漫谈人工智能

从囚徒困境谈起：博弈智能

荣智海

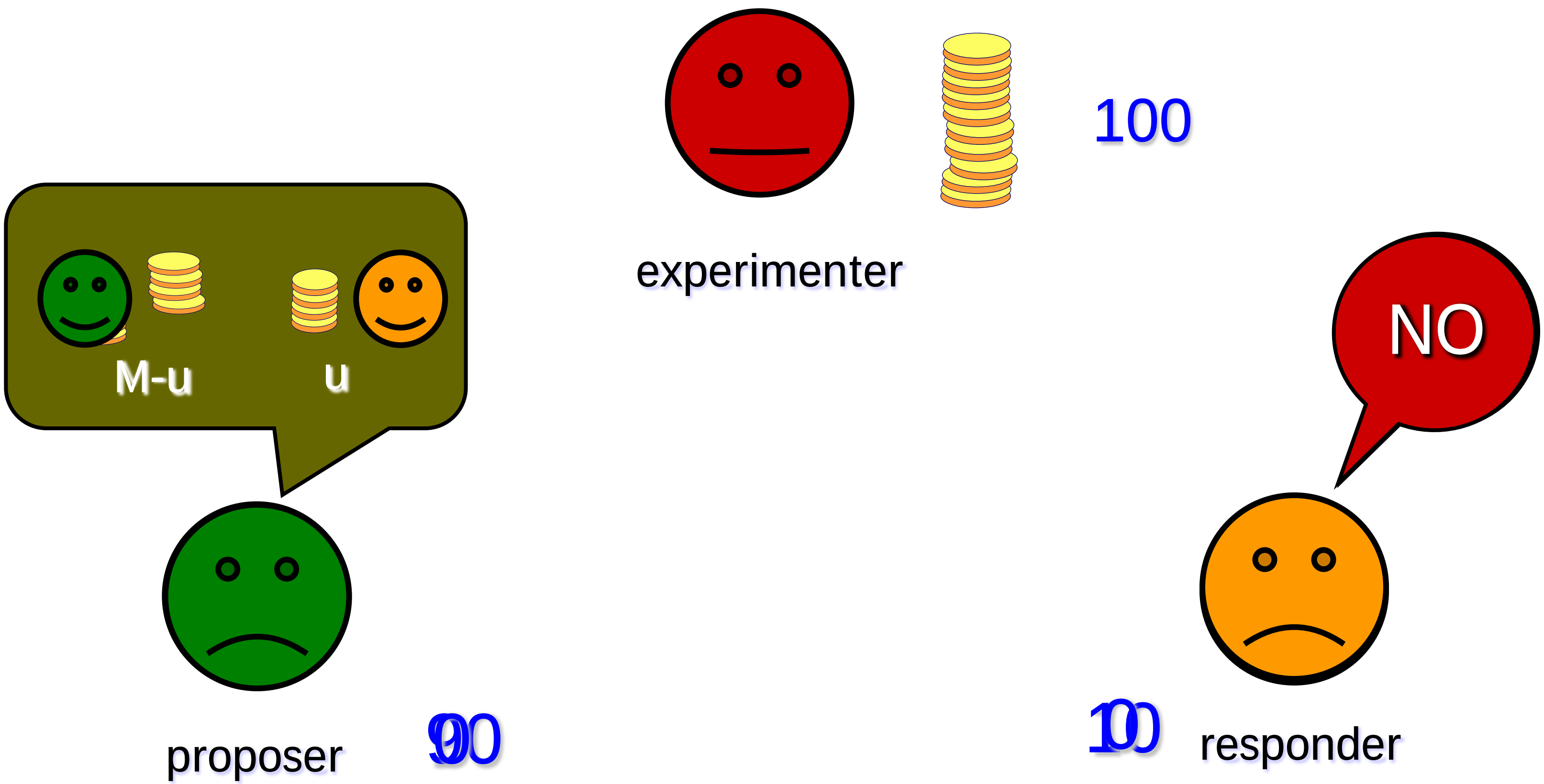
rongzhh@gmail.com

大数据研究中心,电子科技大学,成都

2020.09

Altruistic punishment: Ultimatum Game

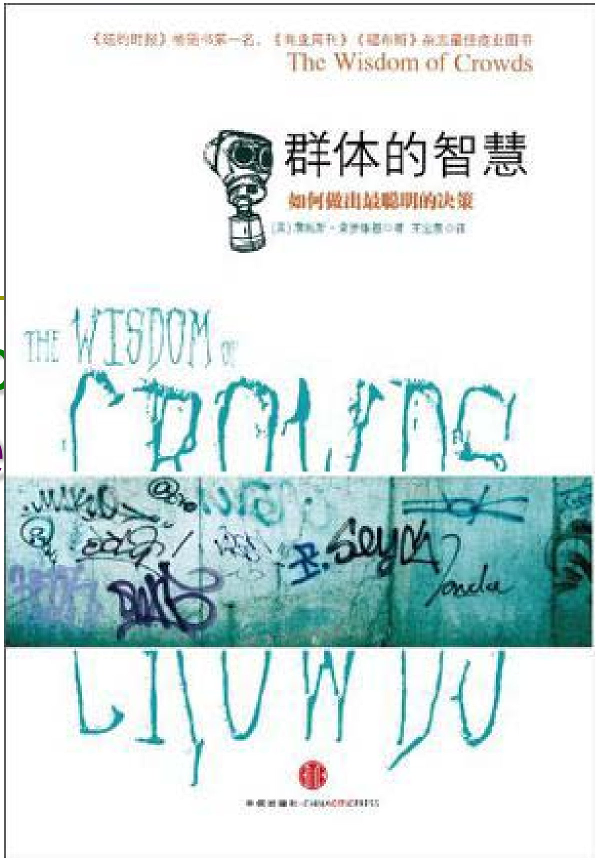
(Güth, Schmittberger & Schwarze, 1982)



Experimental results

Responder’s optimal strategy: ac

Proposer’s optimal strategy: offe

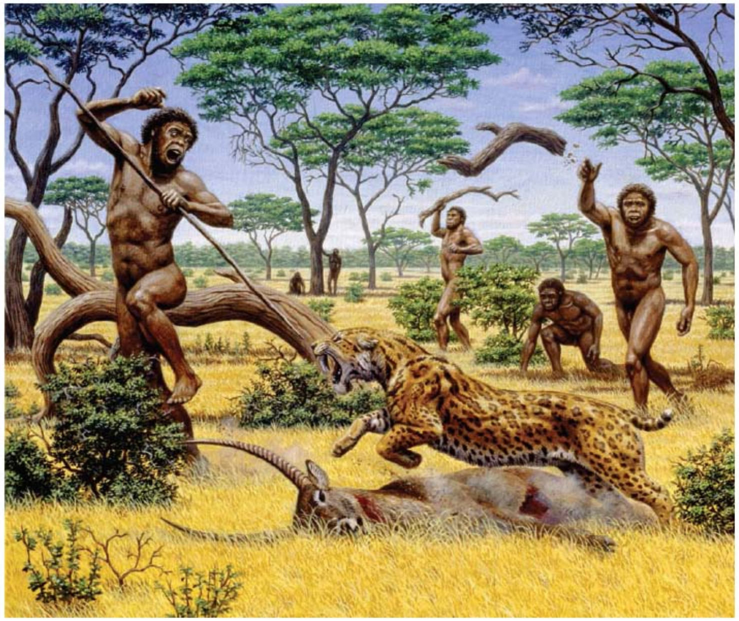


Results:

