Repeated PD game with stochastic reactive strategies Lecture 4

Nowak and Sigmund, *Acta Appl. Math.* 20 (1990) 247; *Nature* 355 (1992) 250.

Like Axelrod's tournament, but with a set of stochastic reactive strategies.

Payoff matrix:

$$\mathbf{A} = \begin{array}{c} \\ D \\ C \end{array} \begin{array}{c} \begin{array}{cc} D & C \end{array} \\ \begin{pmatrix} (P,P) & (T,S) \\ (S,T) & (R,R) \end{pmatrix} \end{array}, \qquad T=5, \quad R=3, \quad P=1, \quad S=0$$

Stochastic reactive strategies:

stochastic decisions that depend on the previous step of the coplayer

similar to TFT strategies:

$$s(p,q,y), \qquad 0 \le p,q,y \le 1$$

p: probability to choose C, if the partner used C in the previous step (otherwise D)

q: probability to choose C, if the partner used D in the previous step (otherwise D)

y: probability to choose C in the first step

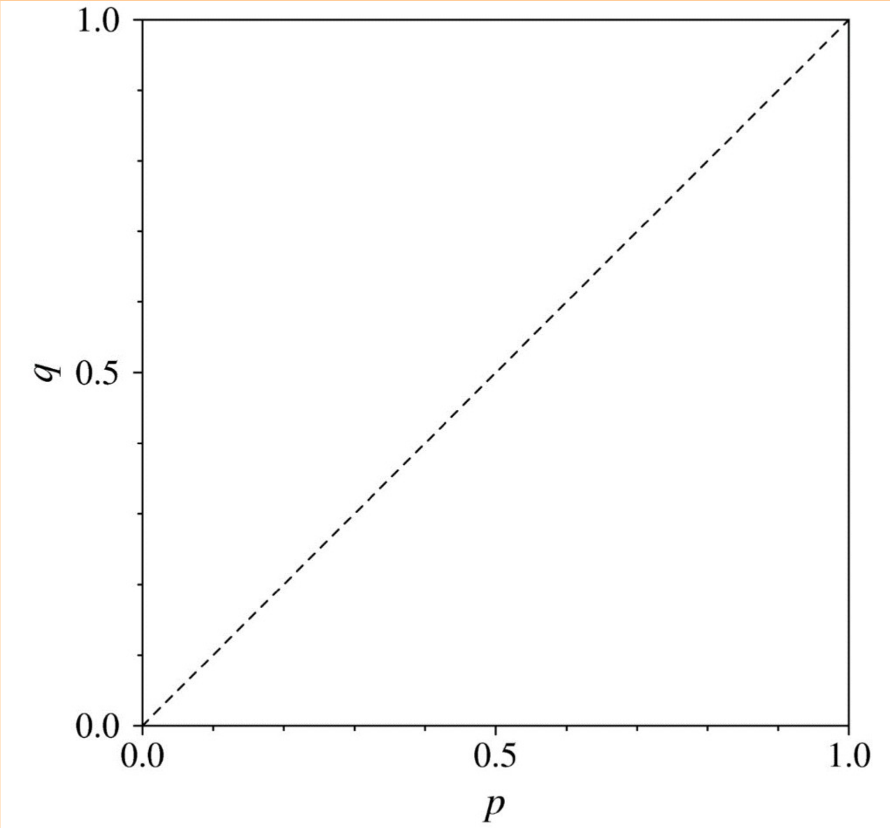
Strategy space

Deterministic strategies: $p, q, y = 0$ or 1

In iterated games, y becomes irrelevant in the long run and

$0 < p, q < 1$ become relevant in the presence of noise.

Each point (p,q) represents a stochastic reactive strategy



$p=q=0$: AllD

$p=q=1$: AllC

$p=q$: independent of the previous step

$p=q=1/2$: coin toss

$p=1$, $q=0$: TFT (if $y=1$)

$p=1$, $q>0$: generous TFT

$p=1$: nice (friendly) strategies

$p=0$, $q=1$: contrarian

Strategy $s(p,q,y)$ plays against $s(p',q',y')$:

