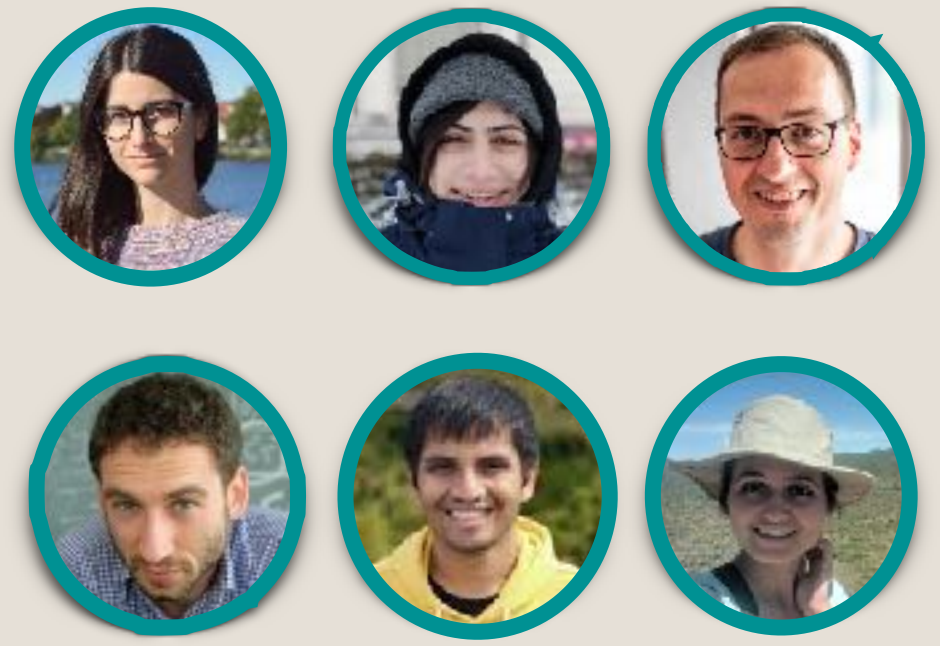
The dynamics of direct reciprocity under memory constraints

Max-Planck-Institut für Evolutionsbiologie, Plön

Christian Hilbe
Max Planck Research Group *Dynamics of Social Behavior*
MPI for Evolutionary Biology, Plön

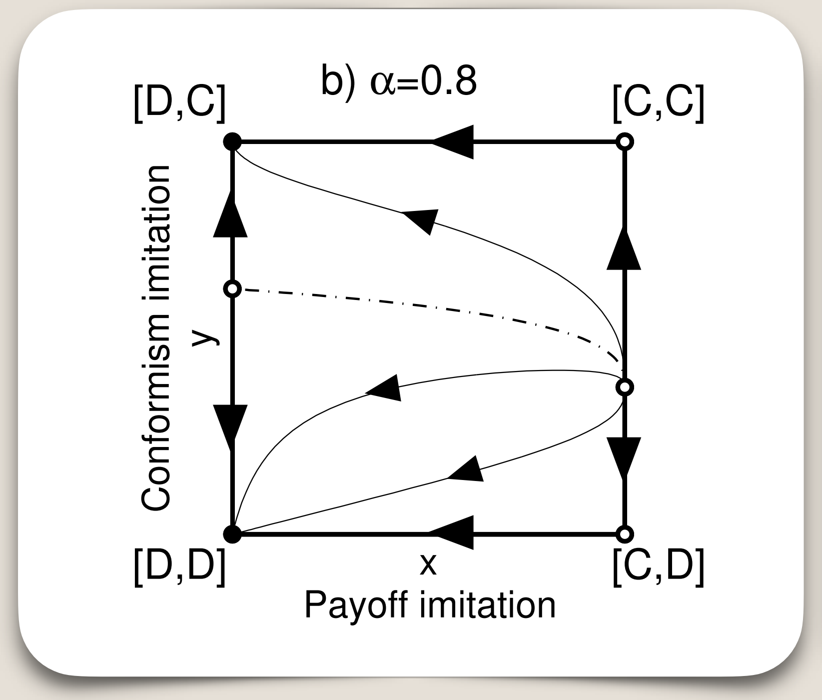
Kolloquium — Vorstellungsvortrag
Universität zu Lübeck

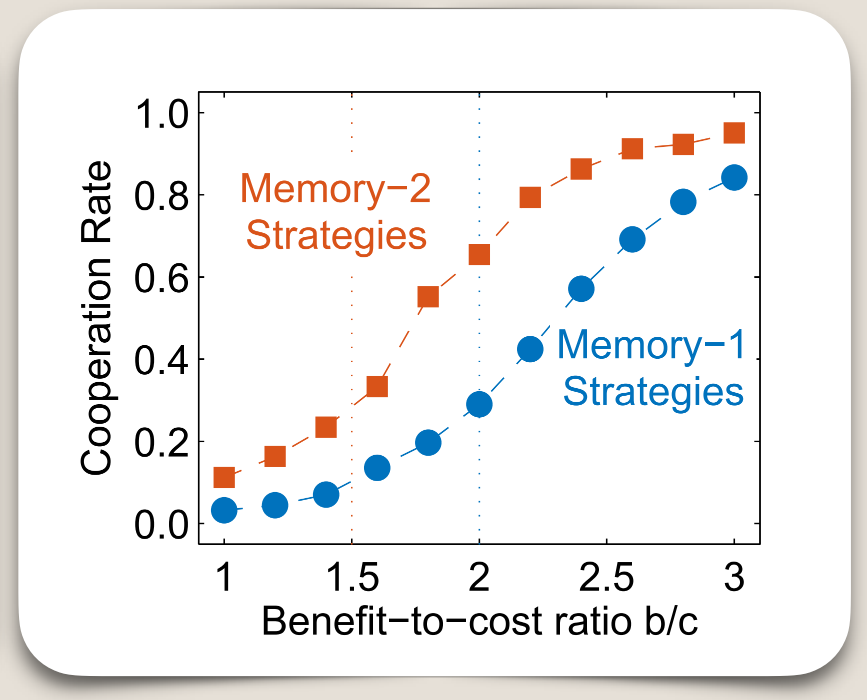
Research Group Dynamics of Social Behavior



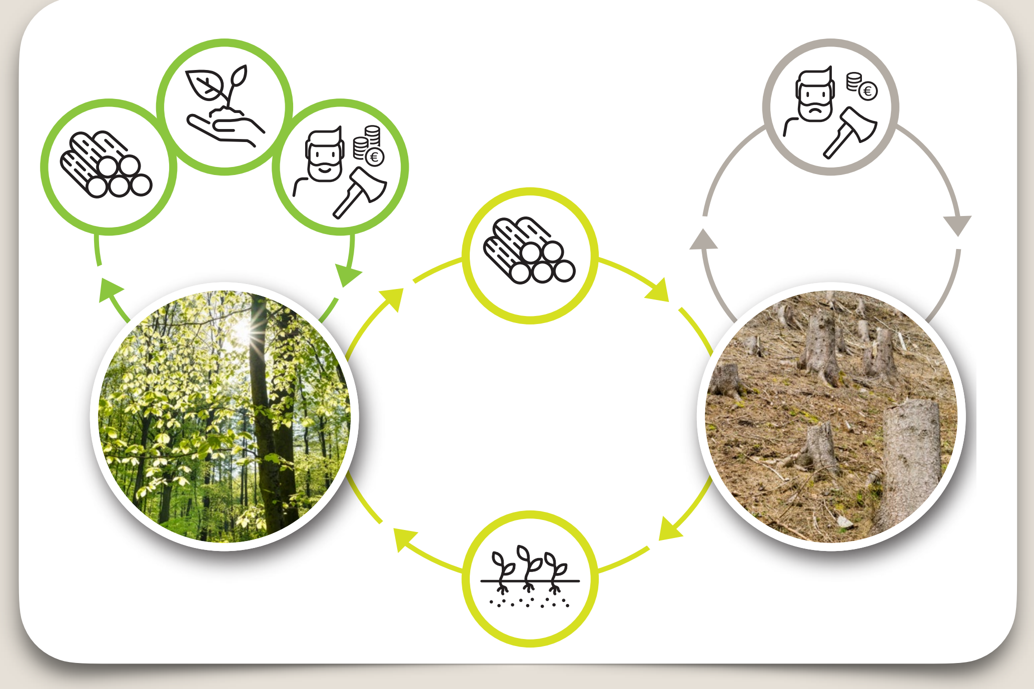
- **Started** in October 2019, currently 6 members

- **Research interest:**
  - **Evolutionary Game Theory**
    How can one describe mathematically how traits or strategic behaviours spread in a population?
  - **Modeling Social Behavior**
    Under which conditions do individuals cooperate?





(Evolutionary) Modeling



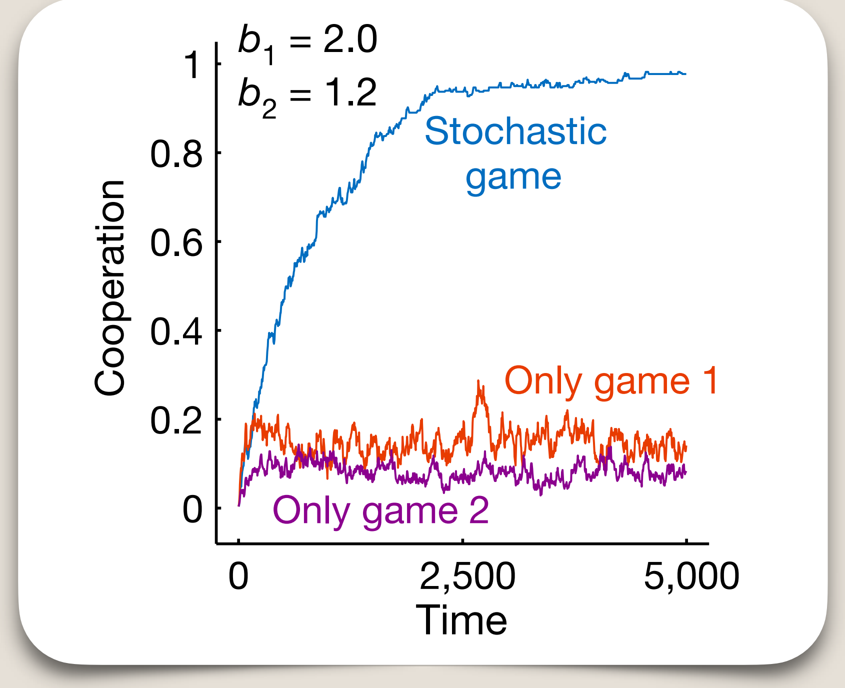
Mathematical analysis

Theorem 1:
Suppose player 1 uses the memory-one strategy

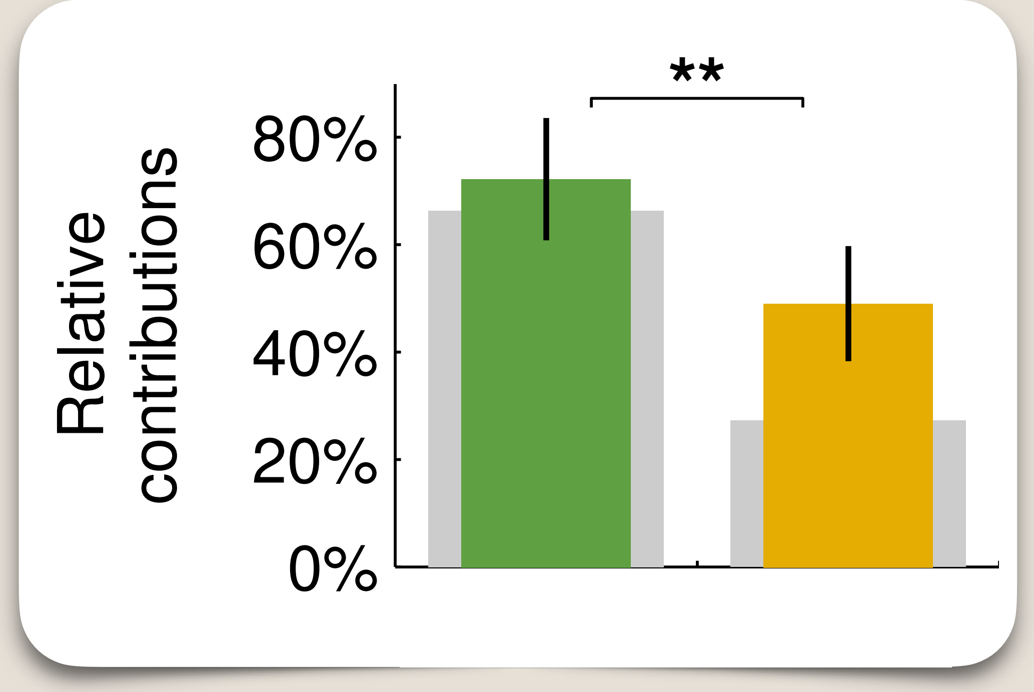
$$
\begin{aligned}
p_{CC} &= 1-\phi(1-s)(b-c-l)\\
p_{CD} &= 1-\phi\big[(1-s)(-c-l)+b+c\big]\\
p_{DC} &= \phi\big[(1-s)(l-b)+b+c\big]\\
p_{DD} &= \phi(1-s)l,
\end{aligned}
$$

for some constants s, l, ϕ with $\phi \neq 0$. Then, no matter what player 2 does, payoffs satisfy $\pi_2 = s\pi_1 + (1-s)l$.

