

# INFO-F-409

## Learning dynamics

Evolutionary game theory, stochastic dynamics and the origins of co-operation



T. Lenaerts and A. Nowé  
MLG, Université Libre de Bruxelles and  
AI-lab, Vrije Universiteit Brussel



1

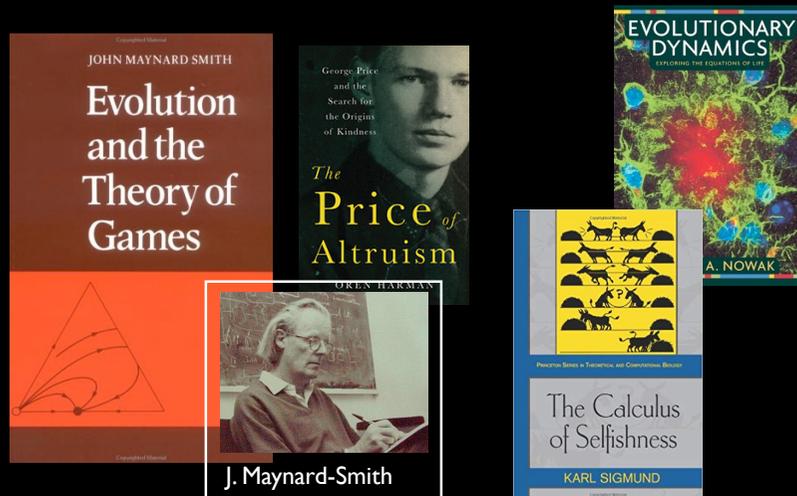
## Summary

- What? Why?
- Rational choice
- Strategic games
- Nash Equilibrium
- Best
- Dominance
- Mixed strategies
- Mixed-strategy Nash Equilibria
- Support finding
- Lemke-Howson algorithm
- Extensive-form games
- sub-game perfect equilibrium
- Simultaneous moves
- Chance moves
- Bayesian games
- Fictitious play and stimulus response learning

2

## An evolutionary perspective

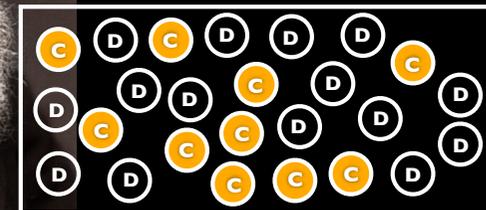
©Tom Lenaerts, 2016



3

Non-rational players: **one player for each action**

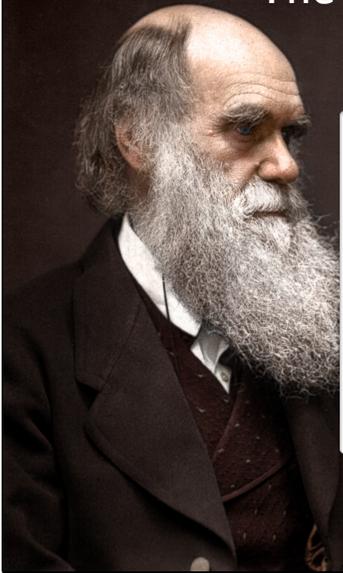
Darwinian competition between simple players within **populations**



Success of a player **depends on the frequencies** of the different types of players

4

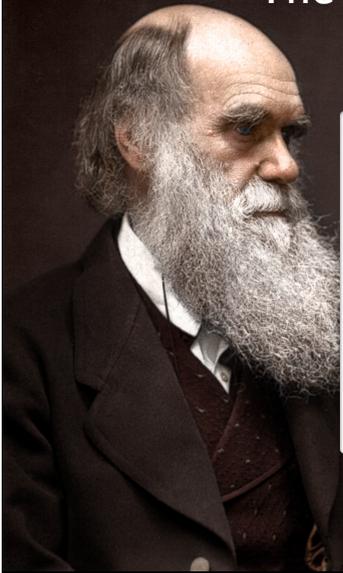
# The question of cooperation



Dilbert's prisoner dilemma

5-1

# The question of cooperation



Dilbert's prisoner dilemma

5-2

## Social dilemmas



6

## Defining the concept

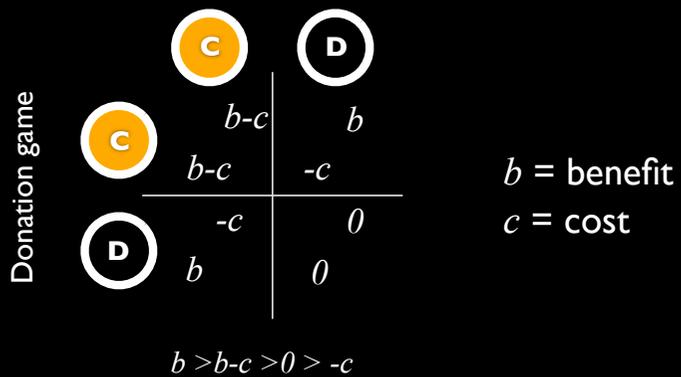
Social dilemmas are situations in which each member of a group has a clear and unambiguous incentive to make a choice that—when made by all members—provides poorer outcomes for all than they would have received if none had made the choice. Thus, by doing what seems individually reasonable and rational, people end up doing less well than they would have done if they had acted unreasonably or irrationally. This paradoxical possibility has emerged in many contexts and it has been

R.M. Dawes and D.M. Messick (2000) Social Dilemmas. International Journal of Psychology 35(2):111-116

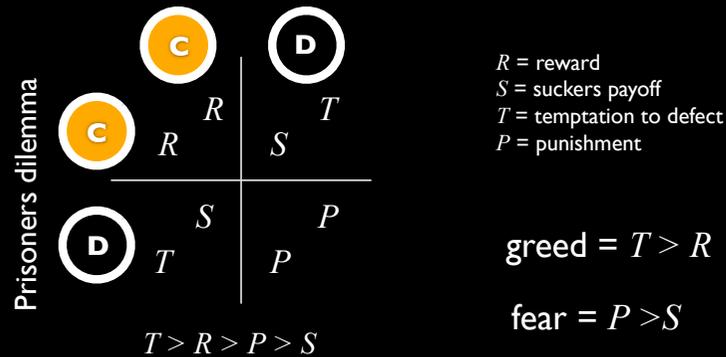
**THE QUESTION OF COOPERATION**  
Social dilemmas are situations in which individual rationality leads to collective irrationality. That is, individually reasonable behavior leads to a situation in which everyone is worse off than they might have been otherwise. Many of the most challenging problems we face, from the international to the internal

P. Kollock (1998) Social Dilemmas: the anatomy of cooperation Ann. Rev. Sociol. 24:183-214

7

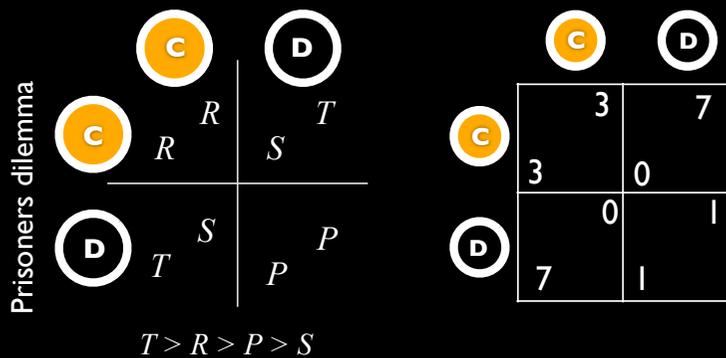


8



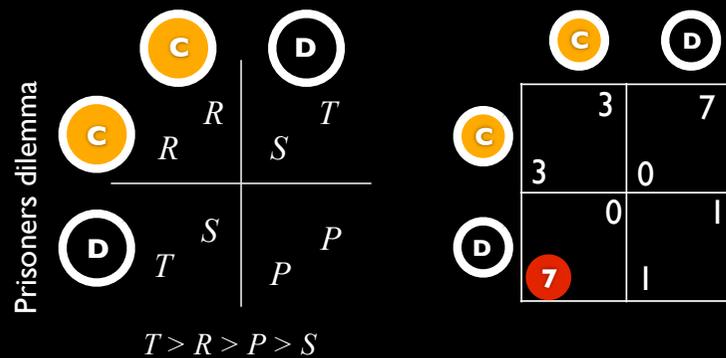
C.H. Coombs (1973) A reparameterization of the prisoner's dilemma game. Behavioral Science 18:424-428

9



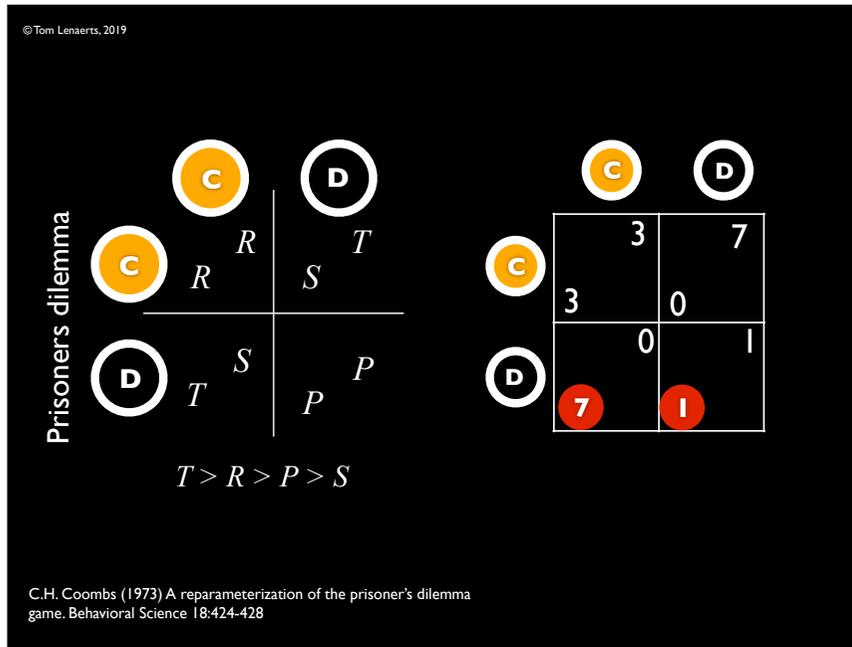
C.H. Coombs (1973) A reparameterization of the prisoner's dilemma game. Behavioral Science 18:424-428

10-1

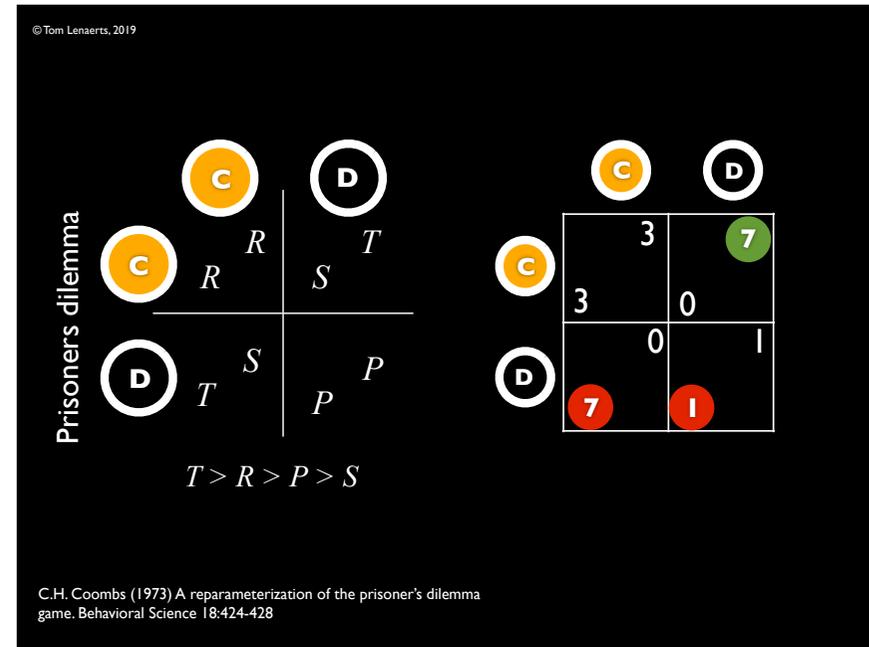


C.H. Coombs (1973) A reparameterization of the prisoner's dilemma game. Behavioral Science 18:424-428