- Responders reject offers below probability
- Large degree of variability of offers among societies (26 - 58%)

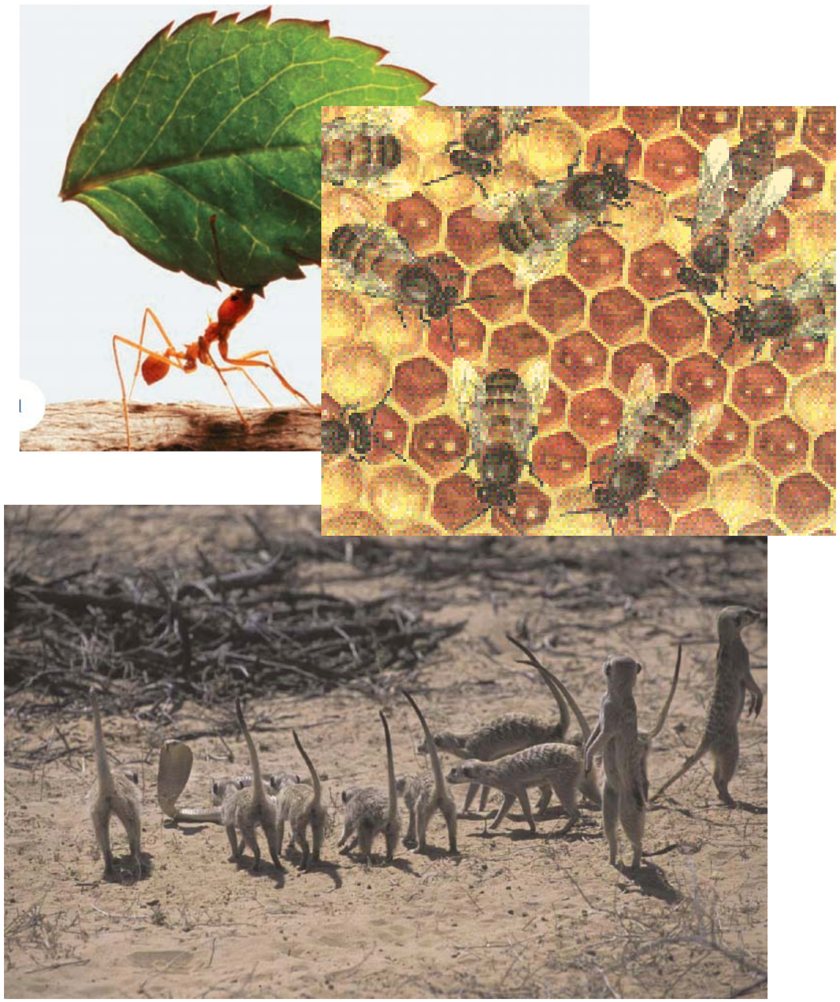
- Paciotti, Brian, Craig Hadley, Christopher Holmes and Monique Borgerhoff Mulder, Grass-roots Justice in Tanzania, American Scientist(2005)
- Sigmund K, Fehr E, Nowak M A, The economics of fair play, Scientific American(2002)

合作: 人类社会的基石



Robert Boyd and Sarah Mathew, *A Narrow Road to Cooperation*, SCIENCE,2007

Cooperation is *ubiquitous* in nature !



Vampire bats share blood

COMMON VAMPIRE BATS frequently engage in acts of mutual help. A bat that feeds successfully on blood from horses or cattle will share its nourishment with an unfed companion by regurgitating a portion of its stomach contents.

- **Nowak MA (2012). Why we help. Sci Am 307 (1): 34-39.**

白腹鹞原想抓八哥美餐，被鸟群团团包围逼退

- 一只单打独斗的白腹鹞原想抓只在湖旁休息的八哥鸟饱餐一顿，不料，八哥鸟群迅速以庞大的队形包围落单的白腹鹞，使得它不得不落荒而逃。
- 八哥鸟群在面对像猎鹰、白腹鹞这样的猛禽捕捉时都有自成一套的:它们不仅会利“团结防御策略”用数量优势将敌人包围起来，还会不断变换队形。

6

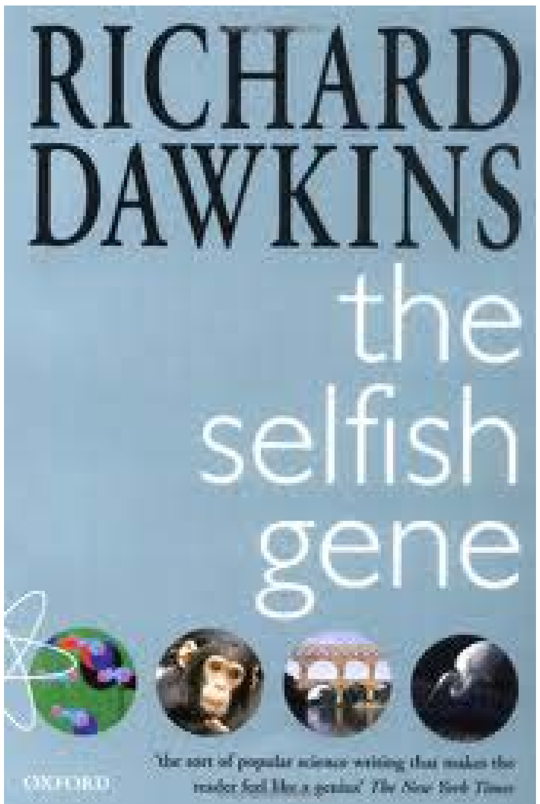
Gene is selfish?

NEWS & VIEWS

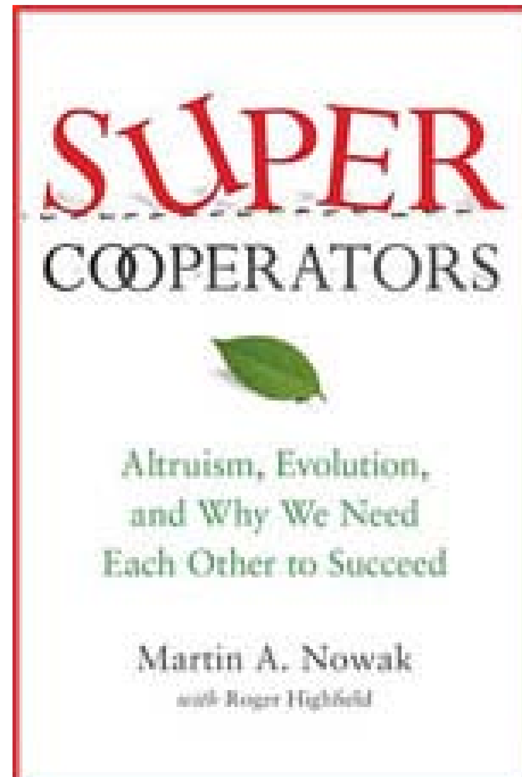
ORIGINS OF LIFE

The cooperative gene

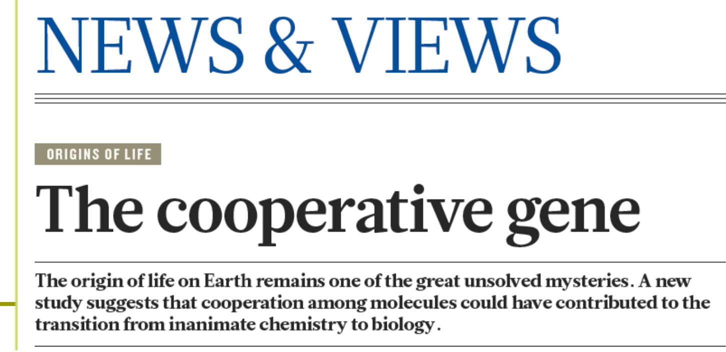
The origin of life on Earth remains one of the great unsolved mysteries. A new study suggests that cooperation among molecules could have contributed to the transition from inanimate chemistry to biology.