Computer simulations



Lab experiments



Outline of the talk (and of the planned thesis)

- **Chapter 0: Introduction**
  Modelling direct reciprocity and previous literature

- **Chapter 1: Of extortion and generosity**
  Theory and evolutionary dynamics of zero-determinant strategies

- **Chapter 2: Of partners and rivals**
  Characterizing the Nash equilibria of the repeated prisoner's dilemma

- **Chapter 3: Beyond the repeated prisoner's dilemma**
  Analysis of stochastic games, and of asymmetric games

Basic question:
Why do individuals cooperate?
That is, why would we pay a cost to the benefit of someone else?

Examples:

- Doing favors to each other
- Reviewing other researcher’s manuscripts
- Organising colloquium talks
- Numerous examples in the animal kingdom

Mechanisms for cooperation:

- Direct reciprocity: Cooperation in repeated interactions

NATURE VOL. 325 29 JANUARY 1987 LETTERS

TIT FOR TAT in sticklebacks and the evolution of cooperation

Manfred Milinski

Arbeitsgruppe für Verhaltensforschun
Ruhr-Universität, Postfach 102148, 4

Reciprocal food sharing in the vampire bat

Gerald S. Wilkinson

Department of Biology, C-016, Univer
San Diego, La Jolla, California 92093,

Matching times of leading and following suggest cooperation through direct reciprocity during V-formation flight in ibis

Bernhard Voelkl[a,b,c,1], Steven J. Portugal[d,2], Markus Unsöld[e,f],
and Johannes Fritz[c,e]

Repeated prisoner's dilemma

- Strategic interaction among two individuals ("Players")

- In each round, players can either cooperate (play C) or defect (play D)
  Cooperation means to pay a cost $c > 0$ for the co-player to get a benefit $b > c$

- After each round, there is another round with probability $\delta \leq 1$

- To play this game, each player chooses some strategy.
  A strategy is a map that take any previous history of the game and outputs an action, $\sigma : \mathscr{H} \rightarrow \{C, D\}$

- Players want to choose strategies that maximise their average payoffs. If player i's payoff in round t is $\pi_i(t)$, the player wishes to maximise

$$\pi_i = (1-\delta)\sum_{t=1}^{\infty}\delta^t \cdot \pi_i(t) \quad \text{[if } \delta < 1\text{]} \quad \text{or} \quad \pi_i = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\pi_i(t) \quad \text{[If } \delta = 1\text{]}$$

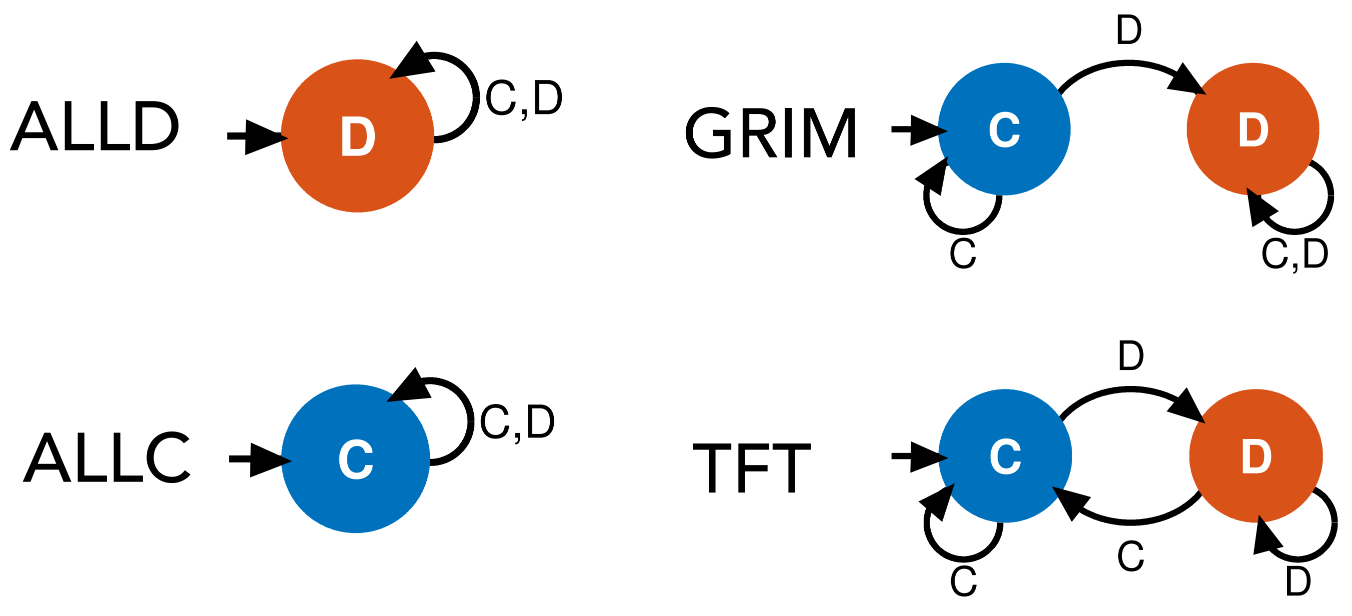
	C	D
C	*b*-*c*	-*c*
D	*b*	0

$\downarrow \delta$

	C	D
C	*b*-*c*	-*c*
D	*b*	0

$\downarrow \delta$

…



There are (uncountably) many other strategies

Memory-1 strategies

- Assumption: When deciding what to do next, players only condition their response to the previous round.
- Such strategies can be represented as vectors, $\mathbf{p} = (p_0, p_{CC}, p_{CD}, p_{DC}, p_{DD}) \in [0,1]^5$
- Examples: ALLD = (0,0,0,0,0); TFT = (1, 1, 0, 1, 0); WSLS = (1,1,0,0,1).
- **Major advantage:** Game between two players with strategies $\mathbf{p}$ and $\mathbf{q}$ can be represented by a Markov chain.
  The states are the possible outcomes of a round, (C,C), (C,D), (D,C), (D,D).

  Initial distribution: $\mathbf{v_0} = \big(p_0\, q_0,\ p_0(1-q_0),\ (1-p_0)q_0,\ (1-p_0)(1-q_0)\big)$

  Transition matrix $M = \begin{pmatrix} p_{CC}q_{CC} & p_{CC}(1-q_{CC}) & (1-p_{CC})q_{CC} & (1-p_{CC})(1-q_{CC}) \\ p_{CD}q_{DC} & p_{CD}(1-q_{DC}) & (1-p_{CD})q_{DC} & (1-p_{CD})(1-q_{DC}) \\ p_{DC}q_{CD} & p_{DC}(1-q_{CD}) & (1-p_{DC})q_{CD} & (1-p_{DC})(1-q_{CD}) \\ p_{DD}q_{DD} & p_{DD}(1-q_{DD}) & (1-p_{DD})q_{DD} & (1-p_{DD})(1-q_{DD}) \end{pmatrix}$

  Expected outcome distribution: $\mathbf{v} = (v_{CC}, v_{CD}, v_{DC}, v_{DD}) := (1-\delta)\mathbf{v_0}(\mathbf{1}-\delta M)^{-1}$

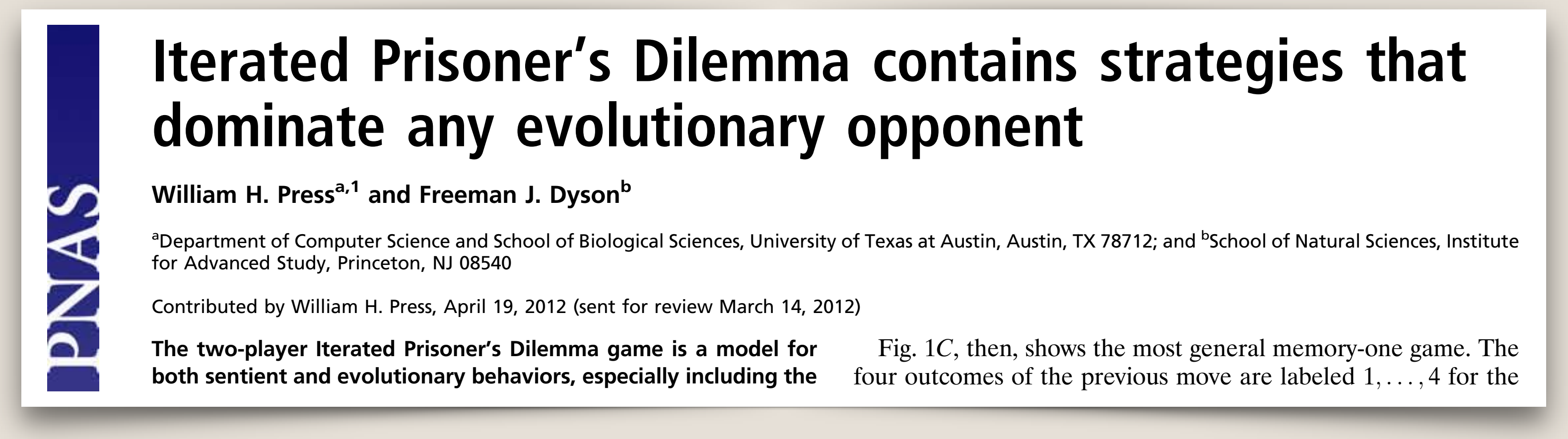
  Expected payoffs: $\pi_1 = (v_{CC}+v_{DC})b - (v_{CC}+v_{CD})c$

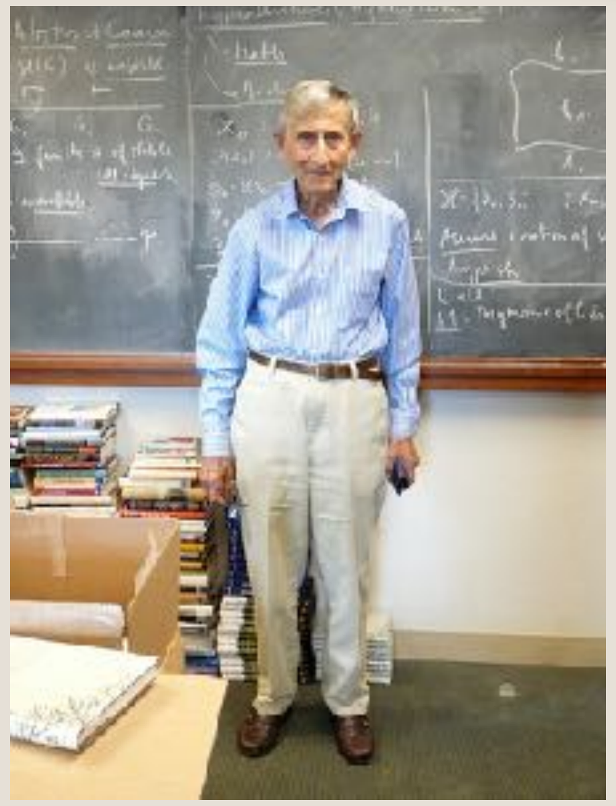
Theorem (Press & Dyson, PNAS 2012)

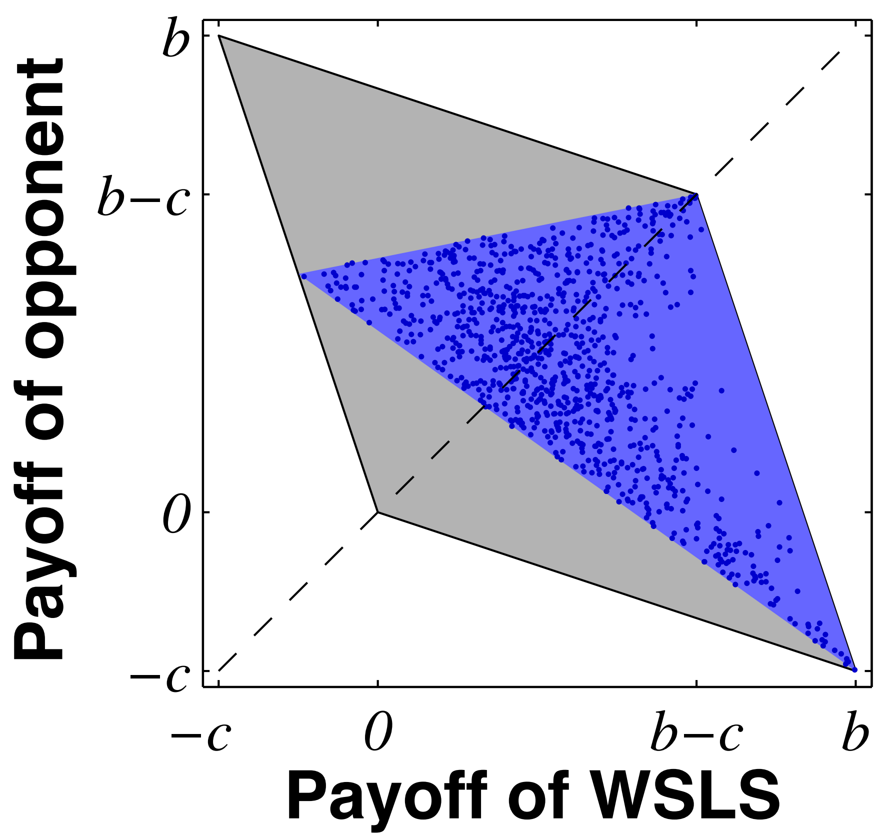
Consider an infinitely repeated prisoner's dilemma (with $\delta = 1$). Suppose there are constants α, β, γ such that player 1 applies a memory-1 strategy $\mathbf{p}$ that satisfies