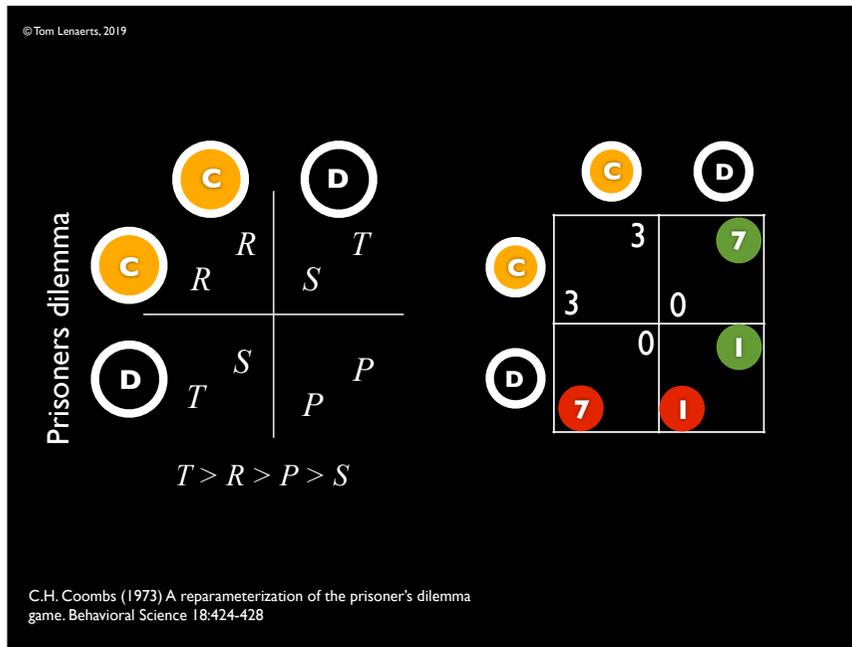
10-2



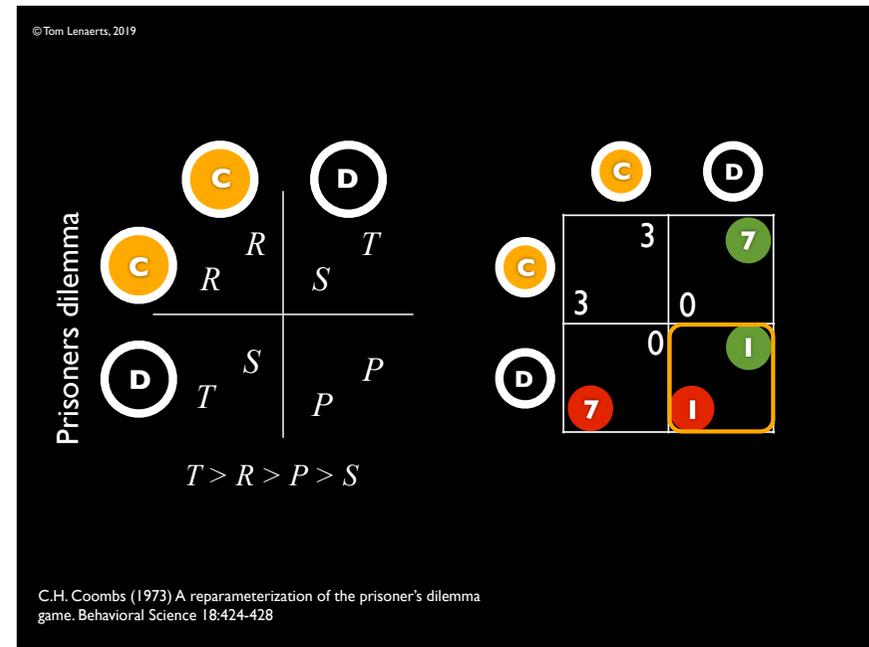
10-3



10-4

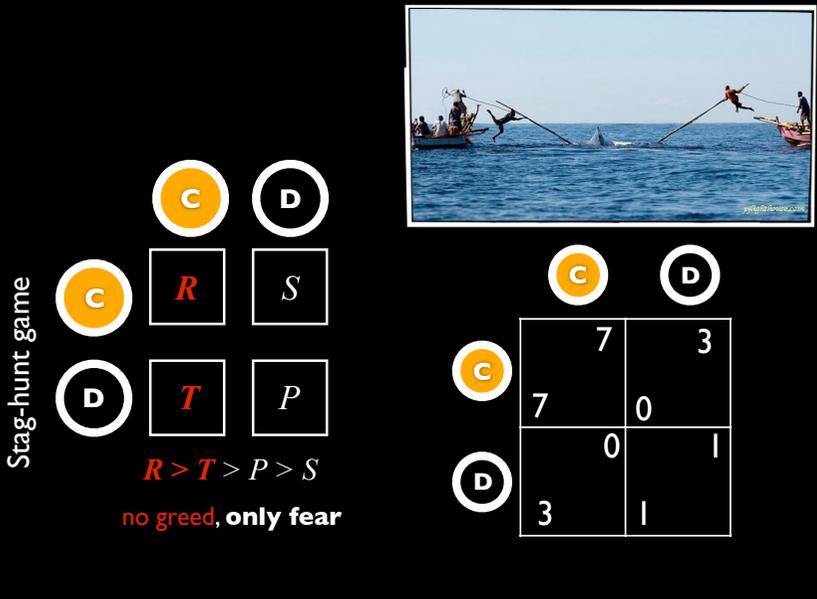


10-5



10-6

© Tom Lenaerts, 2019



Stag-hunt game

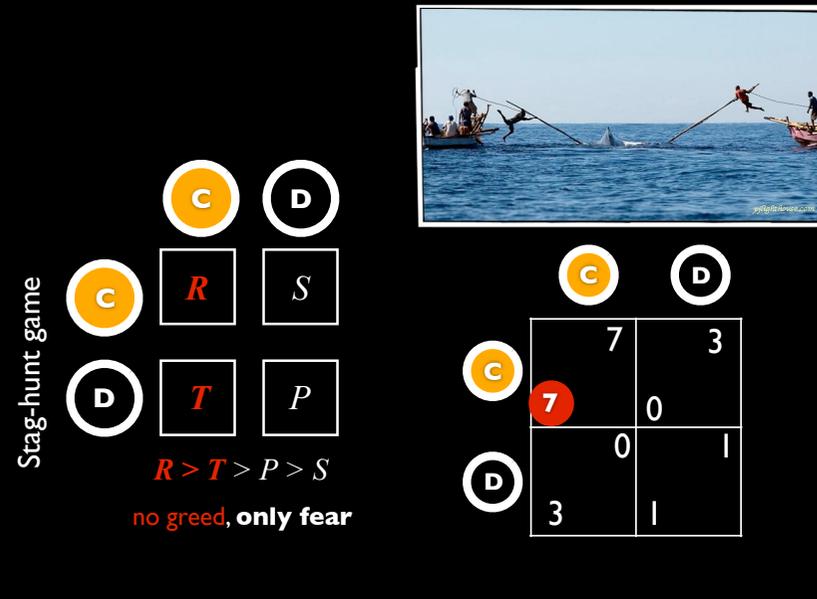
$R > T > P > S$   
 no greed, only fear

	C	D
C	R	S
D	T	P

	C	D
C	7	3
D	0	1

11-1

© Tom Lenaerts, 2019



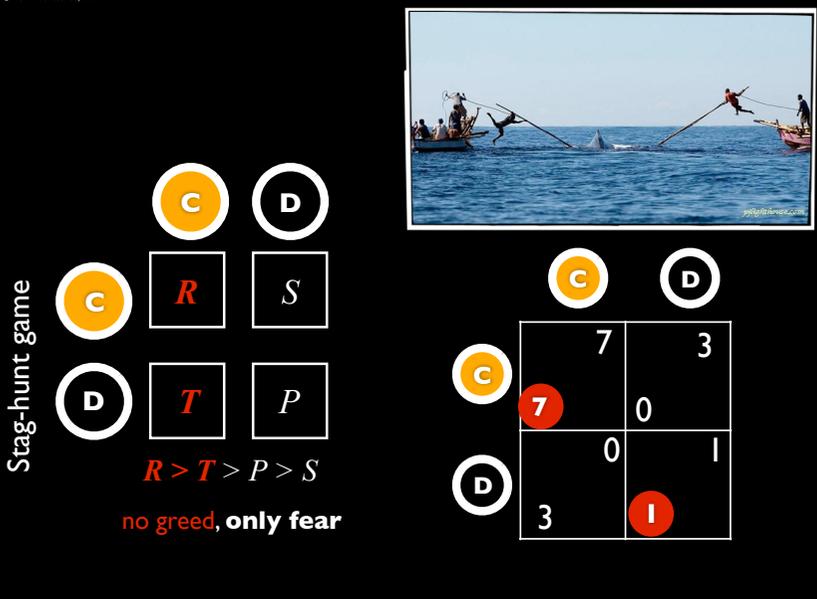
Stag-hunt game

$R > T > P > S$   
 no greed, only fear

	C	D
C	7	3
D	0	1

11-2

© Tom Lenaerts, 2019



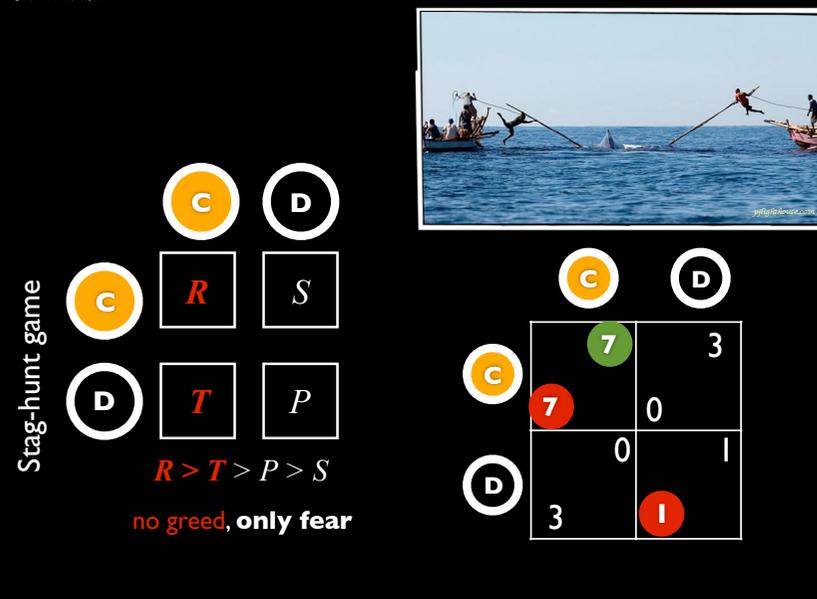
Stag-hunt game

$R > T > P > S$   
 no greed, only fear

	C	D
C	7	3
D	0	1

11-3

© Tom Lenaerts, 2019



Stag-hunt game

$R > T > P > S$   
 no greed, only fear

	C	D
C	7	3
D	0	1

11-4

© Tom Lenaerts, 2019



Stag-hunt game

	C	D
C	R	S
D	T	P

$R > T > P > S$   
no greed, only fear

	C	D
C	7	3
D	3	0

11-5

© Tom Lenaerts, 2019



Stag-hunt game

	C	D
C	R	S
D	T	P

$R > T > P > S$   
no greed, only fear

	C	D
C	7	3
D	3	0

11-6

© Tom Lenaerts, 2019



Stag-hunt game

	C	D
C	R	S
D	T	P

$R > T > P > S$   
no greed, only fear

	C	D
C	7	3
D	3	0

Don't forget the mixed NE!

11-7

© Tom Lenaerts, 2019



Snow-drift game

	C	D
C	R	S
D	T	P

$T > R > S > P$   
no fear, only greed

	C	D
C	3	7
D	7	1

12-1

© Tom Lenaerts, 2019



Snow-drift game

	<b>C</b>	<b>D</b>
<b>C</b>	3	7
<b>D</b>	3	2
	2	1
	<b>7</b>	1

**C** **D**

<b>C</b>	<b>R</b>	<b>S</b>
<b>D</b>	<b>T</b>	<b>P</b>

$T > R > S > P$

no fear, only greed

12-2

© Tom Lenaerts, 2019



Snow-drift game

	<b>C</b>	<b>D</b>
<b>C</b>	3	7
<b>D</b>	3	<b>2</b>
	2	1
	<b>7</b>	1

**C** **D**

<b>C</b>	<b>R</b>	<b>S</b>
<b>D</b>	<b>T</b>	<b>P</b>

$T > R > S > P$

no fear, only greed

12-3

© Tom Lenaerts, 2019



Snow-drift game

	<b>C</b>	<b>D</b>
<b>C</b>	3	<b>7</b>
<b>D</b>	3	<b>2</b>
	2	1
	<b>7</b>	1

**C** **D**

<b>C</b>	<b>R</b>	<b>S</b>
<b>D</b>	<b>T</b>	<b>P</b>

$T > R > S > P$

no fear, only greed

12-4

© Tom Lenaerts, 2019



Snow-drift game

	<b>C</b>	<b>D</b>
<b>C</b>	3	<b>7</b>
<b>D</b>	3	<b>2</b>
	<b>2</b>	1
	<b>7</b>	1

**C** **D**

<b>C</b>	<b>R</b>	<b>S</b>
<b>D</b>	<b>T</b>	<b>P</b>

$T > R > S > P$

no fear, only greed

12-5

© Tom Lenaerts, 2019



Snow-drift game

	C	D
C	3, 3	7, 2
D	2, 7	1, 1

$T > R > S > P$   
 no fear, only greed

12-6

© Tom Lenaerts, 2019



Snow-drift game

	C	D
C	3, 3	7, 2
D	2, 7	1, 1

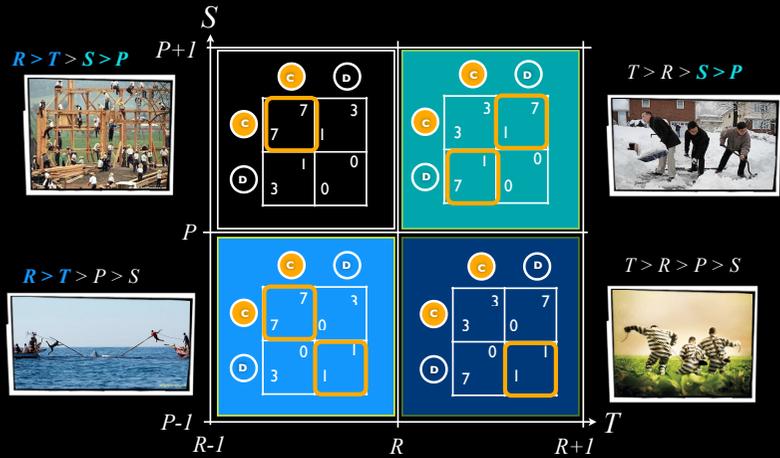
$T > R > S > P$   
 no fear, only greed

Don't forget the mixed NE !

12-7

© Tom Lenaerts, 2016

### Space of social dilemmas



13

© Tom Lenaerts, 2019



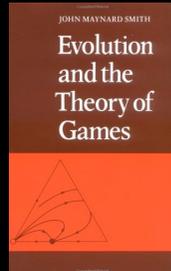
What is the equilibrium notion in populations ?

14-1



What is the equilibrium notion in populations ?