Nature 2012



RNA世界中的合作


NEWS & VIEWS

ORIGINS OF LIFE

The cooperative gene

The origin of life on Earth remains one of the great unsolved mysteries. A new study suggests that cooperation among molecules could have contributed to the transition from inanimate chemistry to biology.

- 如果**RNA**分子能够相互作用、而不是独立发挥功能，生命和演化将会更容易实现。
- 由组装成核酶的**RNA**片段形成的合作性网络能够竞争过自催化的**RNA**片段。
- 合适的策略可以促使不同的**RNA**自组装成核酶，搭建成合作的超循环回路，竞争过自催化的**RNA**片段。
- **RNA**分子具有形成更复杂分子的内在能力，该行为在地球上的生命形成过程的早期已经确立。

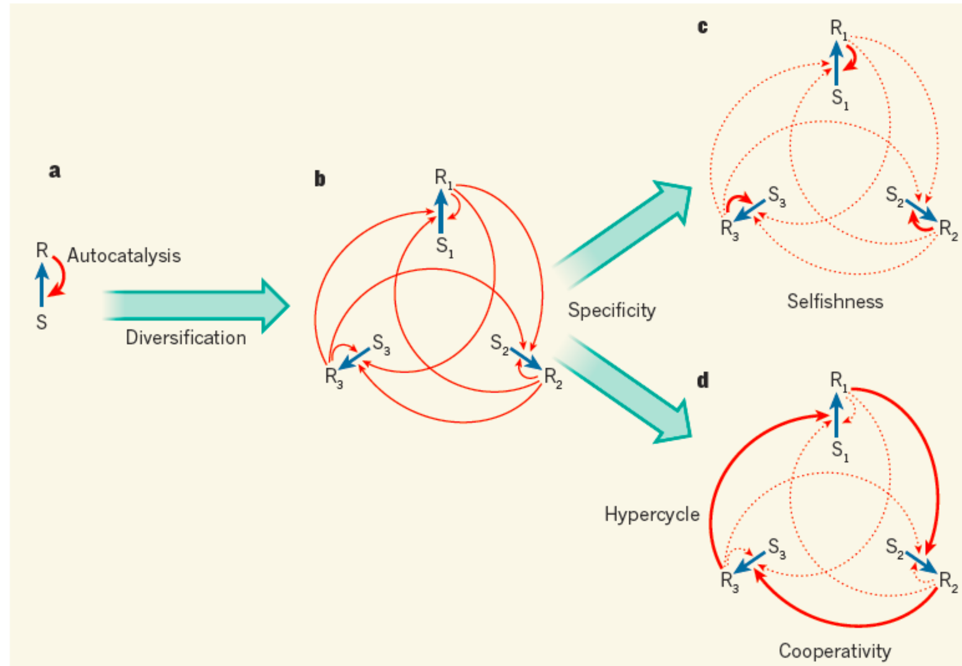


Figure 1 | The emergence of hypercycles. **a,** A primordial replicator molecule (R) enhances its own assembly from substrate molecules (S) in a simple autocatalytic cycle. **b,** Imperfect replication generates a set of related replicators, each promoting the synthesis of all the others. **c, d,** The introduction of biases in replicator specificity gives structure to the network and can lead to selfish subsystems (**c**) or to a cooperative 'hypercycle' (**d**), akin to the system described by Vaidya and colleagues[1]. Such hypercycles remain globally autocatalytic, but are more resistant to the accumulation of mutations, enabling replicators to specialize and to acquire new functions. Thick and dashed red arrows indicate increased and decreased efficacy, respectively, at enhancing replicator assembly.

Vaidya, N. et al. Spontaneous network formation among cooperative RNA replicators, Nature2012

Part II 囚徒困境及其策略

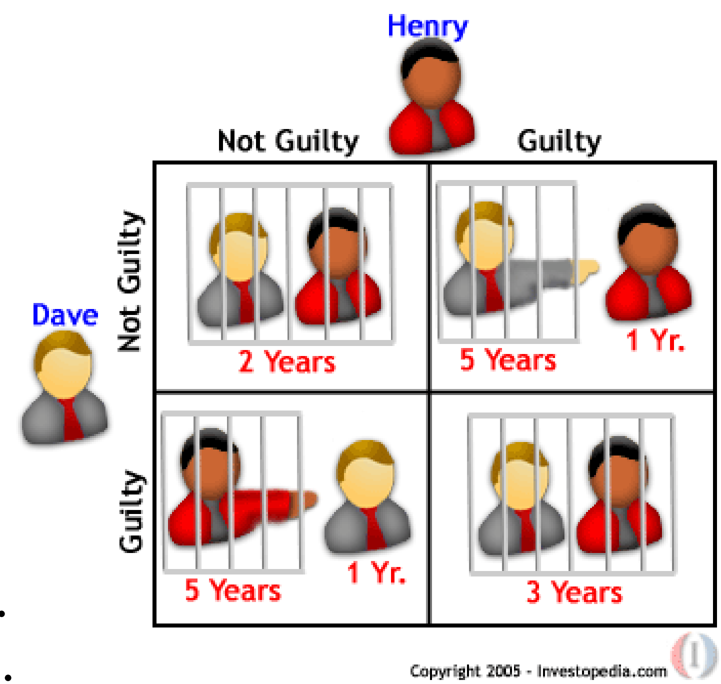
囚徒困境博弈Prisoner's dilemma game,PDG)

- Cooperator (合作者,C): help others at a cost to himself.
- Defector (背叛者D): receive the benefits without providing help.

Henry

Dave		C (Reject)	D (Confess)
	C	(-2,-2)	(-5,-1)
	D	(-1,-5)	(-3,-3)

(Dave gets, Henry gets)



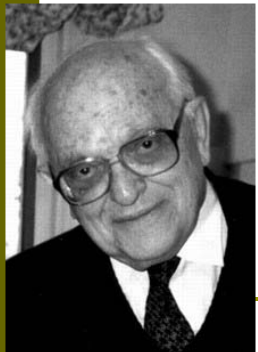
If you opponent plays C: you better play D.
If you opponent plays D: you better play D.

