# Mathematical theories of multi*level* and multi*scale* selection

## with applications to computational models

Rutger Hermsen

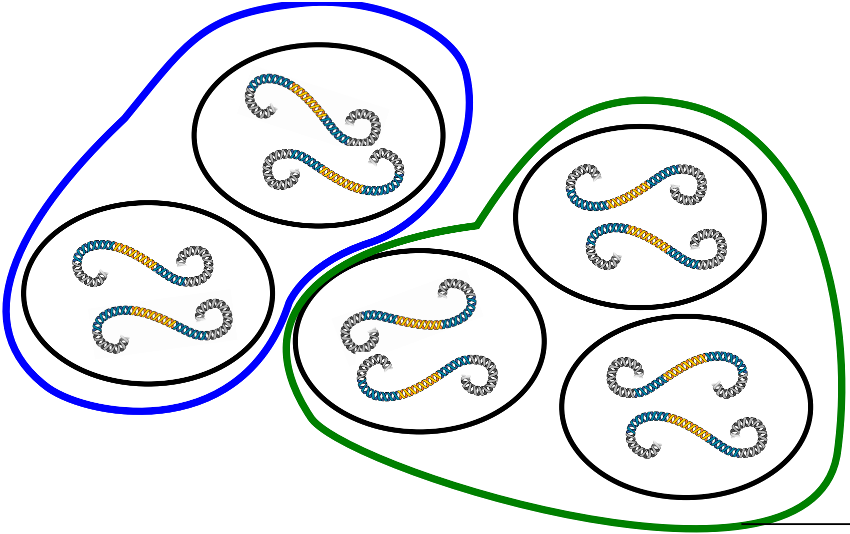
Theoretical Biology

Biology Department, Utrecht University

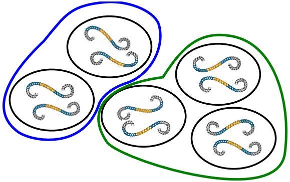
# Outline

1. Introduction: Group and multilevel selection
2. Formal theories of group selection
3. A simple model of the evolution of altruism
4. Spatial structure without groups: Multi*scale* selection
5. Conclusions

# 1.1 Biological systems are hierarchically organized



## 1.2 Natural selection may act at multiple levels



- Entities at every “level” of the hierarchy have their own properties and their own dynamics.
- The idea of adaptation by natural selection *can* be applied at all levels.

  (This does not mean that it is always natural or fruitful to do so…)
- Properties at one level often affect those of another level; selection pressures at different levels may not be aligned.

## 1.3 Group selection is often invoked to explain altruistic behaviors

### What is an altruistic trait in biology?

A trait that reduces the fitness of the actor but increases the fitness of its interaction partners (the recipients).

### Altruism has long been a mystery…

How could natural selection favor a trait that *reduces* fitness?

### …but a wide class of solutions has long been understood.

Altruism can evolve if somehow the benefits of altruism are disproportionately enjoyed by other altruists, thus offsetting their costs.

### But how??

One possibility: competition among *groups* with varying numbers of altruists.

# 1.4 Uncritical group selection came under fire in the late 1960s and 1970s

Naive group selection:
Explanations in terms of benefits to the group.

Example: Wynne-Edwards (1959):

*“A theory is put forward that, for each species, population-densities are limited at a safe level, which will protect the food-supply from long-term depletion and assure its renewal for the future. Instead of competing directly for food, animals compete for conventional substitutes, e.g. territory or social position, which are capable of imposing a ceiling density at the optimum level, and can prevent it from rising to the starvation level which would endanger future resources.”*

Main counter-argument:
instability to cheaters

- *Within* each population, cheaters (also called defectors) have a higher fitness.
- Result: On the long run, cheaters take over. (“Tragedy of the Commons”)

E.g., Maynard Smith (1964), G.C. Williams (1966), Dawkins (1978).

# 1.5 Alternative framework: Inclusive fitness theory and Hamilton's rule (1964)

Inclusive fitness theory, main idea:

- A trait can successfully spread if it promotes the fitness of the organism that possesses it *or that of other organisms that possess that trait.*
- Relatives are likely to share traits. Organisms that mainly interact with relatives may therefore accept fitness costs to benefit others.
- Inclusive fitness effect: effect on the fitness of the bearer *and others weighted by their relatedness.*

Hamilton's rule

A trait experienced positive selection if:

$$c < Rb$$

$c$ = fitness cost of actor,
$b$ = fitness benefit of recipient,
$R$ = relatedness between actors and recipients.

# 1.6 Despite the controversy, new group-selection models emerged steadily

Examples:

- Maynard Smith’s Haystack model (1964)
- D.S. Wilson’s trait group model (1975, 1977, 1979)
- Queller (1992)
- Tarnita, Nowak (2002)
- …

Group selection in these models works roughly if

- it allows altruists to mainly interact with *other* altruists
  (consistent with Hamilton’s rule),
- groups with more altruists have higher mean fitness.

