_i$ → Total frequency of Altruism alleles in the population

Price Equation and Rediscovery of Hamilton's Rule

Ref.: Ch.3 of "Mathematical Models of Social Evolution"

$$\overline{w}\Delta p = E(p_i w_i) - E(p_i)E(w_i) \equiv \mathrm{cov}(w_i, p_i)$$ **Price Equation**

Assumption 1: Assumption of additive fitness

$$w_i = f_0 + y_i b - h_i c$$

Linear approximation for fitness is valid when selection is weak i.e. when benefits accrued from or costs associated with altruism are small

$$b\frac{\mathrm{cov}(y_i, p_i)}{\mathrm{cov}(h_i, p_i)} > c$$

y_i – Probability that the i'th member receives aid (Recipient)

h_i – Probability that the i'th member provides aid (Donor)

$$\mathrm{cov}(h_i, p_i) = k > 0$$

The likelihood ($\mathbf{h_i}$) of being a Donor of an altruistic act has to be positively correlated with the frequency ($\mathbf{p_i}$) of the altruistic gene in the member

Assumption 2: Assumption of Linear mapping between genotype and phenotype

$$y_i = a + kp_j$$

y_i – Phenotype which determines whether the i'th member receives aid

$$h_i = a + kp_i$$

h_i – Phenotype which determines whether the i'th member provides aid

$$b\frac{\mathrm{cov}(p_j, p_i)}{\mathrm{var}(p_i)} \equiv b\beta(p_j, p_i) > c$$

Slope of the regression line of p_j on p_i (i.e. Regression coefficient) determines the spread of altruism

To what extent does the donor's genotype predict the recipient's genotype ?

Altruism can spread if the donor can distinguish between Altruists and Non-altruists and is more likely to help Altruists than Non-altruists

Assumption 3 : Assumption of ***no*** selection for the altruism allele

Key Question: To what extent does interaction between kin aid in spread of altruism ?

$$E(p_j \mid p_i) = E(p_j) + \beta(p_j \mid p_i)\{p_i - p\}$$

Expectation that the altruism allele is present in the j'th member (Donor) ***given*** that it is present in the i'th member (Recipient)

r – fraction of genes identical by descent (IBD) in both members i and j.

$p_{j,IBD}$ – fraction of genes IBD that are altruism genes

$p_{j,\overline{IBD}}$ – fraction of genes that are ***not*** IBD (i.e. not related) but are altruism genes =p

Altruism allele will be present in both i and j if

(i) i and j are related by common descent from a single common ancestor

OR

(ii) i and j are ***not*** related but possess the altruism allele by chance because of the frequency of the altruism allele in the population.

$$E(p_j \mid p_i) = rp_{j,IBD} + (1-r)p_{j,\overline{IBD}}$$

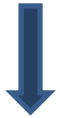
If p_i =0 ➔ $p_{j,IBD}$ =0;

If p_i =1 ➔ $p_{j,IBD}$ =1; for HAPLOID models

$$r = \frac{E(p_j \mid p_i) - p}{p_i - p} \equiv \beta(p_j, p_i) \Rightarrow br > c$$

Hamilton's Rule

Misinterpretations of Hamilton's Rule



Misunderstanding the nature of ***r***

1. Washburn's Fallacy: Interpretation of r as proportion of common genes across the entire genome

Vast majority of human genes are same for all individuals ➔ Washburn's argument : All humans should be universal altruists and only slightly more altruistic with close kin

But humans and chimps also share many common genes more so than humans and dogs

Does that imply humans should be more altruistic towards chimps than towards dogs ?

For evolution of altruism, all gene loci are not relevant, only those where the altruism gene(s) reside

r- Probability that two individuals possess the same allele since they descended from a single *common ancestor*

2. Charlesworth's Paradox : Imagine a particular species where one of the offspring stays back to help save and rear 4 of her later born siblings at the cost of her reproductive success ➔ b=4, c=1, r=0.5

➔ $rb > c$ ➔ such behaviour should evolve, yet it is not observed.

3. Mutations not taken into account when considering relatedness of (altruism) gene loci that are identical by common descent. Individuals related through common descent from an ancestor in which the altruism allele arose may still not be identical as a result of mutations in one of the offspring that changes the altruism phenotype. Total # individuals that are non-identical by common descent = $4N\mu+1$ (from pop gen theory)

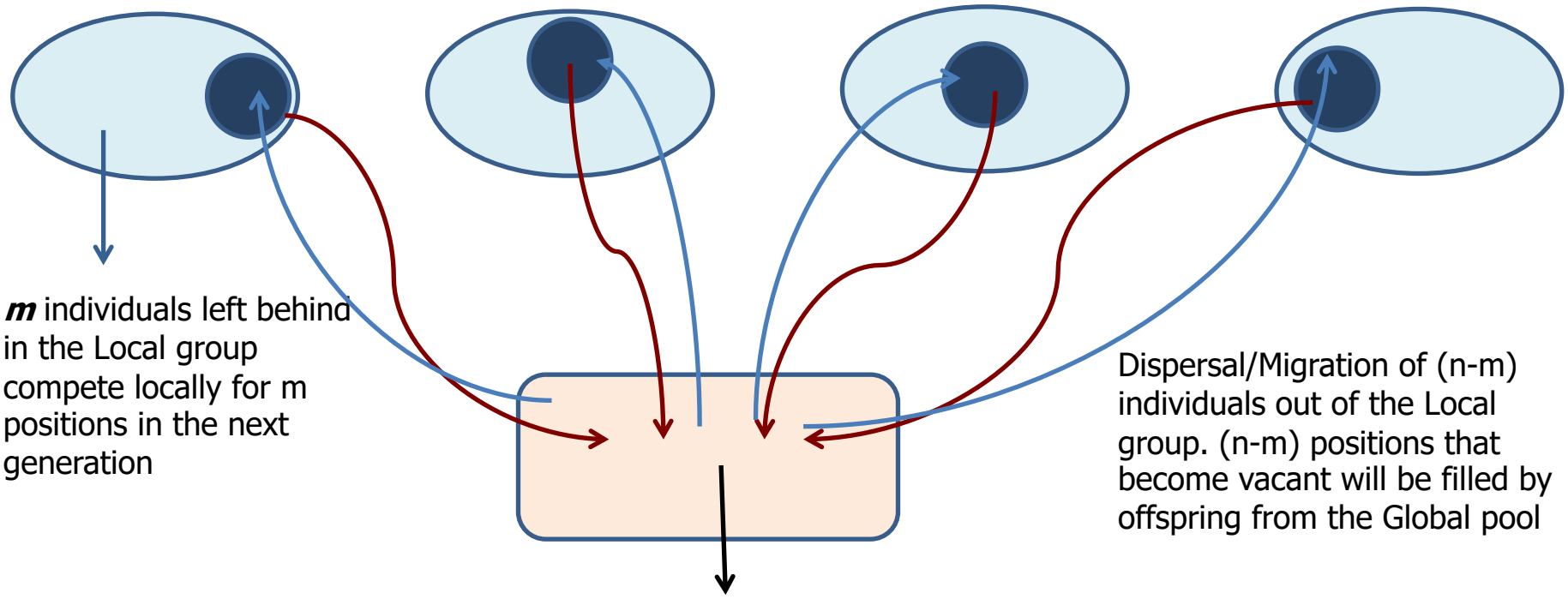
Not significant (for realistic population sizes (N) and mutation rates (μ)) unless N or μ or both are very large