But,

CC is better than DD

Dilemma:

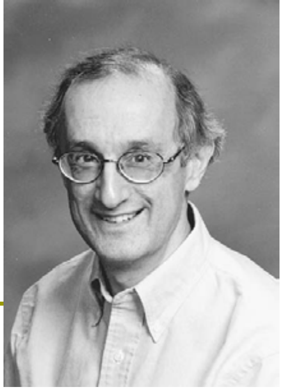
Despite mutual cooperation being the best, individual tends to DD



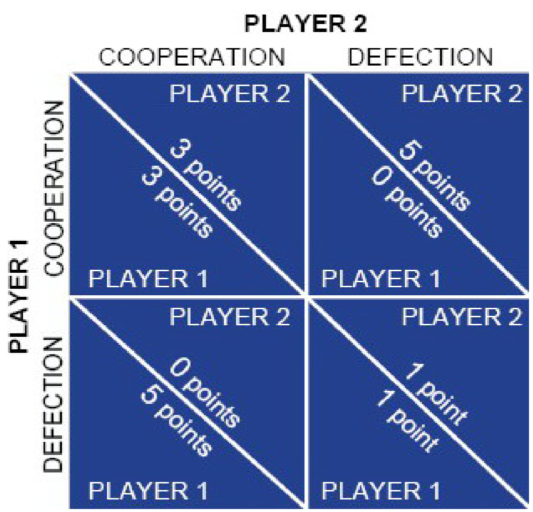
Tit for tat (针锋相对,一还一报,TFT)

A. Rapoport & A. M. Chammah, Prisoner's dilemma: A study in conflict and cooperation, The University of Michigan, 1965

R. Axelrod and W. D. Hamilton, Science 211, 1390 (1981)



- TFT: Strategy tournament: TFT winner
- In general, the best strategies are "nice(善良)", "punishing(惩罚)" and "forgiving(宽容)"

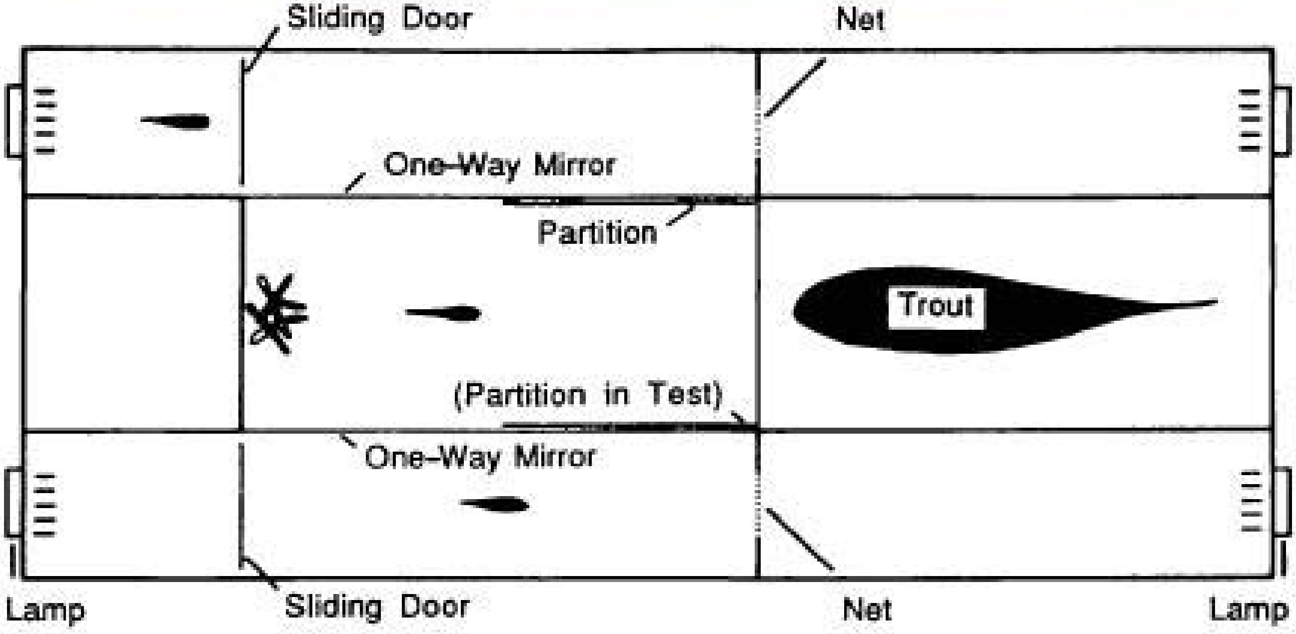


PRIOR PLAY			
Last move	Opponent's move	Outcome	Tit-for-Tat
C	C	"REWARD"	C
C	D	"LOSER'S PAYOFF"	D
D	C	"TEMPTATION"	C
D	D	"PUNISHMENT"	D

11

棘鱼(stickleback)间的TFT

Milinski M.(1987),TIT FOR TAT in sticklebacks and the evolution of cooperation,Nature



TFT vs AllD

$$\begin{array}{c|cc} & C & D \\ C & R=1 & S=1-b \\ D & T=b & P=0 \end{array}$$

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R=1 & S=1-b \\ T=b & P=0 \end{pmatrix} \end{array}$$

- 考虑回合数不确定的重复囚徒困境博弈,进行下一轮博弈的概率为w，则期望回合数m=1/(1-w)

$$\begin{array}{cc} & \begin{array}{cc} \text{TFT} & \text{AllD} \end{array} \\ \begin{array}{c} TFT \\ \text{AllD} \end{array} & \begin{pmatrix} mR & S+(m-1)P \\ T+(m-1)P & mP \end{pmatrix} \end{array}$$

$$\begin{array}{cc} & \begin{array}{cc} \text{TFT} & \text{AllD} \end{array} \\ \begin{array}{c} TFT \\ \text{AllD} \end{array} & \begin{pmatrix} m & 1-b \\ b & 0 \end{pmatrix} \end{array}$$

$$PD: b>1;$$

- 那么，如果m>b，则TFT对于AllD是帕累托占优的。

- 与TFT的博弈实验…

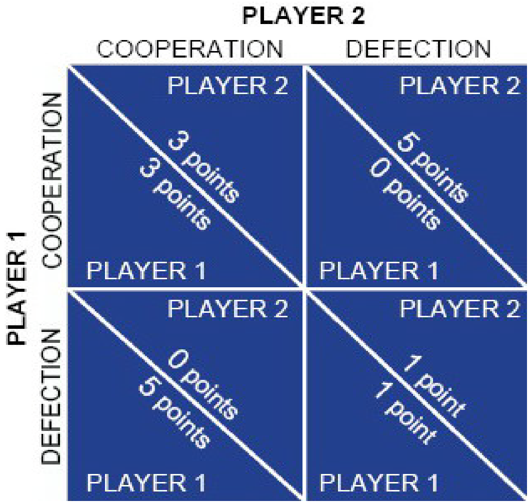
$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix} \end{array}$$

Tit for tat (针锋相对,一还一报,TFT)

*R. Axelrod and W. D. Hamilton, Science **211**, 1390 (1981)*

TIT-FOR-TAT	C	C	C	C	D	C	D
TIT-FOR-TAT	C	C	C	D	C	D	C

- TFT: Strategy tournament: TFT winner
- In general, the best strategies are "nice", "punishing" and "forgiving"
- But there exists noise?