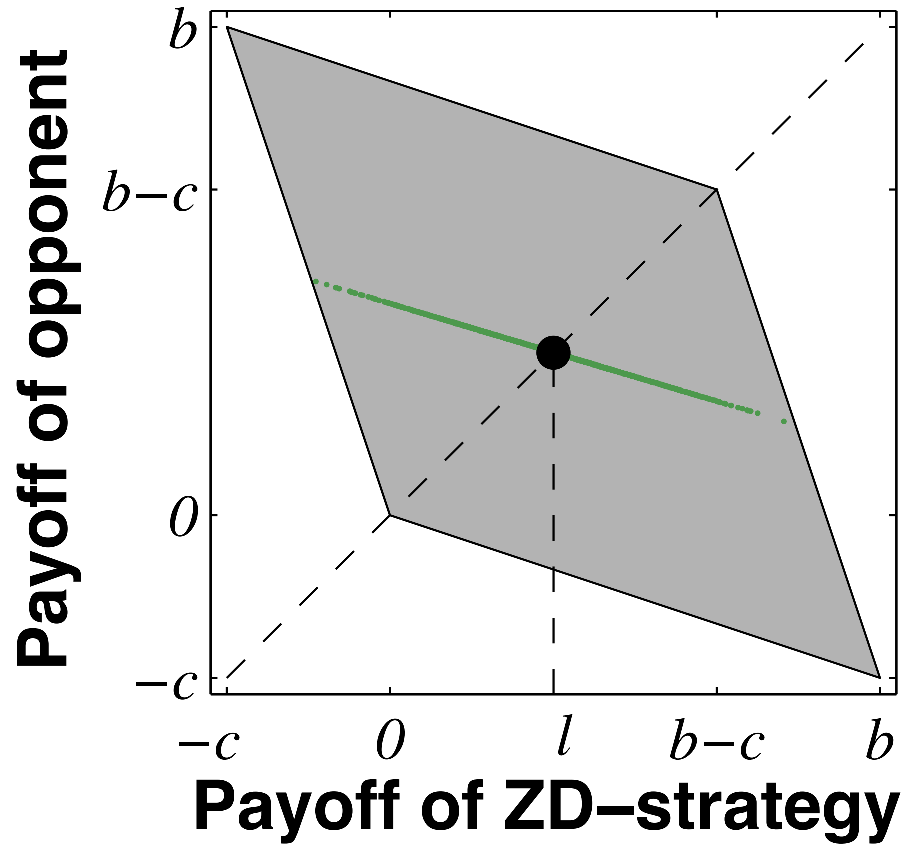
$$
\begin{aligned}
p_{CC} &= (\alpha+\beta)(b-c) + \gamma + 1\\
p_{CD} &= -\alpha c + \beta b + \gamma + 1\\
p_{DC} &= \alpha b - \beta c + \gamma\\
p_{DD} &= \gamma
\end{aligned}
$$

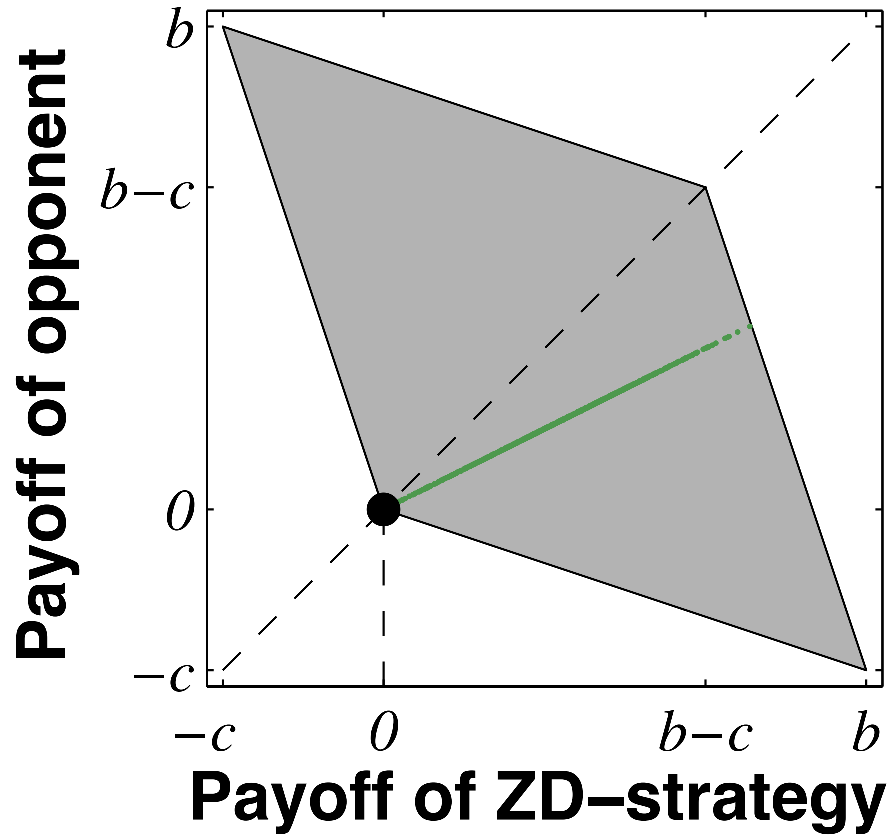
Then, irrespective of player 2's strategy, payoffs satisfy $\alpha\pi_1 + \beta\pi_2 + \gamma = 0$. Such a strategy $\mathbf{p}$ is called a zero-determinant (ZD) strategy.


PNAS
Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent
William H. Press[a,1] and Freeman J. Dyson[b]
[a]Department of Computer Science and School of Biological Sciences, University of Texas at Austin, Austin, TX 78712; and [b]School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540
Contributed by William H. Press, April 19, 2012 (sent for review March 14, 2012)
The two-player Iterated Prisoner's Dilemma game is a model for both sentient and evolutionary behaviors, especially including the
Fig. 1C, then, shows the most general memory-one game. The four outcomes of the previous move are labeled $1, \ldots, 4$ for the









Extortionate ZD strategy

Possible follow-up questions

- If strategies are not consciously chosen, but rather the result of evolution, would (extortionate) ZD strategies ever evolve?
- How do humans react to ZD strategies?

} Chapter 1

- Can we use this mathematical formalism to characterize other interesting sets of memory-1 strategies (e.g., all memory-1 Nash equilibria)?

} Chapter 2

- To which extent can one generalise the theory of direct reciprocity to scenarios that are more complex than the prisoner's dilemma?

} Chapter 3

Outline of the talk

- **Chapter 0: Introduction**
  Modelling direct reciprocity and previous literature

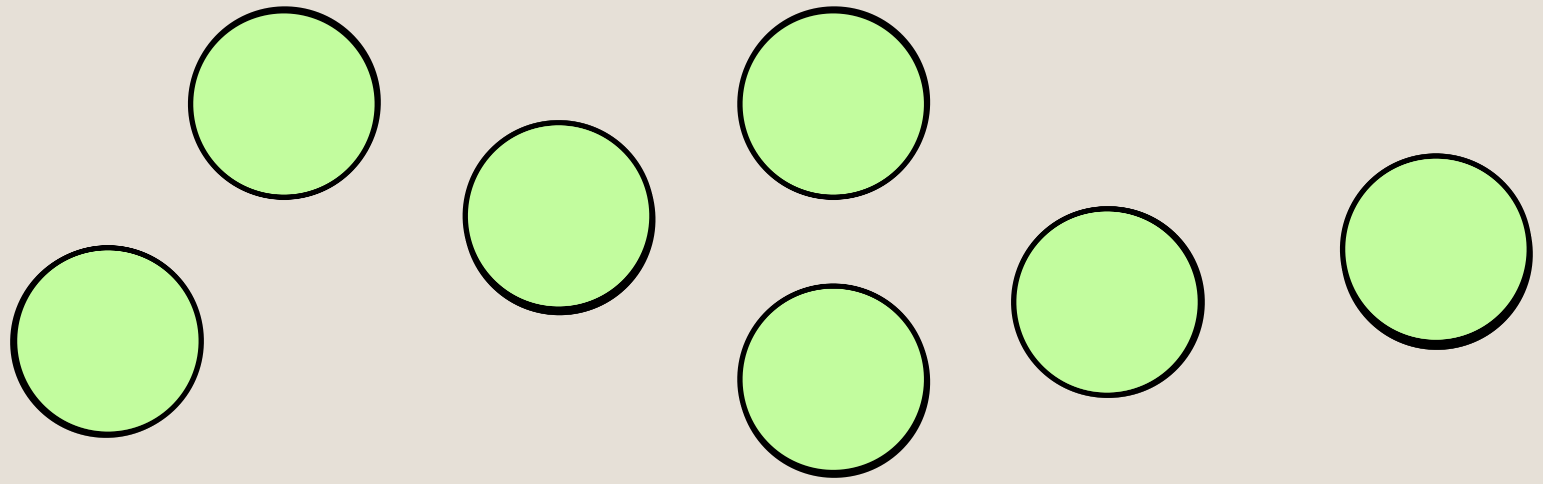
- **Chapter 1: Of extortion and generosity**
  Theory and evolutionary dynamics of zero-determinant strategies

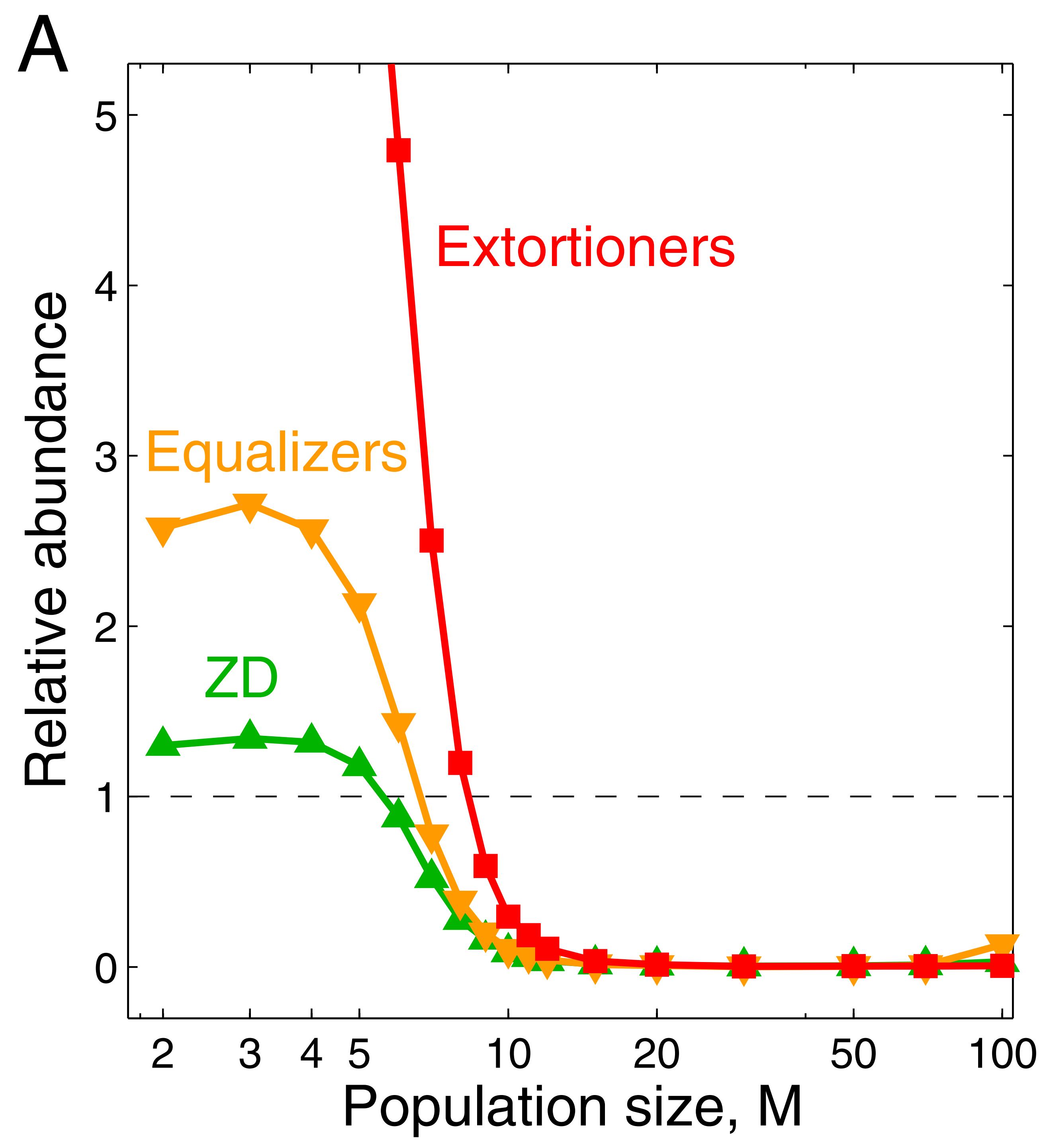
- **Chapter 2: Of partners and rivals**
  Characterizing the Nash equilibria of the repeated prisoner's dilemma

- **Chapter 3: Beyond the repeated prisoner's dilemma**
  Analysis of stochastic games, and of asymmetric games

A model of intraspecific evolution.