But debates are ongoing as to the best framing of these results.

## 1.7 Many controversies remain

Tough questions:

- In those cases where inclusive fitness theory and multilevel selection theory are both applicable, are they mathematically equivalent?
- Even if they do, do they describe different causal/mechanistic situations?
- Do we need to take all levels into account or is the lowest level enough?
- Are the selection pressures at all levels potentially equally important?
- Can we understand how selection pressures at different levels interact?
- Can we apply all this theory to experimental and/or simulation data?
- Can we apply group selection theory if the population is not neatly subdivided into discrete non-overlapping groups?

## 1.8 What will we do today?

We will:

- explain one possible formal definition of natural selection at the group level.
- prerequisites: Price equation, covariance, Simpson's paradox.
- illustrate the theory it by applying it to computational models.
- introduce Multi*scale* selection theory.

# Outline

1. Introduction: Group and multilevel selection
2. Formal theories of group selection
3. A simple model of the evolution of altruism
4. Spatial structure without groups: Multi*scale* selection
5. Conclusions

## 2.1 Formal theories help to resolve confusion

Without formal definitions, it is hard to resolve differences of opinion.

- What is natural selection exactly, and how do we measure it?
- What is group selection exactly, and how do we measure it?
- What is fitness, and how do we measure a fitness cost or benefit?

Analysis based on the Price equation can brings some clarity.

## 2.2 Multiple frameworks exist to formalize and quantify MLS

At least three alternative formalisms exist, called:

1. Multilevel Selection 1 (the "Price approach")
2. Multilevel Selection 2
3. Contextual analysis

In the interest of time, we will focus on MLS 1.

## 2.3 The basic logic of adaptation by natural selection

Necessary and sufficient conditions for natural selection:

1. variation in the trait
2. heritability of the trait
3. the trait is associated with fitness

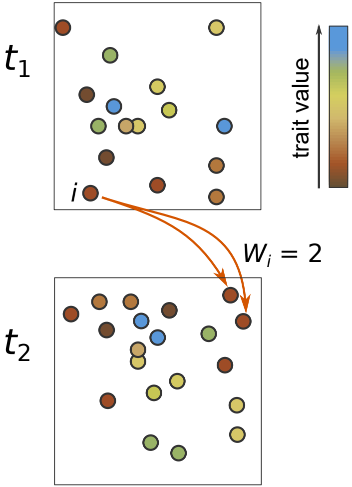
This seems so general, so logical…

Shouldn’t it be possible to derive it as a mathematical theorem?

---

Peter Godfrey-Smith, “Conditions for Evolution by Natural Selection:,” Journal of Philosophy 104, no. 10 (2007): 489–516, https://doi.org/10.5840/jphil2007104103.

## 2.4 The Price Equation: selection measured as a covariance

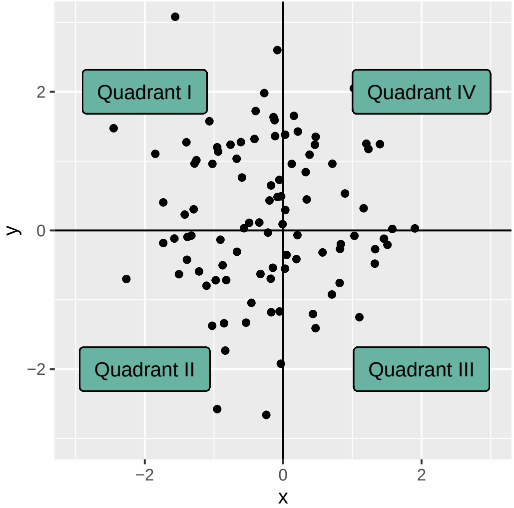


**The standard Price Equation**

- Consider a time step $\Delta t \equiv t_2 - t_1$.
- Change $\Delta\overline{\phi}$ in mean trait value.
- Denote number of offspring of ancestor $i$ as $W_i$. (“Absolute fitness”)
- Define $w_i \equiv W_i/\overline{W}$ (relative fitness).
- Then the Price Equation can be derived as:

$$\Delta\overline{\phi} = \underbrace{\mathrm{Cov}\,(\phi, w)}_{\text{selection differential } S} + \underbrace{\overline{w\Delta\phi}}_{\text{transmission } T} .$$