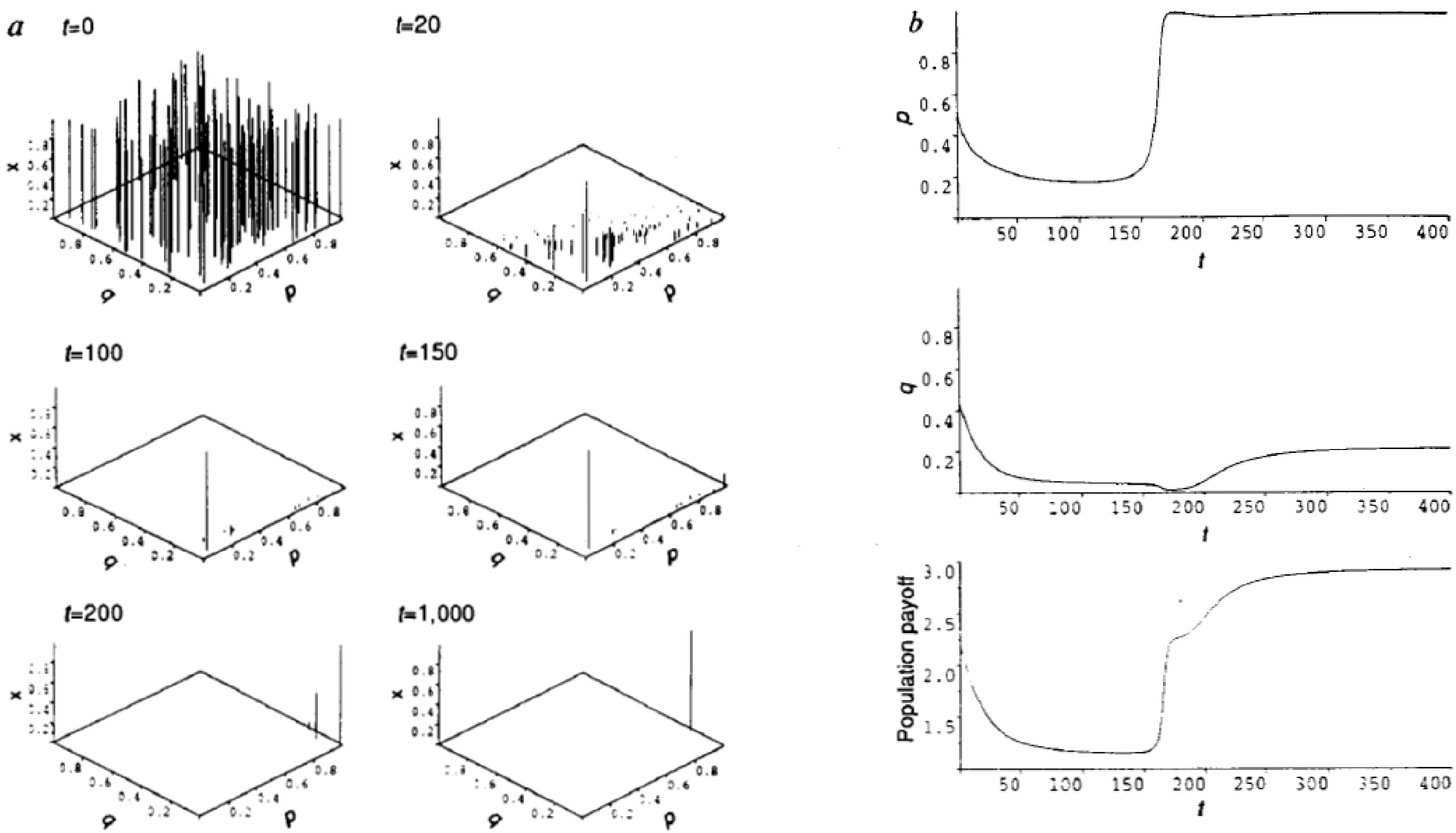
PRIOR PLAY			
Last move	Opponent's move	Outcome	Tit-for-Tat
C	C	"REWARD"	C
C	D	"LOSER'S PAYOFF"	D
D	C	"TEMPTATION"	C
D	D	"PUNISHMENT"	D

14

宽容的TFT (Generous tit-for-tat ,GTFT)

Nowak, Sigmund (1992). Tit for tat in heterogeneous population. Nature

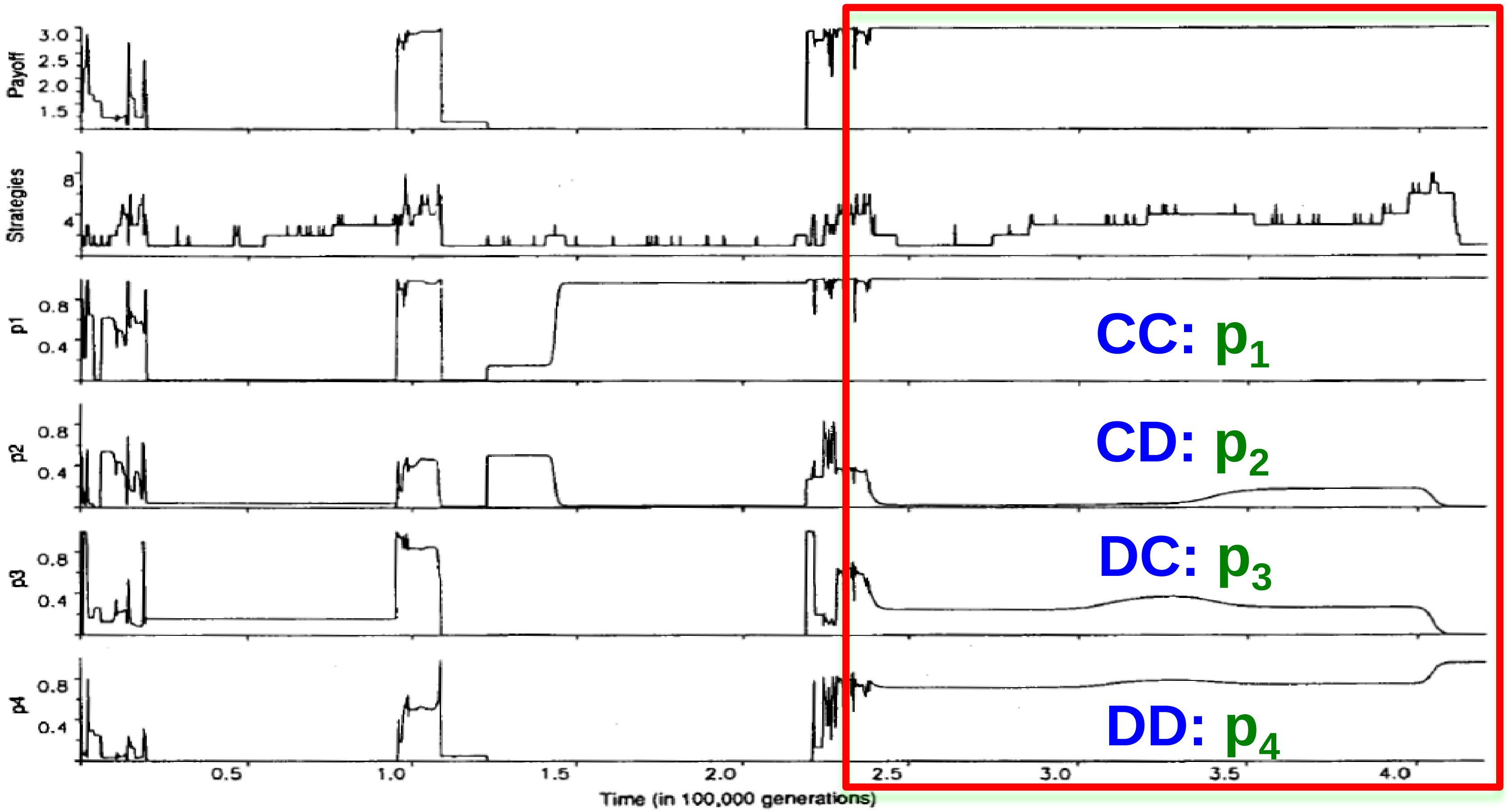
- p: the conditional probability to cooperate after a C
- q: the conditional probability to cooperate after a D
- TFT(p,q)=(1,0), GTFT=(1,0.33)