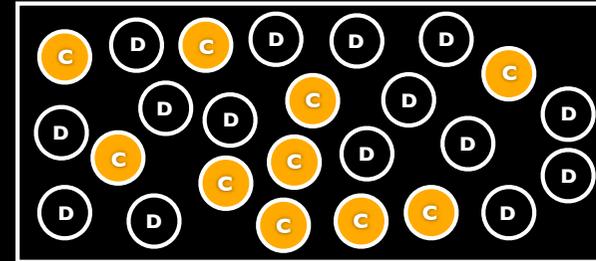
John Maynard Smith and George Price (1973)



The evolutionary stable strategy (ESS) concept

14-2

A strategy is an **ESS** when it cannot be invaded by another strategy



Infinite population assumption

J. Maynard-Smith and G.R. Price (1973) The logic of animal conflict. Nature 246:15-18

15

Can **C** invade a population of **D** players

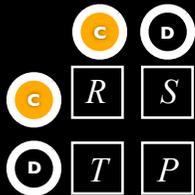
The fraction of **C** (**D**) players is  $\epsilon$  ( $1-\epsilon$ )

success of **C** in a **D** population

success of **D** in a **D** population

$$S(1-\epsilon) + R\epsilon > P(1-\epsilon) + T\epsilon$$

**C** can invade when:



If **C** cannot invade then **D** is an **ESS**

J. Maynard-Smith and G.R. Price (1973) The logic of animal conflict. Nature 246:15-18

16-1

Can **C** invade a population of **D** players

The fraction of **C** (**D**) players is  $\epsilon$  ( $1-\epsilon$ )

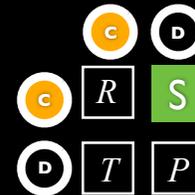
success of **C** in a **D** population

success of **D** in a **D** population

$$S(1-\epsilon) + R\epsilon > P(1-\epsilon) + T\epsilon$$

**C** can invade when:

i)  $S > P$



If **C** cannot invade then **D** is an **ESS**

J. Maynard-Smith and G.R. Price (1973) The logic of animal conflict. Nature 246:15-18

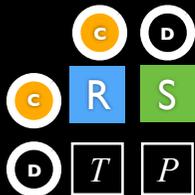
16-2

# Can C invade a population of D players

The fraction of C (D) players is  $\epsilon$  ( $1-\epsilon$ )

success of C in a D population      success of D in a D population

$$S(1-\epsilon) + R\epsilon > P(1-\epsilon) + T\epsilon$$



C can invade when:

- i)  $S > P$
- ii)  $S = P$  and  $R > T$

If C cannot invade then D is an ESS

J. Maynard-Smith and G.R. Price (1973) The logic of animal conflict. Nature 246:15-18

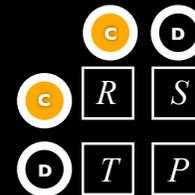
16-3

# Can D invade a population of C players

The fraction of D (C) players is  $\epsilon$  ( $1-\epsilon$ )

success of D in a C population      success of C in a C population

$$T(1-\epsilon) + P\epsilon > R(1-\epsilon) + S\epsilon$$



D can invade when:

If D cannot invade then C is an ESS

J. Maynard-Smith and G.R. Price (1973) The logic of animal conflict. Nature 246:15-18

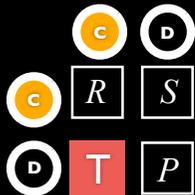
17-1

# Can D invade a population of C players

The fraction of D (C) players is  $\epsilon$  ( $1-\epsilon$ )

success of D in a C population      success of C in a C population

$$T(1-\epsilon) + P\epsilon > R(1-\epsilon) + S\epsilon$$



D can invade when:

- i)  $T > R$

If D cannot invade then C is an ESS

J. Maynard-Smith and G.R. Price (1973) The logic of animal conflict. Nature 246:15-18

17-2

# Can D invade a population of C players

The fraction of D (C) players is  $\epsilon$  ( $1-\epsilon$ )

success of D in a C population      success of C in a C population

$$T(1-\epsilon) + P\epsilon > R(1-\epsilon) + S\epsilon$$



D can invade when:

- i)  $T > R$
- ii)  $T = R$  and  $P > S$

If D cannot invade then C is an ESS

J. Maynard-Smith and G.R. Price (1973) The logic of animal conflict. Nature 246:15-18

17-3

# Can C invade D?



	C	D
C	3	7
D	7	1

	C	D
C	7	3
D	3	1

	C	D
C	3	7
D	2	1

18-1

# Can C invade D?



	C	D
C	3	7
D	7	1

	C	D
C	7	3
D	3	1

	C	D
C	3	7
D	2	1

C no since  $P > S$

18-2

# Can C invade D?



	C	D
C	3	7
D	7	1

	C	D
C	7	3
D	3	1

	C	D
C	3	7
D	2	1

C no since  $P > S$

C no since  $P > S$

18-3

# Can C invade D?



	C	D
C	3	7
D	7	1

	C	D
C	7	3
D	3	1

	C	D
C	3	7
D	2	1

C no since  $P > S$

C no since  $P > S$

C yes since  $P < S$

18-4

### Can D invade C?



	<b>C</b>	<b>D</b>
<b>C</b>	3, 3	7, 0
<b>D</b>	0, 7	1, 1

**C** no since  $P > S$

	<b>C</b>	<b>D</b>
<b>C</b>	7, 7	3, 0
<b>D</b>	0, 3	1, 1

**C** no since  $P > S$

	<b>C</b>	<b>D</b>
<b>C</b>	3, 3	7, 2
<b>D</b>	2, 7	1, 1

**C** yes since  $P < S$

18-5

### Can D invade C?



	<b>C</b>	<b>D</b>
<b>C</b>	3, 3	7, 0
<b>D</b>	0, 7	1, 1

**C** no since  $P > S$

**D** yes since  $R < T$

	<b>C</b>	<b>D</b>
<b>C</b>	7, 7	3, 0
<b>D</b>	0, 3	1, 1

**C** no since  $P > S$

	<b>C</b>	<b>D</b>
<b>C</b>	3, 3	7, 2
<b>D</b>	2, 7	1, 1

**C** yes since  $P < S$

18-6

### Can D invade C?



	<b>C</b>	<b>D</b>
<b>C</b>	3, 3	7, 0
<b>D</b>	0, 7	1, 1

**C** no since  $P > S$

**D** yes since  $R < T$

	<b>C</b>	<b>D</b>
<b>C</b>	7, 7	3, 0
<b>D</b>	0, 3	1, 1

**C** no since  $P > S$

**D** no since  $R > T$

	<b>C</b>	<b>D</b>
<b>C</b>	3, 3	7, 2
<b>D</b>	2, 7	1, 1

**C** yes since  $P < S$

18-7

### Can D invade C?



	<b>C</b>	<b>D</b>
<b>C</b>	3, 3	7, 0
<b>D</b>	0, 7	1, 1

**C** no since  $P > S$

**D** yes since  $R < T$

	<b>C</b>	<b>D</b>
<b>C</b>	7, 7	3, 0
<b>D</b>	0, 3	1, 1

**C** no since  $P > S$

**D** no since  $R > T$

	<b>C</b>	<b>D</b>
<b>C</b>	3, 3	7, 2
<b>D</b>	2, 7	1, 1

**C** yes since  $P < S$

**D** yes since  $R < T$

18-8

©Tom Lenaerts, 2019

## Which one is an ESS?



3	7
3	0
0	1
7	1

**C** is an ESS  
**D** is an ESS

7	3
7	0
0	1
3	1

**C** is an ESS  
**D** is an ESS

3	7
3	2
2	1
7	1

no ESS ?

**D** is an ESS

19-1

©Tom Lenaerts, 2019

## Which one is an ESS?



3	7
3	0
0	1
7	1

**C** is an ESS  
**D** is an ESS

7	3
7	0
0	1
3	1

**C** is an ESS  
**D** is an ESS

3	7
3	2
2	1
7	1

no ESS ?

**D** is an ESS

**Don't forget the mixed NE !**

19-2

©Tom Lenaerts, 2019



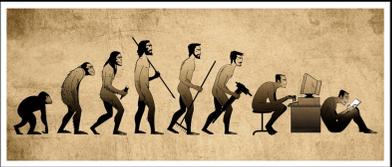
## What is the equilibrium notion in populations ?

20-1

©Tom Lenaerts, 2019



## What is the equilibrium notion in populations ?



## How does learning work in a population?

20-2

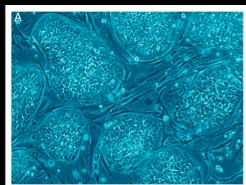
# Change in populations

## Evolutionary dynamics

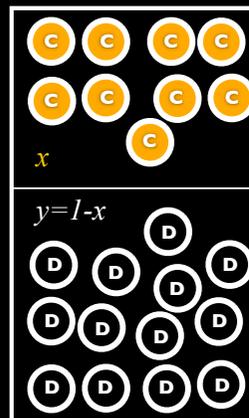


Imitate successful behaviour (a.k.a. social learning)

Genetic evolution of successful properties (a.k.a. survival of the fittest)



# In infinite populations



Replicator equation ...

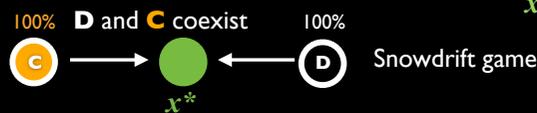
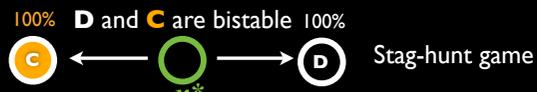
$$\begin{aligned} \frac{dx}{dt} &= x(1-x)[f_C(x) - f_D(x)] \\ &= x(1-x)[(b-c+c-b+0)x - c - 0] \\ &= -cx(1-x) \end{aligned}$$



P.D. Taylor and L.B. Jonker (1978) Evolutionary stable strategies and game dynamics. Mathematical biosciences 40(1-2):145-156

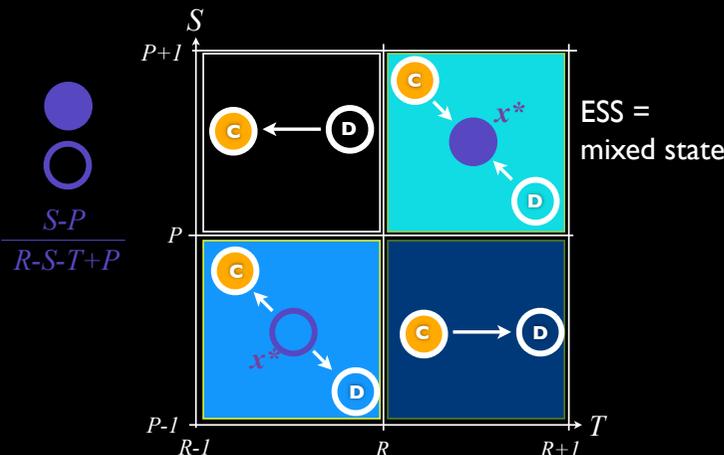
# Change in infinite populations

$$\frac{dx}{dt} = x(1-x)[(R-S-T+P)x + S-P]$$

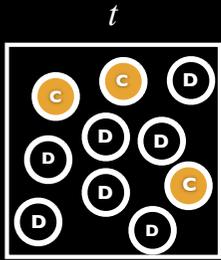


$$x^* = \frac{S-P}{R-S-T+P}$$

# In all social dilemmas



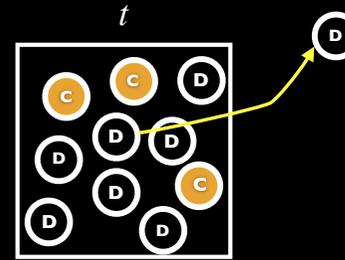
# In **finite** populations



A moran process (birth-death process)

25-1

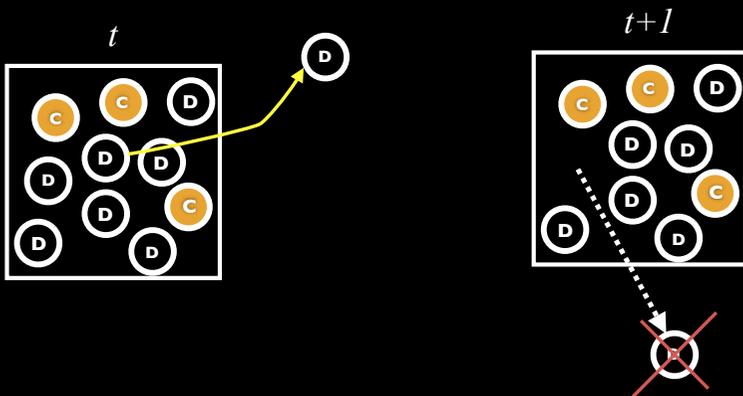
# In **finite** populations



A moran process (birth-death process)

25-2

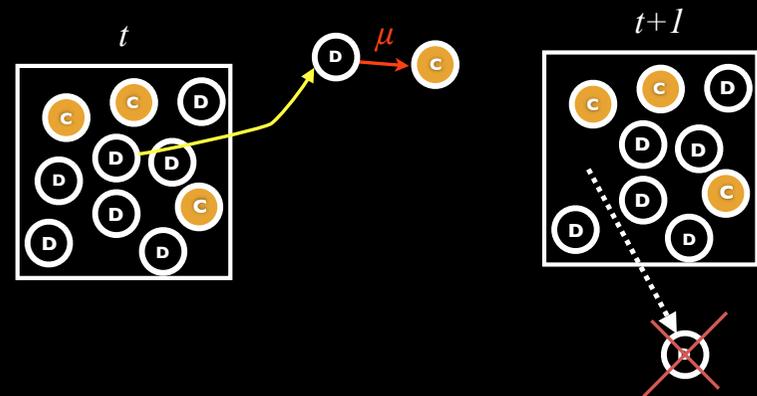
# In **finite** populations



A moran process (birth-death process)

25-3

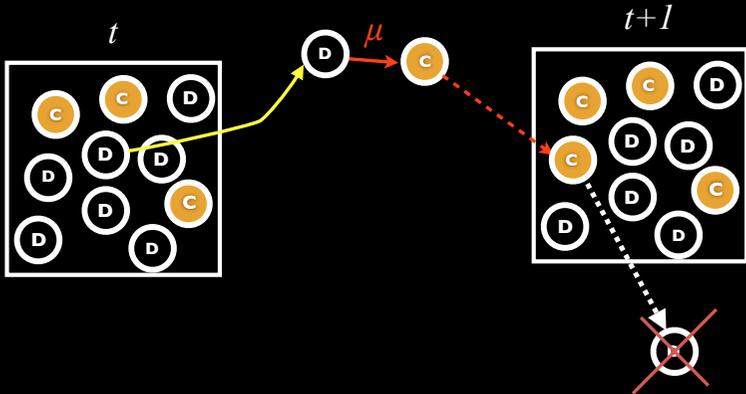
# In **finite** populations



A moran process (birth-death process)

25-4

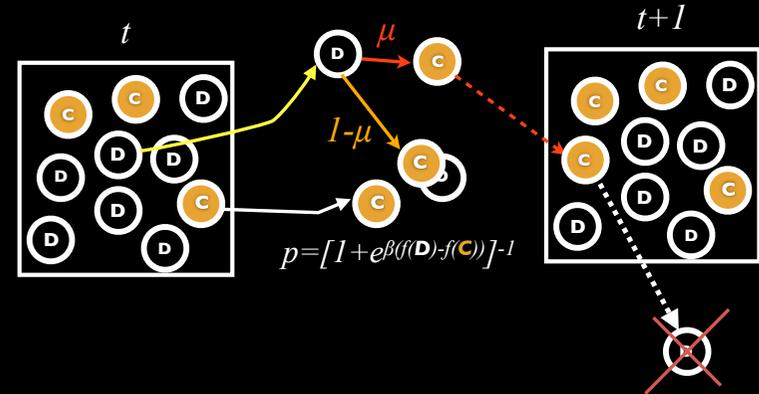
### In **finite** populations



A moran process (birth-death process)

25-5

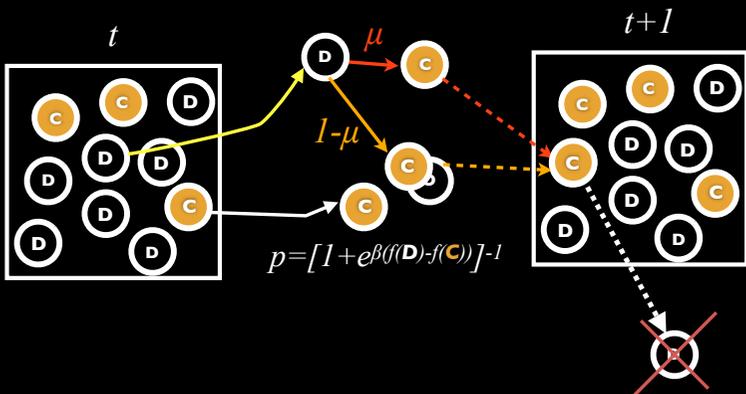
### In **finite** populations



A moran process (birth-death process)