In step n (n=1,2,…) the probability of choosing C is:

$$n=1;\quad \rho_1 = y; \qquad\qquad \rho'_1 = y';$$

$$n=2;\quad \rho_2 = p\rho'_1 + q(1-\rho'_1); \qquad\qquad \rho'_2 = p'\rho_1 + q'(1-\rho_1);$$

$$\vdots$$

$$n+1;\quad \rho_{n+1} = p\rho'_n + q(1-\rho'_n); \qquad\qquad \rho'_{n+1} = p'\rho_n + q'(1-\rho_n);$$

$$n+2;\quad \rho_{n+2} = p\rho'_{n+1} + q(1-\rho'_{n+1}); \qquad\qquad \rho'_{n+2} = p'\rho_{n+1} + q'(1-\rho_{n+1});$$

$$\rho_{n+2} = pp'\rho_n + pq'(1-\rho_n) + q(1-p'\rho_n + q'(1-\rho_n)); \qquad\qquad \rho'_{n+2} = p'p\rho'_n + p'q(1-\rho'_n) + q'(1-p\rho'_n + q(1-\rho'_n));$$

$$\rho_{n+2} = \rho_n(pp' - pq' - qp' + qq') + (pq' + q - qq')$$
$$= \rho_n(p-q)(p'-q') + pq' + q(1-q');$$
$$\rho'_{n+2} = \rho'_n(p'-q')(p-q) + p'q + q'(1-q);$$

The $n\to\infty$ stationary solution becomes independent of ys if $|p-q|$, $|p'-q'|<1$

$$\rho_{n+2} = \rho_n = \rho \quad \text{and} \quad \rho'_{n+2} = \rho'_n = \rho'$$

Solution:

$$\rho = \frac{q+(p-q)q'}{1-(p-q)(p'-q')} \quad \text{and} \quad \rho' = \frac{q'+(p'-q')q}{1-(p-q)(p'-q')}$$

Average payoff of strategy *s* against *s´*:

$$U(s,s') = R\rho\rho' + S\rho(1-\rho') + T(1-\rho)\rho' + P(1-\rho)(1-\rho')$$

or against herself $\quad U(s,s) = \rho^2(R-S-T+P) + \rho(S+T-2P) + P$

In this latter case:

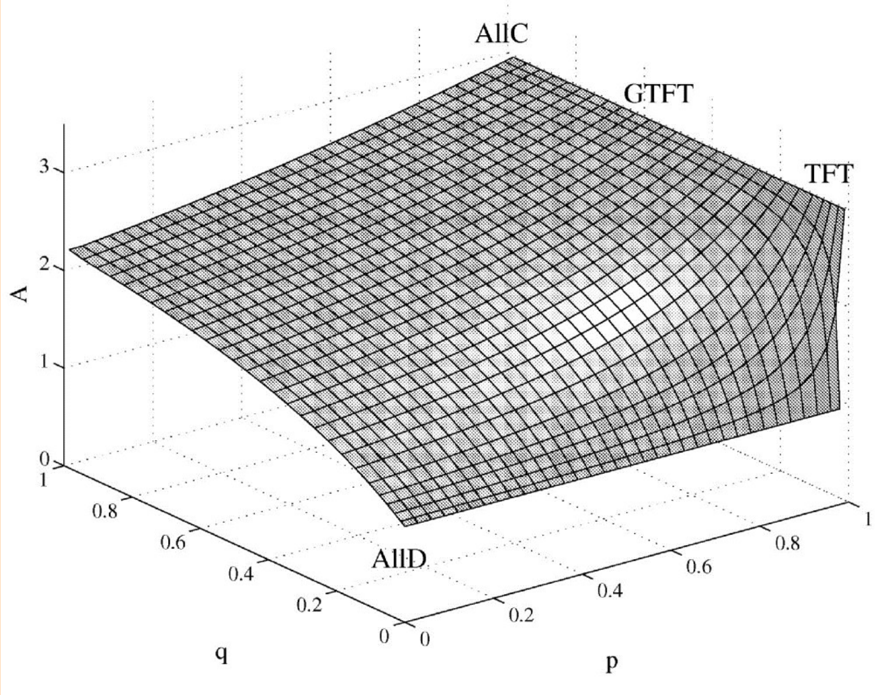
$p = p', \quad q = q',$ and

$$\rho = \frac{q(1+p-q)}{1-(p-q)(p-q)} = \frac{q(1+p-q)}{(1-p+q)(1+p-q)}$$

$$\rho = \frac{q}{1-p+q}$$

if $\quad q = 0, \quad$ then $\quad \rho = 0 \quad$ and $\quad U = P$

if $\quad p = 1, \quad$ then $\quad \rho = 1 \quad$ and $\quad U = R$



Non-analytic behaviour at TFT!