Win stay, lost shift (WSLS，赢存输去)

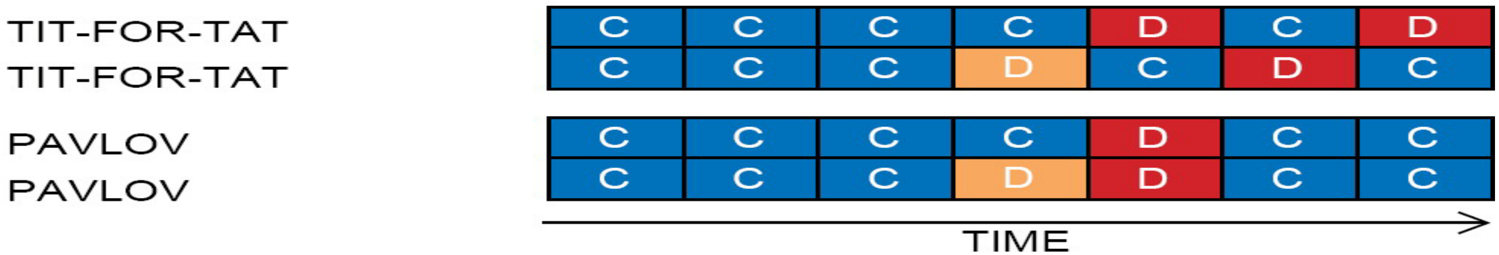
Nowak, Sigmund (1993). A strategy of win-stay, lose-shift that outperforms tit for tat in Prisoner's Dilemma. Nature 364

- (p1,p2,p3,p4) is the conditional probabilities to cooperate, after (CC, CD, DC, DD).
- TFT= (1,0,1,0), GTFT=(1,0.33,1,0.33),WSLS=(1,0,0,1)



Win stay, lost shift (WSLS, 赢存输去)

Nowak, Sigmund (1993). A strategy of win-stay, lose-shift that outperforms tit for tat in Prisoner's Dilemma. Nature 364



- Win stay, lost shift (Pavlov):

$$
\begin{array}{cc}
 & \overbrace{\begin{array}{cc} \mathrm{C} & \mathrm{D} \end{array}}^{\text{player2}} \\
\text{player1}\left\{\begin{array}{c} \mathrm{C} \\ \mathrm{D} \end{array}\right. & \begin{pmatrix} R & S \\ T & P \end{pmatrix}
\end{array}
$$

$$PD: T > R > x > P > S$$

R is **REWARD** for mutual cooperation
S **SUCKER's** payoff
T **TEMPTATION** to defect
P **PUNISHMENT** for mutual defection

PRIOR PLAY				
Last move	Opponent's move	Outcome	Tit-for-Tat	Pavlov
C	C	"REWARD"	C	C
C	D	"LOSER'S PAYOFF"	D	D
D	C	"TEMPTATION"	C	D
D	D	"PUNISHMENT"	D	C

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)

From extortion to generosity, evolution in the Iterated Prisoner's Dilemma

Alexander J. Stewart and Joshua B. Plotkin[1]

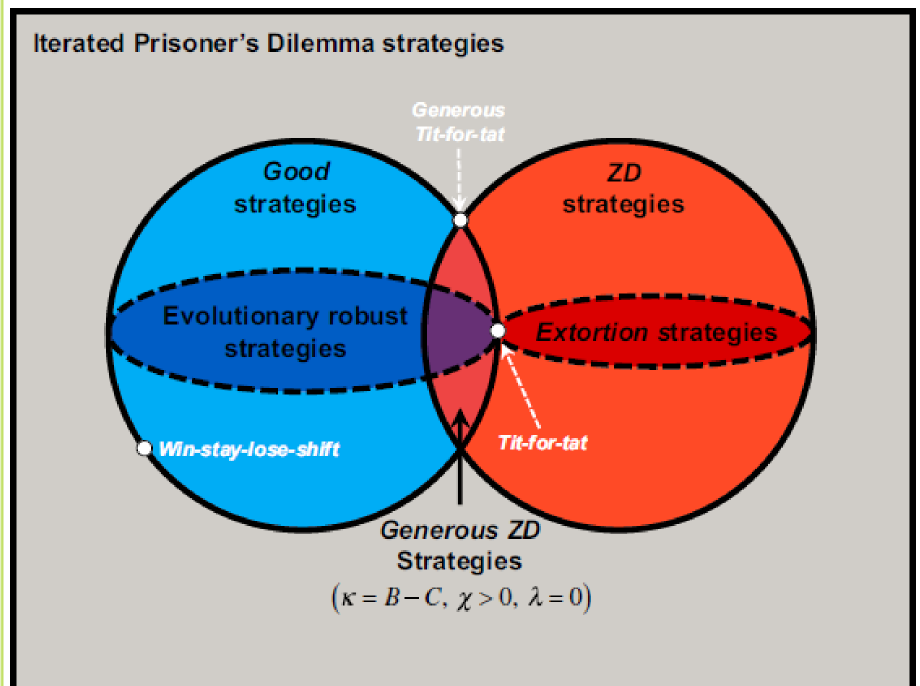
Department of Biology, University of Pennsylvania, Philadelphia, PA 19104

$$\tilde{\mathbf{p}} = \phi[(\mathbf{S}_X - \kappa\mathbf{1}) - \chi(\mathbf{S}_Y - \kappa\mathbf{1})]$$

- κ =P for extortion
- κ=R for generosity

$$\begin{array}{cc} & \overbrace{\begin{array}{cc}\mathrm{C} & \mathrm{D}\end{array}}^{\text{player2}} \\ \text{player1}\left\{\begin{array}{c}\mathrm{C} \\ \mathrm{D}\end{array}\right. & \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array}$$

$$PD: T > R > P > S$$



Part III 其他两人两博弈模型—
雪堆博弈(Snowdrift game ,SG)
猎鹿博弈(Stag hunt game, SH)

Snowdrift game (SG,雪堆博弈)

$$\begin{array}{cc} & \overbrace{\begin{array}{cc} C & \quad\quad\quad D \end{array}}^{\text{player2}} \\ \text{player1}\left\{\begin{array}{c} C \\ D \end{array}\right. & \begin{pmatrix} b-\frac{c}{2}, & b-c \\ b, & 0 \end{pmatrix} \end{array}$$

$C: cooperator; D: defector$

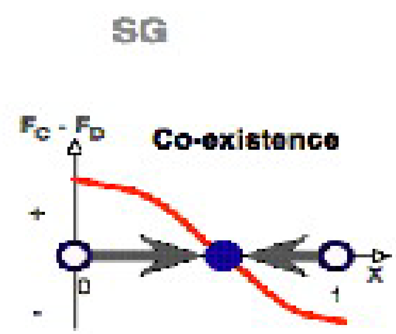
$$\Rightarrow \begin{pmatrix} 1, & 1-\dfrac{c}{2b-c} \\ 1+\dfrac{c}{2b-c}, & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1, & 1-r \\ 1+r, & 0 \end{pmatrix}$$

r=c/(2b-c): cost to benefit of mutual cooperation

$$\begin{array}{cc} & \overbrace{\begin{array}{cc} C & D \end{array}}^{\text{player2}} \\ \text{player1}\left\{\begin{array}{c} C \\ D \end{array}\right. & \begin{pmatrix} 1 & S \\ T & 0 \end{pmatrix} \end{array}$$

C and D coexist: C→1-r←D

The equilibrium frequency of cooperators in SG is 1-r



Chicken game(胆小鬼博弈)& Hawk-Dove game(鹰鸽博弈)

- 一些例子:
  - ✓ 鸟类和动物(蒙哥)的报警
  - ✓ 冷战中的美苏关系

21

Game theory models(几类博弈模型)

- Cooperator(C, 合作者): help others at a cost to themselves.
- Defector(D, 背叛者): receive the benefits without providing help.

