An introduction to evolutionary game theory

Decisions, Games, and Evolution
Bangalore 2025

Christian Hilbe
Interdisciplinary Transformation University IT:U
Linz, Austria

An informal introduction: A few examples

The rock–paper–scissors game and the evolution of alternative male strategies

B. Sinervo & C. M. Lively

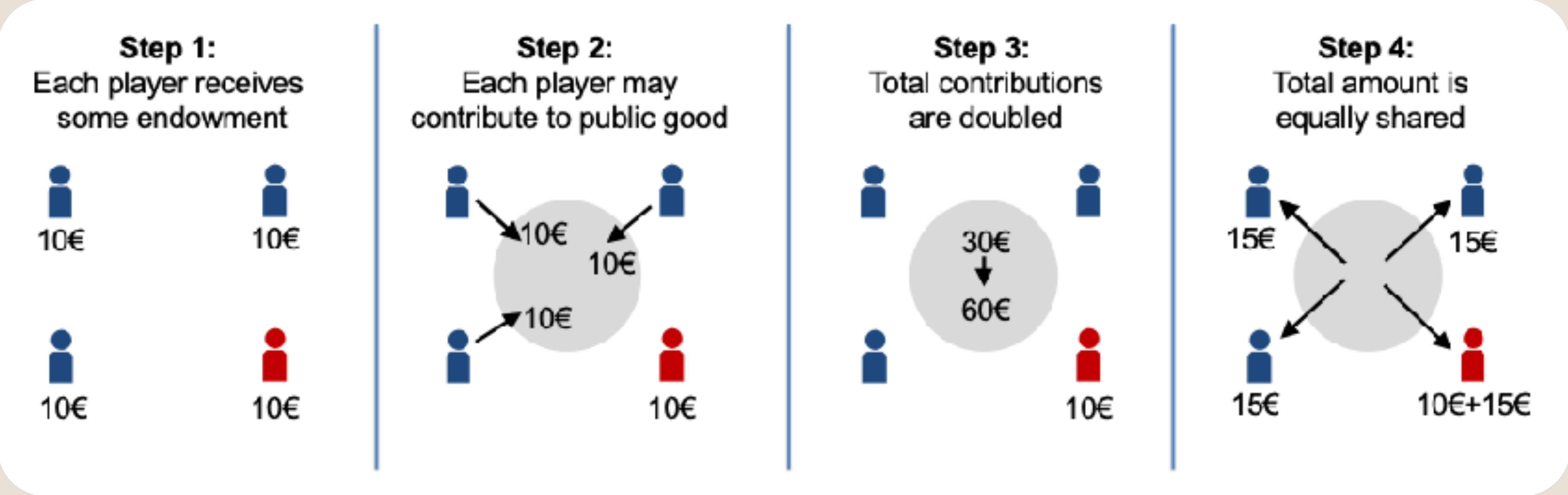
Department of Biology and Center for the Integrative Study of Animal Behavior, Indiana University, Bloomington, Indiana 47405, USA

Example 1: Mating behavior among male lizards (Sinervo & Lively 1996)

- Among side-blotched lizards (*Uta Stansburiana*), there are three male morphs. One can distinguish them by their throat color: yellow, blue, orange.
- These three morphs also differ in their mating behavior
  - Males with orange throats are very aggressive and defend large territories
  - Males with blue throats are less aggressive and defend smaller territories
  - Males with yellow throats resemble female lizards; they are sneakers and do not defend any territory

Question: How can we make sense of this coexistence of different morphs?

An informal introduction: A few examples

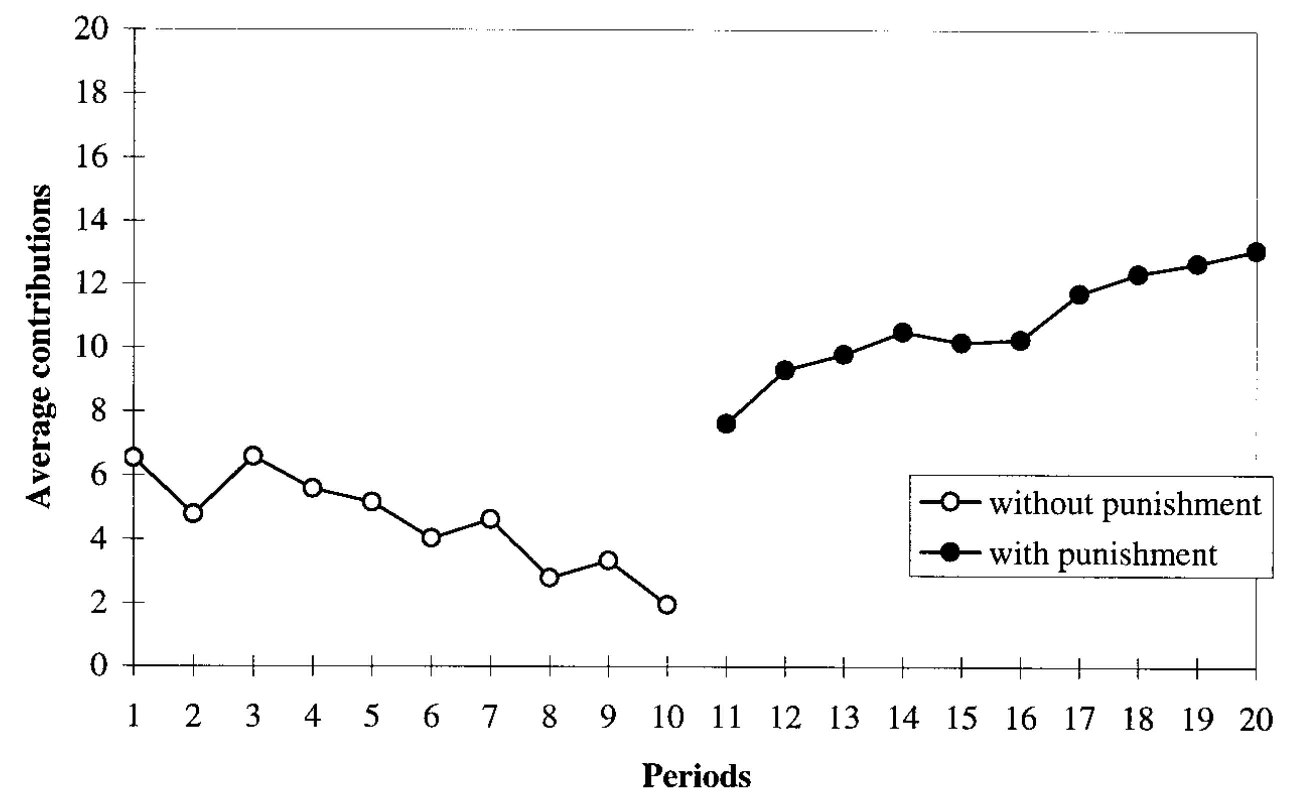


Example 2: Cooperation and punishment among humans (Fehr & Gächter 2000)

- Groups of four individuals interact with each other in a public good game
- Afterwards, participants can punish each other (paying 1 unit to reduce other player's payoff by 3 units).
- Experimental results:
  - Without punishment, cooperation goes down
  - With punishment, cooperation goes up even if participants never meet again (at least in most Western countries)

Cooperation and Punishment in Public Goods Experiments

By Ernst Fehr and Simon Gächter*



Question: How can we make sense of these behaviors?