25-6

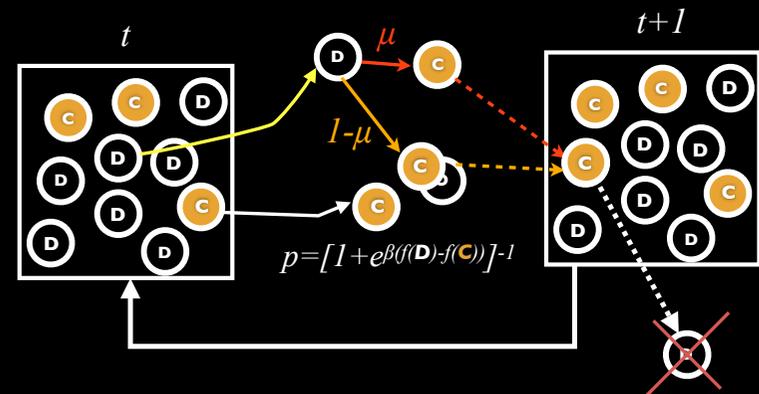
### In **finite** populations



A moran process (birth-death process)

25-7

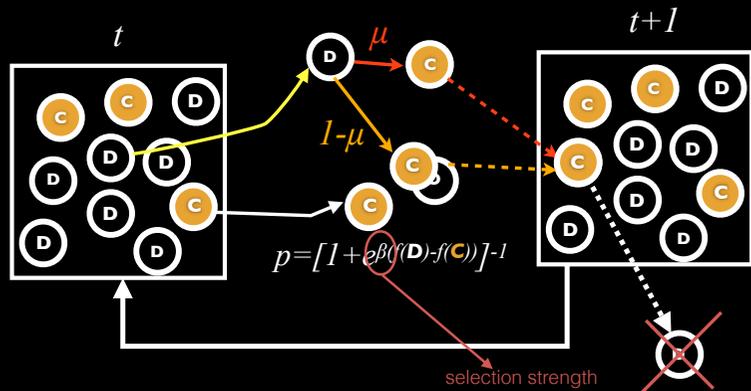
### In **finite** populations



A moran process (birth-death process)

25-8

# In **finite** populations



A moran process (birth-death process)

25-9

# In **finite** populations

## A moran step algorithm

26-1

# In **finite** populations

## A moran step algorithm

First, select randomly two players (with replacement)

26-2

# In **finite** populations

## A moran step algorithm

First, select randomly two players (with replacement)

Second, let each player play the game against all other players (not themselves).

26-3

## In *finite* populations

### A moran step algorithm

First, select randomly two players (with replacement)

Second, let each player play the game against all other players (not themselves).

Third, calculate the average fitness of the player

26-4

## In *finite* populations

### A moran step algorithm

First, select randomly two players (with replacement)

Second, let each player play the game against all other players (not themselves).

Third, calculate the average fitness of the player

**If** a random value is smaller than the fermi probability

26-5

## In *finite* populations

### A moran step algorithm

First, select randomly two players (with replacement)

Second, let each player play the game against all other players (not themselves).

Third, calculate the average fitness of the player

**If** a random value is smaller than the fermi probability

**Then** first player  $\leftarrow$  second player **or** with probability  $\mu$  the first  $\leftarrow$  random strategy

**Else** same but second  $\leftarrow$  first

26-6

## In *pseudo python*

### A moran step algorithm (without mutation):

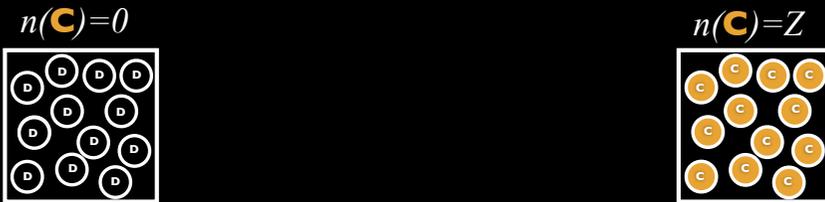
```
def moran-step(beta, population):
    """
    This function implements a birth-death process over
    the population. At time t, two players are randomly
    selected from the population
    """
    selected=select_random_with_replacement(population, 2)
    for i, player in enumerate(selected):
        for j in range(len(population)):
            if j == player: continue
            players_payoffs = play_game(population[player],
                                       population[j])
            fitness[i] += players_payoffs[0]
    fitness = fitness / (Z-1)
    if random() < probab_imitation(beta, fitness):
        population[selected[0]] = population[selected[1]]
    else:
        population[selected[1]] = population[selected[0]]

    return population
```

27

# In finite populations

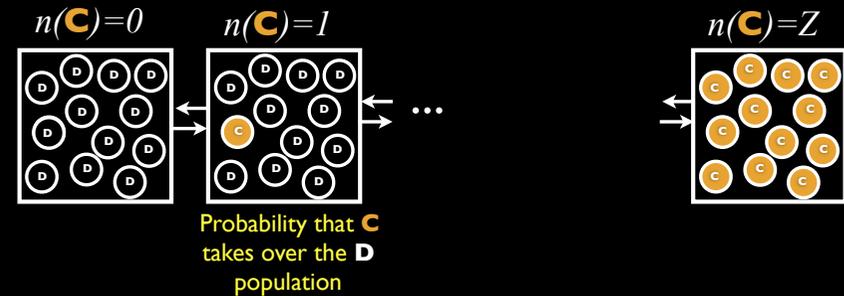
Under the assumption that mutations are rare ( $\mu \rightarrow 0$ ) we either end up with



28-1

# In finite populations

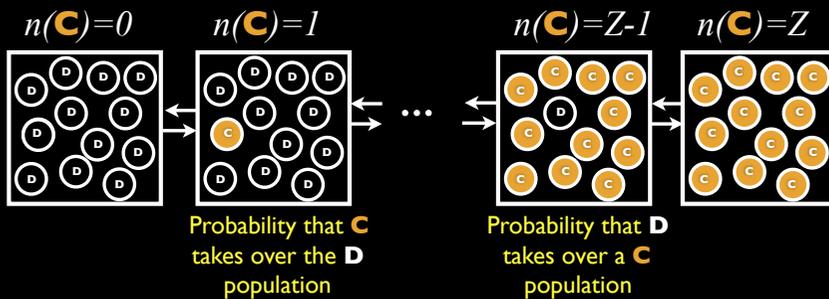
Under the assumption that mutations are rare ( $\mu \rightarrow 0$ ) we either end up with



28-2

# In finite populations

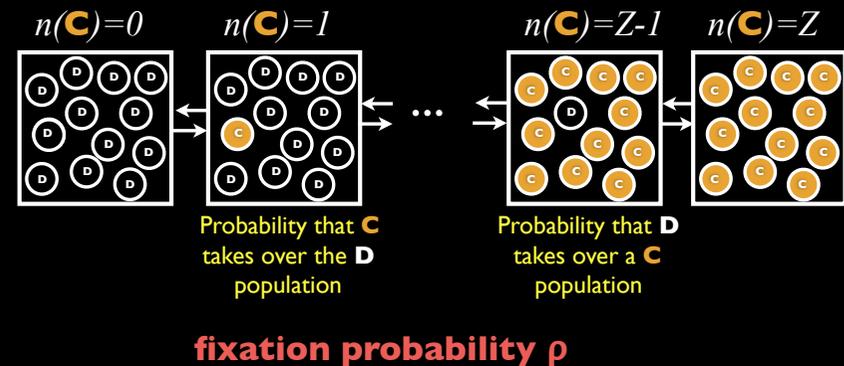
Under the assumption that mutations are rare ( $\mu \rightarrow 0$ ) we either end up with



28-3

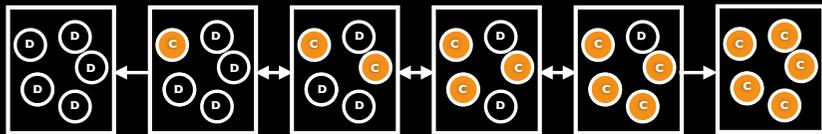
# In finite populations

Under the assumption that mutations are rare ( $\mu \rightarrow 0$ ) we either end up with



28-4

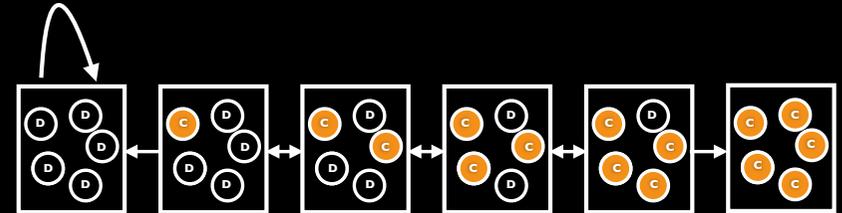
$\rho(i, C)$  = the probability that population fixates to **C** when there are initially  $i$  **C** players in the population



29-1

$\rho(i, C)$  = the probability that population fixates to **C** when there are initially  $i$  **C** players in the population

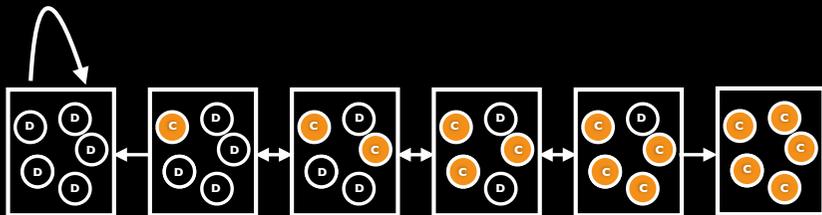
$\rho(0, C) = 0$



29-2

$\rho(i, C)$  = the probability that population fixates to **C** when there are initially  $i$  **C** players in the population

$\rho(0, C) = 0$

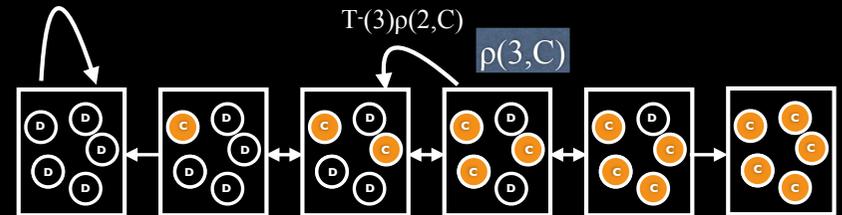


$\rho(5, C) = 1$

29-3

$\rho(i, C)$  = the probability that population fixates to **C** when there are initially  $i$  **C** players in the population

$\rho(0, C) = 0$



$(1 - T^+(3) - T^-(3))\rho(3, C)$

$T^+(3)\rho(4, C)$

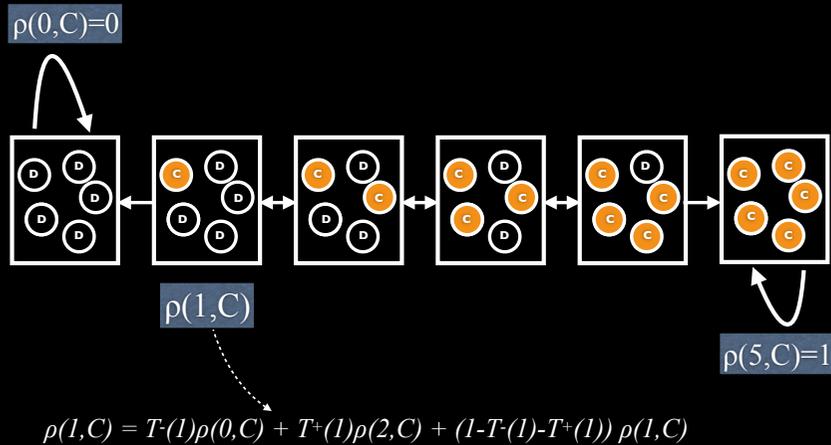
$T^-(3)\rho(2, C)$

$\rho(3, C)$

$\rho(5, C) = 1$

29-4

$\rho(i, C)$  = the probability that population fixates to **C** when there are initially  $i$  **C** players in the population



29-5

$$\rho(1, C) = T^-(1)\rho(0, C) + T^+(1)\rho(2, C) + (1 - T^-(1) - T^+(1))\rho(1, C)$$

Solving recursion gives  $\rho(1, C) = \left( \frac{\sum_{k=0}^{N-1} \prod_{i=1}^k \gamma^i}{\sum_{k=0}^{N-1} \prod_{i=1}^k \gamma^i} \right)^{-1}$

$$\gamma^i = T^-(i)/T^+(i)$$

$T^+(i)$  = the probability that the  $i$  **C** strategists increases by 1

$T^-(i)$  = the probability that the  $i$  **C** strategists decreases by 1

These two probabilities depend on the success (payoff) of each action

30

## In python

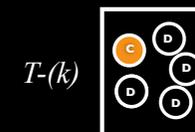
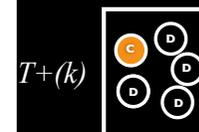
```
def fixation(invader, resident):
    """
    function for calculating the fixation
    probability of the invader
    in a population of residents
    """
    result=0.
    for i in range(1,N):
        sub=1.
        for j in range(1,i+1):
            tmp=probIncreaseDecrease(j,invader, resident)
            sub*=(tmp[1]/float(tmp[0]))
        result += sub
    return np.clip(1./(1. + result), 0., 1.)
```

31

## In python

```
def probIncreaseDecrease(k, invader, resident):
    """
    This function calculates for a give number of invaders
    the probability
    that the number increases or decreases with one.
    """
    fitvalue=fitness(k, invader, resident)
    increase=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(-beta,
        fitvalue[1],fitvalue[0]),0.,1.)
    decrease=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(beta,
        fitvalue[1],fitvalue[0]), 0., 1.)
    return [increase,decrease]
```

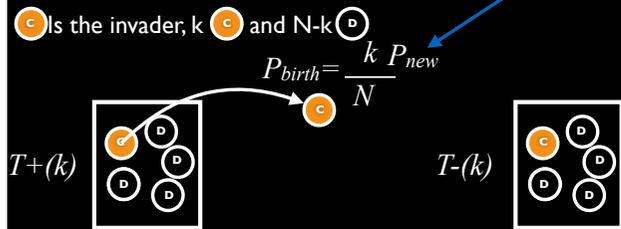
**C** is the invader,  $k$  **C** and  $N-k$  **D**



32-1

# In python

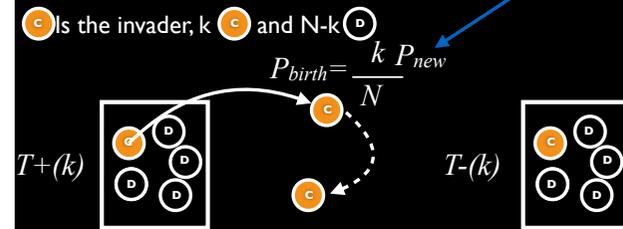
```
def probIncreaseDecrease(k, invader, resident):
    """
    This function calculates for a give number of invaders
    the probability
    that the number increases or decreases with one.
    """
    fitvalue=fitness(k, invader, resident)
    increase=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(-beta,
        fitvalue[1],fitvalue[0]),0.,1.)
    decrease=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(beta,
        fitvalue[1],fitvalue[0]), 0., 1.)
    return [increase,decrease]
```