- Consider a population of size *M*
- Initially, all players use the same resident memory-1 strategy $\mathbf{p}$
- From time to time, one player mutates to a different memory-1 strategy $\mathbf{p}'$
- The other population members may decide to imitate $\mathbf{p}'$ if the new strategy gives a high payoff.
  [Imitation probability $\rho = \dfrac{1}{1+\exp[-s(\pi'-\pi)]}$]
- If mutations are rare, the mutant strategy either fixes in the population or goes extinct before next mutation arises.
- This process generates a sequence of residents $(\mathbf{p}^0, \mathbf{p}^1, \mathbf{p}^2, \ldots)$
- We count how often these residents are in the neighborhood of the extortionate ZD strategies.





Hilbe, Nowak, Sigmund (PNAS 2013)

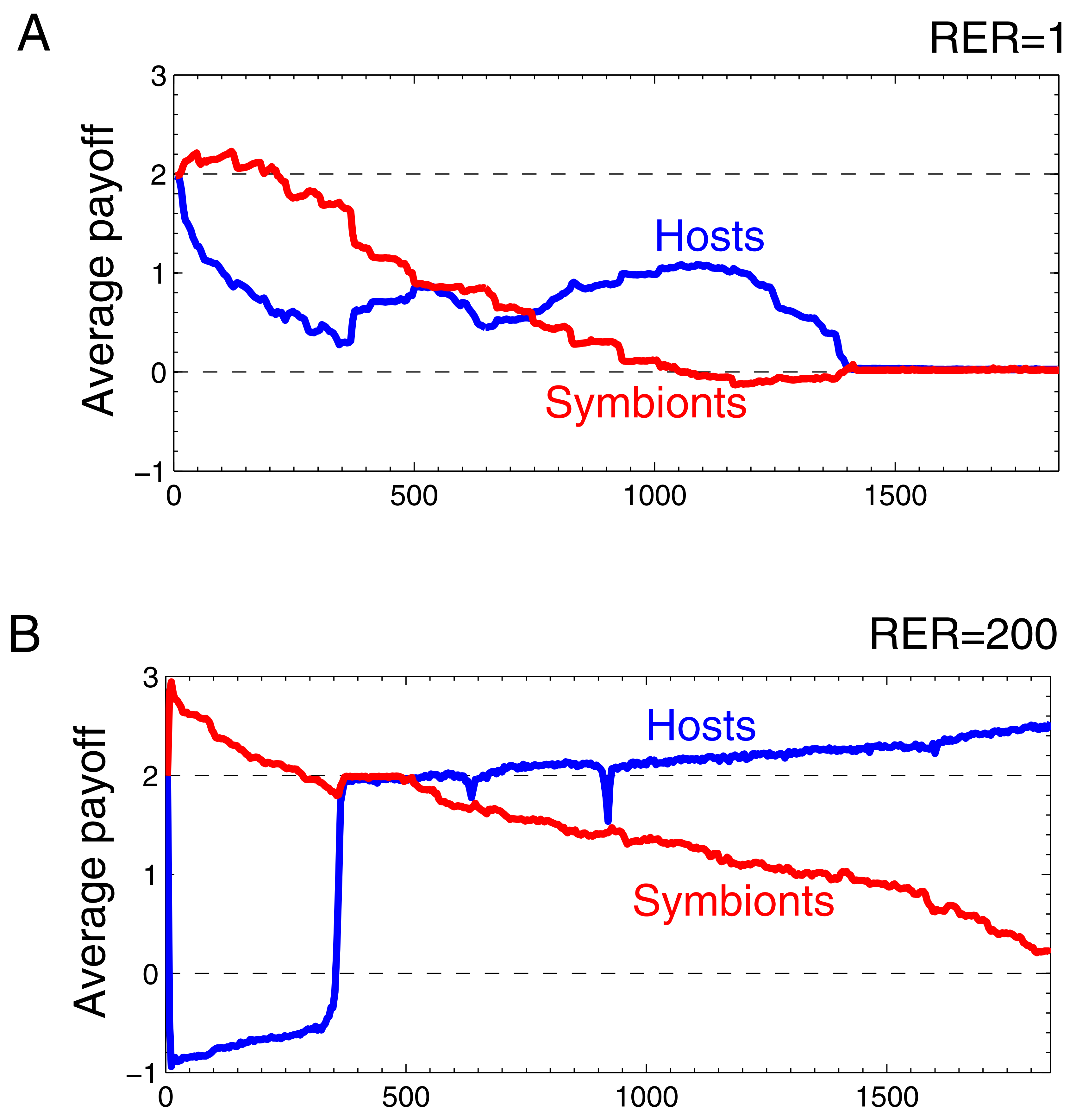
Chapter 1: Of extortion and generosity. Evolution of extortion

A model of interspecific evolution.

- Consider two different populations (e.g., a population of hosts and symbionts).
- The members of both populations use memory-1 strategies
- The evolutionary process is similar to before
- However, the two populations may evolve at different rates, measured by the relative evolutionary rate RER (e.g., symbionts typically evolve faster than their hosts).

Conclusion.

Extortion only evolves in small populations, or when two populations co-evolve at different rates.



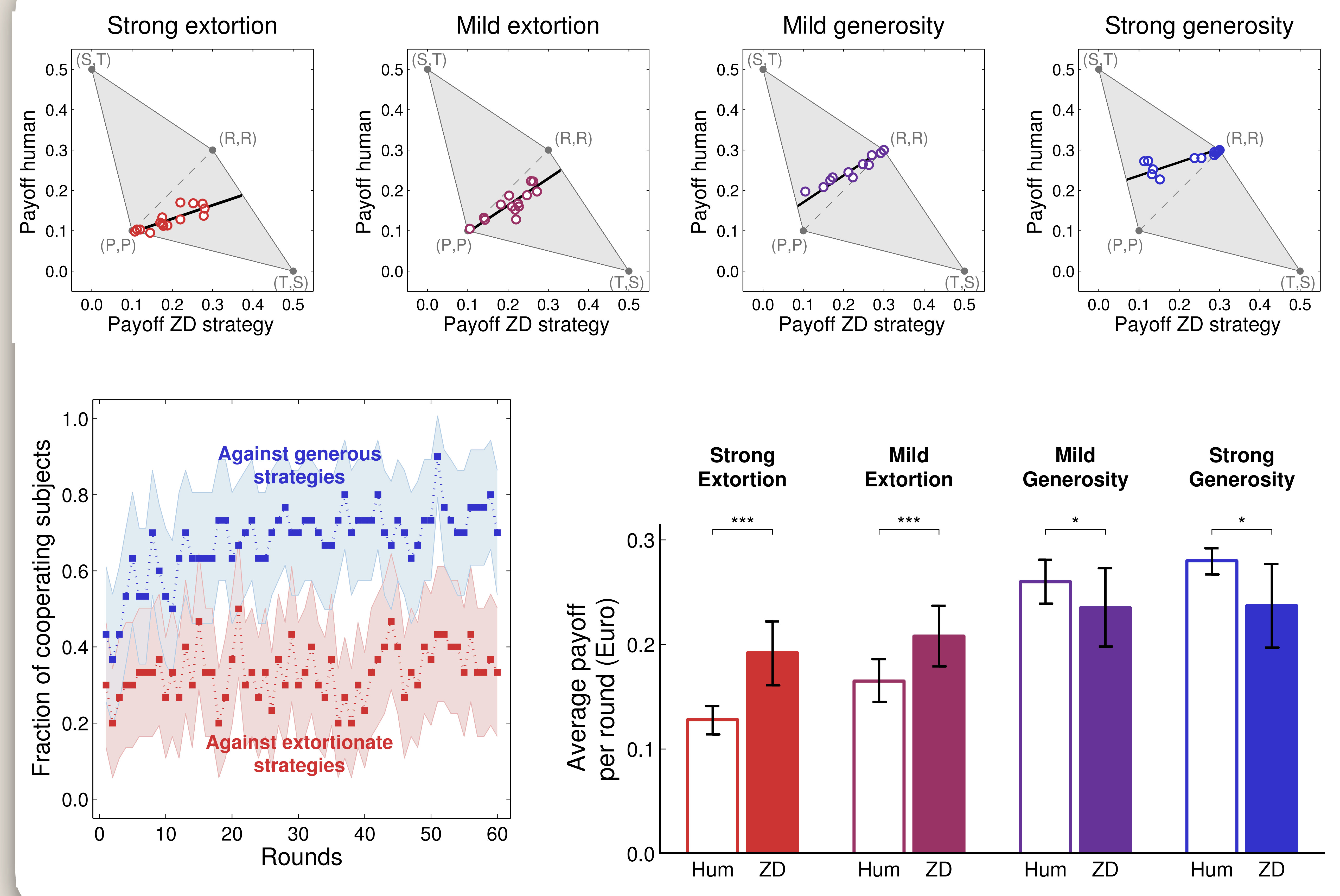
Hilbe, Nowak, Sigmund (PNAS 2013)

Chapter 1: Of extortion and generosity. Extortion in the laboratory

Setup.

- We asked 60 students to participate in an experiment.
- Students played 60 rounds of the repeated prisoner's dilemma (against a computerised opponent).
- Four different treatments (depending on computer strategy: 2 extortionate, 2 generous).
- Payoff matrix:

	C	D
C	0,30 €	0,00 €
D	0,50 €	0,10 €



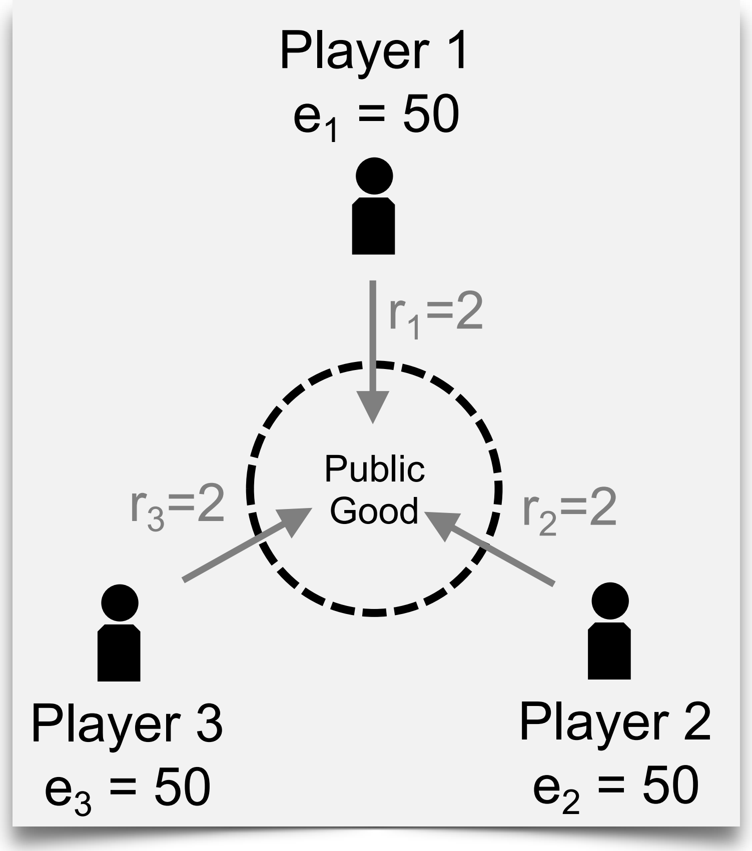
Hypotheses.

1. If people only care about their payoffs, they should equally learn to cooperate in all treatments.
2. In that case, the extortionate computer program should get a higher payoff than the generous program.

Hilbe, Röhl, Milinski (Nature Communications 2014)

Research question.

To which extent do the results of Press & Dyson generalise beyond pairwise social dilemmas?