Ways in which Cooperation can evolve

1. Repeated interactions leading to evolution of cooperation due to behavioral book-keeping
2. Direct Reciprocity as manifest in the success of TFT-like strategies
3. Indirect Reciprocity: Reputation (as a cooperator or defector) determines the likelihood of being helped by others
4. Structured populations can facilitate survival and spread of altruistic behaviour
5. Positive assortment (facilitated by kin-recognition) leading to altruists interacting more often with each-other than with non-altruists
6. Punishment
7. *Limited Dispersal of the offspring from its birthplace*

Can limited dispersal of offspring which increases the probability of interactions with kin lead to spread of altruism?

Dispersal: A pictorial representation



m individuals left behind in the Local group compete locally for m positions in the next generation

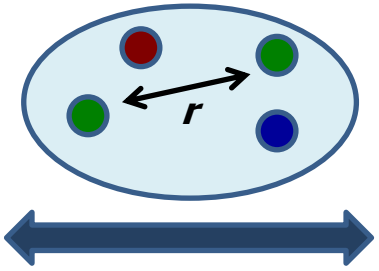
Dispersal/Migration of (n-m) individuals out of the Local group. (n-m) positions that become vacant will be filled by offspring from the Global pool

Contribution of each local group to the Global Pool

l=m/n – Extent to which population regulation is local

m=n ➔ l=1 ➔ NO dispersal; Competition to be selected as a member in the next generation occurs within the group i.e. population regulation is LOCAL

m=0 ➔ l=0 ➔ Rapid dispersal leading to a well-mixed population. Population regulation occurs on a GLOBAL scale



Average relatedness in a local group = $\overline{r}$

Modification of Hamilton's Rule when population regulation is *Local*

When interaction and selection occurs within a group, Altruists compete with other Altruists to get selected to the next generation

Average fitness of the global population is less relevant than the average fitness of the local population in determining change in frequency of the Altruisim allele

$$\Delta p_i = p_i(1-p_i)\frac{w_i(A)-w_i(N)}{\bar{w}_i}$$

$\bar{w}_i$ → Average fitness of the i'th group

Group index ← p_i

$$\Delta p = E(\Delta p_i)$$

Average increase or decrease across ***all*** groups gives the change in frequency of the altruism allele across the *entire* population

Extreme Case: Group consists of a pair of interacting individuals

Only pair-wise interaction that leads to change in Δp is the interaction between A and N

$$\Delta p = \Pr(A \mid A)(0) + \Pr(A \mid N)(\frac{1}{2})(1-\frac{1}{2})\frac{(-b-c)}{(b-c)/2} + \Pr(N \mid N)(0) = -\Pr(A \mid N)(\frac{1}{2})\frac{(b+c)}{(b-c)}$$

$$\beta(p_j, p_i) = \frac{E(p_j \mid p_i) - E(p_j)}{p_i - E(p_i)};$$

$$E(p_j) = (1-l)p + lE(p_k \mid p_j)$$

$$E(p_i) = (1-l)p + lE(p_k \mid p_i)$$

$$E(p_k, p_i) = \bar{r}p_i + (1-\bar{r})p$$

Average relatedness between two individuals within i's group

Probability that an average individual donor (k) in i's local group possess the A-allele given the recipient (i) also possess the A-allele with frequency p_i

$$\beta(p_j, p_i) = \frac{E(p_j \mid p_i) - \{(1-l)p + l(1-\bar{r})p + l\bar{r}p_i\}}{p_i - \{(1-l)p + l(1-\bar{r})p + l\bar{r}p_i\}}$$

For the *specific* individual "***j***" within ***i***'s local group whose *coefficient of relatedness* with ***i*** is **r**

$$E(p_j \mid p_i) = rp_i + (1-r)p$$

If ***i*** and ***j*** are related, both possess the altruism allele with probability proportional to p_i

If they are not related, both can still possess the altruism allele with probability **p**

$$\beta(p_j, p_i) = \frac{r - l\bar{r}}{1 - l\bar{r}}$$

$$b\beta(p_j, p_i) > c \quad \Rightarrow \quad b\left(\frac{r - l\bar{r}}{1 - l\bar{r}}\right) > c$$

Generalized form of Hamilton's rule

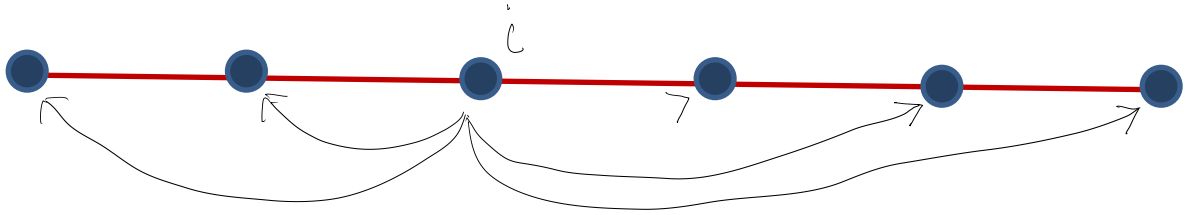
When population regulation is *Global* i.e. $l=0$, the *original* form of Hamilton's rule is recovered : $br > c$

When population regulation is *entirely Local* i.e. $l=1$: Spread of altruism becomes less likely as $r \rightarrow \bar{r}$

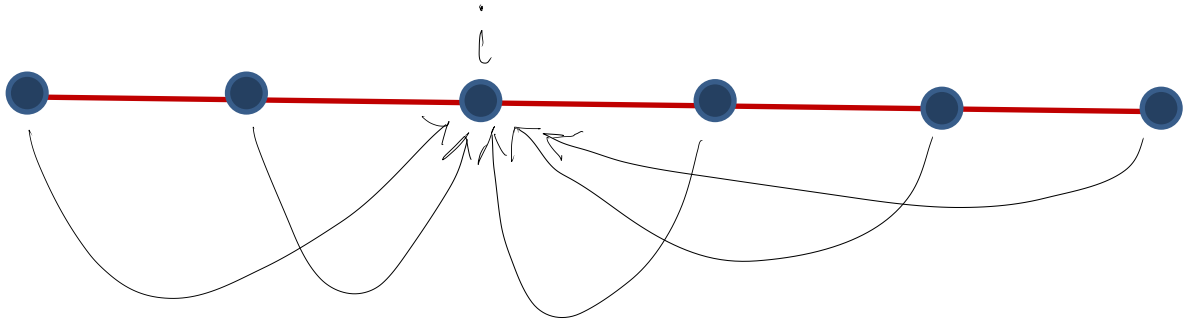
For case of competition between clones i.e. $r = \bar{r} = 1$; spread of altruism possible if $b > c$ for all /