# 2.6 The covariance measures whether two traits (variables) “vary together”



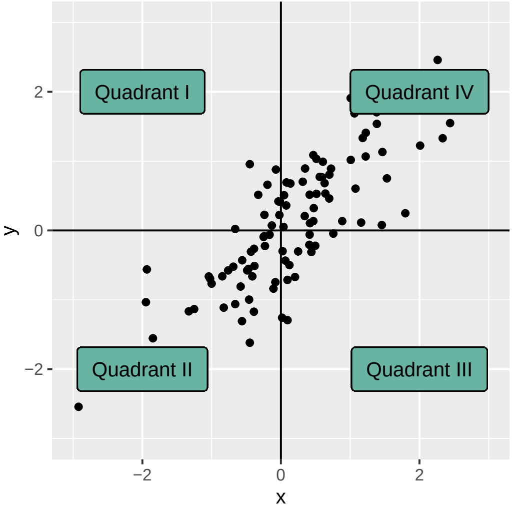
Definition population mean:

$$\overline{x} \equiv \frac{\sum_{i=1}^{n} x_i}{n}.$$

Definition population covariance:

$$\mathrm{Cov}\,(x, y) \equiv \overline{(x - \overline{x})(y - \overline{y})}.$$

# 2.6 Covariance and correlation are intimately related

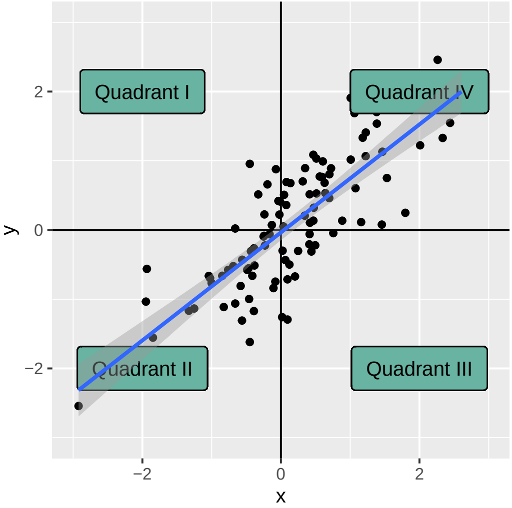


Definition correlation coefficient:

$$\rho_{x,y} \equiv \frac{\mathrm{Cov}\,(x,y)}{\sigma_x \sigma_y}.$$

Here, $\sigma_x$ and $\sigma_y$ are the standard deviations of $x$ and $y$, respectively.

# 2.6 Covariance and slope of regression line are also intimately related



**Regression line:**

Straight line $y = \alpha + \beta x$ that "best fits" the cloud of data points.

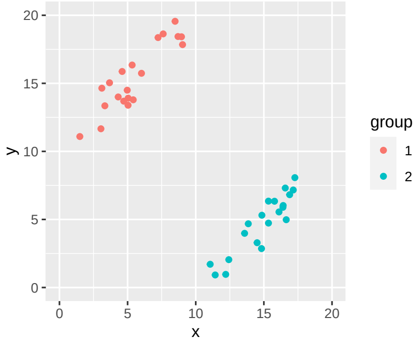
(The one that minimizes the sum of squared vertical distances from the line.)

**It can be derived that:**

$$\beta = \frac{\mathrm{Cov}\,(x, y)}{\mathrm{Var}\,(x)}.$$

Hence, the covariance is also a measure of the *linear relation* between traits.

# 2.7 Simpson's paradox: aggregated data can be confusing



Positive or negative relation?

Suppose $x$ is hours per week playing Minecraft.
Suppose $y$ is the score in a spatial awareness test.
Groups are age groups.
(Blue: age 9–11. Red: age 15–17.)
Is playing Minecraft associated with good spatial awareness?

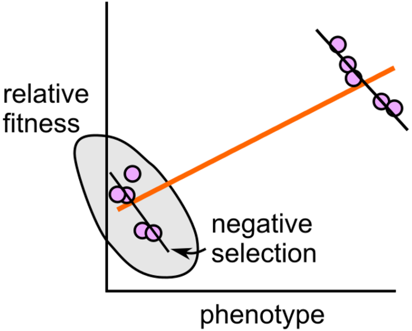
How does the population covariance relate to the covariance in subgroups?

Law of Total Covariance: $\mathrm{Cov}\,(x, y) = \langle \mathrm{Cov}_{\mathrm{w}}\,(x, y \mid g) \rangle_{\mathrm{a}} + \mathrm{Cov}_{\mathrm{a}}\left( \{x \mid g\}_{\mathrm{w}}, \{y \mid g\}_{\mathrm{w}} \right).$

(Weighted averages if groups have unequal size.)

# 2.8 Quantifying MultiLevel Selection 1: the Price approach[1]

George Price's idea: apply the Law of Total Covariance to the selection differential $S$!



- Imagine organisms subdivided into discrete, non-overlapping groups.
- Law of Total Covariance splits $S$ into two parts:

$$S = S_{\text{within}} + S_{\text{among}}.$$

where

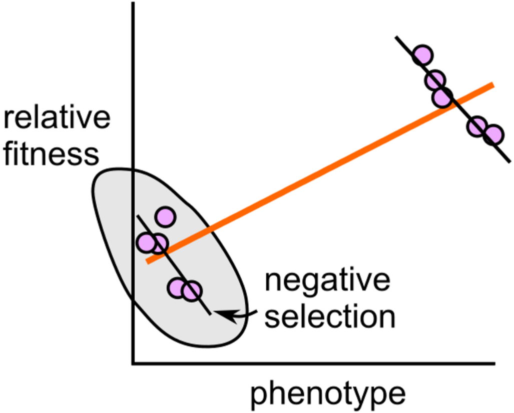
$$S_{\text{within}} = \text{mean covariance } \textit{within} \text{ groups}$$

$$S_{\text{among}} = \text{covariance between mean trait and mean fitness } \textit{among} \text{ groups}$$