An informal introduction: A few examples

Example 3: Signalling (Spence 1973, Zahavi 1973)

- Individuals sometimes invest in something without getting a direct return
- Two examples:
  - Getting an academic degree without using the respective skills in your job
  - Peacock’s tail

Question: It has been suggested that these investments can be worthwhile when they act as (costly) signals. But how exactly do such signals work?

JOB MARKET SIGNALING *

MICHAEL SPENCE

Mate Selection—A Selection for a Handicap

AMOTZ ZAHAVI

An informal introduction: A few examples

What do these examples have in common?

- There are individuals who can exhibit different behaviors / traits
- An individual's optimal behavior / trait depends on what the other individuals do

The appropriate framework to analyze these examples is **game theory**.

Interesting observation

Not in all examples the respective behaviors and traits are consciously chosen.

An overview

Today's class (March 11, 2025)

- An introduction to evolutionary game theory
  (Replicator dynamics, games in finite populations)

Tomorrow's classes (March 12, 2025)

- Evolution of cooperation & direct reciprocity
- Social norms & indirect reciprocity

Thursday's class (March 13, 2025)

- Some current research: Reciprocity in complex environments

A short reminder: Some (classical) game theory

Definition: Normal-form game

A game is referred to as a **normal-form game** if

- It only involves two players (player 1 & player 2)
- Each player can only choose among finitely many actions (player 1: *m* actions, player 2: *n* actions)
- Players move simultaneously and they know all aspects of the game, except for the co-player's decision

Such games can be represented by a (bi)-matrix

	Action 1	…	Action *n*
Action 1	a_{11}, b_{11}	…	a_{1n}, b_{1n}
…	…	…	
Action *m*	a_{m1}, b_{m1}	…	a_{mn}, b_{mn}

$A=(a_{ij})$ and $B=(b_{ij})$ are the payoff matrices of the two players.

Definition: Dominated strategies

A pure strategy $\mathbf{e}_i$ for player 1 is called (strictly) dominated if there is a (possibly mixed) strategy $\mathbf{x}$ for player 1 that yields a better payoff, irrespective of the co-player's strategy $\mathbf{e}_j$, $\pi_1(\mathbf{e}_i, \mathbf{e}_j) < \pi_1(\mathbf{x}, \mathbf{e}_j)$ for all $\mathbf{e}_j$.

Definition: Nash equilibrium

A strategy profile $(\mathbf{x}^*, \mathbf{y}^*)$ is called a Nash equilibrium if the following two conditions hold:

$$\pi_1(\mathbf{x}, \mathbf{y}^*) \leq \pi_1(\mathbf{x}^*, \mathbf{y}^*) \quad \text{for all } \mathbf{x} \in S_m . \qquad (1.13.1)$$

$$\pi_2(\mathbf{x}^*, \mathbf{y}) \leq \pi_2(\mathbf{x}^*, \mathbf{y}^*) \quad \text{for all } \mathbf{y} \in S_n . \qquad (1.13.2)$$

Dominance solvability and the Nash equilibrium concept appear to make strong assumptions on cognitive abilities. In the following, we explore an approach to game theory that avoids these assumptions.

Evolutionary game theory: An example

Example 1.1: A model of animal conflict (Hawk-Dove)

- Conflicts among animals are usually of "limited war" type; serious injuries are rare.
- One example is ritual fighting in deer, that follows well-defined rules. The existence of such rules is surprising. Males who ignore these rules and kill their competitors should have an evolutionary advantage.
- Suppose individuals can choose whether or not to escalate a fight. We consider two possible types of individuals: those who escalate ("Hawks") and those who stick to the ritual, and who escape if the opponent escalates ("Doves").
- Suppose winning the contest leads to a benefit of b, whereas serious injuries lead to a cost of $c > b$. Let x denote the fraction of hawks in a population.

- Expected fitness of the two types:
  $f_H = \frac{b-c}{2}x + b(1-x)$ and $f_D = 0 \cdot x + \frac{b}{2}(1-x)$
- Fitness difference:
  $\Delta f := f_H - f_D = \frac{1}{2}(b - cx)$
- In particular, if hawks are rare ($x < b/c$), we would expect them to spread in the population. Conversely, if hawks are common ($x > b/c$), we would rather expect doves to spread.

- The larger the cost of serious injuries, the more doves we would expect.

Evolutionary game theory: An example

Example 1.2: Hawk-Dove as a classical game

- We could have also interpreted this interaction as a classical game with payoff matrix

	Hawk	Dove
Hawk	*(b-c)/2*, *(b-c)/2*	*b*, *0*
Dove	*0*, *b*	*b/2*, *b/2*

- Note that this bi-matrix (A,B) is symmetric, meaning that $A=B^T$. For symmetric games it is common to only depict the first player's payoff.

	Hawk	Dove
Hawk	*(b-c)/2*	*b*
Dove	*0*	*b/2*

- The game has exactly one symmetric Nash equilibrium $(\boldsymbol{x},\boldsymbol{y})$ with $\boldsymbol{x}=(x_H, x_D)$ and $\boldsymbol{y}=(y_H, y_D)$. In this equilibrium, $x_H = y_H = b/c$.

Remark 1.3: Introducing matrix games for populations

- Consider an infinitely large population
- Individuals in that population can have one of n different traits ("strategies"). Let $\boldsymbol{x} = (x_1, \dots, x_n)^T$ describe the trait distribution in the population.
- When an individual with trait i encounters an individual with trait j, let a_{ij} denote the fitness consequence for individual i. Let $A = (a_{ij})$ be the corresponding matrix.
- If interactions occur randomly, the expected fitness of an individual with trait i is

$$f_i = \sum_{j=1}^{n} a_{ij} x_j = (A\mathbf{x})_i$$

- Similarly, the population's average fitness is

$$\bar{f} = \sum_{i=1}^{n} x_i f_i = \mathbf{x}^T A \mathbf{x}$$

Evolutionary game theory: Replicator equation

Definition 1.4: Replicator equation / Replicator dynamics

The replicator equation is the system of ordinary differential equations

$$\dot{x}_i = x_i\big(f_i(\mathbf{x}) - \bar{f}(\mathbf{x})\big)\,.$$

Proposition 1.5: Properties of replicator dynamics

1. The unit simplex S_n is invariant under replicator dynamics:
$$\sum_{i=1}^{n} x_i(0) = 1 \quad \Rightarrow \quad \sum_{i=1}^{n} x_i(t) = 1 \ \forall t$$
2. All boundary faces of S_n are also invariant:
$$x_i(0) = 0 \quad \Rightarrow \quad x_i(t) = 0 \ \forall t$$
3. Dominated traits go extinct:
$$f_i(\mathbf{x}) < f_j(\mathbf{x}) \ \forall \mathbf{x} \text{ and } \mathbf{x}(0) \in \mathrm{int}(S_n) \ \Rightarrow \ \lim_{t\to\infty} x_i(t) = 0$$

Proof sketch.