A model of dispersal in asexual organisms

Dispersal out of a site



Migration into a site



There are ***n*** sites with each site being occupied by an Asexual adult which can produce multiple offsprings

At each site, only ***one*** offspring can survive into adulthood

All individuals produce **k** offspirngs, all of which are biological clones ➔ no variation within a group of siblings

v – fraction of k offsprings which emigrate (disperse) from its birth-site ➔ # emigrants = kv

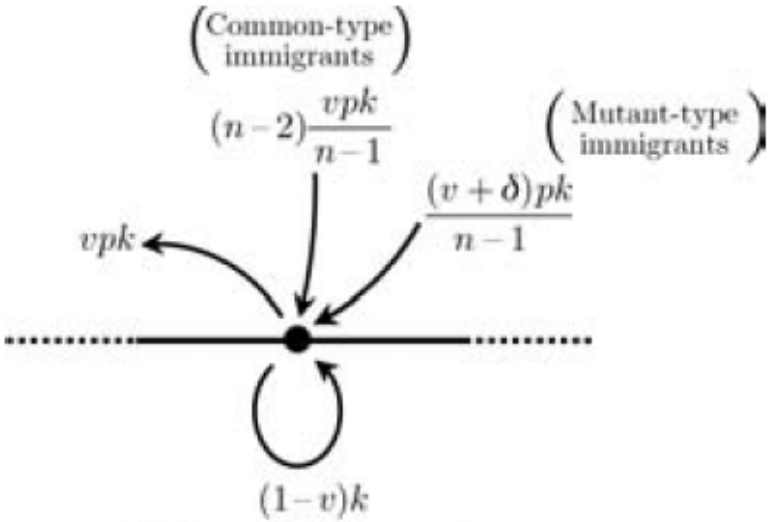
p – fraction of emigrants that survive ➔ # emigrants that survive migration to some other site = pkv

Emigrants from a given site that survive migration are equally likely to reach and compete for any of the remaining ***(n-1)*** sites.

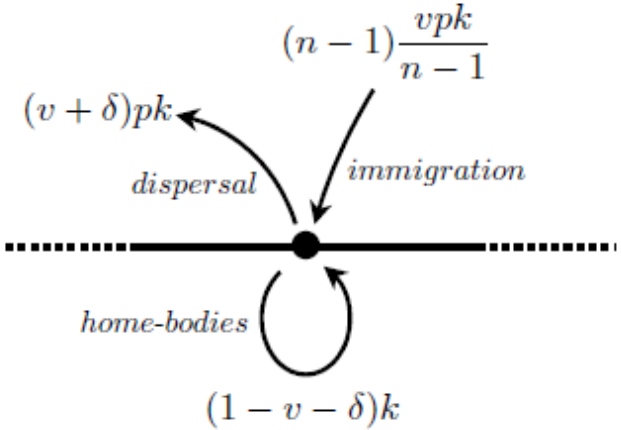
(1-v)k - # offsprings that remain at their birth site

These offsprings compete with each other as well as immigrants from other sites to occupy their birth-site in the next generation

A model of dispersal in asexual organisms



Dispersal, Immigration and competition at a site originally occupied by a WT individual



Dispersal, Immigration and competition at a site originally occupied by a mutant

What is the condition under (which the mutant which initially appears at only one site) invades the population ?

Condition for invasion of mutants

Pr(M|WT): Probability that a mutant takes over a site, originally occupied by a WT, in the next generation

$$\Pr(M \mid WT) = \frac{\{\frac{(v+\delta)pk}{n-1}\}}{\{(1-v)k + \frac{(n-2)vpk}{n-1} + \frac{(v+\delta)pk}{n-1}\}}$$

#mutants reaching that site

Total #(M+WT) competing for that site

Expected number of successful mutant offspring across all (n-1) WT occupied sites

$$E(\#M \mid WT) = (n-1)\Pr(M \mid WT) = \frac{(v+\delta)p}{1-v+vp+\frac{\delta p}{n-1}}$$

Pr(M|M): Probability that a mutant replaces it parent at its home site

$$\Pr(M \mid M) = \frac{(1-v-\delta)k}{vpk + (1-v-\delta)k}$$

Since there is only one such site where the mutant can replace its parent, E(#M|M)=Pr(M|M)

$\mathcal{L}t\ n \to \infty$:

$$E(\#M \mid WT) = \frac{(v+\delta)p}{(1-v)+vp+\delta p} \qquad E(\#M \mid M) = \frac{(1-v-\delta)}{(1-v-\delta)+vp}$$

Total # successful mutants at ALL sites: E(#M)=E(#M|WT) + E(#M|M)

$$E(\#M) = \frac{(v+\delta)p}{1-v+vp} + \frac{1-v-\delta}{1-v-\delta+vp}$$

If the mutant has on an average less than 1 successful offspring across all the n-sites i.e.

E(#M) < 1

it will not be able to invade and replace the WT

For $\delta << 1$ ignoring terms of $O(\delta^2)$ $\quad E(\#M) < 1 \Longrightarrow \frac{\delta}{2-p} < v\delta$

If $\delta > 0$ ➡ Mutant disperses ***more*** offspring than WT $\quad E(\#M) < 1 \Longrightarrow v > \frac{1}{2-p}$

If $\delta < 0$ ➡ Mutant disperses ***less*** offspring than WT $\quad E(\#M) < 1 \Longrightarrow v < \frac{1}{2-p}$

Mutants can *never* invade iff $v = v^* = \frac{1}{2-p}$ regardless of the sign of δ

Connection between the model and spread of altruism

Emigrants  Altruists

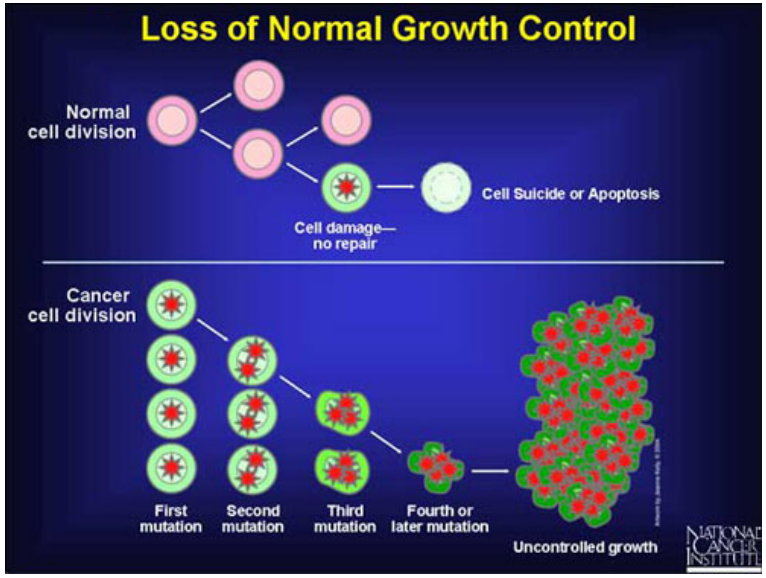
1. Dispersal (emigration) comes at a ***cost*** as it reduces the emigrants' chances of survival.
2. Dispersal has a ***benefit*** : Avoiding competition with (k-1) siblings for survival till the next generation and consequently enhancing the left-behind siblings average fitness.