---

[1] Price (1970)

# 2.8 Positive among-group selection can trump negative within-group selection



$$S = S_{\text{within}} + S_{\text{among}}.$$

where

$S_{\text{within}}$ = mean covariance *within* groups

$S_{\text{among}}$ = covariance between mean trait and mean fitness *among* groups

# Outline

1. Introduction: Group and multilevel selection
2. Formal theories of group selection
3. A simple model of the evolution of altruism
4. Spatial structure without groups: Multi*scale* selection
5. Conclusions

## 3.1 Spatial structure can play an important role

☺

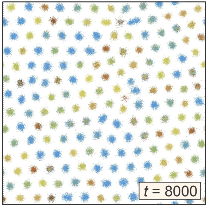
**Assortment can promote altruism**

Interactions happen locally; spatial demographics promote assortment.

☹

**Resource competition is local too**

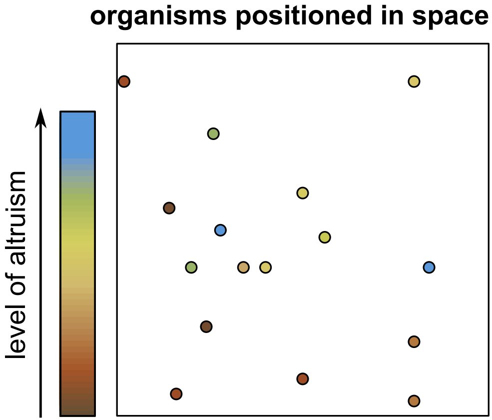
If altruists mainly compete with altruist, altruism does not help.



**This inspired me to study a simple model in which**

- scales of interaction and motility can be tuned directly,
- complex spatial structures emerge.

# 3.2 Essence of the model: simple agents interacting locally



Individuals do three things:

1. Move (by unbiased diffusion)
2. Reproduce (asexually)
3. Die (at a fixed rate)

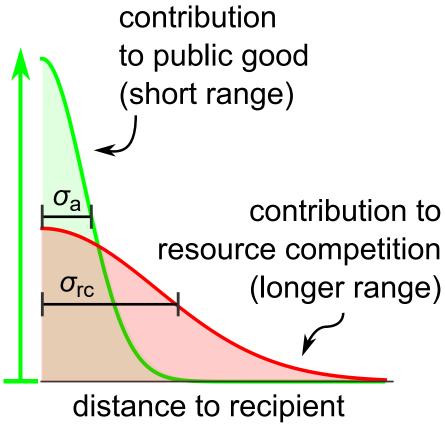
Interactions:

- Resource competition
- Altruistic cooperation

  (Perhaps production of a "public good".)

Single evolvable trait: Level of altruism $\phi$.

# 3.2 Competition and altruism each have their own scale



contribution
to public good
(short range)

contribution to
resource competition
(longer range)

distance to recipient

Interaction ranges:

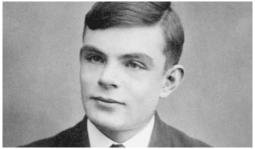
- Range of competition: $\sigma_{rc}$
- Range of altruism: $\sigma_a$

We choose $\sigma_{rc} > \sigma_a$

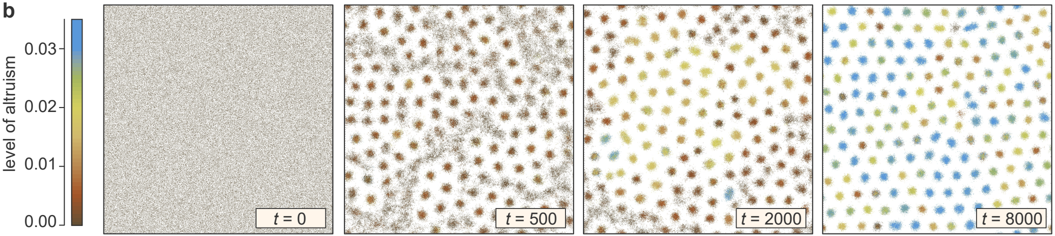
Allow altruism *at the expense of others.*