> Prisoner's Dilemma (PD) : $T>R>P>S$, $T+S\le 2R$

D dominates C : C→D

$$\begin{array}{ccc} & & \overbrace{\begin{array}{cc}\mathrm{C} & \mathrm{D}\end{array}}^{\text{player2}} \\ \text{player1}\left\{\begin{array}{c}\mathrm{C}\\ \mathrm{D}\end{array}\right. & & \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array}$$

R is **REWARD** for mutual cooperation
S **SUCKER's** payoff
T **TEMPTATION** to defect
P **PUNISHMENT** for mutual defection

22

Game theory models(几类博弈模型)

- Prisoner's Dilemma (PD) : T>R>P>S, T+S≤2R

  D dominates C : C→D
- Snowdrift game (SG): T>R>S>P, T+S≤2R

  C and D coexist: C→r←D
- Stag hunt game (SH,猎鹿博弈; Battle of the sexes,性别大战):

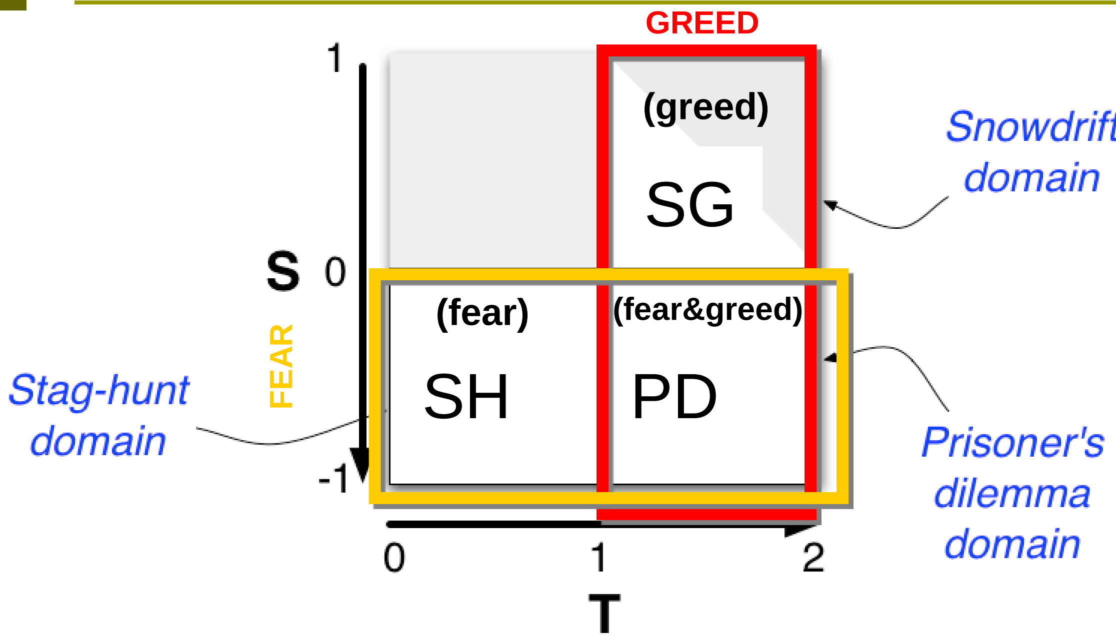
  R>T>P>S, C and D are bistable: C←→D
- C dominates, C←D, R>T>S>P

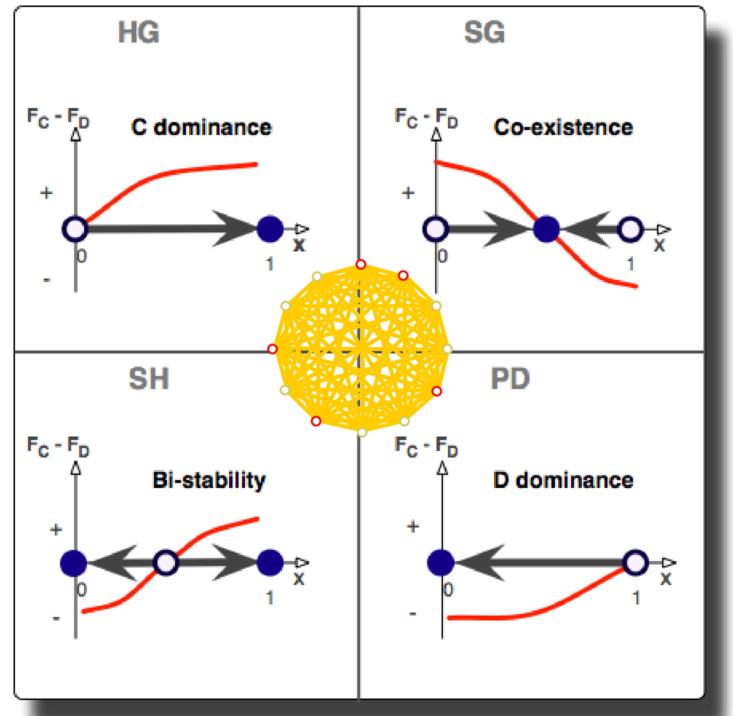
$$
\begin{array}{cc}
 & \overbrace{\begin{array}{cc}\mathrm{C} & \mathrm{D}\end{array}}^{\text{player2}} \\
\text{player1}\left\{\begin{array}{c}\mathrm{C} \\ \mathrm{D}\end{array}\right. & \begin{pmatrix} R & S \\ T & P \end{pmatrix}
\end{array}
$$

23

2D parameter space

$$
\begin{array}{c c c}
 & & \overbrace{\begin{array}{cc} \mathrm{C} & \mathrm{D} \end{array}}^{\text{player2}} \\
\text{player1} & \left\{ \begin{array}{c} \mathrm{C} \\ \mathrm{D} \end{array} \right. & \begin{pmatrix} 1 & S \\ T & 0 \end{pmatrix}
\end{array}
$$

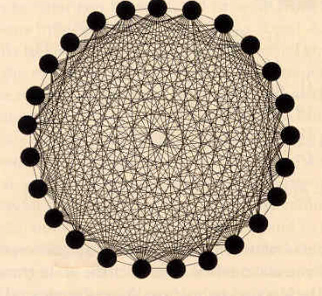




- Formally, these dilemmas span the parameter space of
- **T** (temptation to defect = greed)
- **S** (sucker's payoff = fear).