32-2

# In python

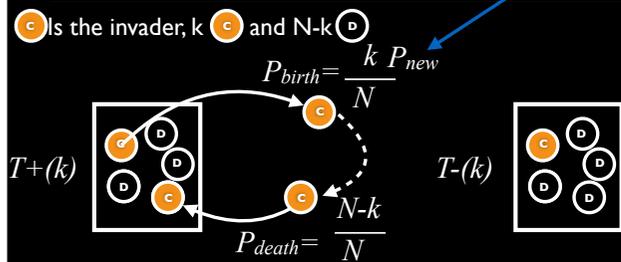
```
def probIncreaseDecrease(k, invader, resident):
    """
    This function calculates for a give number of invaders
    the probability
    that the number increases or decreases with one.
    """
    fitvalue=fitness(k, invader, resident)
    increase=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(-beta,
        fitvalue[1],fitvalue[0]),0.,1.)
    decrease=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(beta,
        fitvalue[1],fitvalue[0]), 0., 1.)
    return [increase,decrease]
```



32-3

# In python

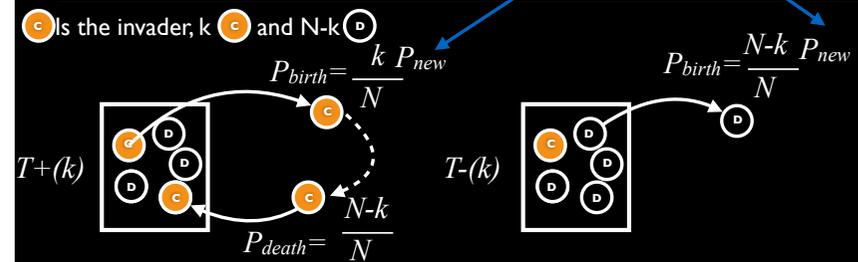
```
def probIncreaseDecrease(k, invader, resident):
    """
    This function calculates for a give number of invaders
    the probability
    that the number increases or decreases with one.
    """
    fitvalue=fitness(k, invader, resident)
    increase=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(-beta,
        fitvalue[1],fitvalue[0]),0.,1.)
    decrease=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(beta,
        fitvalue[1],fitvalue[0]), 0., 1.)
    return [increase,decrease]
```



32-4

# In python

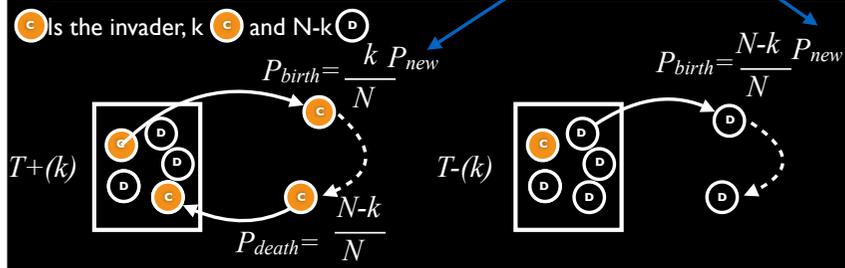
```
def probIncreaseDecrease(k, invader, resident):
    """
    This function calculates for a give number of invaders
    the probability
    that the number increases or decreases with one.
    """
    fitvalue=fitness(k, invader, resident)
    increase=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(-beta,
        fitvalue[1],fitvalue[0]),0.,1.)
    decrease=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(beta,
        fitvalue[1],fitvalue[0]), 0., 1.)
    return [increase,decrease]
```



32-5

# In python

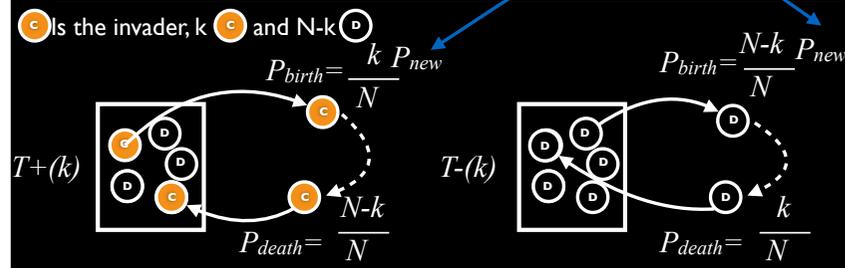
```
def probIncreaseDecrease(k, invader, resident):
    """
    This function calculates for a give number of invaders
    the probability
    that the number increases or decreases with one.
    """
    fitvalue=fitness(k, invader, resident)
    increase=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(-beta,
        fitvalue[1],fitvalue[0]),0.,1.)
    decrease=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(beta,
        fitvalue[1],fitvalue[0]), 0., 1.)
    return [increase,decrease]
```



32-6

# In python

```
def probIncreaseDecrease(k, invader, resident):
    """
    This function calculates for a give number of invaders
    the probability
    that the number increases or decreases with one.
    """
    fitvalue=fitness(k, invader, resident)
    increase=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(-beta,
        fitvalue[1],fitvalue[0]),0.,1.)
    decrease=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(beta,
        fitvalue[1],fitvalue[0]), 0., 1.)
    return [increase,decrease]
```



32-7

# In python

```
def probIncreaseDecrease(k, invader, resident):
    """
    This function calculates for a give number of invaders
    the probability
    that the number increases or decreases with one.
    """
    fitvalue=fitness(k, invader, resident)
    increase=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(-beta,
        fitvalue[1],fitvalue[0]),0.,1.)
    decrease=np.clip(((N-k)/float(N))*(k/float(N))*fermifunc(beta,
        fitvalue[1],fitvalue[0]), 0., 1.)
    return [increase,decrease]
```

Remember Fermi  $p = [1 + e^{\beta(f(D)-f(C))}]^{-1}$

```
def fermifunc(b,first, second):
    """
    The probability that the first type imitates the second
    """
    return 1./(1. + np.exp(-b*(first-second)))
```

33

# In python

```
def fitness(k, invader, resident):
    """
    The fitness function determines the average payoff of k
    invaders and N-k residents
    in the population of N players.
    """
    resultA=((k-1)*payoff[invader][invader])+
        ((N-k)*payoff[invader, resident])/float(N-1)
    resultB=((k*payoff[resident][invader])+
        ((N-k-1)*payoff[resident, resident])/float(N-1)
    return [resultA, resultB]
```

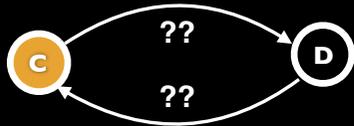
**resultA** = average payoff of an invader playing against his own type A or the other type B

**resultB** = average payoff of an invader playing against his own type B or the other type A

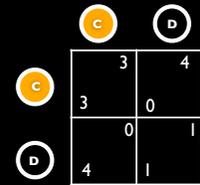
34

# In python

```
def transition_and_fixprob_matrix():
    transitions=np.zeros((q,q))
    fixprobs=np.zeros((q,q))
    for first in range(q):
        transitions[first,first]=1.
        for second in range(q) :
            if second != first :
                fp=fixation(second, first)
                fixprobs[first][second]=(fp/drift)
                transitions[first][second]=fp/float(q-1)
                transitions[first][first]=
                    transitions[first][first]-(fp/float(q-1))
    return [transitions,fixprobs]
```



35



with  $\beta=0.01$   
 $\rho_N = \text{drift} = 1/N$

```
N=100 #Population size
T=3. #Temptation to defect
R=4. #Reward for mutual cooperation
P=0. #Punishment for mutual defection
S=1. #Suckers payoff for unilateral cooperation
q=2 #Number of strategies
drift=1.0/N

strats=['C','D']
payoff=np.array([R,S,T,P]).reshape(2,2)
beta=0.01

t,f=transition_and_fixprob_matrix()
print ("transition probabilities (first %s,
second %s and read as row invading column) \n
%s" %(strats[0], strats[1],f))
```

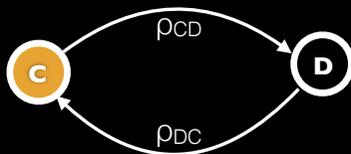
```
transition probabilities (first
C, second D and read as row
invading column)
[[ 0.          0.5951908 ]
 [ 1.55445772  0.        ]]
```



36

# Stationary distribution

Which produces a **reduced Markov chain**



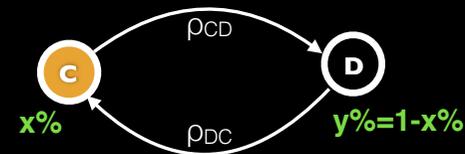
Fudenberg, D., & Imhof, L.A. (2006). Imitation processes with small mutations. J. Econ.Theo, 131,251–262.

Imhof, L.A., Fudenberg, D., & Nowak, M.A. (2005). Evolutionary cycles of cooperation and defection. Proc Nat Acad Sci USA, 102(31), 10797–10800.

37-1

# Stationary distribution

Which produces a **reduced Markov chain**



For which the **stationary distributions** can be calculated = *how likely it is to end up in either monomorphic state*

Fudenberg, D., & Imhof, L.A. (2006). Imitation processes with small mutations. J. Econ.Theo, 131,251–262.

Imhof, L.A., Fudenberg, D., & Nowak, M.A. (2005). Evolutionary cycles of cooperation and defection. Proc Nat Acad Sci USA, 102(31), 10797–10800.

37-2

```
#calculate stationary distributions using eigenvalues and eigenvectors
w,v=np.linalg.eig(t.transpose())

#look for the element closest to 1 in the list of eigenvalues
j_stationary=np.argmax(abs(w-1.0))

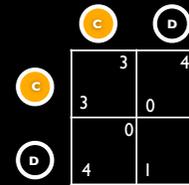
# the, is essential to access the matrix by column
p_stationary=abs(v[:,j_stationary].real)

#normalize
p_stationary /= p_stationary.sum()

print ("stationary distribution (first %s, second %s) \n %s" %(strats[0],
strats[1], p_stationary))
```

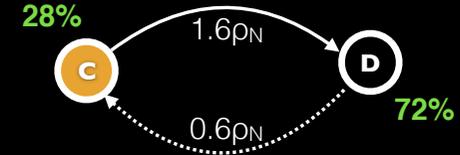
The stationary distribution is the **minimum left eigenvalue** of the transition matrix.

38



stationary distribution (first D, second C)  
[ 0.72312181 0.27687819]

with  $\beta=0.01$   
 $\rho_N = \text{drift} = 1/N$



Python **code for numerical approximation**  
(random walks in Markov chain) see 2IPD-  
numerical.ipynb (see assignment I)

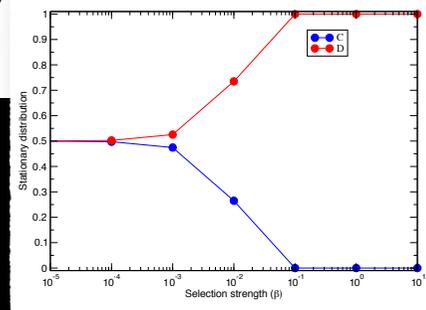
39

For varying  $\beta$

```
#calculate stationary distributions using eigenvalues and eigenvectors
distribution=[]
betas=[0.00001,0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0]
for beta in betas:
    t,f=transition_and_fixprob_matrix()
    w,v=np.linalg.eig(t.transpose())
    j_stationary=np.argmax(abs(w-1.0))
    p_stationary=abs(v[:,j_stationary].real)
    p_stationary /= p_stationary.sum()
    distribution.append(p_stationary)

darray=np.asarray(distribution)

# and plot the curve
```



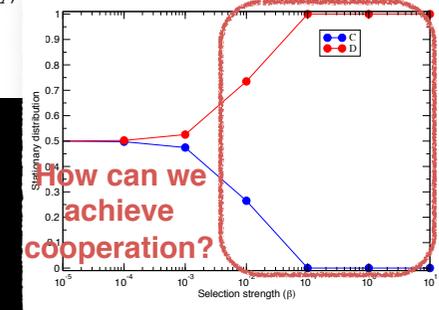
40-1

For varying  $\beta$

```
#calculate stationary distributions using eigenvalues and eigenvectors
distribution=[]
betas=[0.00001,0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0]
for beta in betas:
    t,f=transition_and_fixprob_matrix()
    w,v=np.linalg.eig(t.transpose())
    j_stationary=np.argmax(abs(w-1.0))
    p_stationary=abs(v[:,j_stationary].real)
    p_stationary /= p_stationary.sum()
    distribution.append(p_stationary)

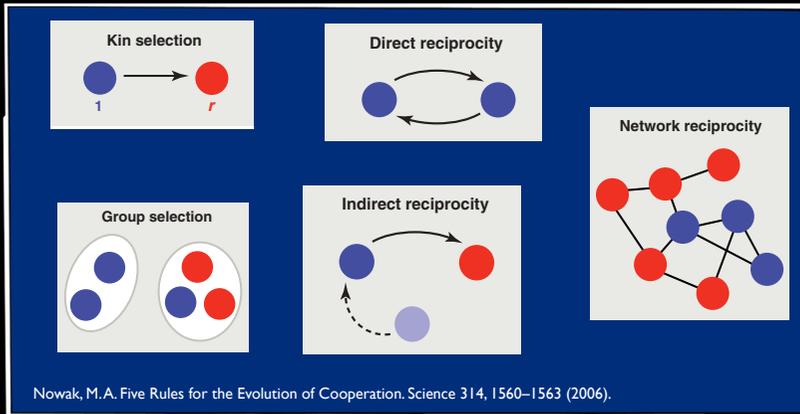
darray=np.asarray(distribution)

# and plot the curve
```



40-2

# How to reach cooperation?



Nowak, M.A. Five Rules for the Evolution of Cooperation. Science 314, 1560–1563 (2006).