# 3.3 Dynamics: Movie time!

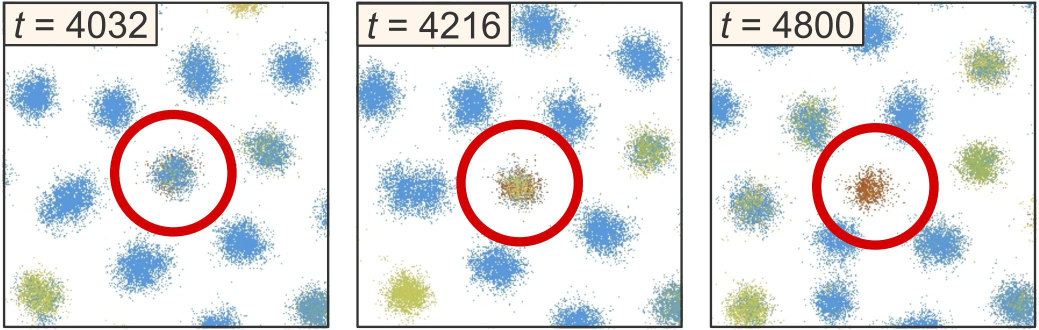
# 3.3 Hexagonal pattern of colonies emerges due to a Turing-like instability



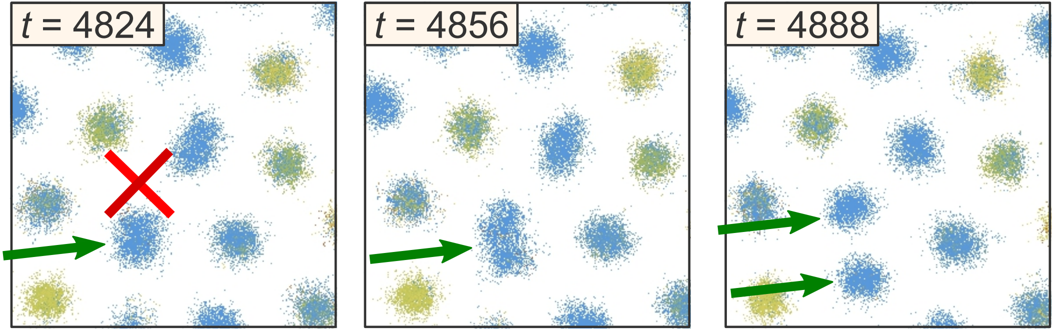
"Short-range activation plus long-range inhibition"



## 3.3 Colonies are unstable and die: "tragedy of the commons"

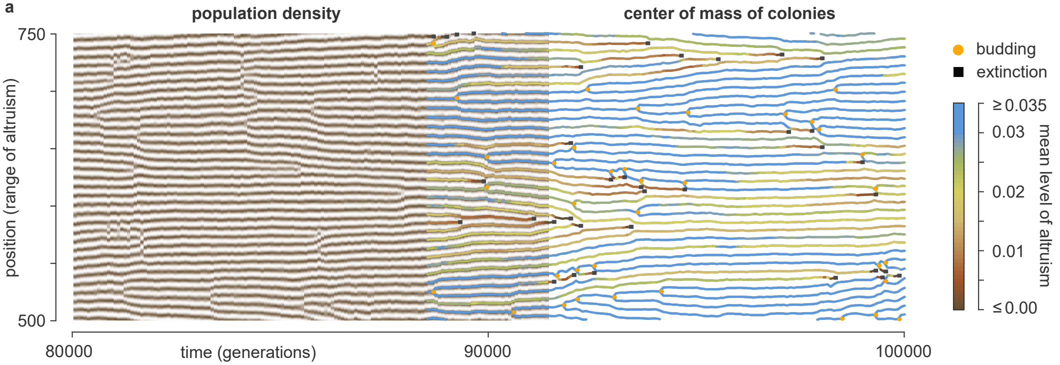


## 3.3 Colonies reproduce (binary fission) to replace the dead



# 3.3 In a 1D version of the model, colonies can be tracked

Within-colony tragedy of the commons,
but among-colony selection for altruism.

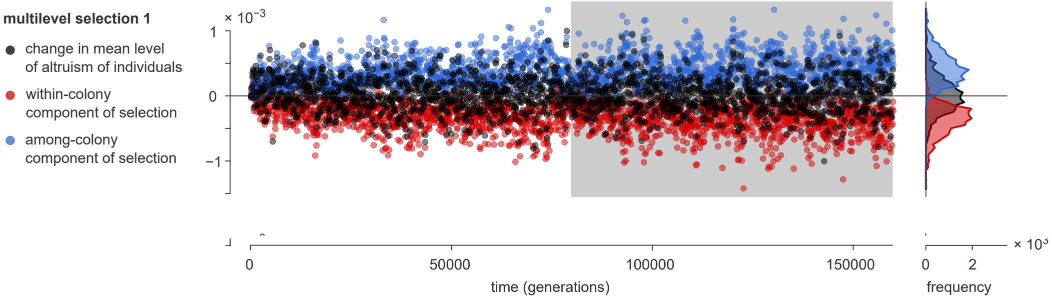


## 3.4 Conclusions so far on the model of altruism

- Altruism can evolve (both in the 1D and 2D version).
- Spontaneous emergence of colonies that reproduce and die.
- Colonies are intrinsically unstable against corruption by cheaters: "Tragedy of the commons".
- A colony that dies is replaced through growth and division of a colony nearby.
- Altruistic colonies seem to reproduce more frequently.