A mutant (M) which produces ***more*** emigrants ($\delta>0$) is more altruistic than the WT

A mutant (M) which produces ***less*** emigrants ($\delta<0$) is less altruistic than the WT

Evolutionary Dynamics of Cancer

Cancer: Uncontrolled growth of abnormal (DNA damaged) cells.



Cancer cells manage to evade apoptosis

Apoptosis: mechanism of programmed cell death which ensures damaged cells are eliminated

Evolution and Cancer: How can cancer be understood in terms of evolution ?

In multi-cellular organisms, different types of cells must act in coordination with each other to ensure growth and development of the entire organism.

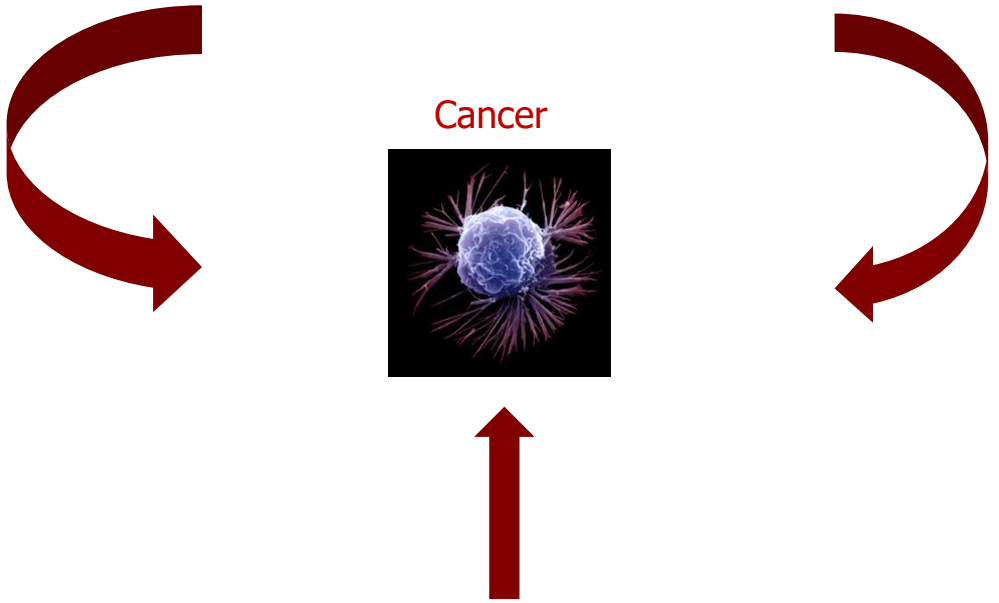
Cells should divide when needed ➔ cooperation between cells essential for normal functioning

Each cell has numerous mechanisms to prevent uncontrolled cell-division. Failure of one or more such mechanisms lead to cancerous growth ➔ Cancer occurs when certain cells turn into *selfish agents* by replicating at abnormally high rates.

Cancer: Genetic Disease caused by mutations in certain types of genes

Mutations in Tumor Suppressor Genes (TSG)

Mutations in Oncogenes



Mutations in genes causing chromosomal instability (CIN)

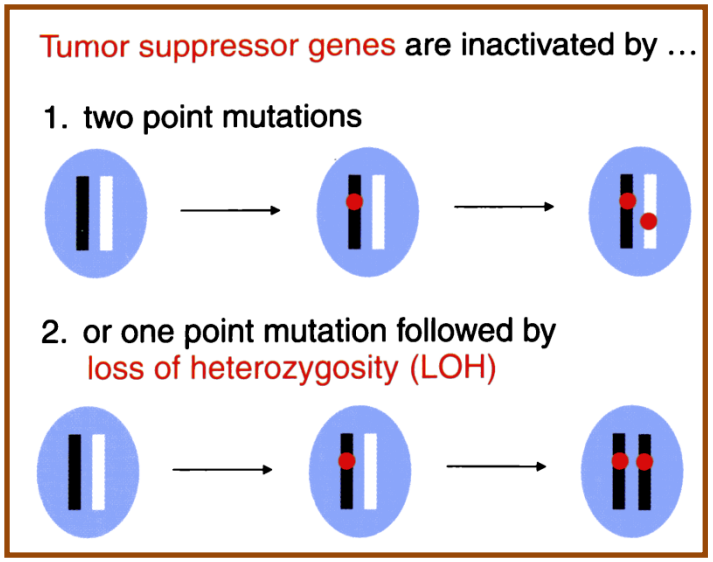
Tumor Suppressor Genes (TSG): Class of genes that prevent cancerous tumor formation and growth.

Example: p53 – mutated in more than half of all human cancers.

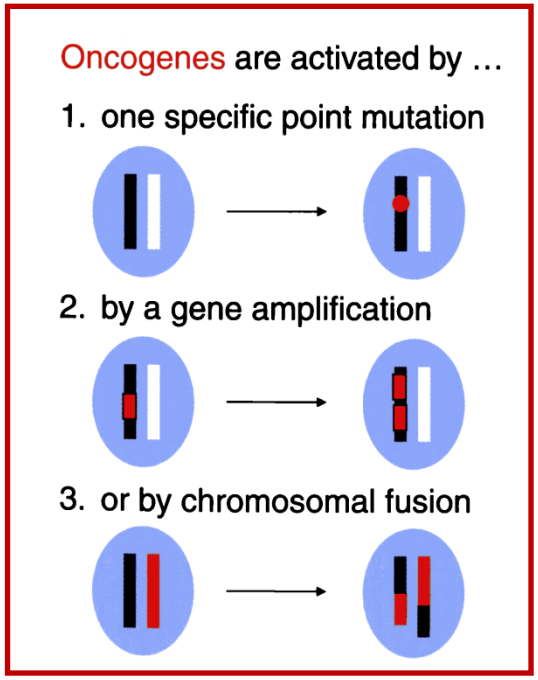
Oncogenes: Class of genes which when mutated leads to cancer

Chromosomal Instability (CIN): Mutations in certain oncogenes lead to increase/decrease in the number of chromosomes. *Examples:* MAD2, BRCA2, hBUB1, hCDC4