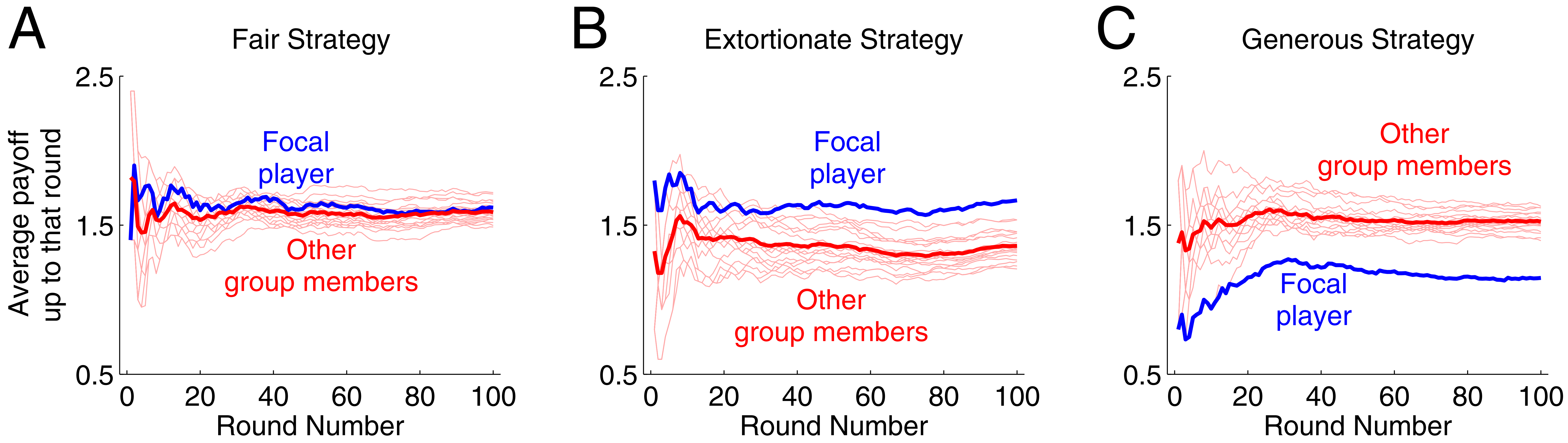


Theorem (Press and Dyson). *Let α, β_j, and γ be parameters such that $\sum_{j\neq i}\beta_j \neq 0$. If player i applies a memory-one strategy of the form*

$$\mathbf{p} = \mathbf{p}^{Rep} + \alpha\mathbf{g}^i + \sum_{j\neq i}^{n} \beta_j \mathbf{g}^j + \gamma \mathbf{1}, \qquad \text{[S8]}$$

then, irrespective of the strategies of the remaining $n-1$ group members, payoffs obey the equation

$$\alpha\pi^i + \sum_{j\neq i} \beta_j \pi^j + \gamma = 0. \qquad \text{[S9]}$$



Hilbe, Wu, Traulsen, Nowak (PNAS 2014)

Outline of the talk

- **Chapter 0: Introduction**
  Modelling direct reciprocity and previous literature

- **Chapter 1: Of extortion and generosity**
  Theory and evolutionary dynamics of zero-determinant strategies

- **Chapter 2: Of partners and rivals**
  Characterizing the Nash equilibria of the repeated prisoner’s dilemma

- **Chapter 3: Beyond the repeated prisoner’s dilemma**
  Analysis of stochastic games, and of asymmetric games

Chapter 2: Of partners and rivals. Characterizing alternative strategy sets

Motivation.

Can the formalism of Press & Dyson (and its extensions by Akin and ourselves) be used to characterize more general strategy classes?

Example 1.

Extortionate strategies $\mathbf{p}$ are those memory-1 strategies that satisfy

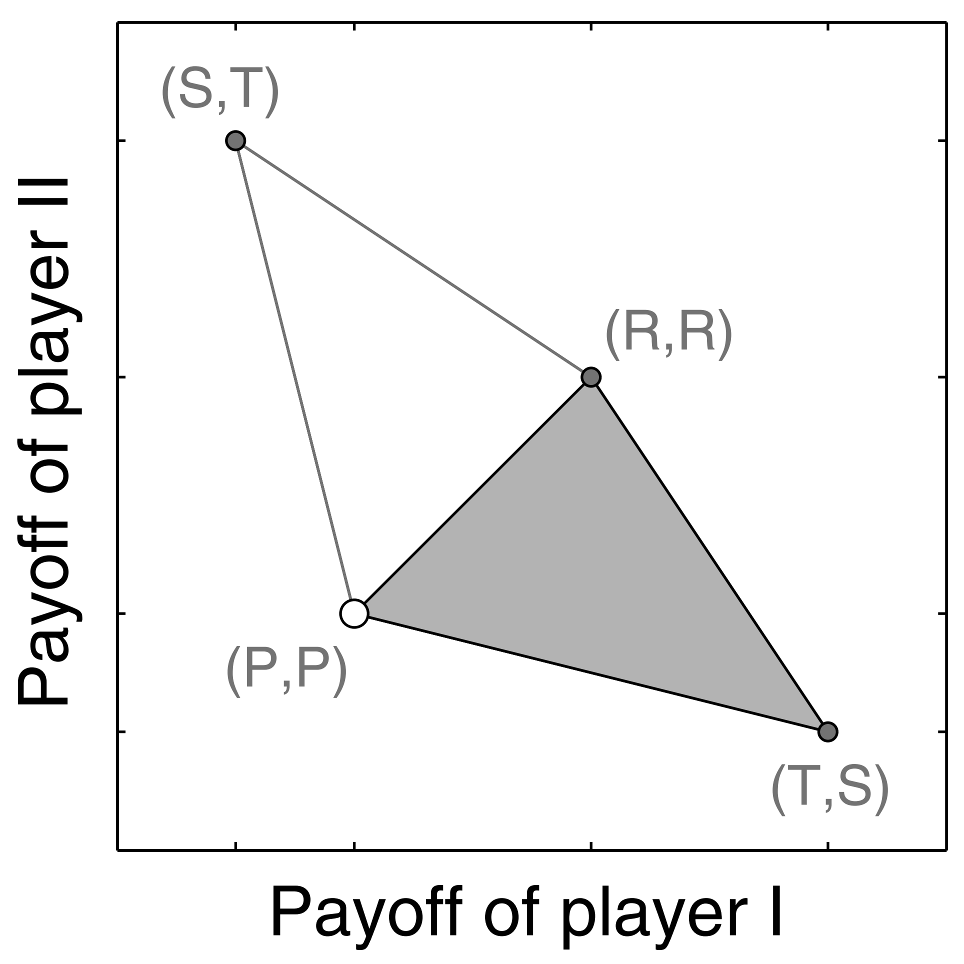
"competitive"

1. $\mathbf{p}$-player obtains at least co-player's payoff
2. ~~Co-player's best response is to always cooperate~~
3. ~~Payoffs are on a line.~~

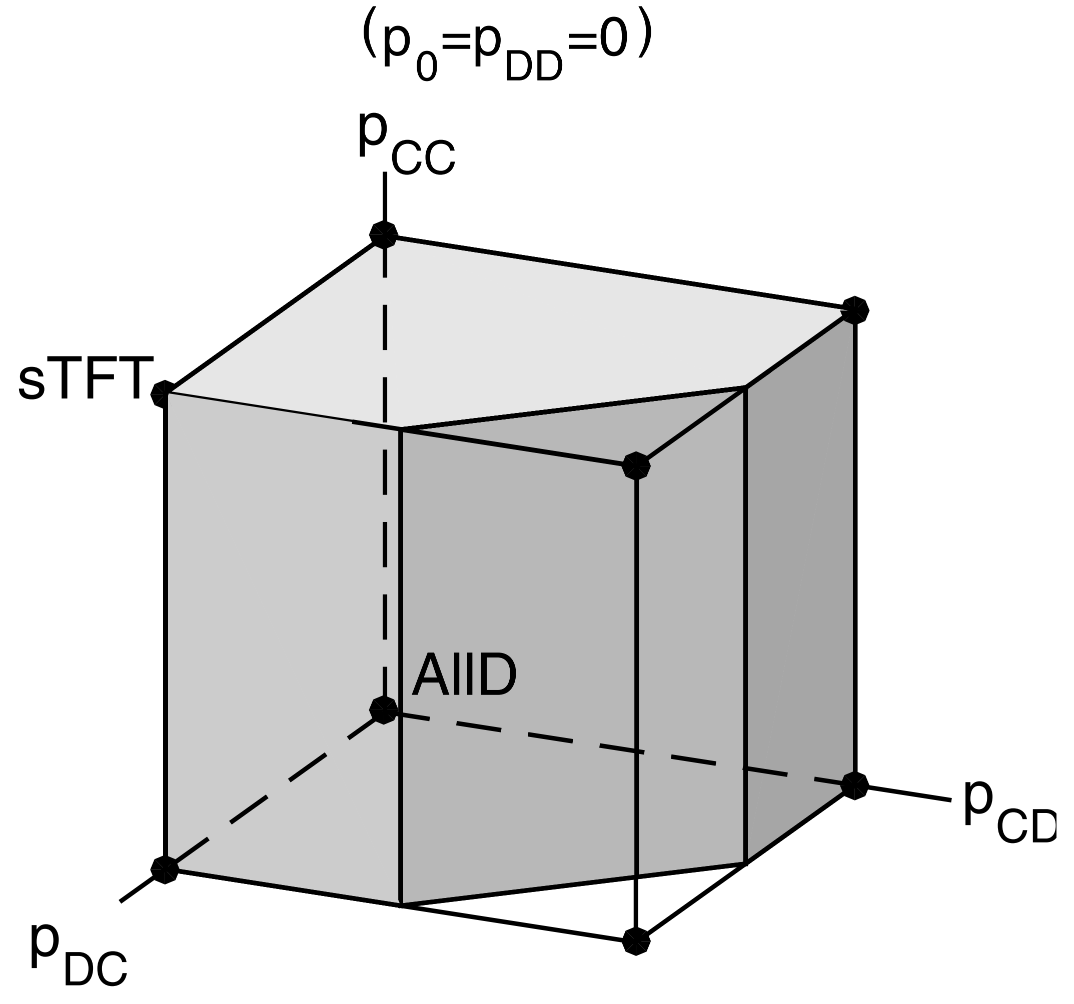
Proposition 2. *Suppose player I applies the memory-one strategy* $\mathbf{p}$. *Then the following are equivalent:*

(i) $\mathbf{p}$ *is competitive.*
(ii) *If the co-player uses either AllD or the strategy* $(0, 0, 0, 1; 0)$, *then* $\pi_I \geq \pi_{II}$.
(iii) *The entries of* $\mathbf{p}$ *satisfy* $p_0 = p_P = 0$ *and* $\delta(p_{CD} + p_{DC}) \leq 1$.

Competitive strategy



Competitive strategies



Hilbe, Traulsen, Sigmund
(Games and Economic Behavior 2015)

Example 2.

Generous strategies $\mathbf{p}$ are those memory-1 strategies that satisfy

"Partner"

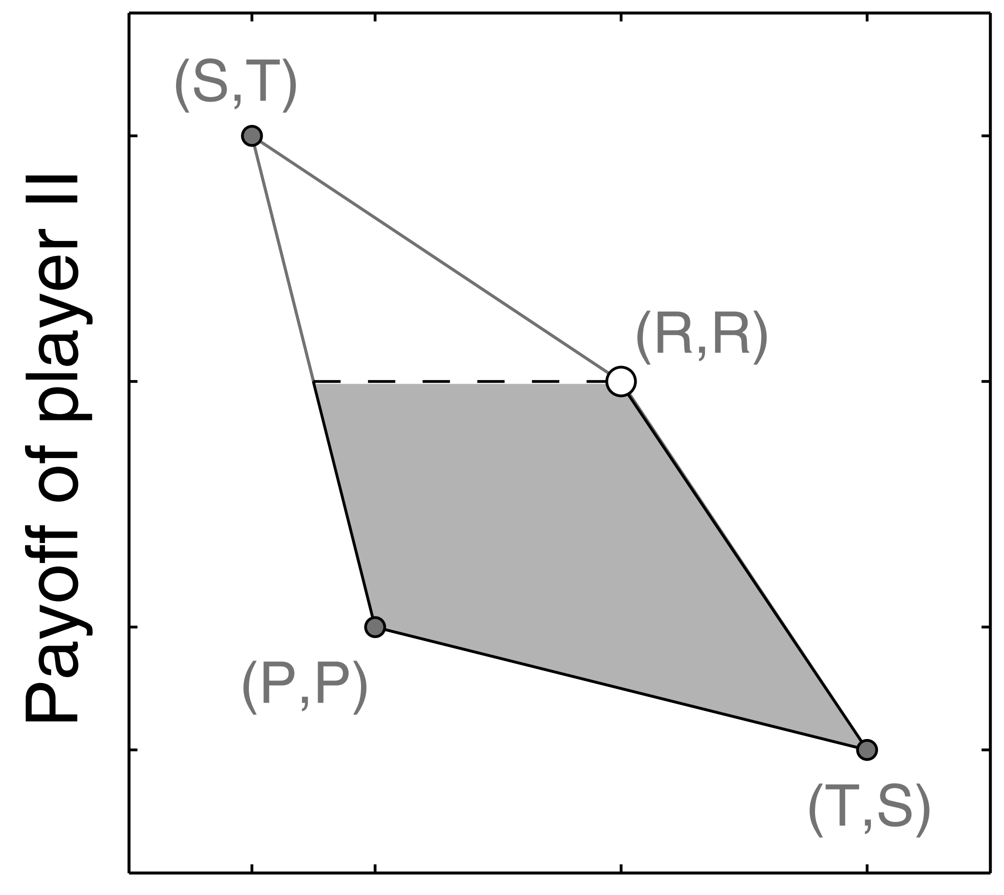
1. Mutually cooperative against themselves
2. Impossible for co-player to get more than the mutual cooperation payoff
3. ~~Payoffs are on a line.~~