Can we use theory to quantitatively measure selection within and among colonies?

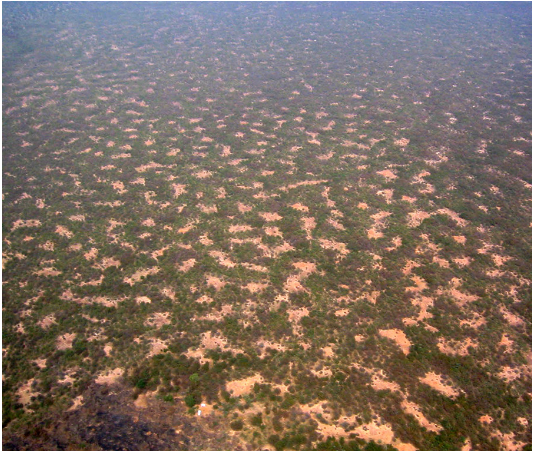
# 3.5 Applying MLS 1 to the simulations



# Outline

1. Introduction: Group and multilevel selection
2. Formal theories of group selection
3. A simple model of the evolution of altruism
4. Spatial structure without groups: Multi*scale* selection
5. Conclusions

# 4.1 Natural populations tend to be highly structured in space

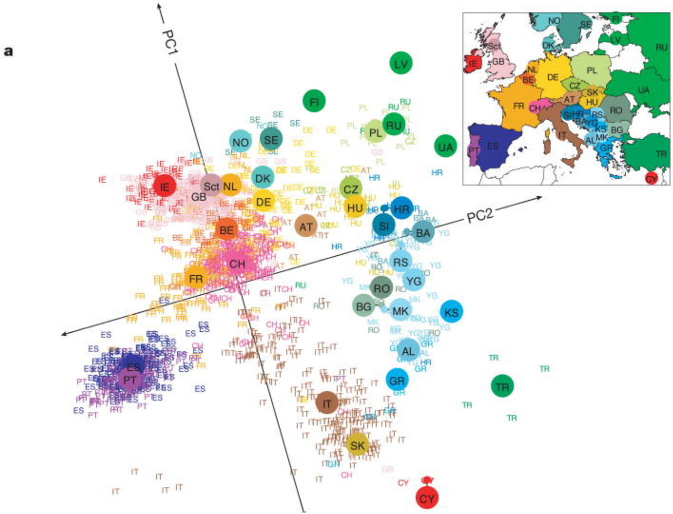


Gapped bush vegetation patterns, Niger
(photo: Nicolas Barbier, WikiMedia)

# 4.1 Natural populations tend to be highly structured in space

Colonies of *B. subtilis* strains (Ben Jacob, 1997)

# 4.1 Natural populations also tend to be *genetically* structured in space



PCA of genetic variation in Europeans reveals map of Europe
(Novembre *et al*, Nature, 2009)

# 4.2 MLS theory not directly applicable to such "viscous" populations

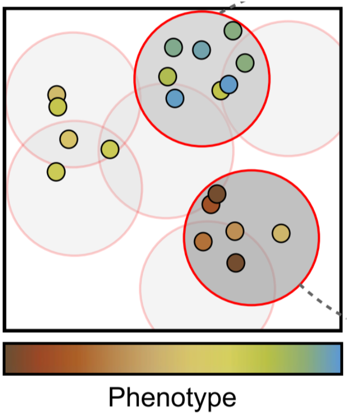
## Local environments are somewhat (un)like groups

- Interactions (competition, cooperation, ...) and reproduction are *local*; emergent spatial patterns affect who interacts with whom.
- Spatial assortment can result in relatedness among interacting individuals.
- But, MLS theory not applicable if no discrete groups can be discerned.
- Yet, we intuit: sign of *local selection* may be different from *global selection*.

# 4.3 Our idea to formalize this: multi*scale* selection

*Multiscale selection in spatially structured populations*
Hilje M. Doekes, Rutger Hermsen
bioRxiv 2021.12.21.473617

# 4.3 Multi*scale* selection instead of multi*level* selection



**Local selection measured as Local Selection Differential**

- Define a local environment as a disk of radius $r$.
- Local Selection Differential (LSD) of that environment: $\mathrm{Cov}(\phi, w)$ among population within the disk.

**Then we can derive: $S = S_{\text{local}}(r) + S_{\text{interlocal}}(r)$**

where

$$S_{\text{local}}(r) = \text{mean of LSD over all possible disks}$$

$$S_{\text{interlocal}}(r) = \text{covariance between } \textit{mean} \text{ phenotype and } \textit{mean} \text{ fitness over all disks}$$

(Environments must be weighted by local density.)