1. $$\sum_{i=1}^{n} \dot{x}_i = \sum_{i}^{n} x_i\big(f_i(\mathbf{x}) - \bar{f}(\mathbf{x})\big) = \sum_{i}^{n} x_i f_i(\mathbf{x}) - \sum_{i}^{n} x_i \bar{f}(\mathbf{x})$$
$$= \bar{f}(\mathbf{x}) - \bar{f}(\mathbf{x}) = 0$$

3. Consider the fraction x_i/x_j.
$$\dot{\left(\frac{x_i}{x_j}\right)} = \frac{\dot{x}_i x_j - x_i \dot{x}_j}{x_j^2}$$
$$= \frac{x_i x_j\big(f_i(\mathbf{x}) - \bar{f}(\mathbf{x})\big) - x_i x_j\big(f_j(\mathbf{x}) - \bar{f}(\mathbf{x})\big)}{x_j^2}$$
$$= \frac{x_i}{x_j}\big(f_i(\mathbf{x}) - f_j(\mathbf{x})\big) < -\delta\left(\frac{x_i}{x_j}\right)$$

Therefore, the fraction x_i/x_j decreases exponentially.

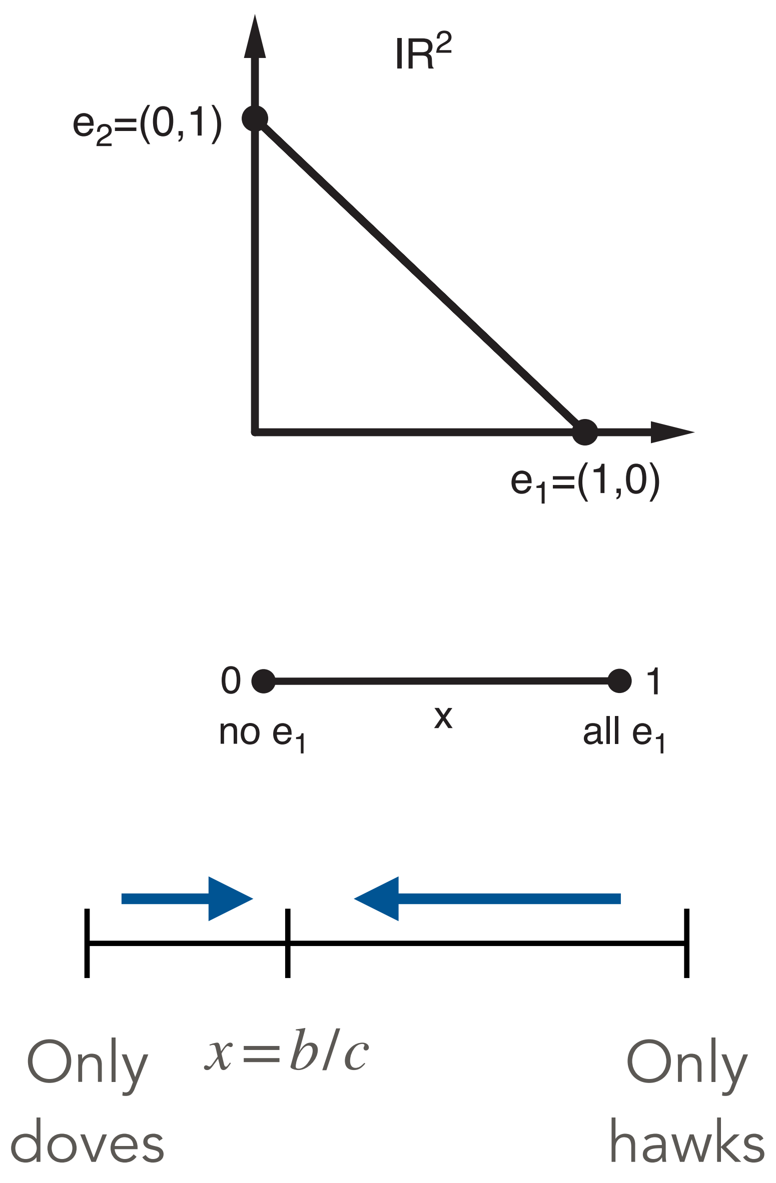
Evolutionary game theory: Classification of 2x2 games

Remark 1.6: On representing the unit simplex

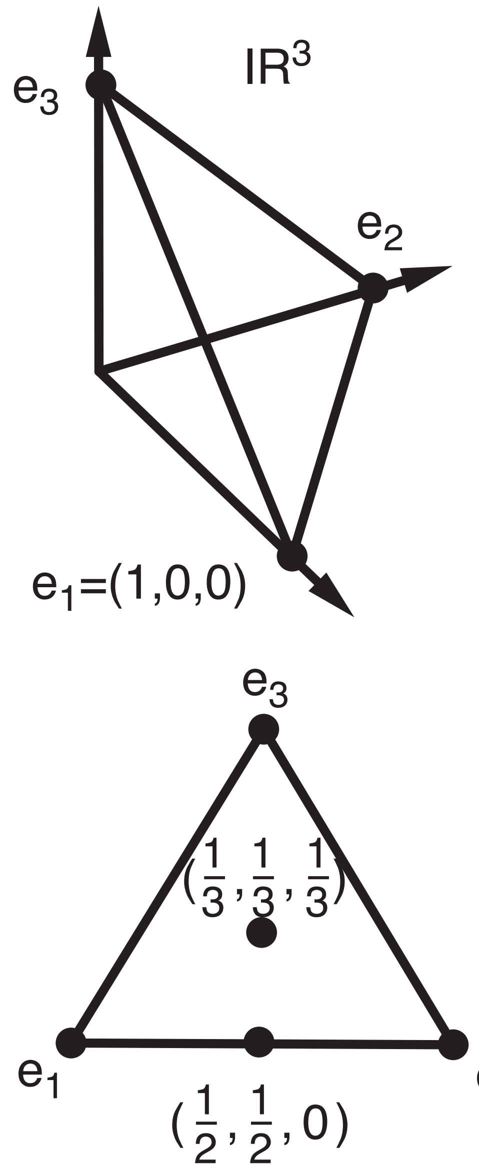
Consider the replicator equation $\dot{x}_i = x_i\big(f_i(\mathbf{x}) - \bar{f}(\mathbf{x})\big)$.
For a game with n strategies in total, this is, in principle, an *n*-dimensional system. However, we are only interested in those orbits on the unit simplex:

$$S_n = \left\{ \mathbf{z} \in \mathbb{R}^n : \; z_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{n} z_i = 1 \right\}$$

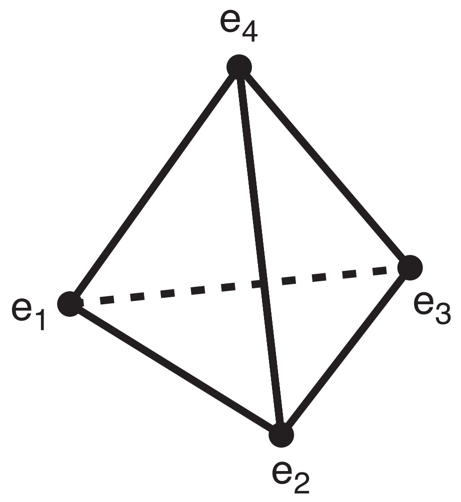
n = 2



n = 3



n = 4



Remark 1.7: A classification of 2x2 games

To get some intuition, let us analyze the simplest non-trivial case: a symmetric game with two strategies:

	Action 1	Action 2
Action 1	*a*	*b*
Action 2	*c*	*d*

We can represent the replicator equation as a 1-dim. system. Let x be the proportion of individuals who use Action 1, and 1-x is the proportion of individuals who use Action 2.