Proposition 1. *For a player I with a nice memory-one strategy* $\mathbf{p}$*, the following are equivalent:*

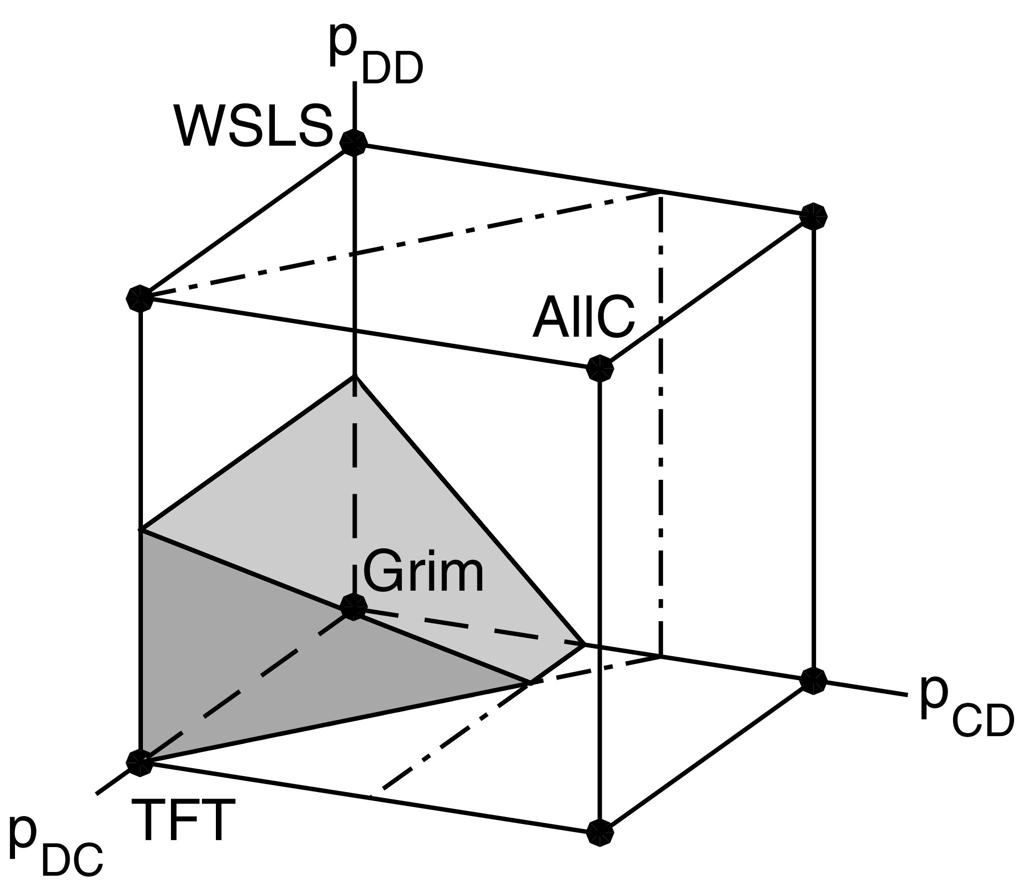
(i) $\mathbf{p}$ *is a partner strategy;*
(ii) *If the co-player uses either AllD or the strategy* $(0, 1, 1, 1; 0)$*, then* $\pi_{II} < R$*;*
(iii) *The two inequalities* $B_1 < 0$ *and* $B_2 < 0$ *hold, with*

$$B_1 = \delta(T-R)p_{DD} - \delta(R-P)(1-p_{CD}) + (1-\delta)(T-R)$$
$$B_2 = \delta(T-R)p_{DC} - \delta(R-S)(1-p_{CD}) + (1-\delta)(T-R).$$

Partner strategy



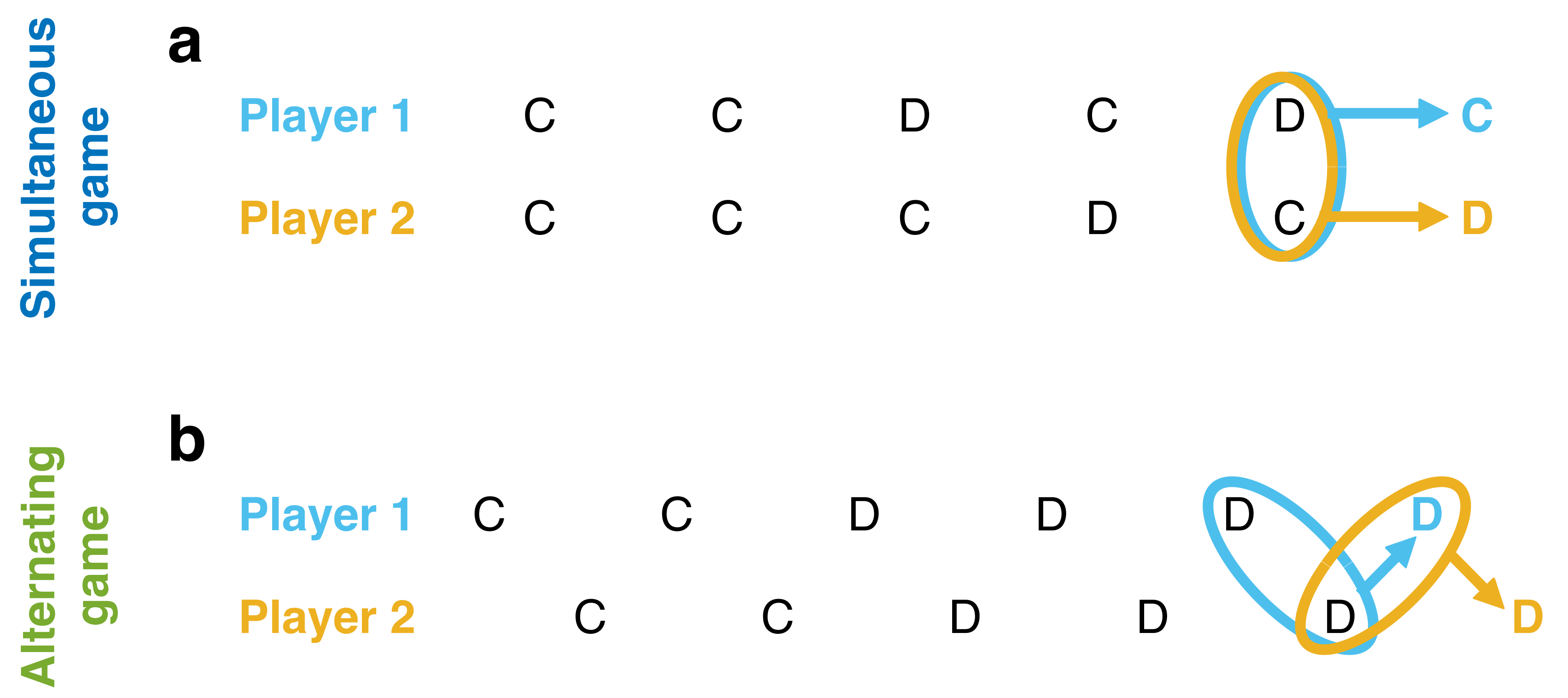
Partner strategies
$(p_0 = p_{CC} = 1)$



Hilbe, Traulsen, Sigmund
(Games and Economic Behavior 2015)

Motivation.

In many of the repeated games in nature, players do not move simultaneously. How do these techniques apply to alternating games?



Proposition: There are three classes of Nash equilibria among the memory-1 strategies of the alternating game:

Partners.

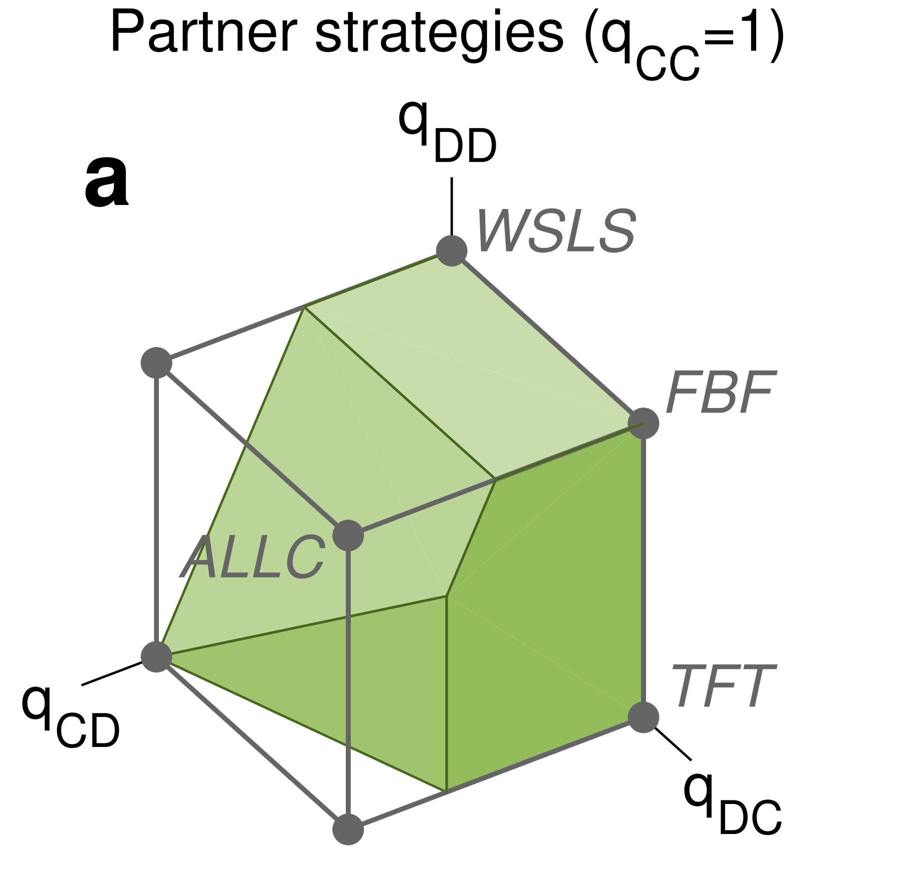
$$q_{CC}=1$$
$$(b-c)(1-q_{CD})\geq c\,q_{DD}$$
$$b(1-q_{CD})\geq c\,q_{DC}.$$

Defectors.

$$q_{DD}=0$$
$$b\,q_{DC}\leq c\,(1-q_{CD})$$
$$(b-c)\,q_{DC}\leq c\,(1-q_{CC}).$$

Equalizers.

$$q_{CD}=\frac{b\,q_{CC}-c\,(1+q_{DD})}{b-c}$$
$$q_{DC}=\frac{b\,q_{DD}+c(1-q_{CC})}{b-c}.$$



Park, Nowak, Hilbe (under review, 2021)

Outline of the talk

- **Chapter 0: Introduction**
  Modelling direct reciprocity and previous literature

- **Chapter 1: Of extortion and generosity**
  Theory and evolutionary dynamics of zero-determinant strategies

- **Chapter 2: Of partners and rivals**
  Characterizing the Nash equilibria of the repeated prisoner's dilemma

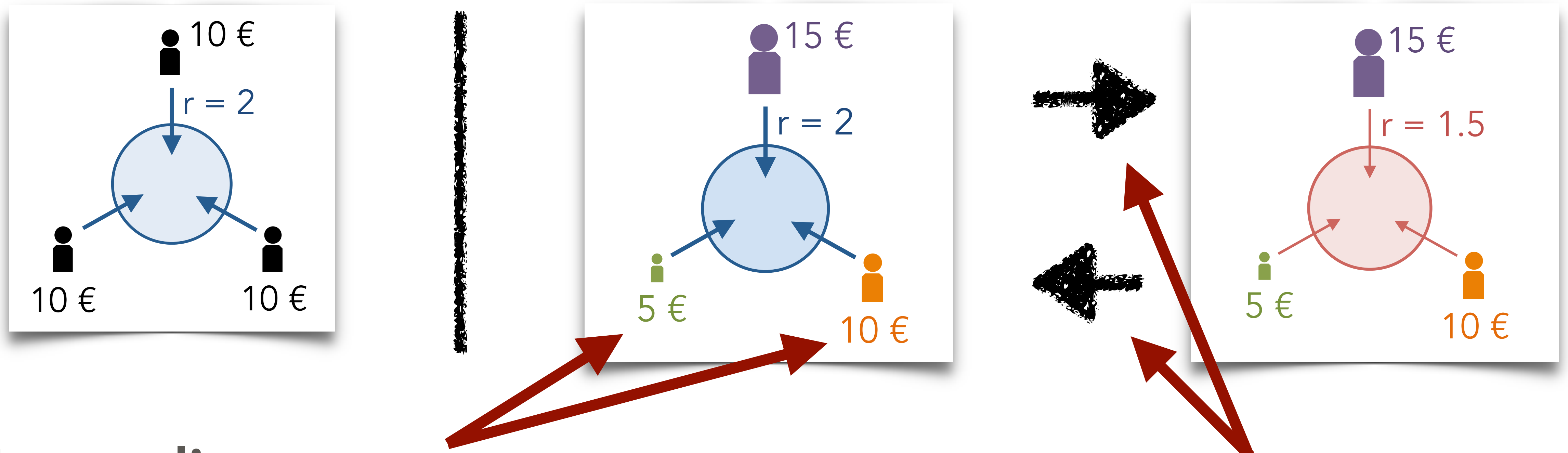
- **Chapter 3: Beyond the repeated prisoner's dilemma**
  Analysis of stochastic games, and of asymmetric games

Motivation.