## 4.3 Properties of $S_{\text{local}}(r)$ and $S_{\text{interlocal}}(r)$

**Limit of small scales**

All variance is lost; hence

$$S_{\text{local}}(r) \longrightarrow 0,$$

$$S_{\text{interlocal}}(r) \longrightarrow S.$$

**Limit of large scales**

Local environments become representative of full population; hence

$$S_{\text{local}}(r) \longrightarrow S,$$

$$S_{\text{interlocal}}(r) \longrightarrow 0.$$

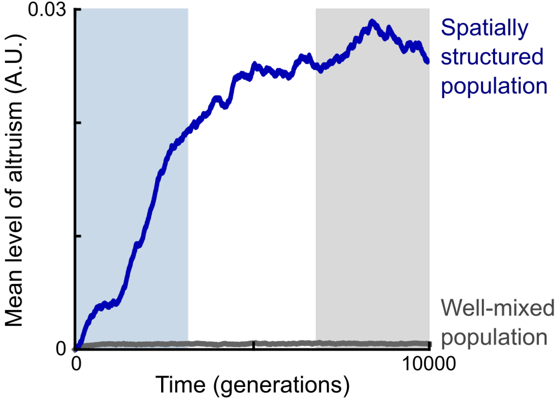
## 4.3 The contribution to selection of scale $r$

Define the Contribution to selection of scale $r$ as

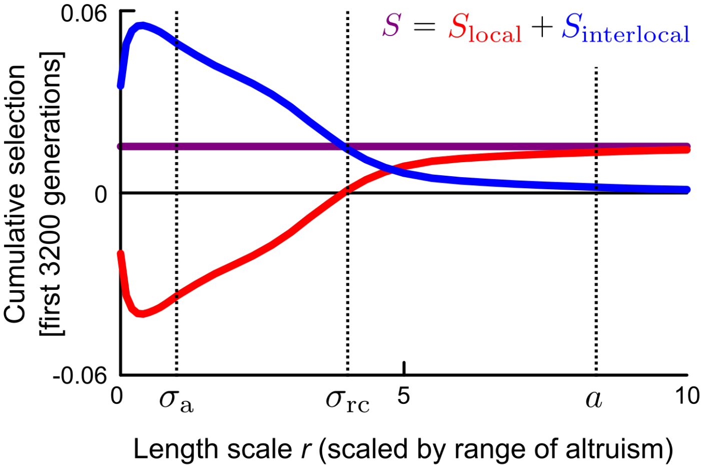
$$s(r) = \frac{dS_{\text{local}}(r)}{dr}.$$

If $S_{\text{local}}(r)$ increases if we increase $r$ to $r + dr$, the scale $r$ contributes positively.

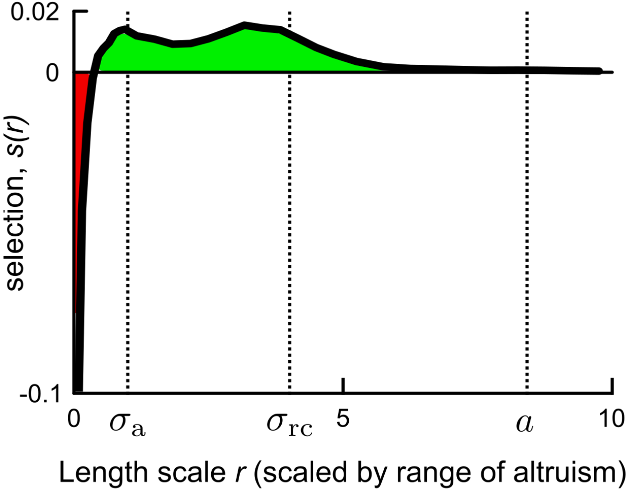
# 4.4 Selection is positive over the initial part of the simulation



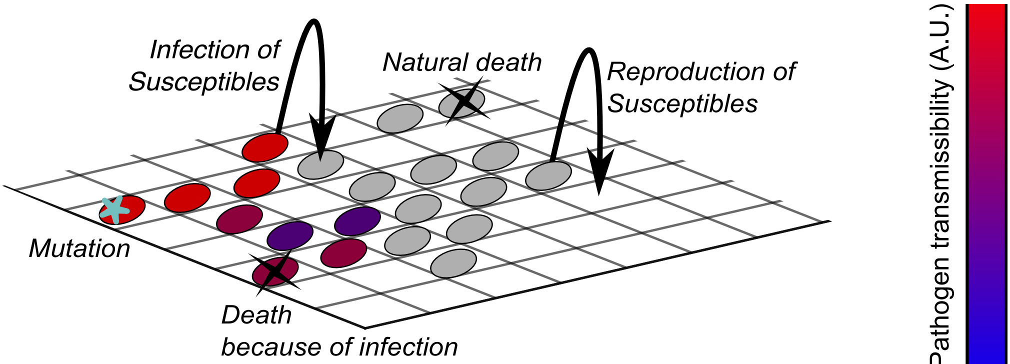
## 4.4 Yet, locally selection is negative