Santos, Pacheco, Lenaerts, PNAS 103 (2006) 3490-3494

24

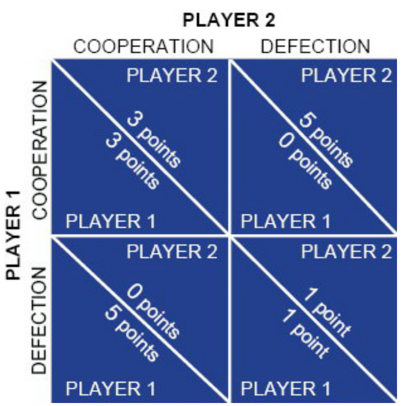


Defectors can diffuse in well-mixed populations

An Austrian holiday led to iterated cooperation between Karl Sigmund (left) and Martin Nowak.

Prisoners of the dilemma: When mathematics and biology met on a mountain. Nature 2004

- Cooperator (C): help others at a cost to himself.
- Defector (D): receive the benefits without providing help.



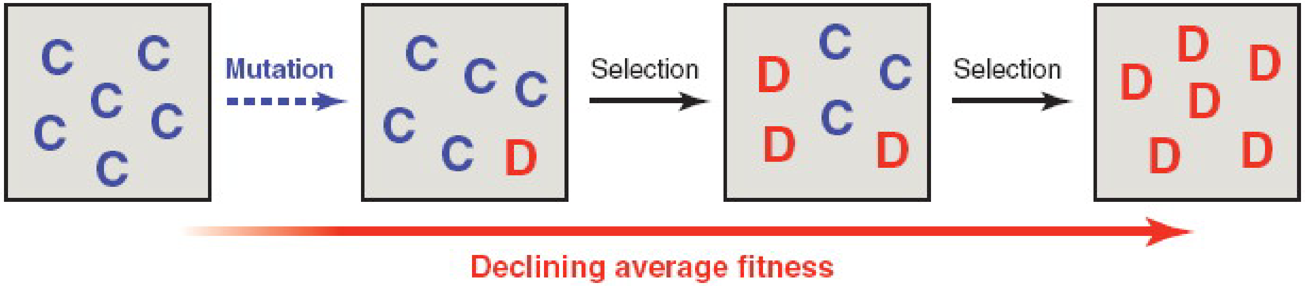
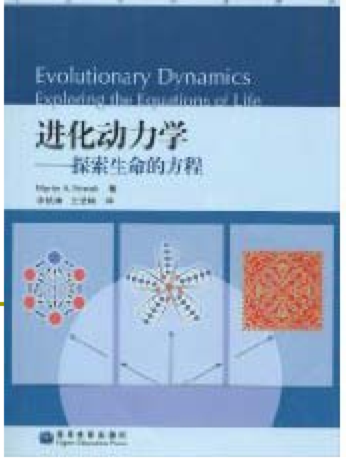


Fig. 1. Without any mechanism for the evolution of cooperation, natural selection favors defectors. In a mixed population, defectors, D, have a higher payoff (= fitness) than cooperators, C. Therefore, natural selection continuously reduces the abundance, i, of cooperators until they are extinct. The average fitness of the population also declines under natural selection. The total population size is given by N. If there are i cooperators and $N - i$ defectors, then the fitness of cooperators and defectors, respectively, is given by $f_C = [b(i-1)/(N-1)] - c$ and $f_D = bi/(N-1)$. The average fitness of the population is given by $\bar{f} = (b-c)i/N$.

Nowak MA (2006). Five rules for the evolution of cooperation. Science

Natural cooperation(自然合作)

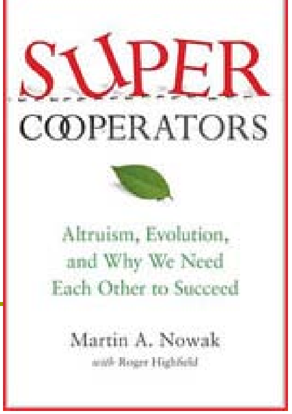
Nowak MA (2006). Five rules for the evolution of cooperation. Science



- “The two fundamental principles of evolution are mutation and natural selection. But evolution is constructive because of cooperation. New levels of organization evolve when the competing units on the lower level begin to cooperate. ...

- Thus, we might add “natural cooperation” as a third fundamental principle of evolution beside mutation and natural selection.”

Some rules for evolutions cooperation (一些演化合作的规则)



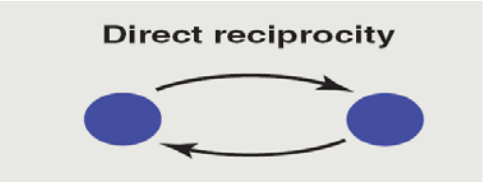
Nowak MA (2006). Five rules for the evolution of cooperation. Science

- Kin selection: relative

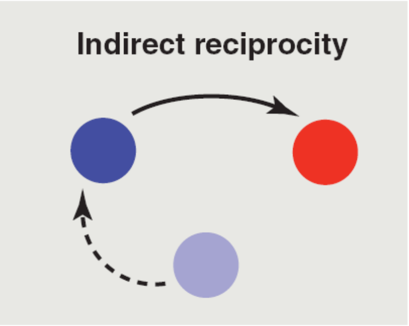
Cooperators   Defectors

“我会跳进河里去救我的两个弟弟或者八个堂兄”

- Direct reciprocity: unrelated individuals



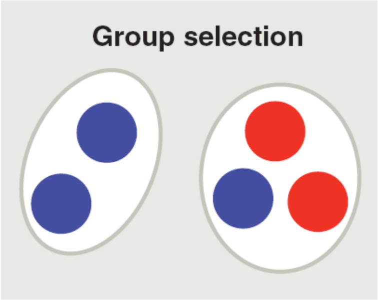
“如果你帮我挠挠后背，我也帮你挠”



- Indirect reciprocity: reputation

“为了得到大家的回报，我现在得多半住别人，为自己赢得一个好的名声”

- Group (Multi-level) selection



“一群相互协作的人往往比一群相会背叛的人更能获得成功”

- Network reciprocity

27

人工智能 vs 博弈论 vs 机器行为

- 自然(系统)哲学的数学原理→人造(系统)哲学的数学原理

REVIEW　**Nature 2019. 04. 25**

https://doi.org/10.1038/s41586-019-1138-y

Machine behaviour

Iyad Rahwan[1,2,3,34]*, Manuel Cebrian[1,34], Nick Obradovich[1,34], Josh Bongard[4], Jean-François Bonnefon[5], Cynthia Breazeal[1], Jacob W. Crandall[6], Nicholas A. Christakis[7,8,9,10], Iain D. Couzin[11,12,13], Matthew O. Jackson[14,15,16], Nicholas R. Jennings[17,18], Ece Kamar[19], Isabel M. Kloumann[20], Hugo Larochelle[21], David Lazer[22,23,24], Richard McElreath[25,26], Alan Mislove[27], David C. Parkes[28,29], Alex 'Sandy' Pentland[1], Margaret E. Roberts[30], Azim Shariff[31], Joshua B. Tenenbaum[32] & Michael Wellman[33]

Human–machine co-behaviour

Although it can be methodologically convenient to separate studies into the ways that humans shape machines and vice versa, most AI systems function in domains where they co-exist with humans in complex hybrid systems[39,67,125,128]. Questions of importance to the study of these systems include those that examine the behaviours that characterize human–machine interactions including cooperation, competition and coordination—for example, how human biases combine with AI to alter human emotions or beliefs[14,55,56,129,130], how human tendencies