The fitnesses are

$$f_1(x) = ax + b(1-x) \quad \text{and} \quad f_2(x) = cx + d(1-x)$$

Replicator dynamics takes the form:

$$\begin{aligned} \dot{x} &= x\big(f_1(x) - \bar{f}(x)\big) = x\big(f_1(x) - xf_1(x) - (1-x)f_2(x)\big) \\ &= x(1-x)\big(f_1(x) - f_2(x)\big) = x(1-x)\Big((b-d) + (a-b-c+d)x\Big) \end{aligned}$$

Fixed points: (1) Corners: $x = 0, x = 1$

(2) Interior: $x = \dfrac{d-b}{a-b-c+d}$, if $x \in (0,1)$

Evolutionary game theory: Classification of 2x2 games

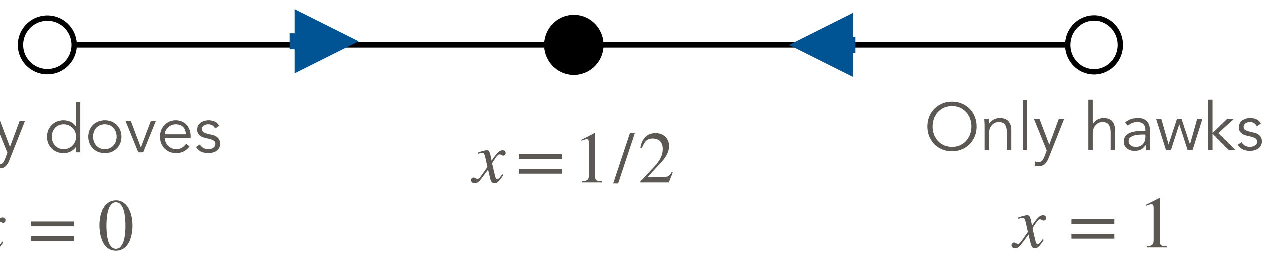
Examples 1.8: Some 2x2 games

1. The hawk-dove game (with b=2, c=4)

	Hawk	Dove
Hawk	*-1*	*2*
Dove	*0*	*1*

Stable coexistence

$$\dot{x} = x(1-x)(1-2x)$$

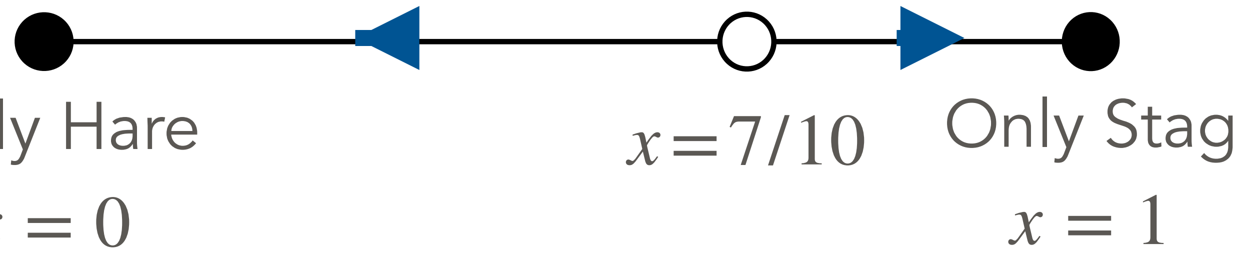


2. Stag-hunt game (coordination game)

	Stag	Hare
Stag	*10*	*0*
Hare	*7*	*7*

Bistability

$$\dot{x} = x(1-x)(10x-7)$$

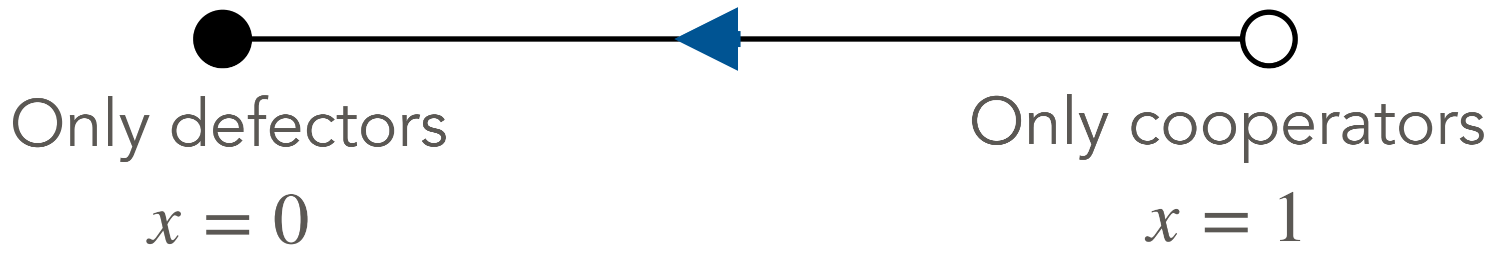


3. Prisoner's dilemma

	Cooperate	Defect
Cooperate	*2*	*-1*
Defect	*3*	*0*

Dominance

$$\dot{x} = x(1-x)(-1)$$

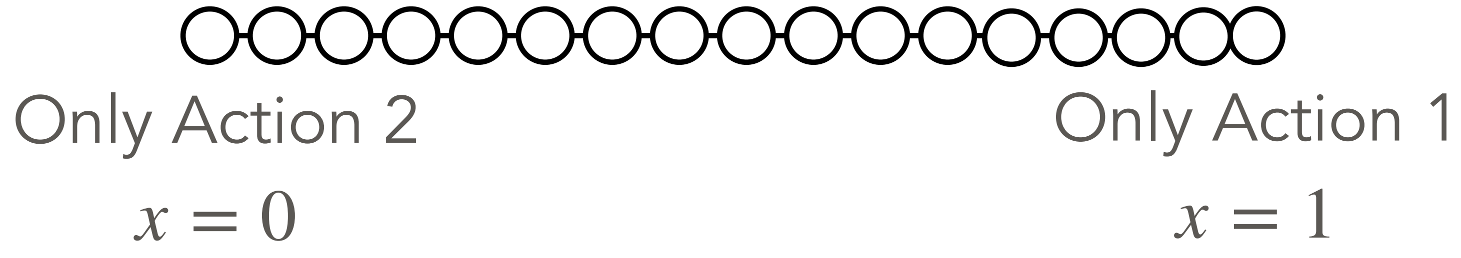


4. A trivial game

	Action 1	Action 2
Action 1	*3*	*1*
Action 2	*3*	*1*

Neutrality

$$\dot{x} = x(1-x) \cdot 0$$



Qualitatively, these are all possible cases.

Evolutionary game theory: An example of a 3x3 game

Example 1.9. A 3x3 game: The volunteer's timing dilemma

Consider the following variant of a so-called volunteer's dilemma. There are two players; at least one of them should volunteer to do a task that benefits both of them.

However, we assume that players can either volunteer early or volunteer late. If someone volunteers early, this creates a high benefit of 5. If no one volunteers early, but someone volunteers late, this creates a smaller benefit of 4. Volunteering always comes with a cost of 3.

We distinguish three strategies. A cooperator volunteers early. A wait & see player volunteers late, unless the co-player has already volunteered early. A defector never volunteers.