**Assortment** between cooperators is key to success!

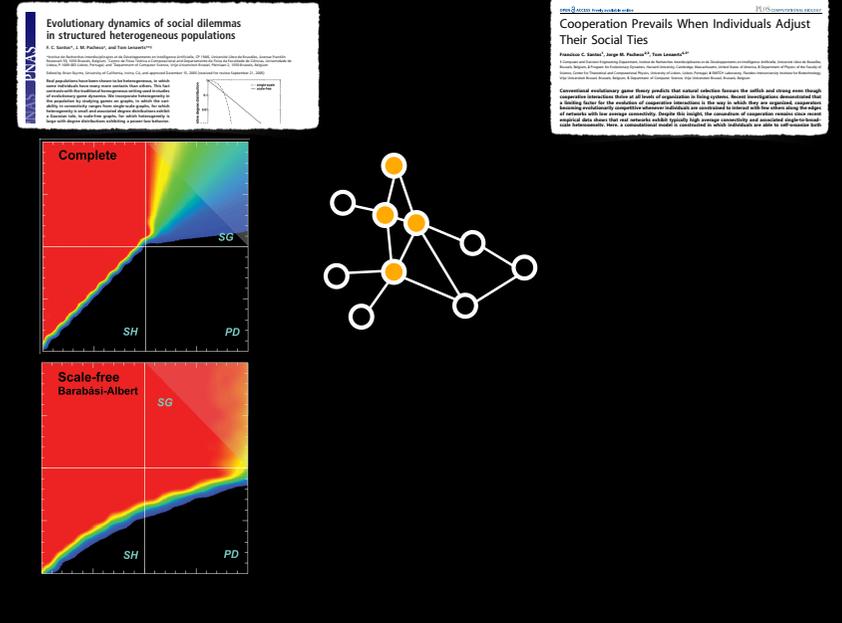
41

# Networks and the evolution of cooperation

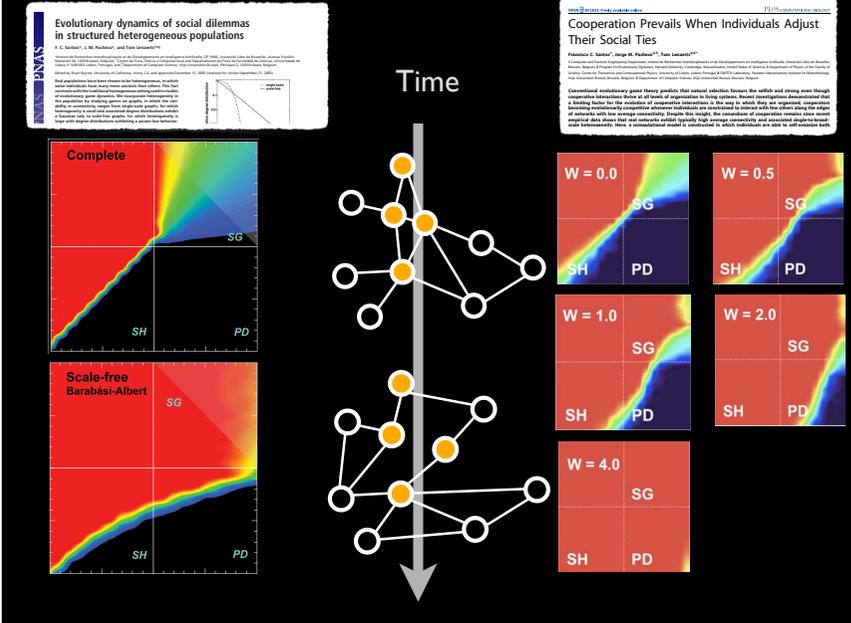
**Evolutionary dynamics of social dilemmas in structured heterogeneous populations**  
F. C. Santos<sup>1</sup>, J. M. Pacheco<sup>2</sup>, and Tom Lenaerts<sup>1,3</sup>

**Cooperation Prevails When Individuals Adjust Their Social Ties**  
Francisco C. Santos<sup>1</sup>, Jorge M. Pacheco<sup>2,3</sup>, Tom Lenaerts<sup>1,3</sup>

42



43-1



43-2

# Or more elaborate behaviours

## Mechanisms of Social Cognition

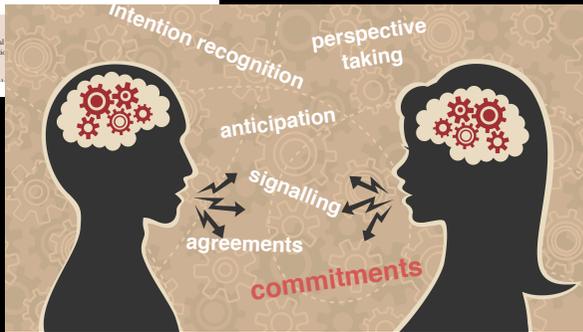
Chris D. Frith<sup>1,3</sup> and Uta Frith<sup>2,3</sup>

<sup>1</sup>Wellcome Trust Centre for Neuroimaging and <sup>2</sup>Centre for Cognitive Neuroscience, University College London, WC1E 6BT, United Kingdom, and <sup>3</sup>Centre of Functionally Integrative Neuroscience, Aarhus University, 8000 Aarhus, Denmark, email: c.d.frith@ucl.ac.uk, u.f.frith@ucl.ac.uk

doi:10.1016/j.neurosci.2015.08.011

Ann. Rev. Psychol. 2015.66:207-231. First published online as a Review in Advance on August 11, 2015. The Annual Review of Psychology is online at <http://www.annualreviews.org>.