Evolution of the population of stochastic reactive strategies

Repeated competition between N stochastic reactive strategies at times t=0,1,2, …

ρ_i portion of players follow strategy $s_i=s(p_i,q_i)$ (i=1, …, N)

Initially: $\rho_i(t=0)=1/N$

Payoff of strategy s_i (or strategy i):

$$U_i(t)=\sum_j U(s_i,s_j)\rho_j(t), \quad \text{where} \quad \sum_j \rho_j(t)=1$$

$$t=0,1,2,\ldots$$

At time t+1 the players can change their strategies:

Successful populations expand at the expense of the others (Nowak & Sigmund 1992).

$$\rho_i(t+1)=\frac{\rho_i(t)U_i(t)}{\sum_k \rho_k(t)U_k(t)}$$

the new size of population i is proportional to its share of the total income ($U_i>0$!)

This rule preserves the normalization $\sum_j \rho_j(t)=1.$

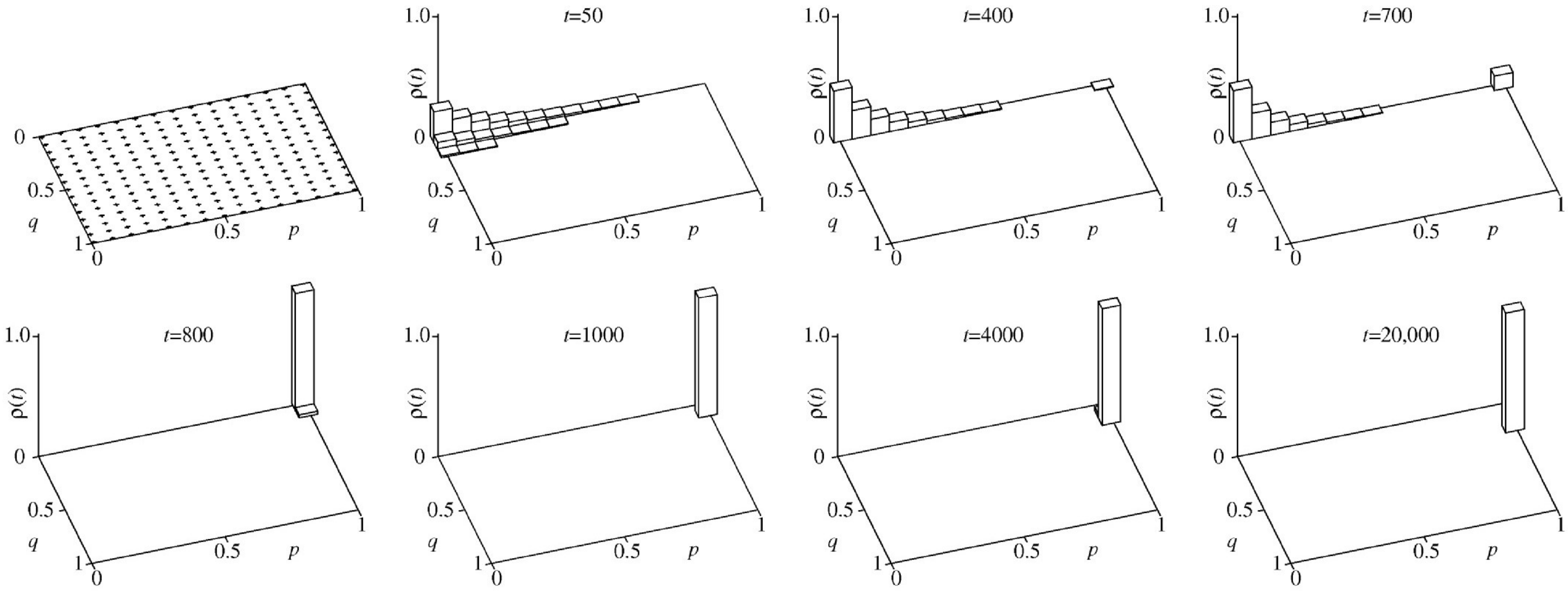
Nowak and Sigmund chose N=100 strategies with random (p_i,q_i) pairs and investigated the model numerically.

Numerical solution

i=1, …, N=15×15=225 (with an equidistantly sampled strategy distribution)

Initially the strategies are present with a frequency of $\rho_i(t=0)=1/N$.