Previous models make the following assumptions.

- **Homogeneity:** All individuals are the same
- **Stationarity:** The individuals' environment does not change

Example: Public goods game

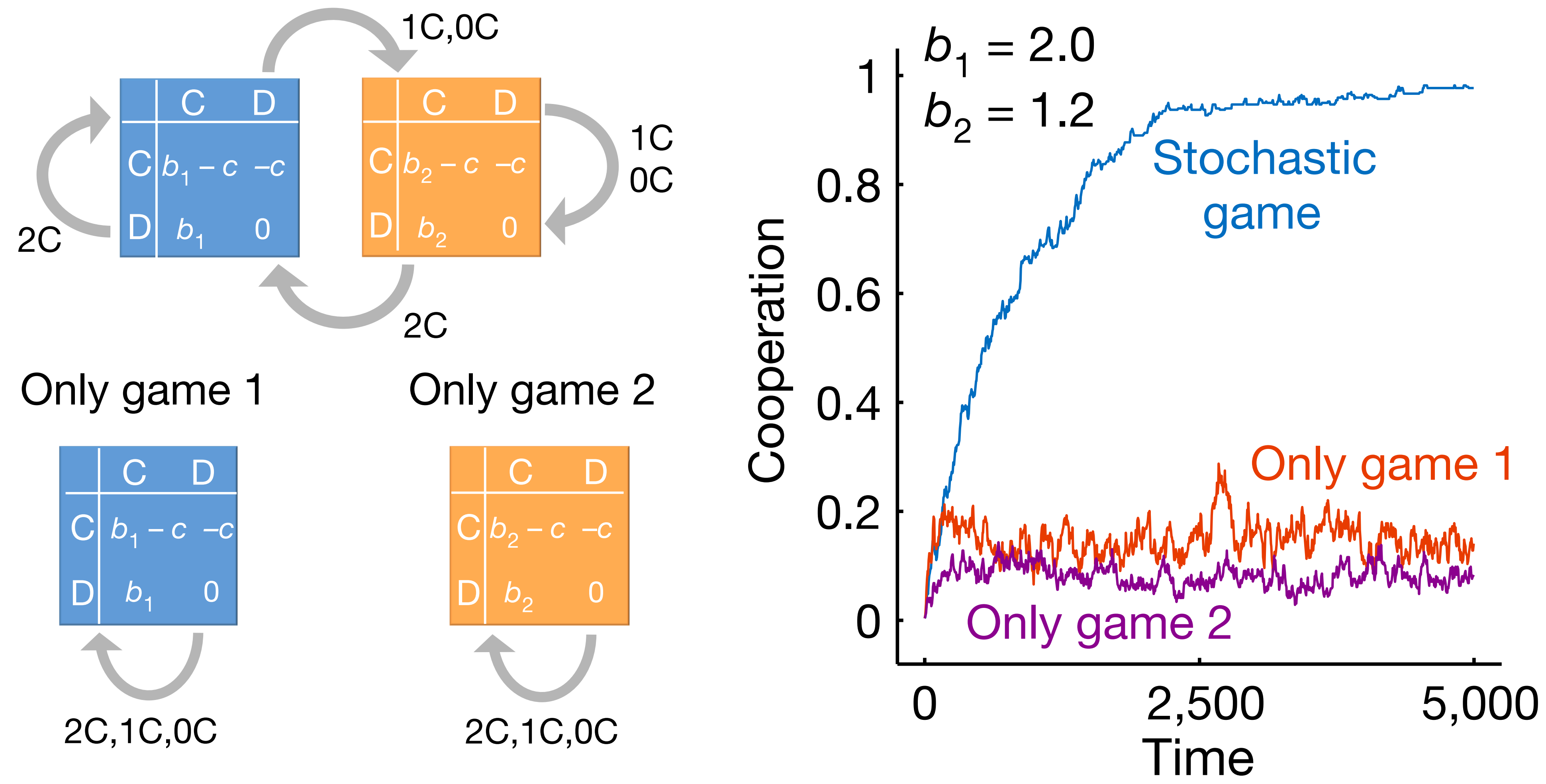


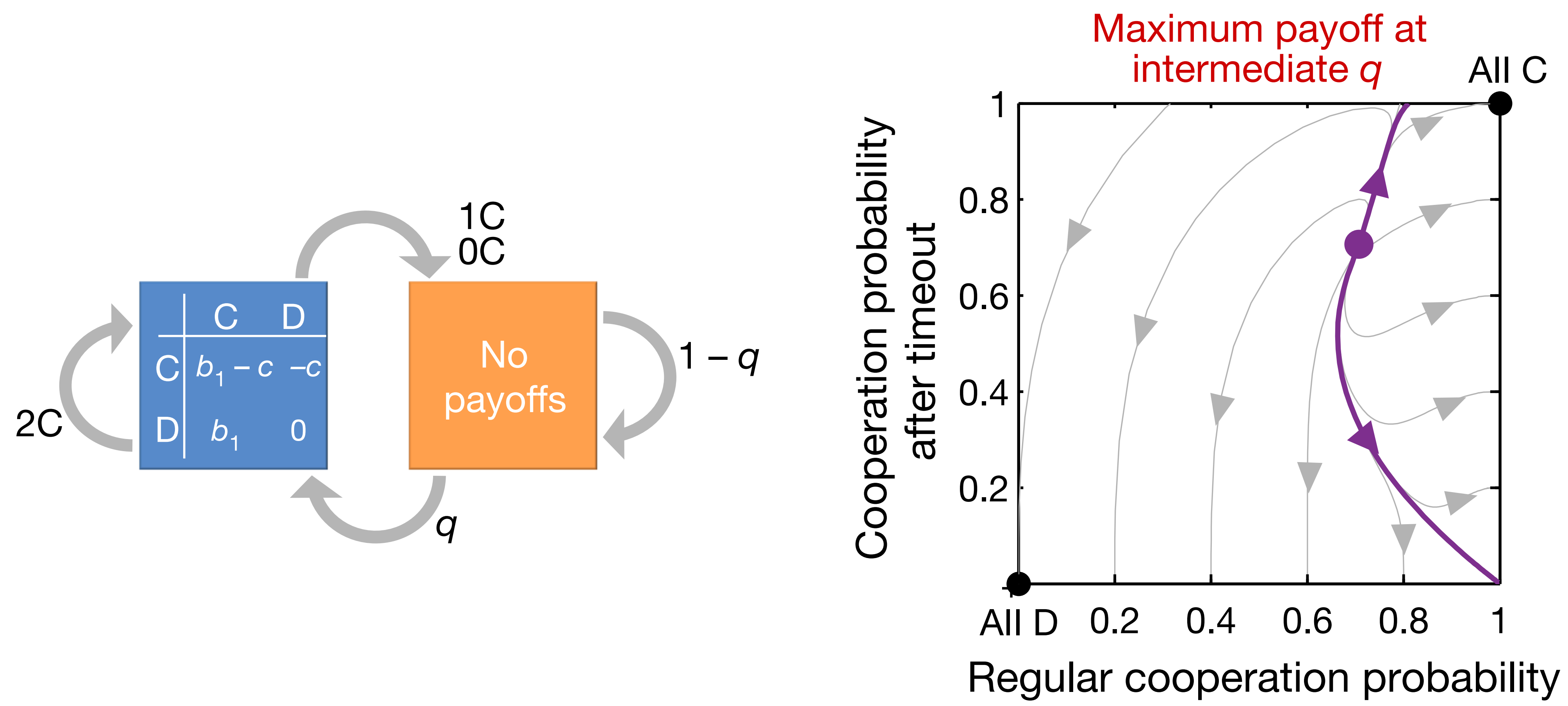
Inequality:
How does inequality influence cooperation? How do people learn from each other?

Environmental feedback:
How do people cooperate when they know their actions today affect the future?

Games with environmental feedback: Stochastic games

- Repeated interactions among 2 players
- In each round, players can find themselves in one of k possible environments $E_1, \ldots, E_k$
- In each environment, players can cooperate or defect. However, the cost of cooperation c_i and the benefit b_i depends on the players' present environment E_i
- The next round's environment depends on the previous environment and the players' previous actions
- Again, when players adopt memory-1 strategies, this type of game can be represented by a Markov chain
- One can explore the equilibria among the memory-1 strategies and their dynamics with similar methods





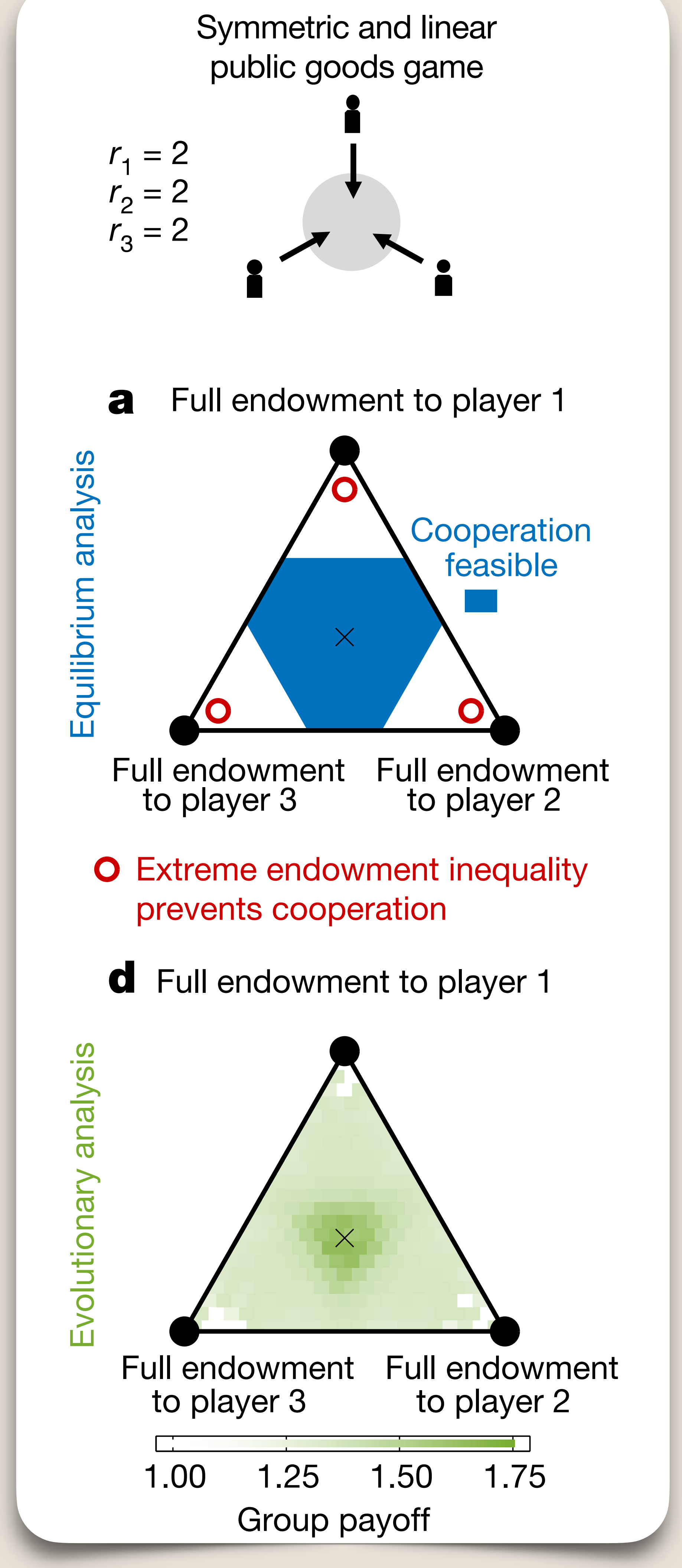
Hilbe, Simsa, Chatterjee, Nowak (Nature 2018)