Pathways to Cancer

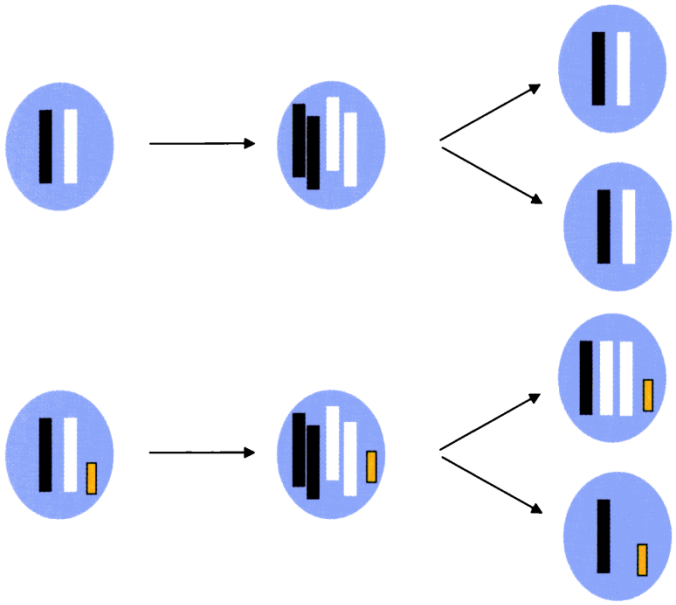


- Mutations in the first allele is *neutral*; no effect on cell division rate.
- LOH occurs when a chromosome is lost eliminating the corresponding allele of the TSG. Duplication of the remaining chromosome leads to two identical (homozygous) alleles of the TSG
- Mutations in the second allele *increase* the rate of cell division of damaged cells leading to cancer



- A single mutation is sufficient to activate an oncogene.
- Activating mutations can occur in different ways such as point mutation, gene amplification or chromosome fusion.
- Either of the above processes *increase* the rate of cell division leading to uncontrolled proliferation of damaged (cancerous) cells.

Chromosomal instability (CIN):
mutations in CIN genes increase the rate of gaining or losing whole chromosomes



3 classes of CIN genes

Onco-CIN genes

Class I CIN genes trigger CIN if one allele is mutated or lost.
Example: MAD2

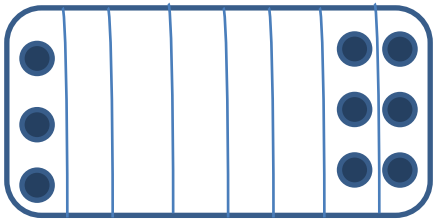
Class II CIN genes trigger CIN if one allele is mutated in a dominant negative fashion.
Example: hBUB1

Class III CIN genes trigger CIN if both alleles are mutated.
Example: BRCA2

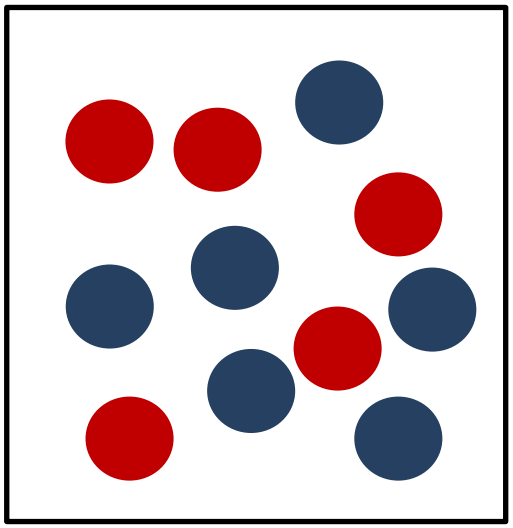
CIN suppressor genes

- Rate of loosing a chromosome in a cell due to mutations in CIN gene = 10^{-2} per chromosome per cell division event.
- Rate of LOH in cells in which CIN genes have not undergone specific mutations ~ 10^{-7} - 10^{-6} .
- Mutations in CIN genes increase the replication rate of damaged cells which have more or less than 46 chromosomes ➔ proliferation of such cancerous cells.

Tissue architecture and spatial organization on cancer progression

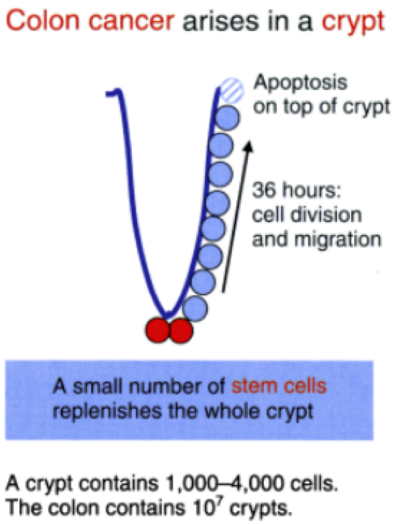


Tissues divided into compartments each containing a population (Ne) of cells

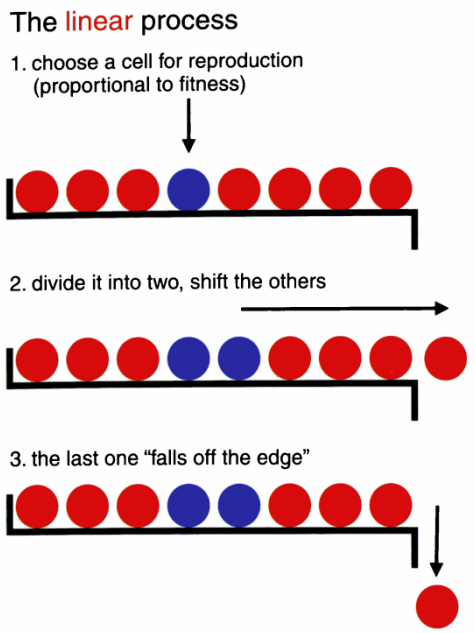


Well-mixed population of cells in a compartment

(Unstructured network)

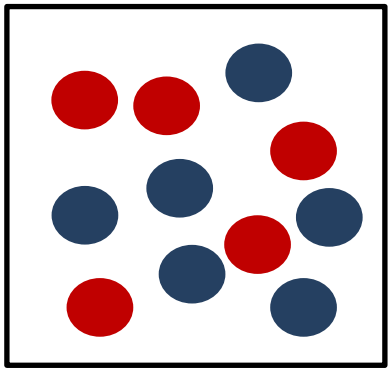


Tissue architecture and colon cancer



Linear network

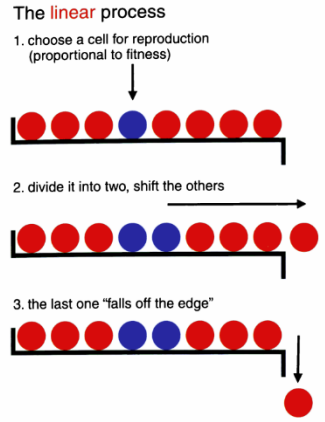
Fixation of a mutant within a tissue compartment: Effect of tissue architecture and spatial organization