Time-dependence of $\rho_i(t)$ [only if $\rho_i(t)>\rho_i(0)$] movie



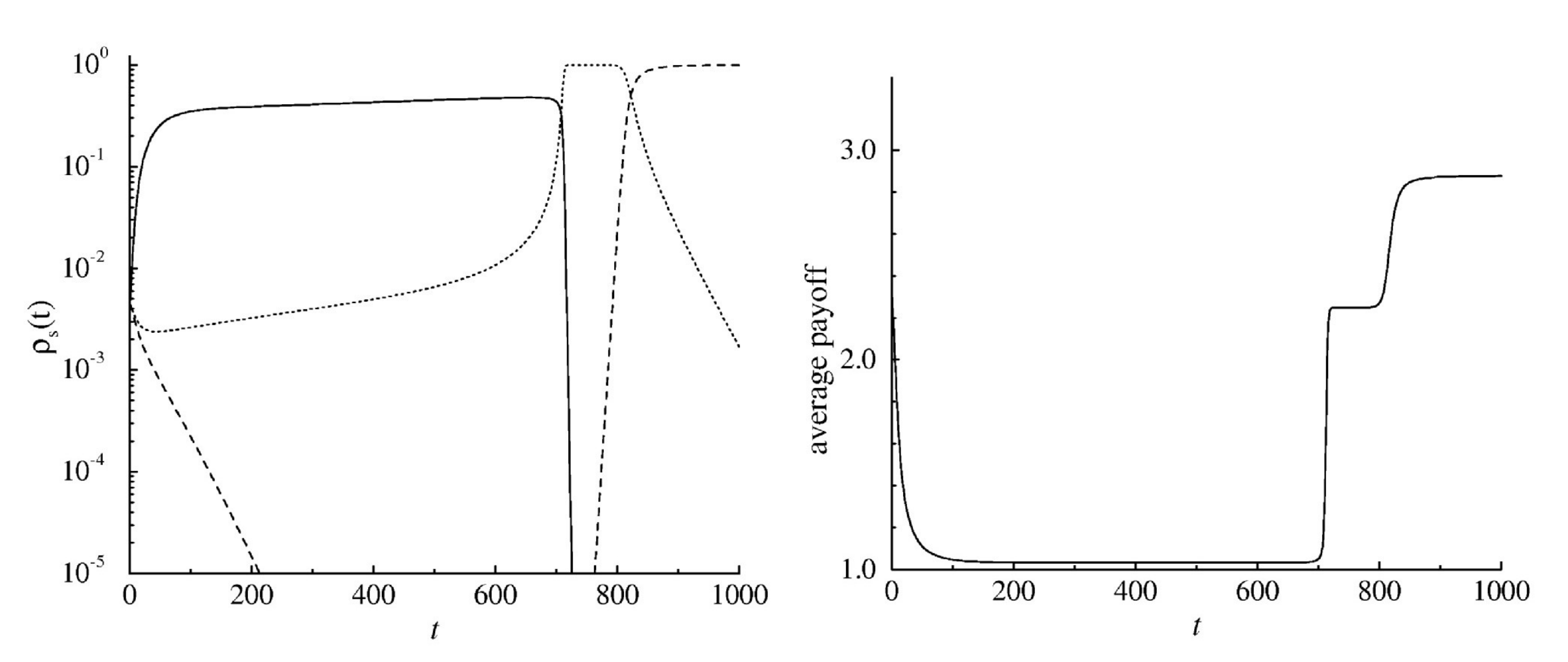
Time-dependence of $\rho_i(t)$ for some relevant strategies:

AllD: (solid line) after its initial success, AllD fails

AllC: (dashed line) becomes extinct due to its exploitation by AllD

TFT: (dotted line) rises after a transient time, when AllD runs out of exploitees

GTFT: variants with higher and higher q (generosity) successively replace each other