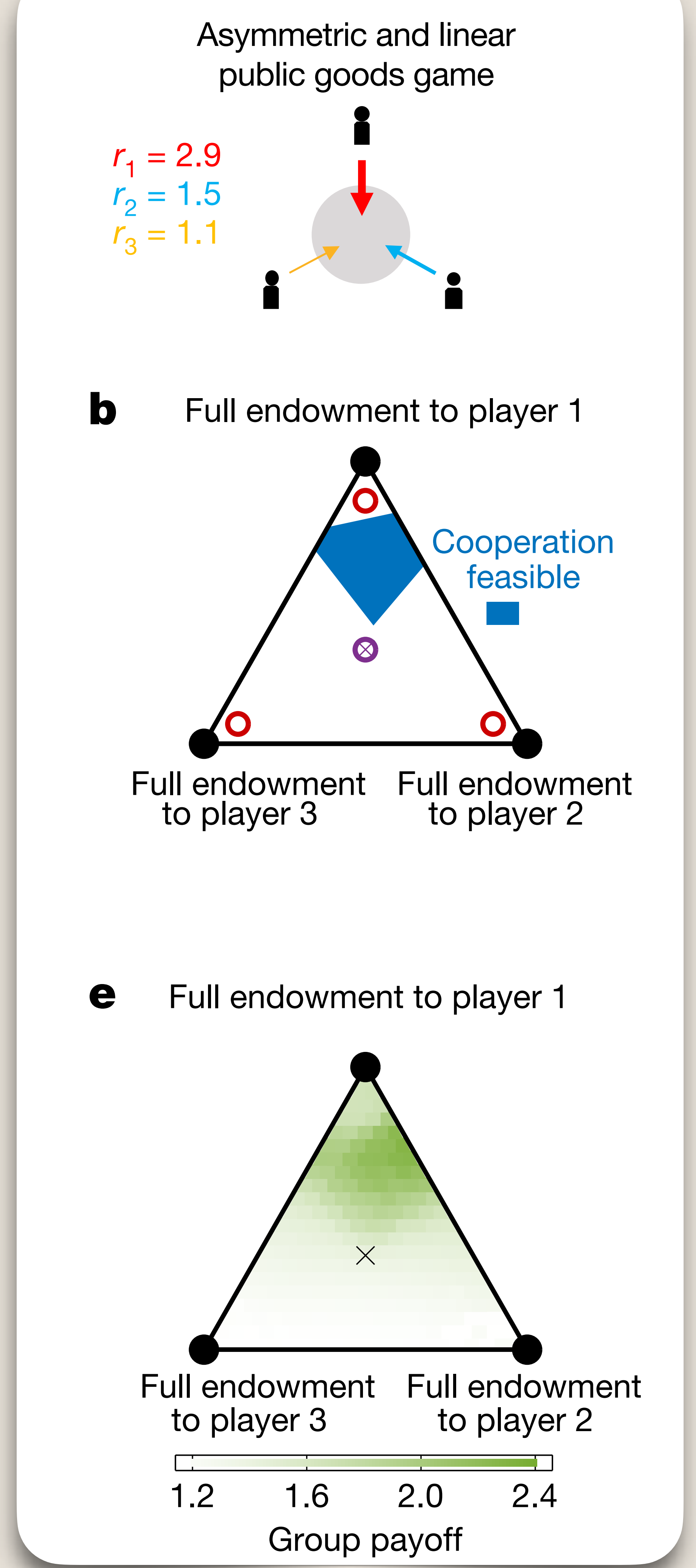
Games with inequality: Asymmetric public good game

- Repeated interactions among n players, with continuation probability δ
- In the beginning, each player receives an endowment $e_i > 0$. Endowments are normalised $e_1 + \ldots + e_n = 1$
- Each player i decides which fraction x_i of the endowment to contribute to a public good. A players' contributions are multiplied by some productivity factor $r_i > 1$, and then equally divided among all group members.

$$\text{Payoff: } \pi_i = \frac{1}{n}\sum_{j=1}^{n} r_j x_j e_j + (1-x_i)e_i$$

- **Question:** What is the best way to allocate endowments among players to maximise cooperation?



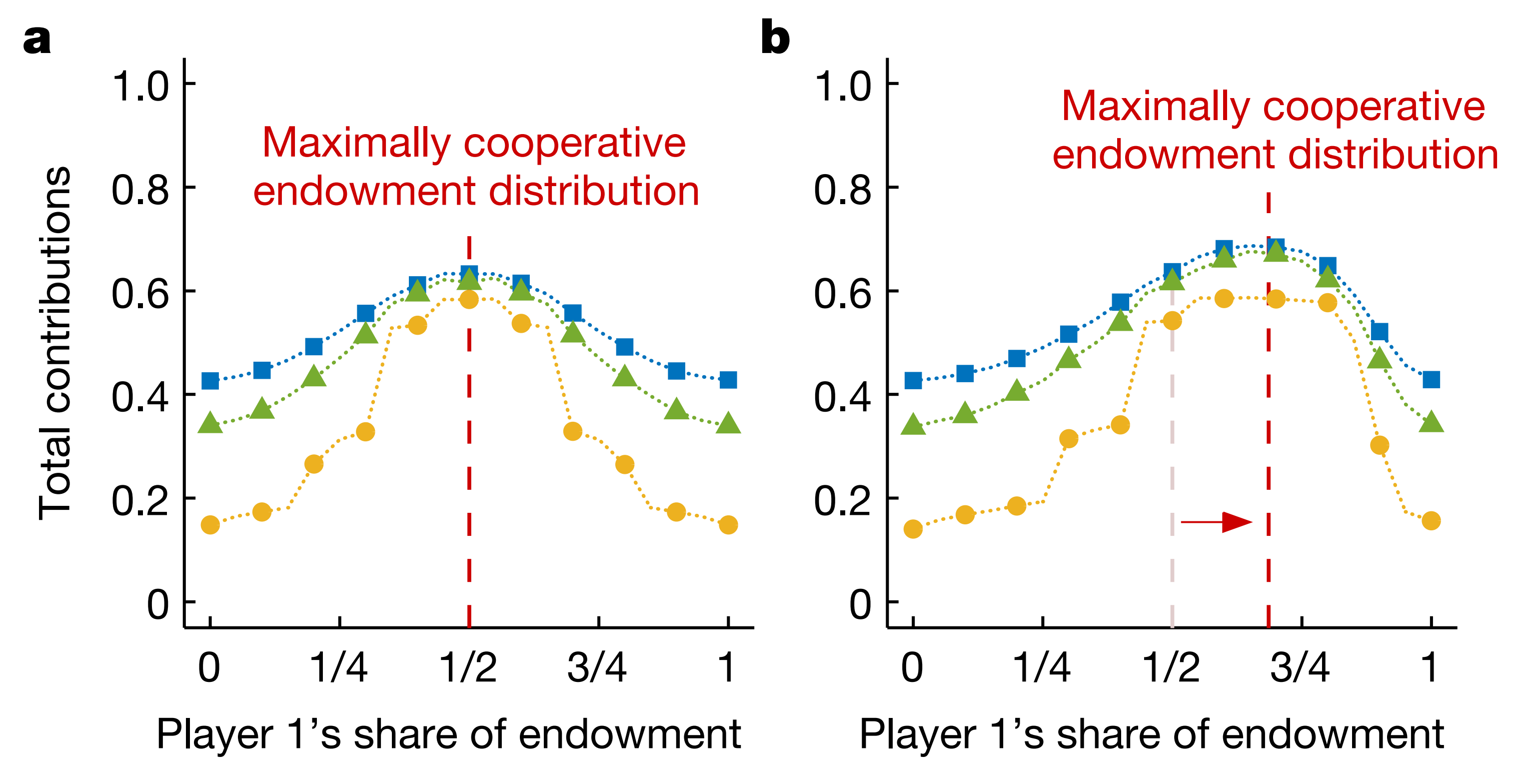


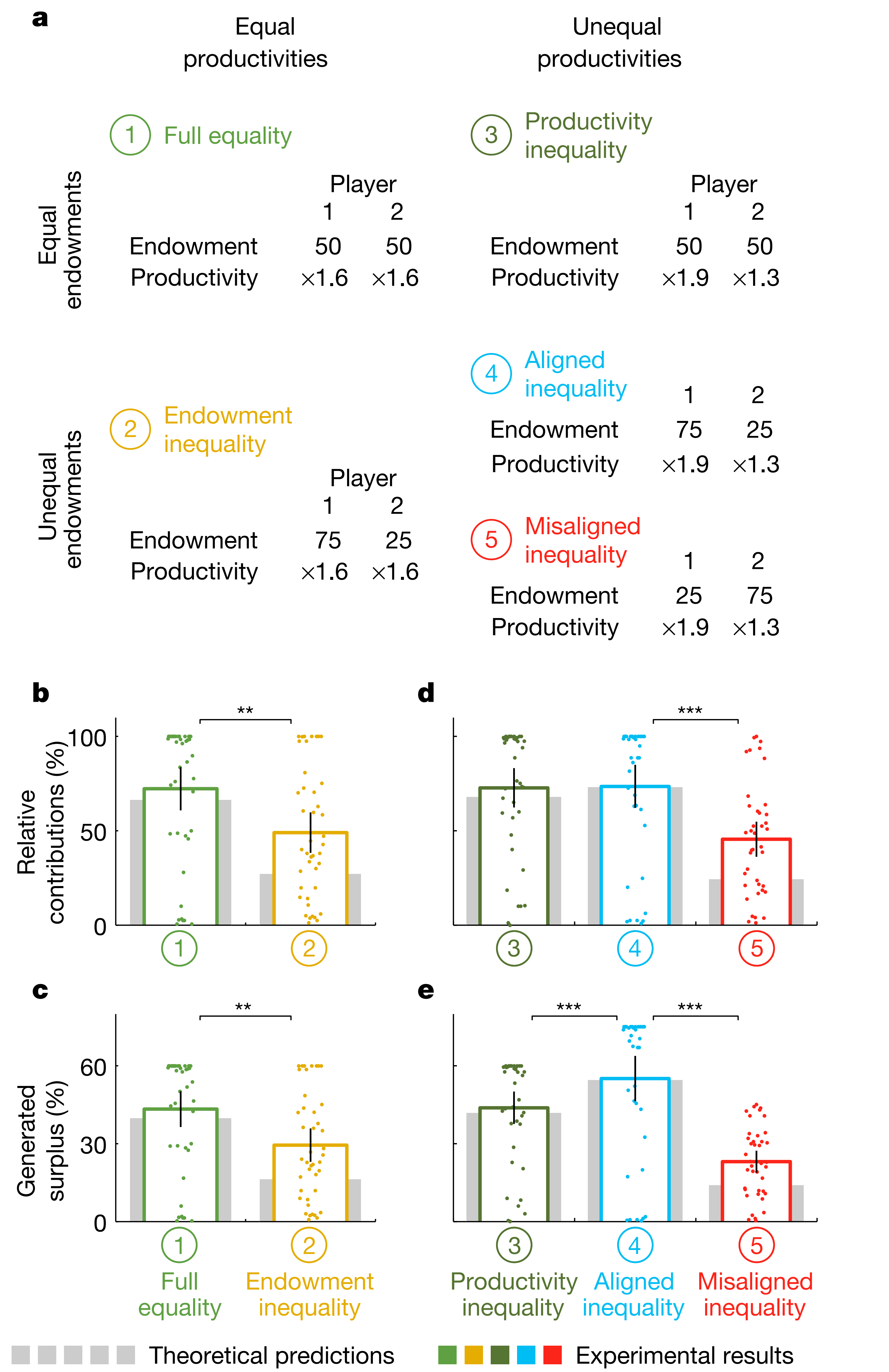
Hauser, Hilbe, Chatterjee, Nowak (Nature 2019)

Definition: An endowment distribution $(e_1, \dots, e_n)$ is said to be *most conducive to cooperation* if it allows full cooperation to be stable with the smallest possible continuation probability δ.

Proposition: In the special case of $n = 2$ players, the endowment distribution most conducive to cooperation satisfies

$$\frac{e_1}{e_2} = \sqrt{\frac{r_2(2-r_2)}{r_1(2-r_1)}}.$$





Hauser, Hilbe, Chatterjee, Nowak (Nature 2019)

Summary

Max-Planck-Institut für
Evolutionsbiologie, Plön

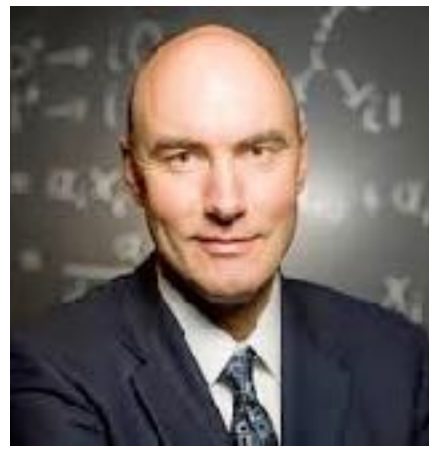
Background:

Direct reciprocity is an important mechanism for cooperation based on repeated interactions. With models of direct reciprocity, one can explore which environments are most favorable to the evolution and stability of cooperation.

Some conclusions:

Already with the most simple strategies of direct reciprocity, memory-1 strategies, players can have remarkable control on the game dynamics.
Cooperation is feasible (and sometimes even promoted) if there is environmental feedback or endowment inequality among players.

Some (generous) co-players:






Thank you for your attention!