	Cooperation	Wait&See	Defection
Cooperation	*2*	*2*	*2*
Wait&See	*5*	*1*	*1*
Defection	*5*	*4*	*0*

Replicator dynamics, $\mathbf{x} = (x_C, x_W, x_D)$

1. **Dynamics at the edges:**
   - No defectors ($x_D = 0$): Coexistence among cooperators and wait & see, $\mathbf{x}^*_{CW} = (1/4, 3/4, 0)$
   - No wait&see ($x_W = 0$): Coexistence among cooperators and defectors, $\mathbf{x}^*_{CD} = (2/5, 0, 3/5)$
   - No cooperators ($x_C = 0$): Coexistence among wait&see and defectors, $\mathbf{x}^*_{WD} = (0, 1/4, 3/4)$

2. **Fixed point in the interior:**

   If $\dot{x}_i = x_i\big(f_i(\mathbf{x}) - \bar{f}(\mathbf{x})\big)$ has a fixed point with $x_i > 0\ \forall i$, it must hold that $f_i(\mathbf{x}) = \bar{f}(\mathbf{x})\ \forall i$.
   Equivalently, it must hold that $f_1(\mathbf{x}) = f_2(\mathbf{x}) = f_3(\mathbf{x})$.
   This is a simple linear system (and either has 0, 1, or infinitely many solutions).
   In our case, solution: $\mathbf{x}^*_{\text{int}} = (1/4, 3/16, 9/16)$.

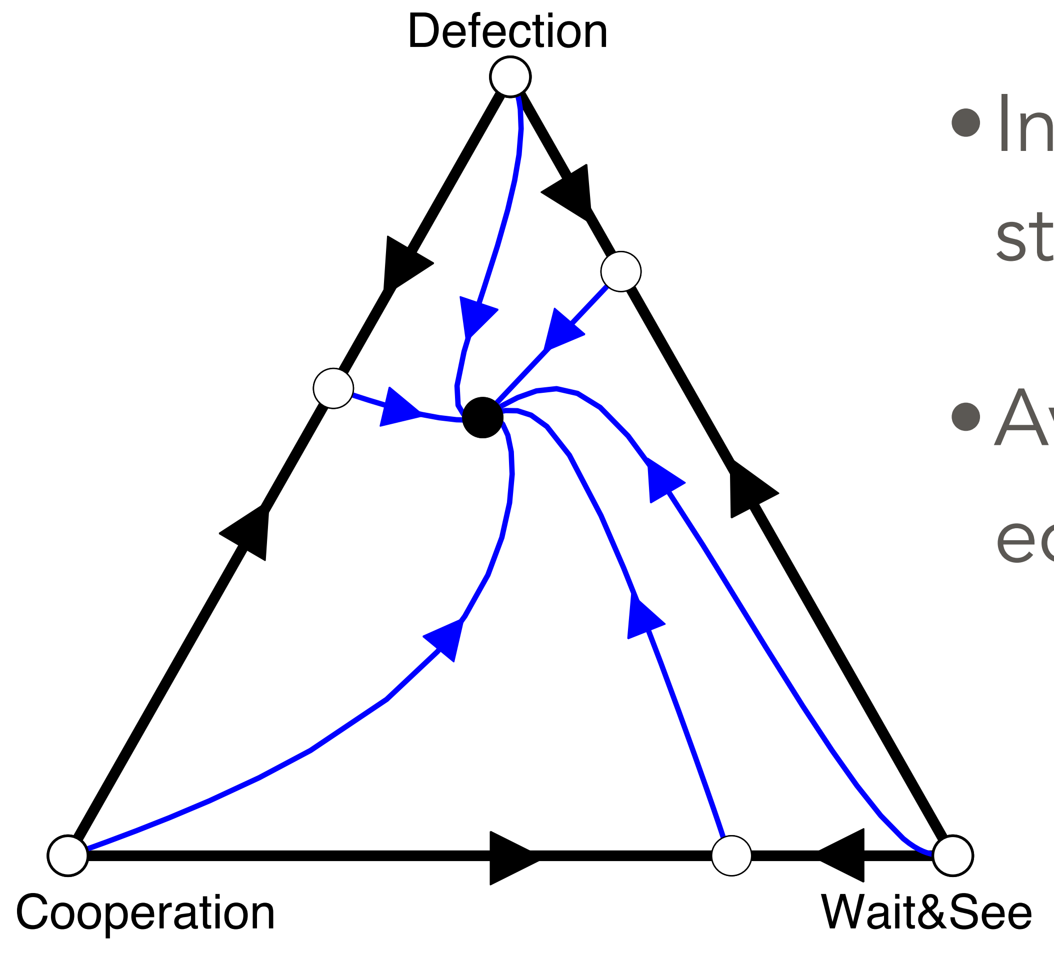
Evolutionary game theory: An example of a 3x3 game

Example 1.9. The volunteer's timing dilemma (continued)

3. Local stability analysis for the fixed points

- For the equilibria on the edges, in each case it is true that the missing strategy can invade when rare.
- The interior equilibrium is stable.

4. Plotting some orbits numerically



- In the end, all three strategies coexist.
- Average fitness in this equilibrium: $\bar{f} = 2$

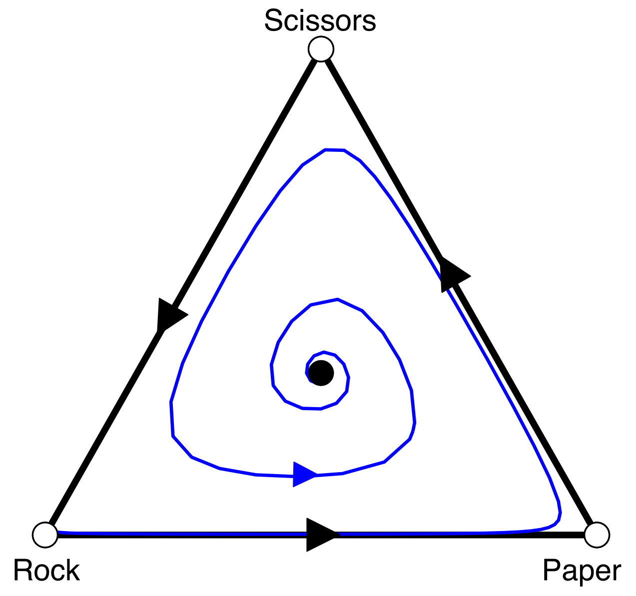
Example 1.10. Rock Paper Scissors

Consider the following generalised version of rock paper scissors.

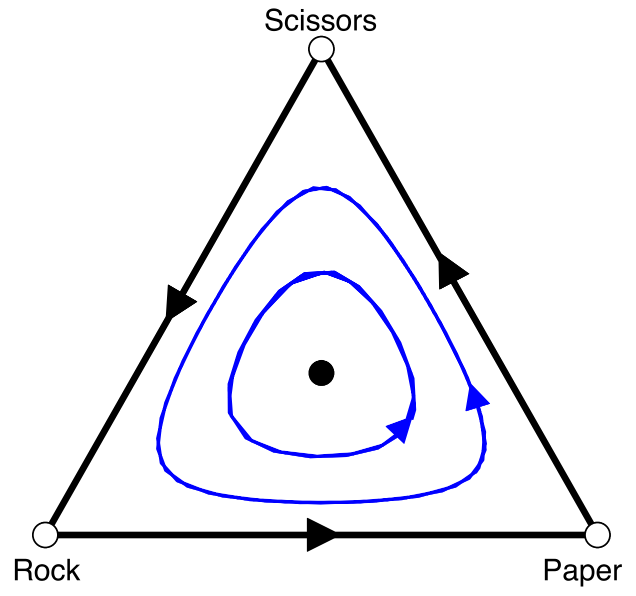
	Rock	Paper	Scissors
Rock	0	$-a_2$	b_3
Paper	b_1	0	$-a_3$
Scissors	$-a_1$	b_2	0

It turns out there are three possible dynamics.