**Keywords:** observational; meta-cognitive; Abstract



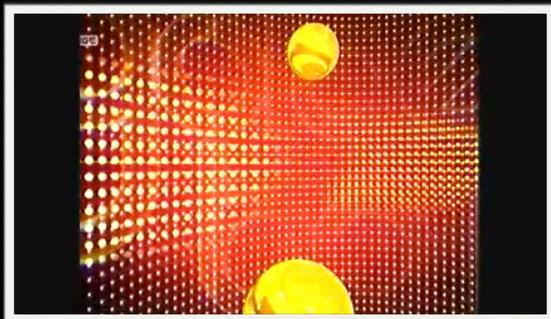
44

# Trust drives social interactions ...

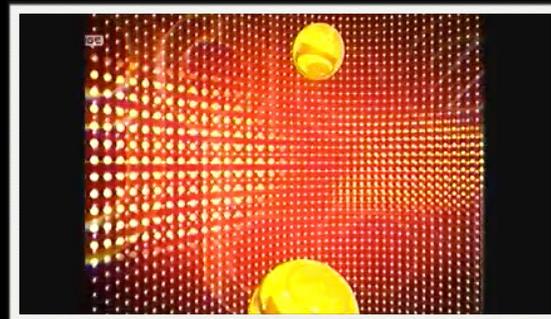


**How to make sure we can trust a partner?**

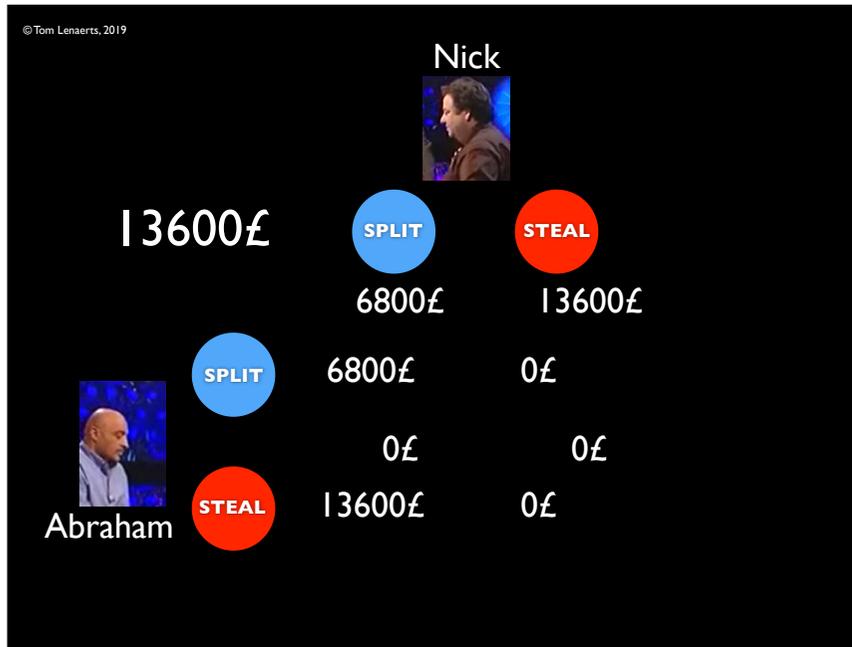
45



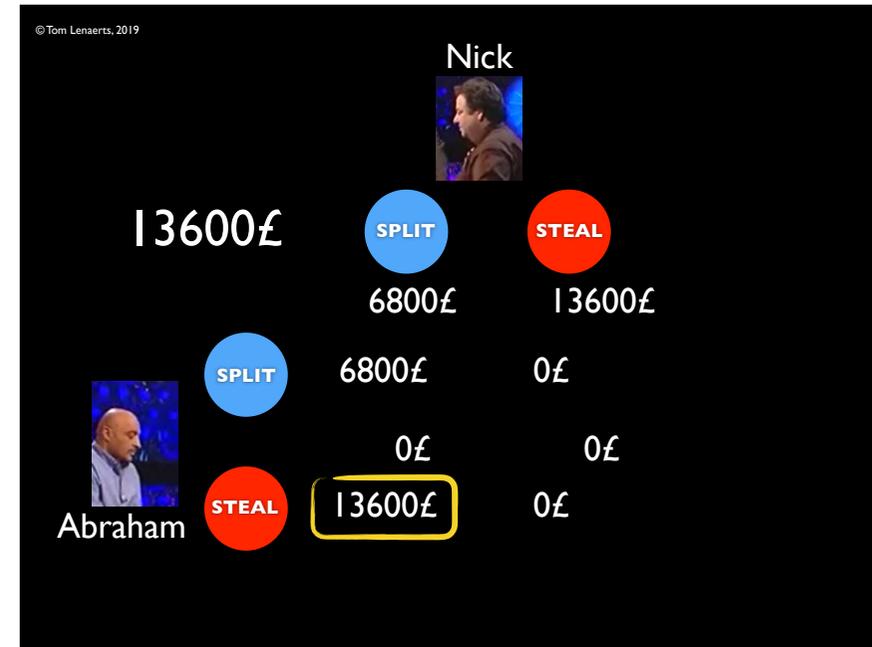
46-1



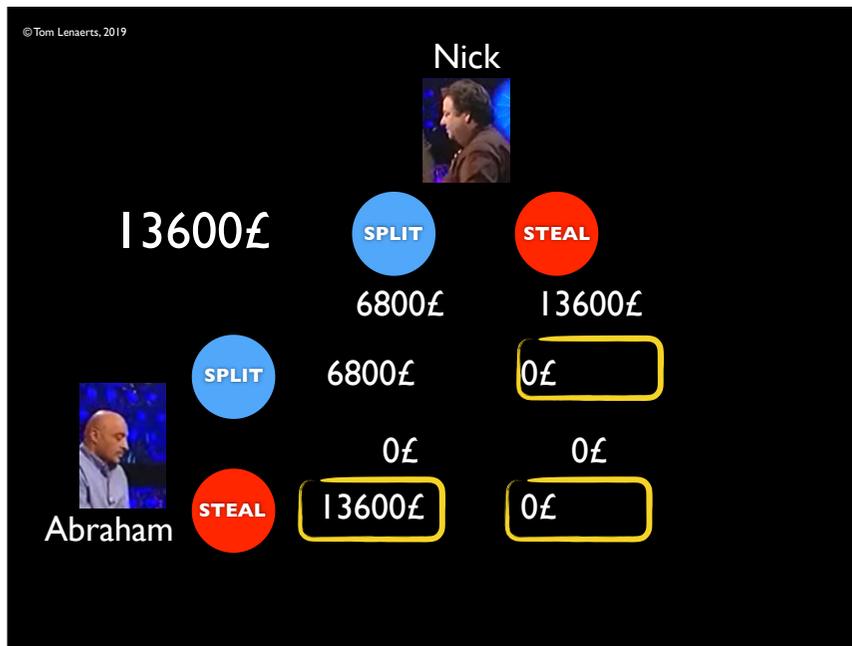
46-2



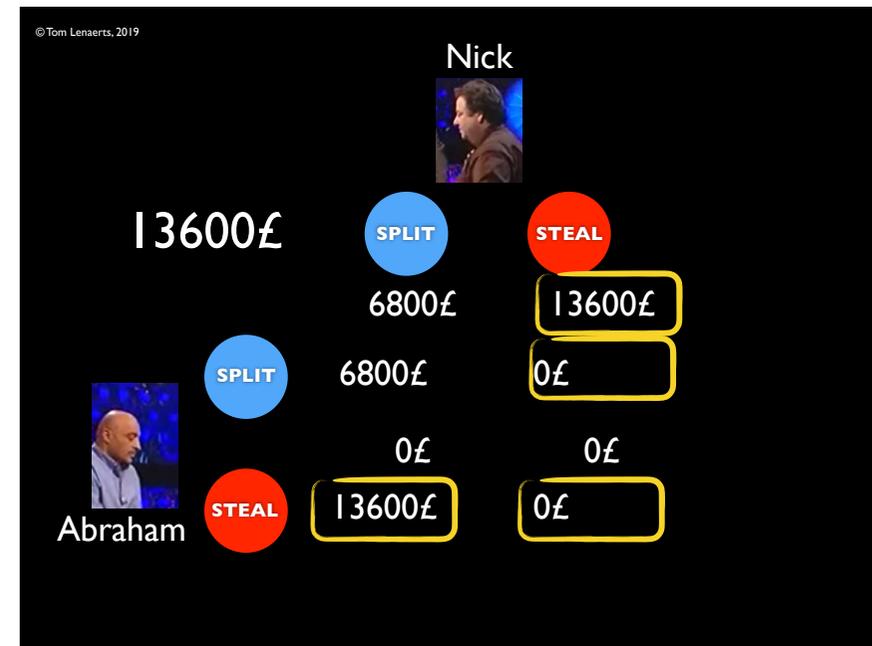
47-1



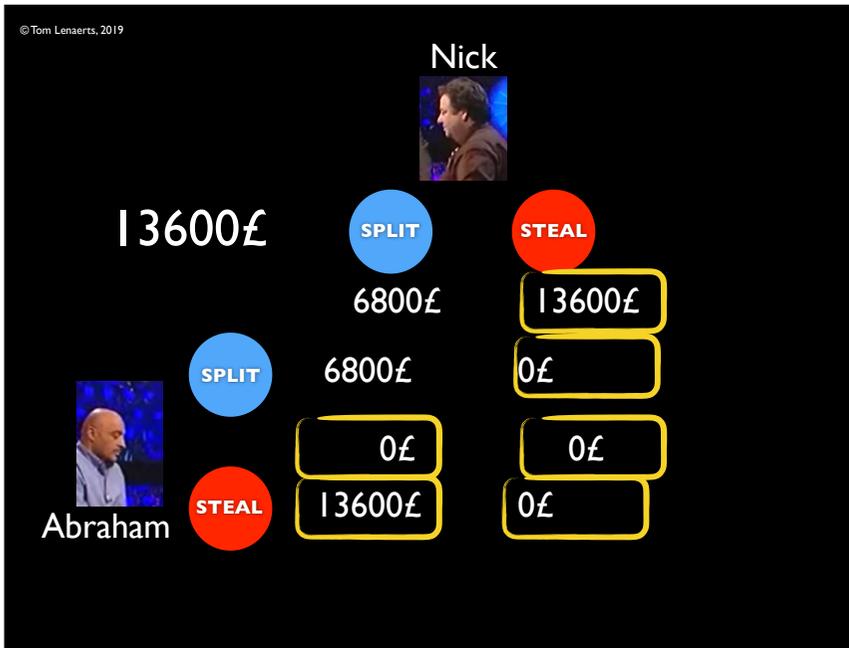
47-2



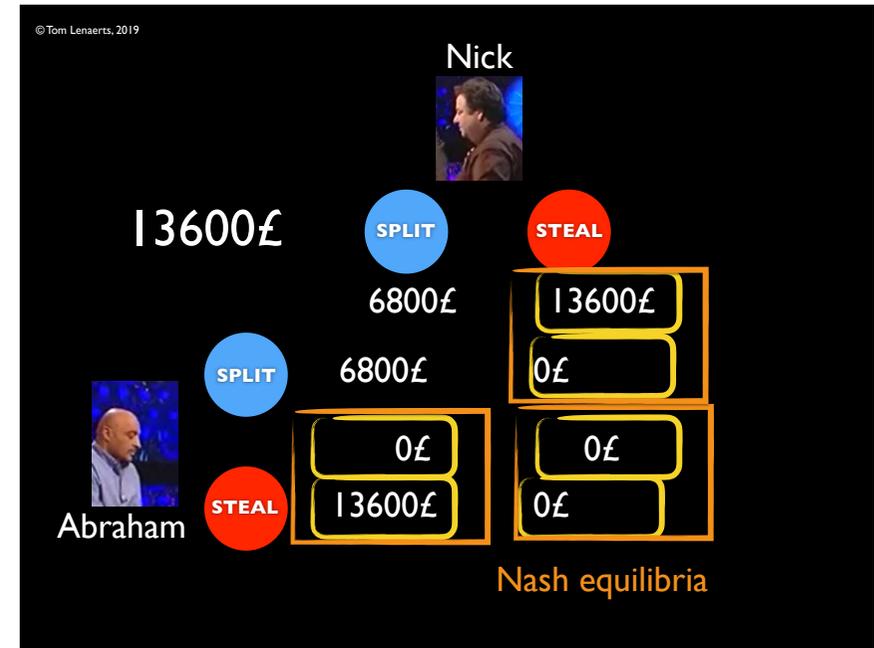
47-3



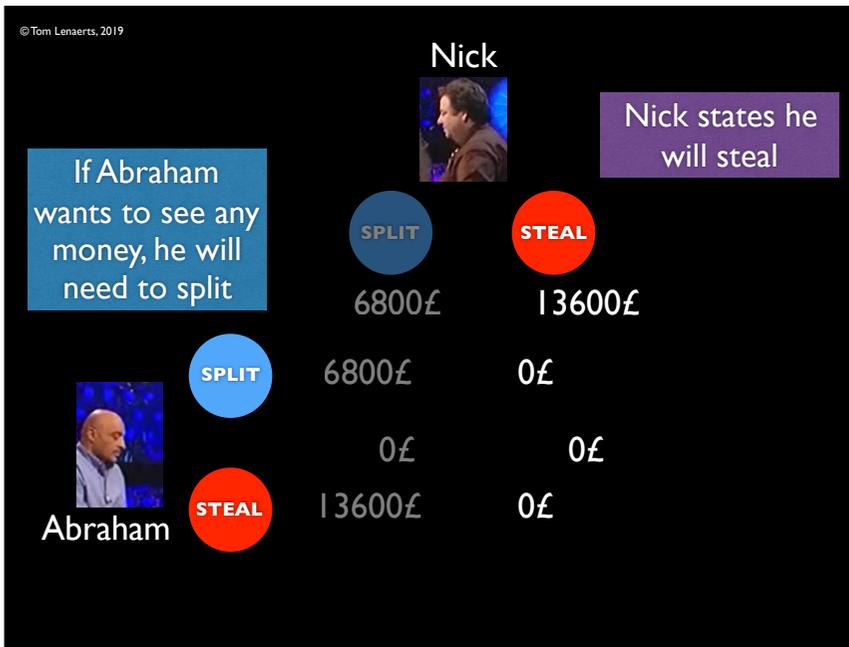
47-4



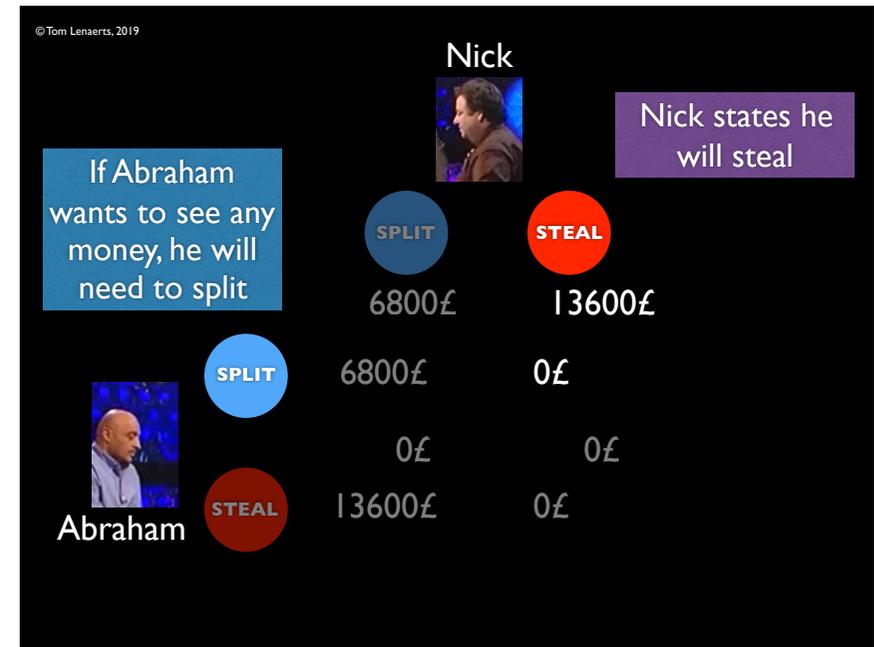
47-5



47-6



48-1



48-2

© Tom Lenaerts, 2019

Nick

Nick states he will steal

If Abraham wants to see any money, he will need to split

SPLIT 6800£ STEAL 13600£

SPLIT 6800£ STEAL 0£

Abraham

Giving Nick control of the outcome

48-3

© Tom Lenaerts, 2019

Nick

Nick states he will steal

If Abraham wants to see any money, he will need to split

SPLIT 6800£ STEAL 13600£

SPLIT 6800£ STEAL 0£

Abraham

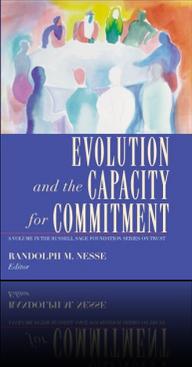
Giving Nick control of the outcome

Nick forced Abraham to **commit** to Split

48-4

**Commitment ?**

“A **commitment** is an act or signal that gives up options in order to **influence someone’s behaviour** by changing incentives and expectations”



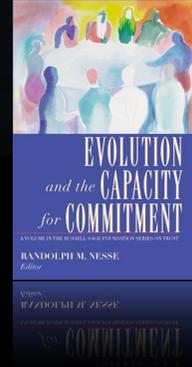
© Tom Lenaerts, 2018 FutureICT2.0 Tallinn

49-1

**Commitment ?**

“A **commitment** is an act or signal that gives up options in order to **influence someone’s behaviour** by changing incentives and expectations”

“Commitments can be **promises to help**, or **threats to harm**”



© Tom Lenaerts, 2018 FutureICT2.0 Tallinn

49-2

# Commitment ?

“A **commitment** is an act or signal that gives up options in order to **influence someone’s behaviour** by changing incentives and expectations”

“Commitments can be **promises to help**, or **threats to harm**”

“They can be **enforced** by external incentives, but also by some combination of **reputation** and **emotion**”

© Tom Lenaerts, 2018 FutureCT2.0 Tallinn

49-3

# Commitment ?

“A **commitment** is an act or signal that gives up options in order to **influence someone’s behaviour** by changing incentives and expectations”

“Commitments can be **promises to help**, or **threats to harm**”

“They can be **enforced** by external incentives, but also by some combination of **reputation** and **emotion**”

“Our (cognitive) capacity for commitment may have **evolved by natural selection**”

© Tom Lenaerts, 2018 FutureCT2.0 Tallinn

49-4

## SCIENTIFIC REPORTS

**OPEN** Good Agreements Make Good Friends

The Anh Han<sup>1</sup>, Luis Morin Perea<sup>1</sup>, Francisco C. Santos<sup>1\*</sup> & Tom Lenaerts<sup>1\*</sup>

**SUBJECT AREAS:**  
 BIOLOGICAL PHYSICS  
 BEHAVIOURAL SCIENCES  
 EVOLUTIONARY THEORY  
 SOCIAL EVOLUTION

Received: 2 July 2013  
 Accepted: 30 August 2013  
 Published: 18 September 2013

When starting a new collaborative endeavour, it pays to establish upfront how strongly your partner commits to the common goal and what consequences can be expected in case the collaboration is violated. Online examples in biological and social contexts have demonstrated the pervasiveness of making prior agreements and promise commitments, suggesting that this behaviour could have been shaped by natural selection. Here, we analyse the evolutionary relevance of such commitment strategies and reveal that the costly punishment strategy, when no prior agreements are made. We show that when the cost of arranging a commitment deal is within certain limits, substantial levels of cooperation can be achieved. Moreover, these levels are higher than that achieved by simple costly punishment, especially when one invests on sharing the arrangement cost. Our study also shows that good agreements make good friends, agreements based on shared costs result in even better outcomes.

Our (cognitive) capacity for **commitment** may have evolved by natural selection

© Tom Lenaerts, 2018 FutureCT2.0 Tallinn

50-1

## SCIENTIFIC REPORTS

**OPEN** Good Agreements Make Good Friends

The Anh Han<sup>1</sup>, Luis Morin Perea<sup>1</sup>, Francisco C. Santos<sup>1\*</sup> & Tom Lenaerts<sup>1\*</sup>

**SUBJECT AREAS:**  
 BIOLOGICAL PHYSICS  
 BEHAVIOURAL SCIENCES  
 EVOLUTIONARY THEORY  
 SOCIAL EVOLUTION

Received: 2 July 2013  
 Accepted: 30 August 2013  
 Published: 18 September 2013

When starting a new collaborative endeavour, it pays to establish upfront how strongly your partner commits to the common goal and what consequences can be expected in case the collaboration is violated. Online examples in biological and social contexts have demonstrated the pervasiveness of making prior agreements and promise commitments, suggesting that this behaviour could have been shaped by natural selection. Here, we analyse the evolutionary relevance of such commitment strategies and reveal that the costly punishment strategy, when no prior agreements are made. We show that when the cost of arranging a commitment deal is within certain limits, substantial levels of cooperation can be achieved. Moreover, these levels are higher than that achieved by simple costly punishment, especially when one invests on sharing the arrangement cost. Our study also shows that good agreements make good friends, agreements based on shared costs result in even better outcomes.

Emotions manage **mutually beneficial relationships**

© Tom Lenaerts, 2018 FutureCT2.0 Tallinn

50-2

## SCIENTIFIC REPORTS

**OPEN** Apology and forgiveness evolve to resolve failures in cooperative agreements

Received: 13 February 2013  
 Accepted: 21 April 2013  
 Published: 05 May 2013

Luis A. Martínez-Vejerano<sup>1</sup>, The Anh Han<sup>1</sup>, Luis Morin Perea<sup>1</sup> & Tom Lenaerts<sup>1\*</sup>

Making agreements on how to behave has been shown to be an evolutionarily viable strategy in one-shot social dilemmas. However, in many situations agreements aim to establish long-term mutually beneficial interactions. Our analytical and numerical results reveal for the first time under which conditions revenge, apology and forgiveness can evolve and deal with mistakes within ongoing agreements in the context of the Iterated Prisoner's Dilemma. We show that, when the agreement fails, participants prefer to take revenge by defecting in the subsequent encounter. Incorporating costly apology and forgiveness reveals that, even when mistakes are frequent, there exists a strategy through which mistakes will not lead to the destruction of the agreement, inducing even higher levels of cooperation. In short, even when to err is human, revenge, apology and forgiveness are evolutionarily viable strategies which play an important role in inducing cooperation in repeated dilemmas.