N-population size within a tissue compartment

u-mutation rate per cell-division event

M-#compartments



Well-mixed population of cells in a compartment

(Unstructured network)

Linear network

$$\frac{dP}{dt} = Nu(1-P) \qquad P(t) = 1 - e^{-Nut}$$

P(t) - Probability that a mutant arises by time t

Fixation probability of a single mutant: $\rho = \frac{1-1/r}{1-1/r^N}$ (well-mixed case) $\rho = \frac{1}{N}$ (Linear network case)

$P_{fix}(t) = 1 - e^{-Nu\rho t}$: Probability that a mutant gets fixed by time t $N=10^3$; $u=10^{-9}$; $M=10^7$; $r=1.1$; $t=70$ yrs

$\rho \approx 0.09$ (well-mixed case) $\rho \approx 0.001$ (Linear network case)

Expected # mutated compartments after 70 years: P(t)M = 23000 (well-mixed); 26 (linear)

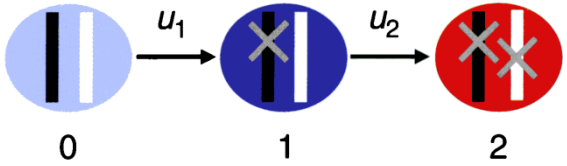
Architecture without compartments: $N=10^7$ P(t) ~ 0.28 in t<1 year

Linear architecture of tissue compartments significantly reduce rate of cancer progression

Cancer progression and TSG

What are the quantitative evolutionary questions that one can ask and answer in this context ?

How long does it take for a population of replicating cells to inactivate a TSG ?



- Starting from a population of N WT cells, how long does it take for two mutations (one in each allele) to appear in at least one cell ?
- How does this time depend on the population size and the mutation rates u1 and u2 ?

Assumption: The first mutation in one of the alleles that occur with rate u1 is neutral and it has no effect on the replication rate of cell where that mutation has occurred.

Let **A+/+** represent a "Type 0" cell with two functional (unmutated) alleles

Let **A+/-** represent a "Type 1" cell in which one allele of the TSG has been mutated but the remaining allele is still functional

Let **A-/-** represent a "Type 2" cell in which both alleles have been mutated.

Question: What is the probability that a single **A-/-** cell arises by time t in a population of replicating cells ?

Markovian Analysis of TSG Inactivation

Consider a Markov process with N+2 states. States $i=0,1,2,\ldots,N$ are either Type 0 or Type 1

Suppose there are i cells of Type 1 and $(N-i)$ cells of Type 0

State $i=N+1$ is the only absorbing state and indicates that a cell of Type 2 has been produced.

The *Transition Probabilities* are given by the following equations

$$P_{i,i-1} = \left[\frac{i}{N}\right]\left[\frac{N-i}{N}\right](1-u_1)$$

- $\frac{i}{N}$: Prob. of picking A+/- for death
- $\frac{N-i}{N}$: Prob. of picking A+/+ for reproduction
- $(1-u_1)$: Prob. that A+/+ picked for reproduction does not mutate to A+/-

$$P_{i,i+1} = \left[\frac{N-i}{N}\right]\left[\frac{i}{N}(1-u_2)+\frac{N-i}{N}u_1\right]$$

- $\frac{N-i}{N}$: Prob. of picking A+/+ for death
- $\frac{i}{N}$: Prob. of picking A+/- for reproduction
- $(1-u_2)$: Prob. that A+/- picked for reproduction does not mutate to A-/-
- $\frac{N-i}{N}$: Prob. of picking A+/+ for reproduction
- u_1: Prob. that A+/+ picked for reproduction mutates to A+/-

Markovian Analysis of TSG Inactivation

$$P_{i,i-1} = \frac{i}{N}\frac{N-i}{N}(1-u_1) \qquad i = 1, \ldots, N$$

$$P_{i,i} = \frac{i}{N}\left[\frac{i}{N}(1-u_2) + \frac{N-i}{N}u_1\right] + \left(\frac{N-i}{N}\right)^2(1-u_1) \qquad i = 0, \ldots, N$$

$$P_{i,i+1} = \frac{N-i}{N}\left[\frac{i}{N}(1-u_2) + \frac{N-i}{N}u_1\right] \qquad i = 0, \ldots, N-1$$

$$P_{i,N+1} = \frac{i}{N}u_2 \qquad i = 0, \ldots, N$$

$$P_{N+1,i} = 0 \qquad i = 0, \ldots, N$$

$$P_{N+1,N+1} = 1$$

Let t_i be the time required to generate at least one Type 2 cell starting from a cell in state i.

$$t_0 = 1 + P_{0,0}t_0 + P_{0,1}t_1$$

$$t_i = 1 + P_{i,i-1}t_{i-1} + P_{i,i}t_i + P_{i,i+1}t_{i+1} \qquad i = 1, \ldots, N$$

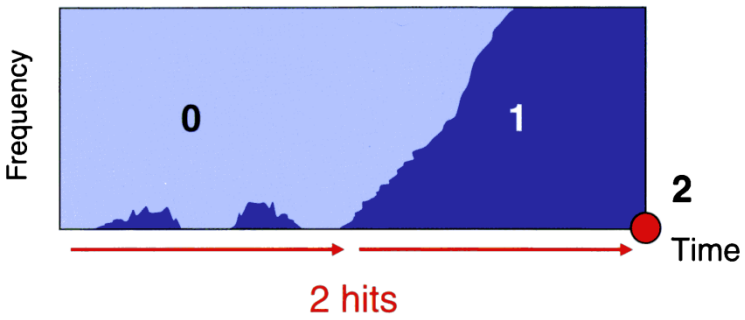
$$t_{N+1} = 0$$

This set of equations can be solved analytically in the limit of *small* and *large* population size N

Small Population Size: Allows for separation of time-scales

$\tau_1 = N \qquad \tau_2 = \frac{1}{Nu_2} \qquad \tau_1 << \tau_2 \Longrightarrow N << \frac{1}{\sqrt{u_2}}$

Small populations: $N < 1/\sqrt{u_2}$



N+2 state Markov process reduces to a 3 state Markov process

$X_0(t), X_1(t), X_2(t)$ – probabilities of being in states 0,1,2 respectively

State 0 : All cells are Type 0 i.e. **A+/+**

State 1 : All cells are Type 1 i.e. **A+/-**

State 2 : At least one cell is Type 2 i.e. **A-/-**

$X_i(t)$- Stochastic variable

$X_2(t)=P(t)$

Rate of producing Type 1 cells from Type 0

Fixation probability of Type 1 cells

Rate of producing Type 2 cells from Type 1

$$\dot{X}_0 = -(Nu_1)(\frac{1}{N})X_0 \qquad \dot{X}_1 = u_1 X_0 - Nu_2 X_1 \qquad \dot{X}_2 = Nu_2 X_1$$

Assumption: State 1 can be attained only after ***all*** A+/+ cells mutates to A+/- cells