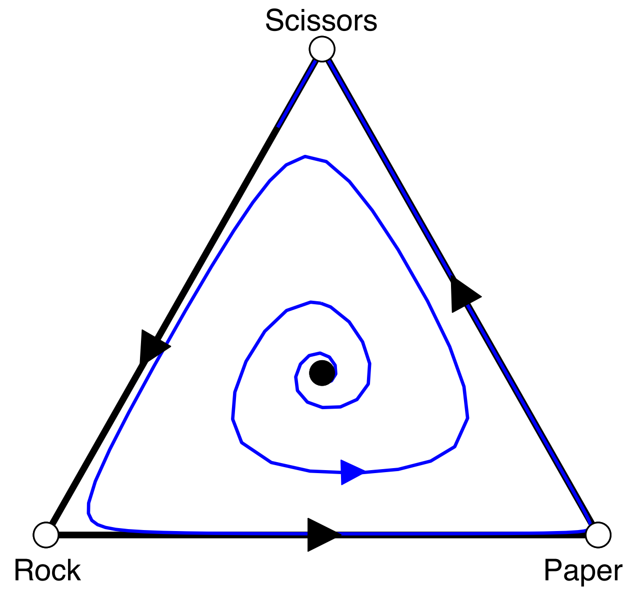
$a_1a_2a_3 < b_1b_2b_3$



$a_1a_2a_3 = b_1b_2b_3$



$a_1a_2a_3 > b_1b_2b_3$



Evolutionary game theory: Non-transitive game in nature

Example 1.11. Non-transitive games in nature

1. Mating behavior in lizards (Sinervo & Lively 1996)

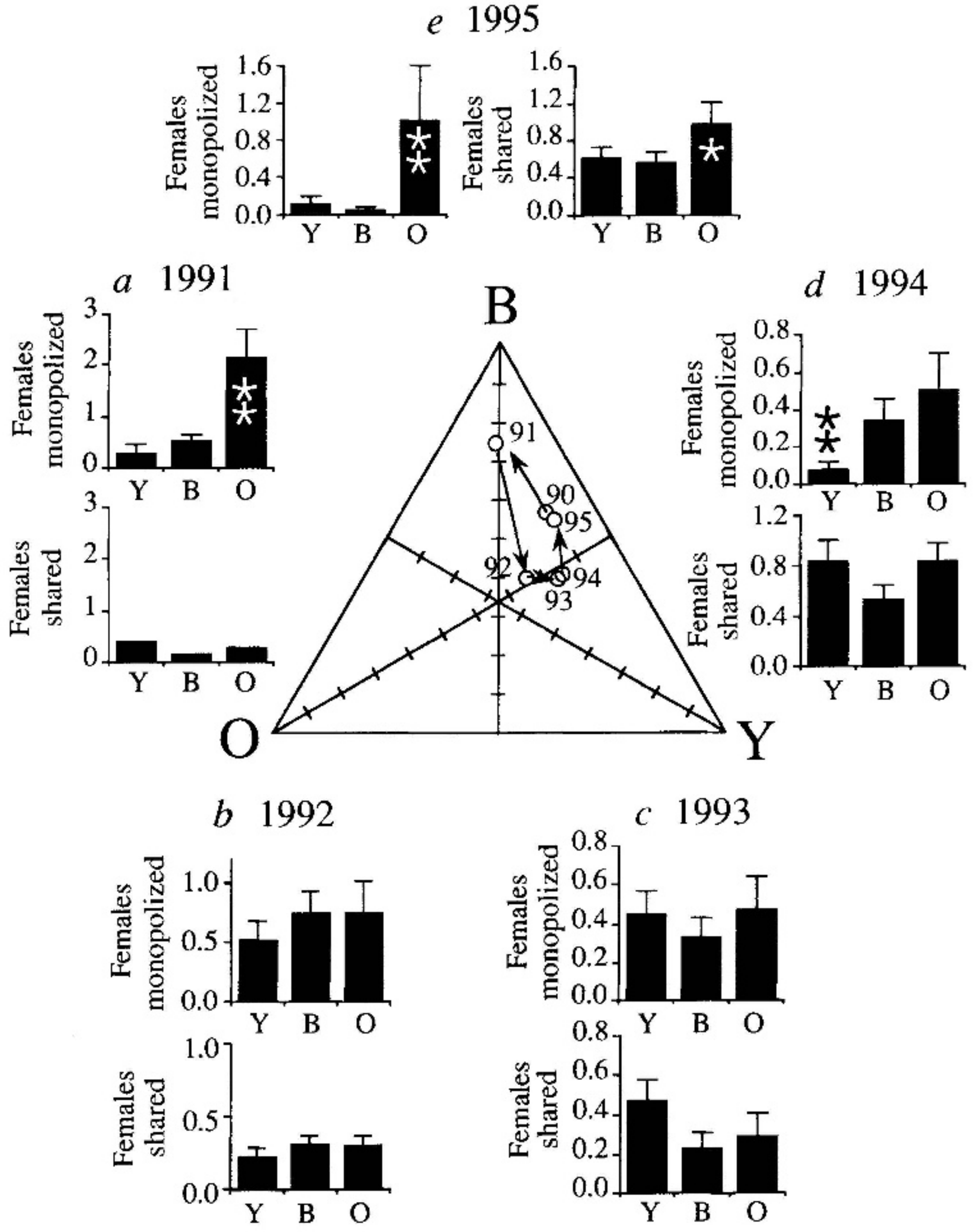
Three male morphs in side-blotched lizards:

- Males with orange throats defend large territories
- Males with blue throats defend smaller territories
- Males with yellow throats are sneakers without territory

The rock–paper–scissors game and the evolution of alternative male strategies

B. Sinervo & C. M. Lively

Department of Biology and Center for the Integrative Study of Animal Behavior, Indiana University, Bloomington, Indiana 47405, USA



Evolutionary game theory: Non-transitive game in nature

Example 1.11. Non-transitive games in nature (continued)