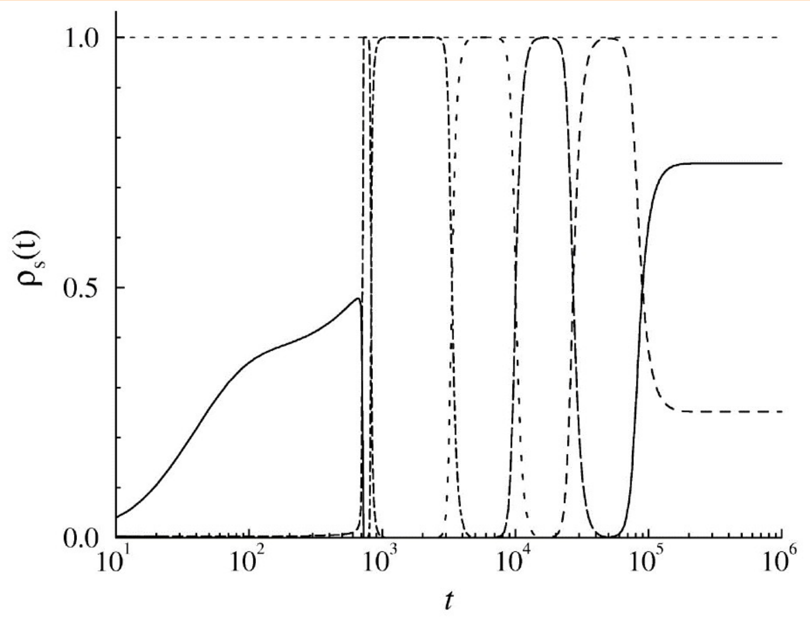


Successive evolution stops at a particular value of q. Why?

Analytical calculations

1.) AllD conquers a homogeneous population of (p,q) strategy within the grey area.

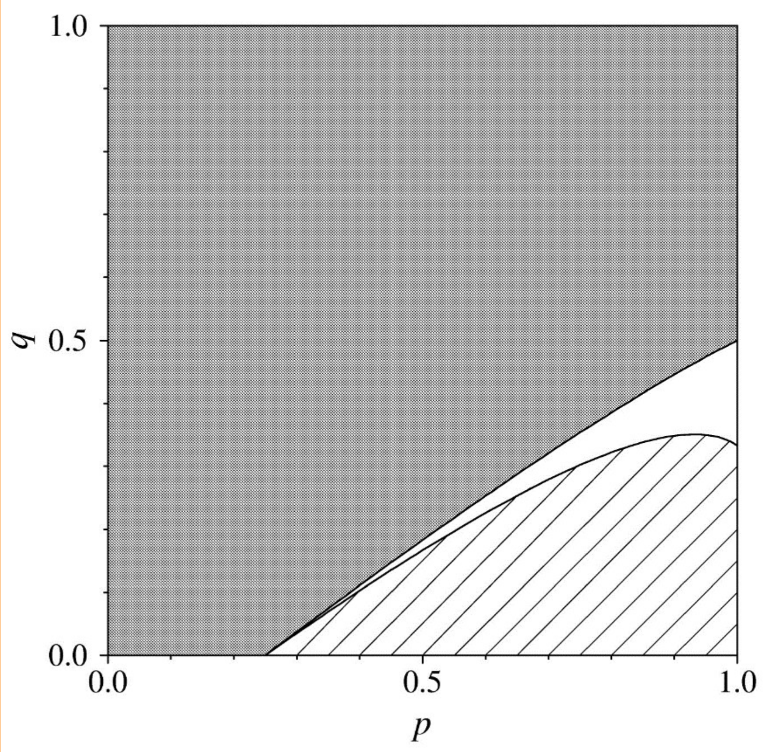
(Home assignment 4.1.)



2.) Assume weak mutations are possible in the form of slow variations from $p \to p+\Delta p$ and $q \to q+\Delta q$ (Δp and $\Delta q \to 0+$) in a homogeneous population if the mutants receive a higher payoff. The direction of this evolution via weak mutation is given by the following partial derivatives:

$$\dot{p} = \left.\frac{\partial U(s,s')}{\partial p}\right|_{s=s'} \quad \text{and} \quad \dot{q} = \left.\frac{\partial U(s,s')}{\partial q}\right|_{s=s'}$$

$\dot{p}, \dot{q} > 0$ denoted by hatching



Optimal value of forgiveness (generosity)

$$\downarrow$$

TFT strategies punish each other in the presence of noise.

The optimal value of q is determined by the $p\to 1$ limit:

(Home assignment 4.2.)

$$\mathrm{TFT}_1(t) = C\,C\,C\,C\,D\,C\,D\cdots$$

$$\mathrm{TFT}_2(t) = C\,C\,C\,C\,C\,D\,C\cdots$$

$$q_{\mathrm{opt}} \cong \min\left(1-\frac{T-R}{R-S}, \frac{R-P}{T-P}\right)$$

The previous dynamical rule drives the system towards p=1 and q_{opt}, where the average payoff is maximal.

The Old Testament favours TFT, while the New Testament favours AllC.