Assumption: Fixation of Type 1 occurs before Type 2 cells can be produced

$$P(t) = 1 - \frac{Nu_2 e^{-u_1 t} - u_1 e^{-Nu_2 t}}{Nu_2 - u_1} \approx Nu_1 u_2 \frac{t^2}{2}$$

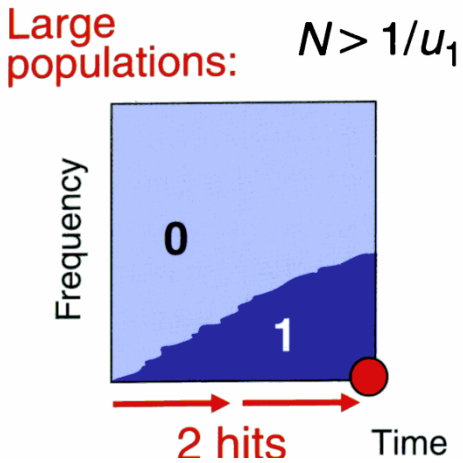
$t << \frac{1}{Nu_2}$

$$P(t) \approx 1 - e^{-u_1 t}$$

$t > \frac{1}{Nu_2}$

Large Population Size

Time scale of appearance of Type 1 cells : $\tau \sim 1/Nu_1 < 1$



For large population size ($N>1/u_1$) we can assume frequency (x_1) of Type 1 (A+/-) cells grows *linearly* with time

$$x_1(t) \approx Nu_1 t$$

P(t) – Probability that a mutation in the second allele occurs after time t

$$\frac{dP}{dt} = (1-P)x_1(t)u_2$$

$$P(t) = 1 - \exp(-Nu_1 u_2 t^2/2)$$

$Nu_1>1$ ➔P(t) ➔ 1 ➔ A-/- cell is produced in a shorter time scale compared to that in small populations

Time taken to inactivate a TSG in *Large* populations << Time taken to inactivate a TSG in *small* populations

What is the primary factor in cancer initiation ?

Inactivation of a TSG?

OR

CIN due to mutations in Onco-CIN genes?

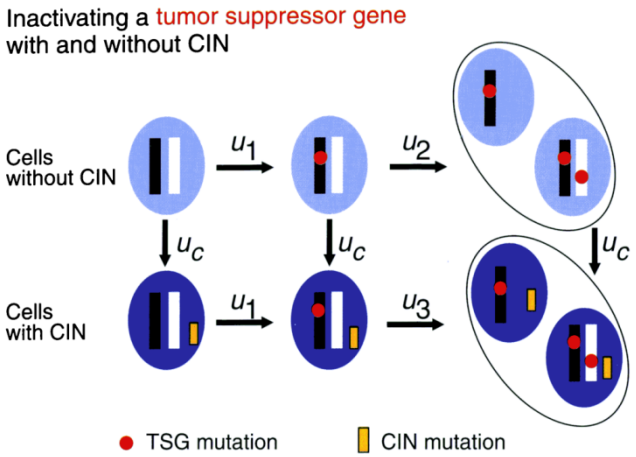
Which mechanism is responsible for cancer initiation ?

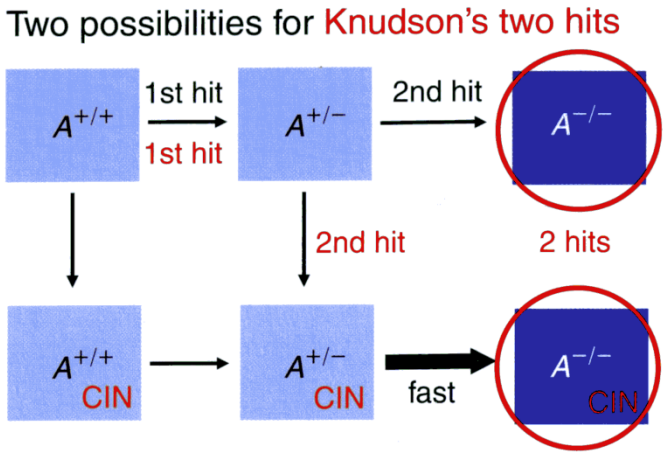
Can inactivation of a TSG occur before CIN brought about by mutations in CIN genes ?

Does CIN speed up the process of inactivating a TSG ?

➔ Faster initiation of cancer in CIN cells than in cells without CIN

} Imp. question for cancer treatment





Rapid LOH for A+/- CIN cells ➔ A-/- CIN

Relation between mutation rates

$$u_1 < u_2$$
$$u_1 = 10^{-7}; u_2 = 10^{-6}$$

Cells without CIN

$$N << u_1^{-1}, u_2^{-1}, u_c^{-1}$$

$$u_1 << u_3$$
$$u_1 = 10^{-7}; u_3 = 10^{-2}$$

CIN cells

Consider a stochastic process with 6 states

In state X_0, all cells are of type $A^{+/+}$.
In state X_1, all cells are of type $A^{+/-}$.
In state X_2, all cells are of type $A^{-/-}$.
In state Y_0, all cells are of type $A^{+/+}CIN$.
In state Y_1, all cells are of type $A^{+/-}CIN$.
In state Y_2, all cells are of type $A^{-/-}CIN$.

X_i also represent probabilities of finding the system in the corresponding state i at time t

If

Probability of finding the population in state Y2 (all A-/- CIN cells)

>

Probability of finding the population in state X2 (all A-/- cells without CIN)

➔ CIN is more likely to be responsible for cancer initiation than TSG inactivation

If

Probability of finding the population in state Y2 (all A-/- CIN cells)

<

Probability of finding the population in state X2 (all A-/- cells without CIN)

➔ TSG is more likely to be responsible for cancer initiation than CIN mutations

$$\dot{X}_0 = -(u_1 + u_c)X_0$$

$$\dot{X}_1 = u_1 X_0 - (u_c + Nu_2)X_1$$

$$\dot{X}_2 = Nu_2 X_1$$

$$\dot{Y}_0 = u_c X_0 - u_1 Y_0$$

$$\dot{Y}_1 = u_c X_1 + u_1 Y_0 - Nu_3 Y_1$$

$$\dot{Y}_2 = Nu_3 Y_1$$

Assumptions

- The mutation in the first allele is selectively neutral ➔ prob. Of fixation of the A+/- mutant is 1/N
- The CIN mutation which occurs at the rate u_c is also selectively neutral.
- A-/- cells have a strong selective advantage ➔ A single A-/- cell is very rapidly fixed in the population. This allows us to consider transitions from state X_1 (or Y_1) to the state X_2 (or Y_2 in which **all** the cells are A-/-)