2. Competition among E. Coli (Kerr et al, 2002)

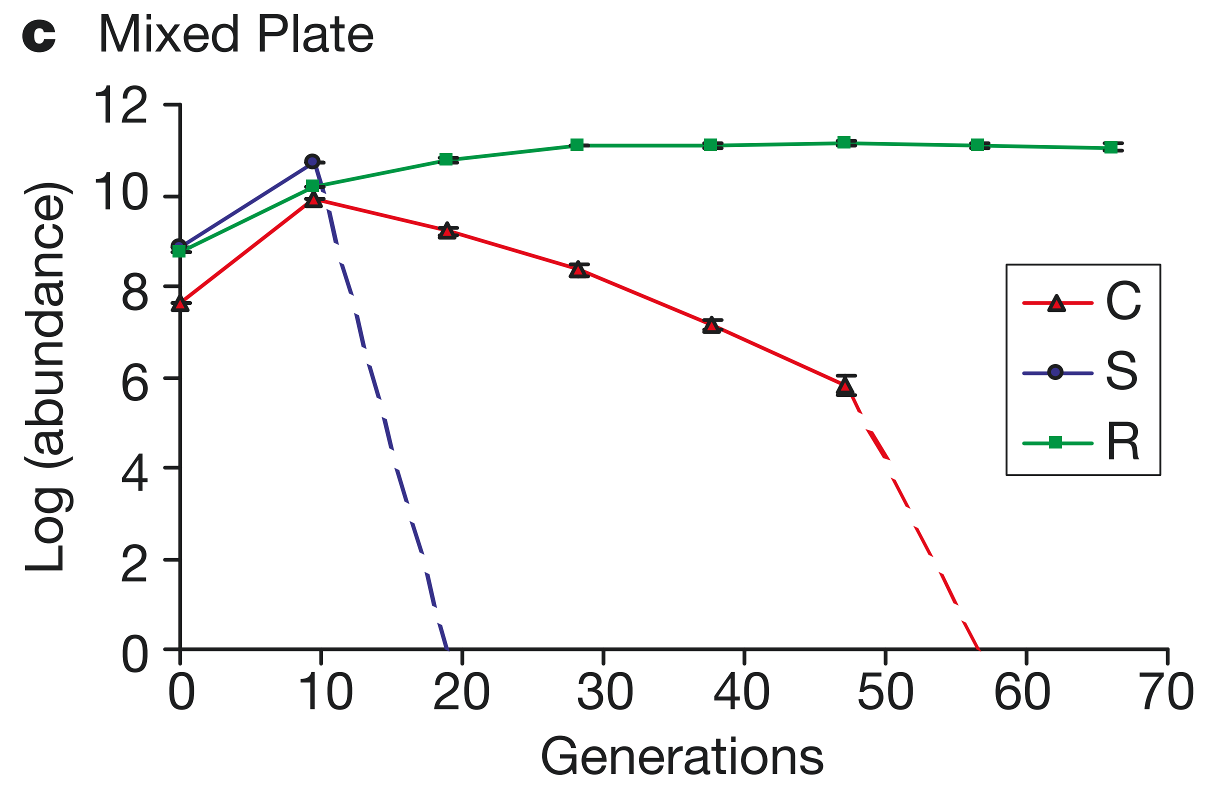
Three strains of E. Coli

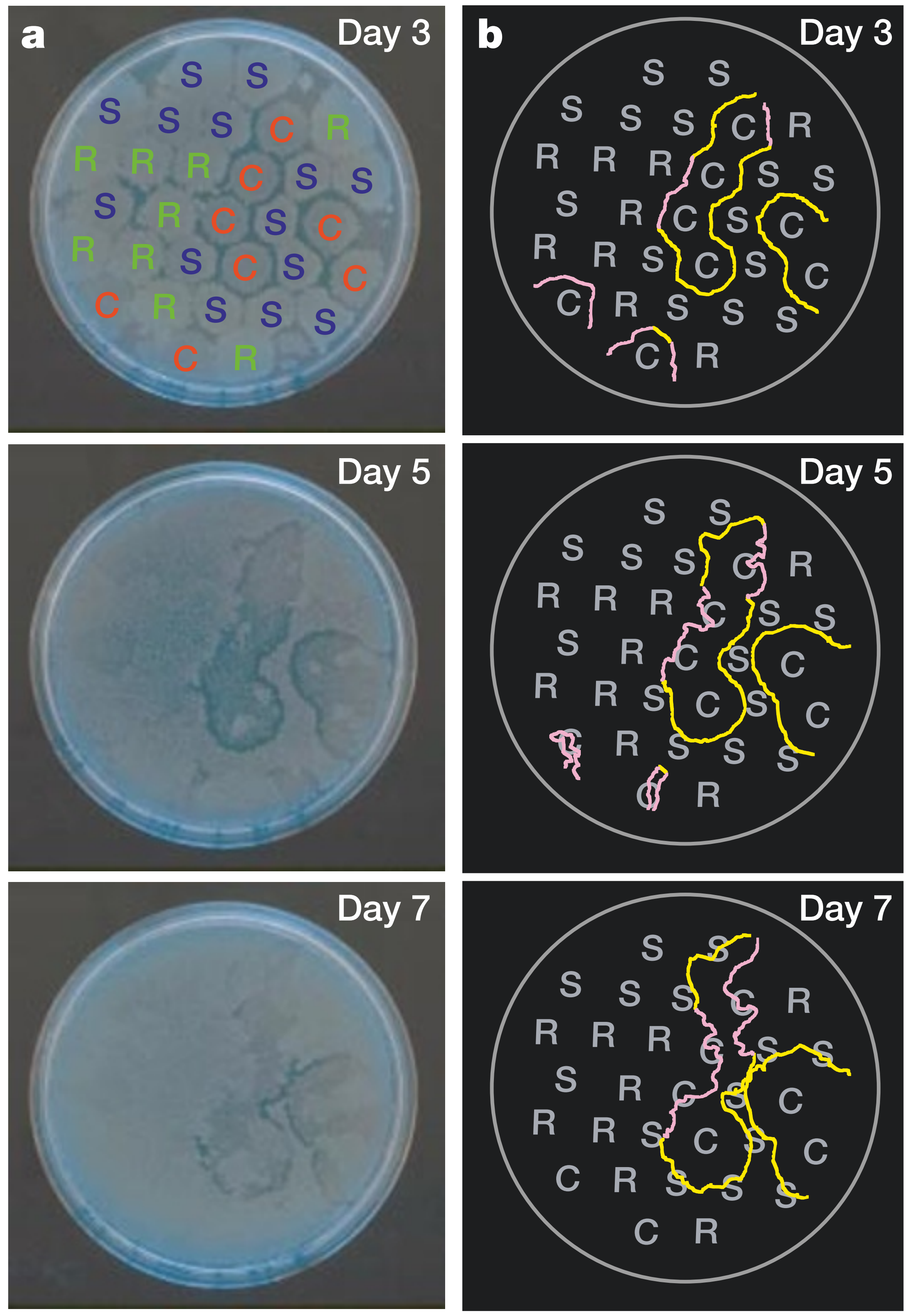
- Colicin-producing strain (C)
- Sensitive strain (S)
- Resistant strain (R)

S is invaded by C, C is invaded by R, R is invaded by S.

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*





Evolutionary game theory: Replicator dynamics versus classical game theory

Remark 1.12. Replicator dynamics vs Nash equilibrium

In the examples we have seen so far, the outcome "predicted" by replicator dynamics often had a close relationship to the (symmetric) Nash equilibria of the game. This is not a coincidence; instead one can show the following results.

1. If $\mathbf{x}^*$ is a symmetric Nash equilibrium of the symmetric game with payoff matrix A, then $\mathbf{x}^*$ is a fixed point of the replicator equation.
2. If $\mathbf{x}^*$ is the ω-limit of an orbit $\mathbf{x}(t)$ in $\mathrm{int}(S_n)$, then it is a Nash equilibrium.
3. If $\mathbf{x}^*$ is Lyapunov stable, then it is a Nash equilibrium.