Evolutionary game theory suggests using gTFT with q_{opt}.

The outcome of evolution depends on the payoffs and the available strategy set.

(1, q_{opt}) or the closest available strategy might be inside AllD's domain, which allows AllD to rise again, only to again be replaced by TFT, which is in turn replaced by its more generous variants, and this cycle keeps repeating.

Result with three strategies (AllC, AllD, TFT) for weak noise limit: movie

Further connections to ancient philosophy

- *I Ching (Book of Changes)* (~1000 BC): ”Here the climax of the darkening is reached. The dark power at first held so high a place that it could wound all who were on the side of good and of the light. But in the end it perishes of its own darkness, for evil must itself fall at the very moment when it has wholly overcome the good, and thus consumed the energy to which it owed its duration.”

- The parable of long chopsticks (or spoons): ” In each location, the inhabitants are given access to food, but the utensils are too unwieldy to serve oneself with. In hell, the people cannot cooperate, and consequently starve. In heaven, the diners feed one another across the table and are sated.”

Complex behaviour when varying payoffs and the number of (p,q) strategies

To avoid numerical difficulties, we introduce a weak mutation, that is, after each step strategies that go extinct are allowed to replenish via a weak mutation (with a small probability μ).

Results: different outcomes.

E.g., coexistence of several strategies or TFT fails to repress AllD.

This analysis of these systems is far from complete.

Examples if μ=0.000001:

(T,R,P,S)=(5,3,1,0) Axelrod	5310
(7,3,1,0)	7310
(5,2,1,0)	5210
(6,4,3,0)	6430

Generalization of stochastic reactive strategies

Players (*x* and *y*) take into consideration both their own and their co-player's decision in the *n*th step when deciding what to choose in the upcoming (*n*+1)th step.

In this case, the strategy of player *x* (and *y*) is determined by 4 parameters (if and after the first steps become irrelevant), namely,

$$d_x(p_1, p_2, p_3, p_4), \qquad 0 < p_1, p_2, p_3, p_4 < 1$$

$$d_y(q_1, q_2, q_3, q_4), \qquad 0 < q_1, q_2, q_3, q_4 < 1$$

Accordingly, player *x* chooses

C with prob. p_1, or *D* with prob. $(1-p_1)$, if the previous strategy profile was *CC*

C with prob. p_2, or *D* with prob. $(1-p_2)$, if the previous strategy profile was *CD*

C with prob. p_3, or *D* with prob. $(1-p_3)$, if the previous strategy profile was *DC*

C with prob. p_4, or *D* with prob. $(1-p_4)$, if the previous strategy profile was *DD*

The same rules are used by player *y* with probabilities q_1, q_2, q_3, q_4
(from her own point of view)

If the components of the 4-dimensional vector $\mathbf{v}(n)$ describe the probabilities of the strategy profiles *CC*, *CD*, *DC*, and *DD* in the *n*th step, then $\mathbf{v}(n+1)$ can be described as:

$$\mathbf{v}(n+1) = \mathbf{M} \cdot \mathbf{v}(n)$$

when the players use the extended stochastic reactive strategies d_x and d_y.

In the stationary state:

$$\mathbf{v}(n+1) = \mathbf{v}(n)$$

Remark: Press and Dyson [*PNAS* 109 (2012) 10409] found that, under the right circumstances, the players can both unilaterally set a linear relation between their scores in this game with the help of so-called zero-determinant strategies. This means that the players can set their opponent's score or demand an extortionate share of the payoffs, turning the iterated prisoner's dilemma into a sort of ultimatum game. TFT is a fair ”extortionate” zero-determinant strategy.

Home assignments

4.1. Determine the (p,q) strategies whose homogeneous population can be occupied by AllD! (The grey area on slide 8.)

4.2. Determine the boundary in the (p,q) strategy space where the evolution via weak mutations stops, that is, where the formulae on page 8 predict

$$\dot{p} = \left.\frac{\partial U(s,s')}{\partial p}\right|_{s=s'} = 0 \quad \text{and} \quad \dot{q} = \left.\frac{\partial U(s,s')}{\partial q}\right|_{s=s'} = 0 \quad !$$

Plot the results for the payoff parameters used by Axelrod (T=5, R=3, P=1, S=0)!

What is the optimal value of forgiveness (q_{opt}) in the limit $p \to 1$?

4.3. Determine the matrix **M** from the previous slide and evaluate the stationary solution!