In the time-scale (t) of a human lifetime ~ 70-80 yrs

$$u_1 t, Nu_2 t, u_c t << 1$$

$$X_0(t) \approx 1$$

$$X_1(t) \approx u_1 t$$

$$X_2(t) \approx Nu_1 u_2 t^2/2$$

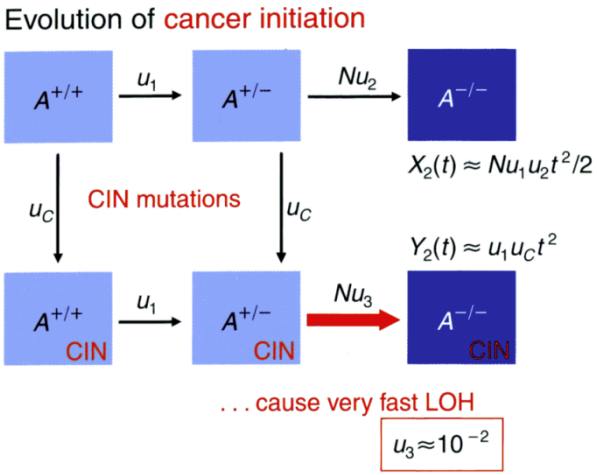
$$Y_0(t) \approx u_c t$$

$$Y_1(t) \approx u_1 u_c t^2$$

$$Y_2(t) \approx u_1 u_c t^2$$

If CIN is to initiate cancer earlier than TSG

$$Y_2(t) > X_2(t) \Rightarrow u_c > Nu_2/2$$



If rate of LOH = p_0

and

point mutation rate per gene = u

$u_2 = u + p_0$ Combination of point mutation rate for mutating the second allele and LOH

If there exists n_1 Class 1 CIN genes and n_2 Class 2 CIN genes

$$u_c = 2n_1(u + p_0) + 2n_2u$$

$$Y_2(t) > X_2(t) \Rightarrow 4n_1(u + p_0) + 4n_2u > N(u + p_0)$$

If u~p0, N=4, n1=n2=2 ➔ $\frac{Y_2(t)}{X_2(t)} = 3$ ⇒

~75% of all cancers initiated by CIN

~25% of all cancers initiated by TSG inactivation

Likelihood of cancer initiation by CIN in ***small*** populations when CIN mutations have a selective disadvantage

$$\dot{X}_0 = -(u_1 + N\rho u_c)X_0$$

$$\dot{X}_1 = u_1 X_0 - N(\rho u_c + u_2)X_1$$

$$\dot{X}_2 = N u_2 X_1$$

$$\dot{Y}_0 = N\rho u_c X_0 - u_1 Y_0$$

$$\dot{Y}_1 = N\rho u_c X_1 + u_1 Y_0 - N u_3 Y_1$$

$$\dot{Y}_2 = N u_3 Y_1$$

$$X_0(t) \approx 1$$

$$X_1(t) \approx u_1 t$$

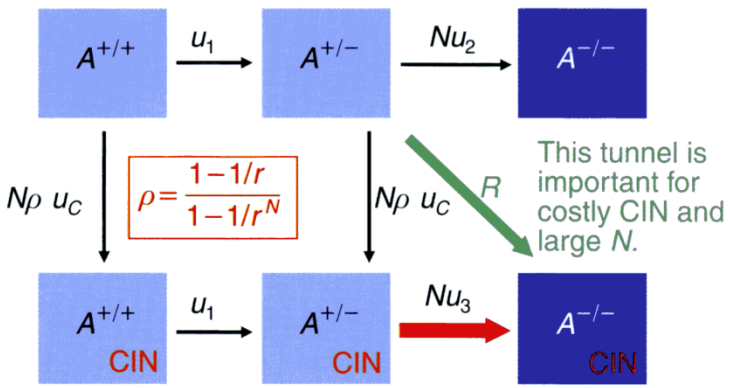
$$X_2(t) \approx N u_1 u_2 t^2/2$$

$$Y_0(t) \approx N\rho u_c t$$

$$Y_1(t) \approx N\rho u_1 u_c t^2$$

$$Y_2(t) \approx N\rho u_1 u_c t^2$$

CIN can imply a fitness gain, $r > 1$, or a fitness cost, $r < 1$.



$$\frac{Y_2(t)}{X_2(t)} = \frac{2\rho u_c}{u_2} = \frac{4\rho[n_1(u+p_0)+n_2 u]}{u+p_0}$$

If u~p0, N=4, n1=n2=2, r=0.8 ➔ $\rho = 0.1734$

$$\frac{Y_2(t)}{X_2(t)} = 2.0813$$

~68% of all cancers would be initiated by CIN

~32% of all cancers would be initiated by TSG inactivation

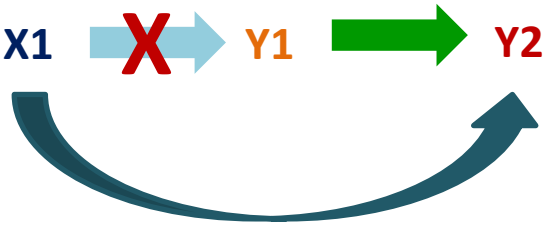
Likelihood of cancer initiation by CIN in ***large*** populations when CIN mutations have a selective disadvantage

If N=100, r=0.7 ➔ $\rho = 1.39 \times 10^{-16}$ ➔ Fixation of a CIN mutant is extremely unlikely

A+/- CIN cells are produced at the rate Nu_c but they are never fixed in the population

➔ Transition from state X1 (A+/- non-CIN cells only) ***to*** state Y1 (A+/- CIN cells only) does not occur.

X1➔ Y2 transition is still possible.



Average frequency of A+/- CIN cells at mutation/selection eq. $= Nu_c/(1-r)$

Rate at which A-/- CIN cells are produced from A+/- CIN cells $= ru_3$

Tunneling Rate from state X1 to state Y2: $R = \dfrac{Nu_c ru_3}{1-r}$

$$\dot{X}_0 = -u_1 X_0$$
$$\dot{X}_1 = u_1 X_0 - (R + Nu_2) X_1$$
$$\dot{X}_2 = Nu_2 X_1$$
$$\dot{Y}_2 = R X_1$$

$$X_0(t) \approx 1$$
$$X_1(t) \approx u_1 t$$
$$X_2(t) \approx Nu_1 u_2 t^2/2$$
$$Y_0(t) \approx 0$$
$$Y_1(t) \approx 0$$
$$Y_2(t) \approx Ru_1 t^2/2$$

If r=0.8, u3~0.01,u~p0,n1=n2=5

$$\frac{Y_2(t)}{X_2(t)} = \frac{ru_c u_3}{u_2(1-r)} \quad \Rightarrow \quad \frac{Y_2(t)}{X_2(t)} = 0.6$$

~38% of all cancers initiated by CIN

~62% of all cancers initiated by TSG inactivation

Is CIN is more likely to be responsible for cancer initiation than TSG inactivation ?

Depends on

(i) Whether the CIN mutation is neutral or selectively costly for the cell

(ii) The effective population size of the compartment


IT'S A MAGICAL WORLD, HOBBES, OL' BUDDY...
...LET'S GO EXPLORING!