These results are sometimes referred to as the "Folk theorem of evolutionary game theory". In this way, evolutionary dynamics has also become important for economics ("equilibrium selection")

Remark 1.13. On the status of the replicator equation

Today, the replicator equation is one of the standard models of evolutionary game theory, for various reasons:

1. Historically, it is (one of) the first dynamics that has been considered for biological games (Taylor and Jonker, 1978).
2. The equation is comparably simple and well-understood.
3. There is a beautiful connection to the Lotka-Volterra equation.
4. There are beautiful connections to the concepts of classical game theory, without making any strong assumptions on the rationality of individuals.
(How is this possible?)

Evolutionary game theory: Replicator dynamics versus classical game theory

Remark 2.14. Beyond replicator dynamics

Replicator dynamics might be both considered as a model of biological evolution, or of cultural evolution (imitation). However, it is also important to stress that replicator dynamics is one out of many evolutionary dynamics to consider. The optimal model depends on the applications one has in mind.

For example, one property of replicator dynamics is that strategies that are absent $(x_i = 0)$ are not introduced by the evolutionary process (i.e., replicator-dynamics is *non-innovative*). One could interpret replicator dynamics as a model of selection without mutations.

One might want to consider other models, for example, if one is interested in games in finite populations (e.g., Nowak et al, Nature 2004), games in structured populations (e.g., Ohtsuki et al, Nature 2006), or games with continuous traits (e.g., Geritz et al, Evol Ecol Res 1998).

To provide some intuition for how other models look like, I briefly discuss in the following the case of finite populations.

Evolutionary game theory: Games in finite populations

Remark 2.15. Basic Setup of the Moran process

- We consider a population of finite size N.
- Individuals are engaging in a symmetric game with two possible strategies, represented by the payoff matrix

	Strategy 1	Strategy 2
Strategy 1	a	b
Strategy 2	c	d

- Everyone is equally likely to interact. As a result, if there are i individuals with strategy 1, the players' expected payoffs are given by:

$$\pi_1(i) = \frac{i-1}{N-1}a + \frac{N-i}{N-1}b$$

$$\pi_2(i) = \frac{i}{N-1}c + \frac{N-i-1}{N-1}d\,.$$

- Payoffs are mapped into fitnesses by the map

$$f_1(i) = 1 - w + w\pi_1(i) \quad \text{and} \quad f_2(i) = 1 - w + w\pi_2(i)$$

  The parameter $w \geq 0$ is called the strength of selection. It measures the importance of the game for an individual's fitness.

- Evolution occurs in discrete time steps. In each time step, one individual is randomly drawn (proportional to fitness) to reproduce. To keep the population size constant, we also randomly choose an individual to die (independent of fitness).
- This assumption leads to the following transition probabilities in a single time step:

$$T_i^+ = \frac{if_1(i)}{if_1(i) + (N-i)f_2(i)} \cdot \frac{N-i}{N}$$

$$T_i^- = \frac{(N-i)f_2(i)}{if_1(i) + (N-i)f_2(i)} \cdot \frac{i}{N}$$

Evolutionary game theory: Games in finite populations

Remark 2.16. Computing a strategy's fixation probability

- If we rule out mutations, one of the two strategies will eventually reach fixation. One important question is: given that only one individual currently plays strategy 1, what is the probability that eventually everyone will adopt it?
- Let φ_i be the probability that eventually everyone adopts strategy 1, given that there are currently i players with this strategy.
- We can derive a recursion for φ_i

  $$\varphi_i = T_i^+ \varphi_{i+1} + T_i^- \varphi_{i-1} + (1 - T_i^+ - T_i^-)\varphi_i \,.$$

  with $\varphi_0 = 0$ and $\varphi_N = 1$.
- Setting $y_i := \varphi_i - \varphi_{i-1}$, we can write this as

  $$0 = T_i^+ y_{i+1} - T_i^- y_i, \quad \text{or} \quad y_{i+1} = \frac{T^-(i)}{T^+(i)} y_i$$

  with $y_1 = \varphi_1$.

- In particular,

  $$y_i = \prod_{k=1}^{i-1} \frac{T^-(k)}{T^+(k)} \varphi_1 \qquad (2.16.1)$$

- Now, we use two different methods to sum up over all y_i. By its definition, we have

  $$\sum_{i=1}^{N} y_i = (\varphi_1 - \varphi_0) + (\varphi_2 - \varphi_1) + \ldots + (\varphi_N - \varphi_{N-1}) = \varphi_N = 1.$$

  On the other hand, using (2.16.1), we obtain:

  $$\sum_{i=1}^{N} y_i = \varphi_1 \cdot \sum_{i=1}^{N} \prod_{k=1}^{i-1} \frac{T^-(k)}{T^+(k)} = \varphi_1 \cdot \left(1 + \sum_{i=1}^{N-1} \prod_{k=1}^{i} \frac{T^-(k)}{T^+(k)}\right)$$

- Because the two expressions need to coincide, we get

  $$\varphi_1 = \frac{1}{1 + \sum_{i=1}^{N-1} \prod_{k=1}^{i} \frac{T^-(k)}{T^+(k)}}$$

Evolutionary game theory: Games in finite populations

Remark 2.17. Fixation under weak selection

- To further simplify the expression for φ_1, we can do a Taylor expansion when selection is weak, $w \to 0$.

- One can show (Nowak et al, 2004):

$$\varphi_1 \approx \frac{1}{N} + \frac{6}{N}\Big(N(a+2b-c-2d) - (2a+b+c-4d)\Big)w$$

- We say that strategy i is favored to invade if its fixation probability φ_i is larger than the neutral $1/N$.

- For large N, we conclude that strategy i is favored if

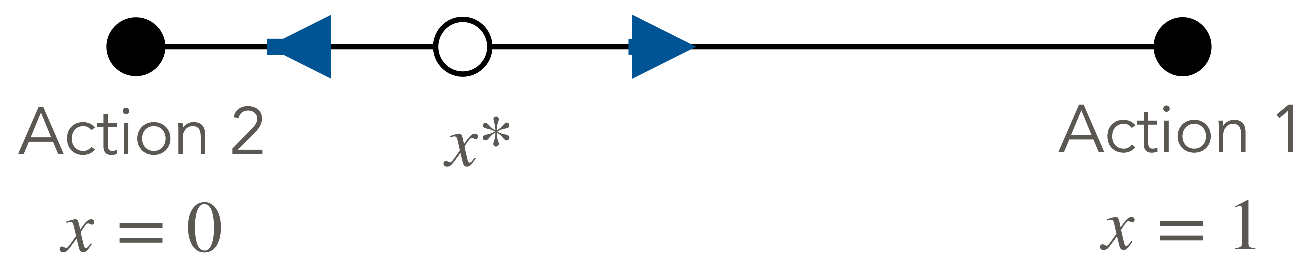
$$a + 2b > c + 2d \qquad (2.17.1)$$

Remark 2.18. One-third rule

- In the special case that the game is a coordination game like stag-hunt ($a > c$, $d > b$), condition (2.17.1) is equivalent to $x^* < 1/3$, where

$$x^* = \frac{d-b}{a-b-c+d}.$$

  This x^* is precisely the interior fixed point according to replicator dynamics.



Summary

Some things you should have learned today:

1. To illustrate the logic of evolutionary game theory, we looked at two well-known dynamics, the replicator equation (infinite population) and the Moran process (finite population).

2. Both dynamics have interesting mathematical properties, and they are well-connected to each other (and to the concepts of classical game theory; without making any a priori assumptions on the rationality of players).

3. Tomorrow, we will use such models of evolutionary dynamics to address one particular problem in evolutionary biology: why do individuals cooperate?