# 4.4 The contribution to selection is only negative for very small scales

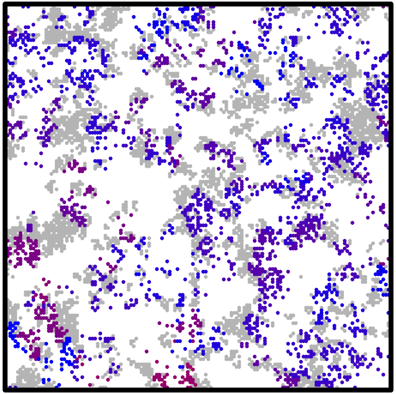


# 4.5 An SI model of the evolution of infectivity

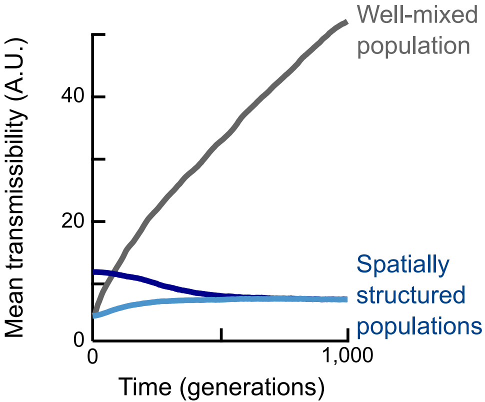


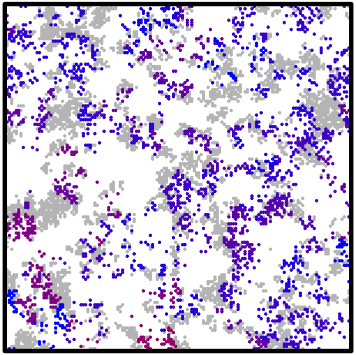
Note: Could be a model of an organism consuming a resource.

# 4.6 Dynamics: Movie time!



# 4.6 In spatial models, transmissibility stays limited. Self-shading?





Hypothesis: *locally*, higher transmissibility is favored, *globally* it is not.

Can we define and quantify this?

## 4.7 Multi*level* selection framework is not applicable

Patches are not groups

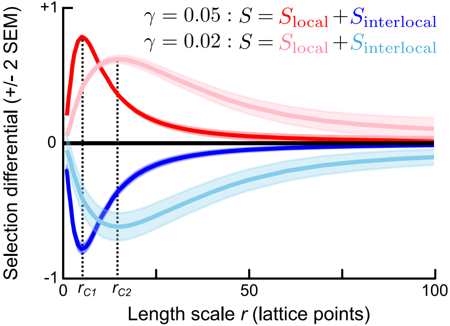
- Patches are too ephemeral to be conceptualized as “groups”
- No relation between scale of patch and range of interaction
- (Patches do not reproduce in any clear-cut sense)

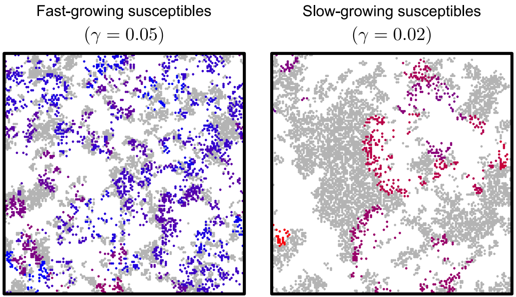
Intuition:

in *local* environments, high transmissibility is favored;
*globally*, low transmissibility is favored.

Proof: apply multi*scale* selection theory.

# 4.8 Multi*scale* selection: differences between local and global selection




Fast-growing susceptibles ($\gamma = 0.05$)

Slow-growing susceptibles ($\gamma = 0.02$)

Self-shading: rapid reproduction locally depletes susceptible hosts and hence locally reduces mean fitness.

Selection for fast transmission within patches, selection for restraint among them.

# Outline

1. Introduction: Group and multilevel selection

2. Formal theories of group selection

3. A simple model of the evolution of altruism

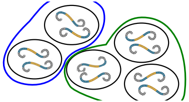
4. Spatial structure without groups: Multi*scale* selection

5. Conclusions

## 5.1 Take-away messages

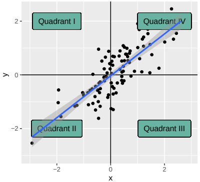
### 1. Hierarchy

Natural selection can act at any level of organization.



### 2. The Price Equation

Natural selection can be measured as a covariance.



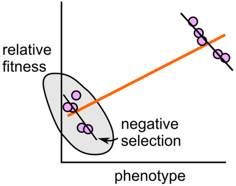
### 3. The Law of Total Covariance

In group-structured populations, the Law of Total Covariance can split up selection into within- and among-group parts.

$$S = S_{\text{within}} + S_{\text{among}}$$

### 4. Simpson's Paradox

Aggregated data can conceal relations within subgroups.



### 5. Solving conundrums

Because of Simpson's paradox, selection on a trait *can* be positive in the population even if it is negative in each group.
Generally, rigorous mathematical definitions help to solve confusing questions on evolution.

### 6. Application to simulations

Such formalisms can be used to measure quantities in simulations.