51-1



51-2



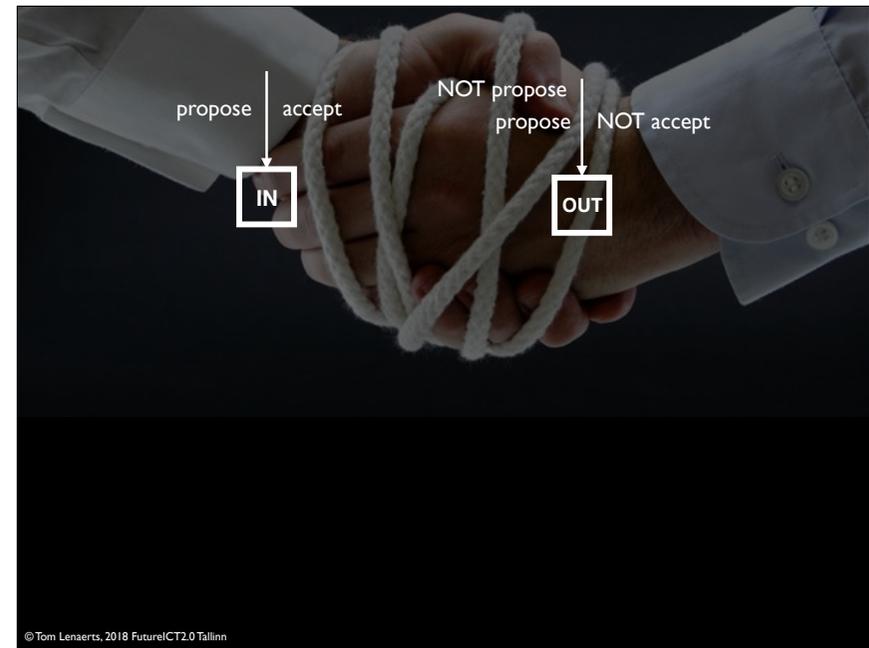
51-3



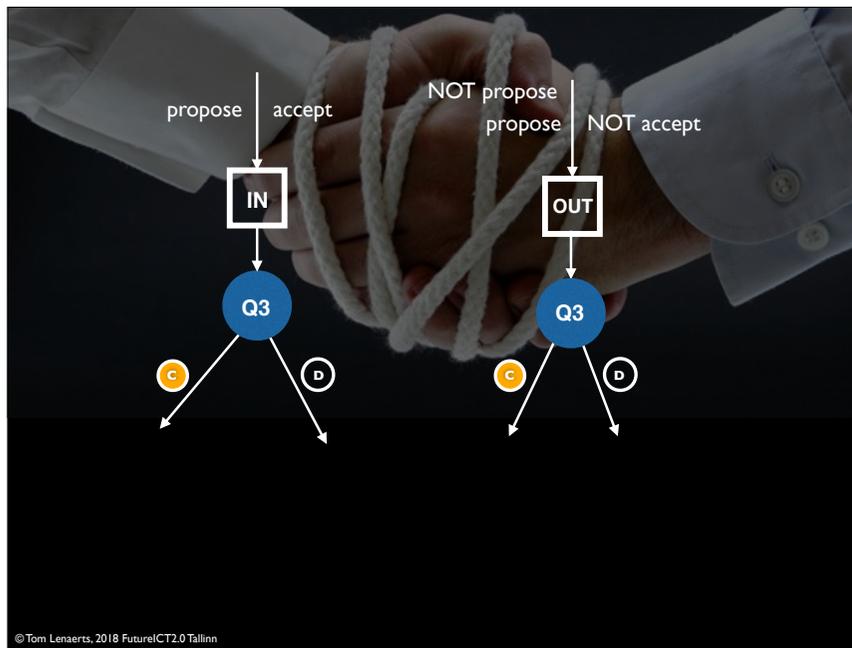
51-4



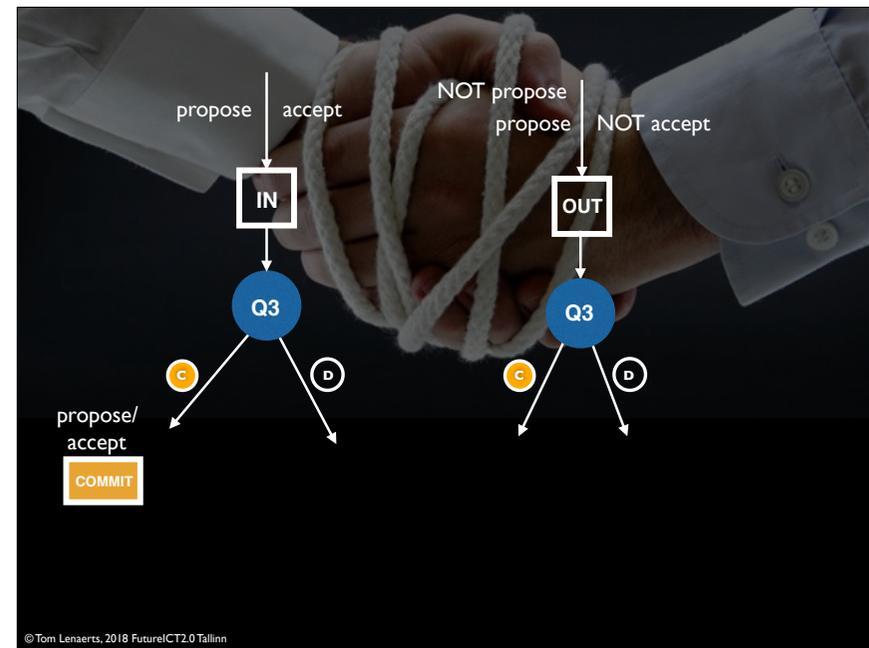
51-5



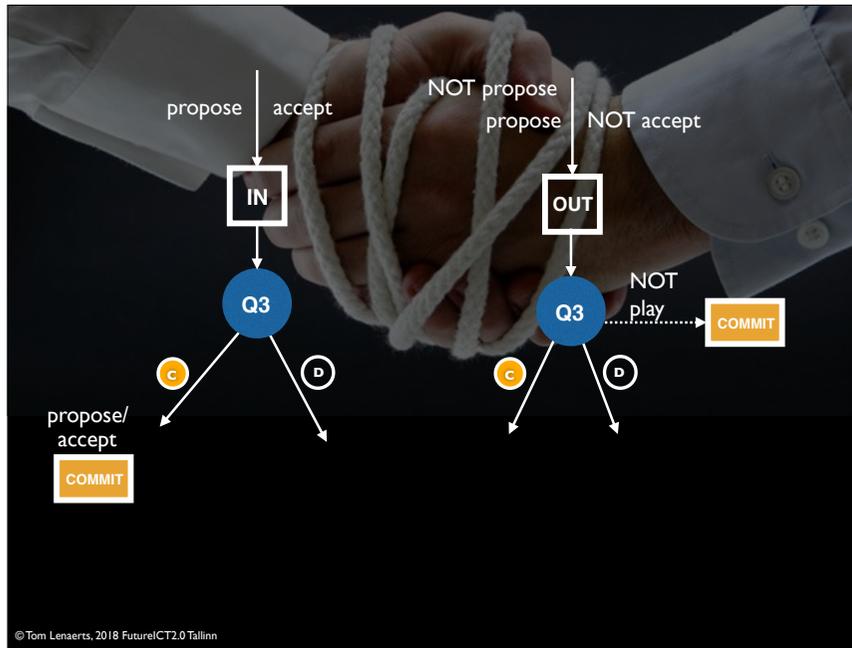
52-1



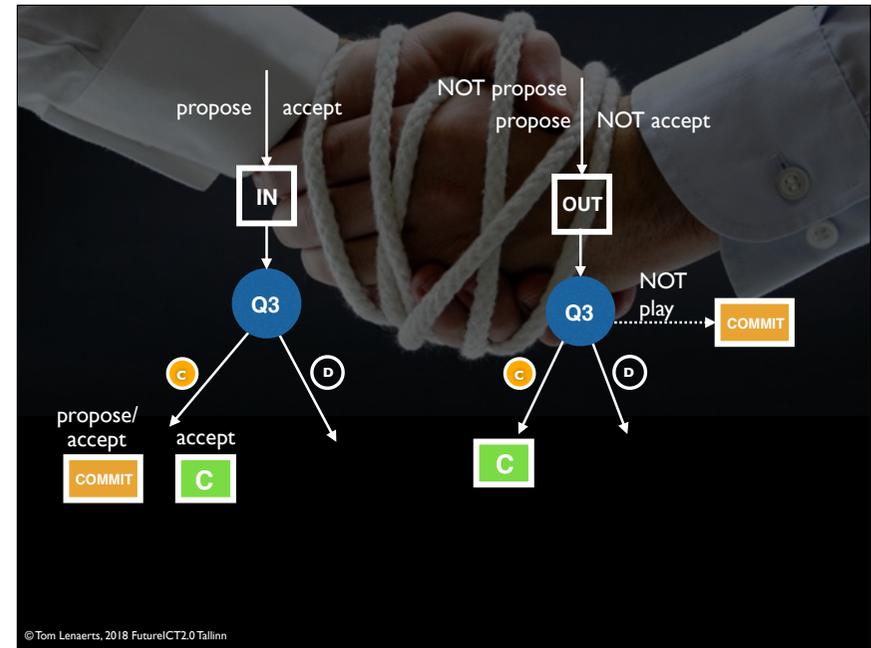
52-2



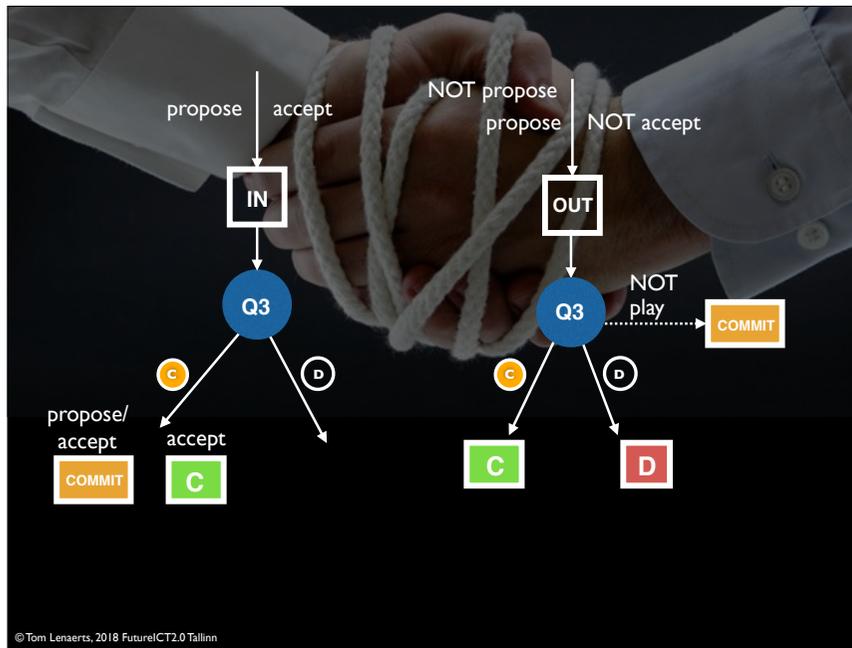
52-3



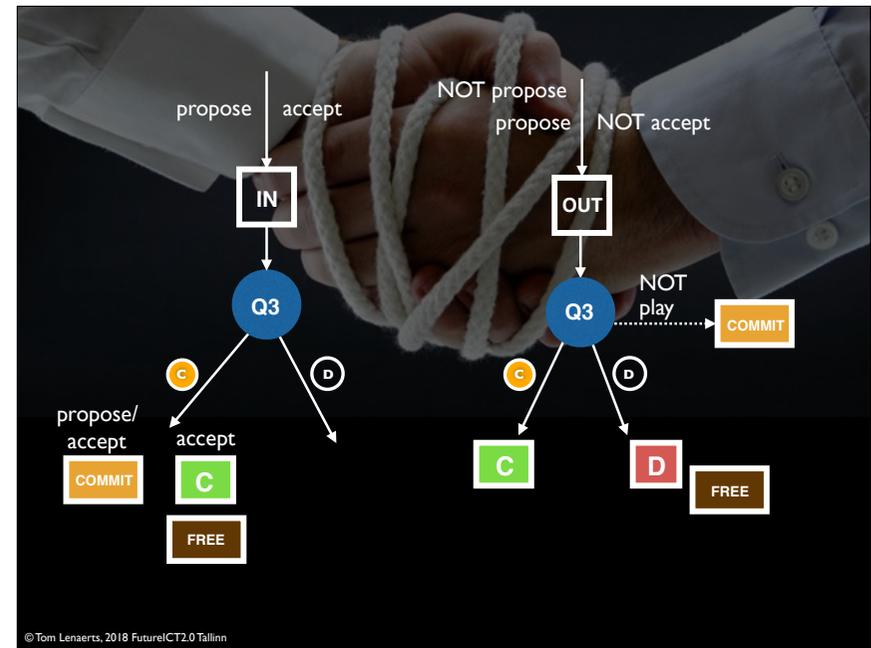
52-4



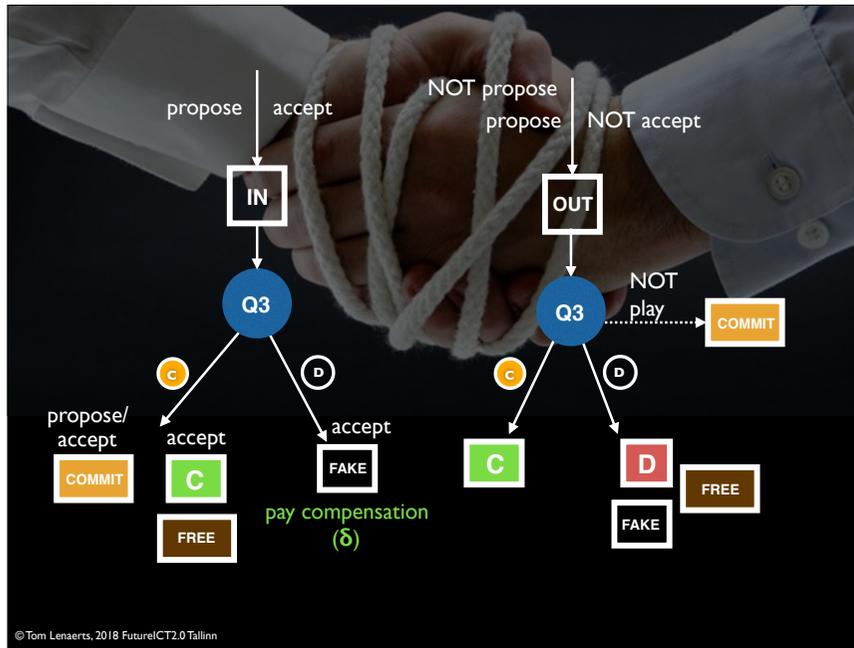
52-5



52-6



52-7



52-8

	COMMIT	C	D	FAKE	FREE
COMMIT	$R-\epsilon/2$	$R-\epsilon$	0	$S+\delta-\epsilon$	$R-\epsilon$
C	R	R	S	S	S
D	0	T	P	P	P
FAKE	$T-\delta$	T	P	P	P
FREE	R	T	P	P	P

© Tom Lenaerts, 2018 FutureICT2.0 Tallinn

53-1

	COMMIT	C	D	FAKE	FREE
COMMIT	$R-\epsilon/2$	$R-\epsilon$	0	$S+\delta-\epsilon$	$R-\epsilon$
C	R	R	S	S	S
D	0	T	P	P	P
FAKE	$T-\delta$	T	P	P	P
FREE	R	T	P	P	P

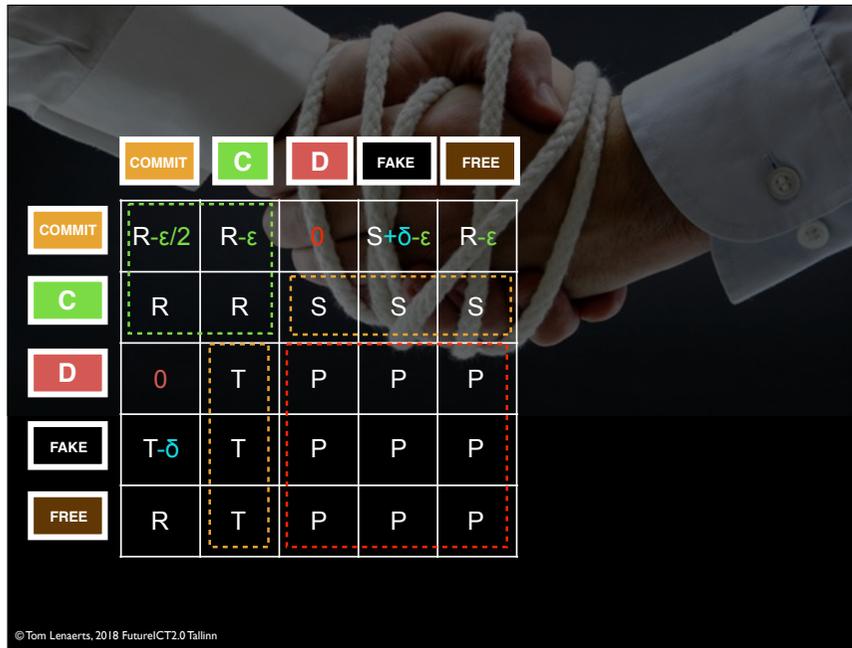
© Tom Lenaerts, 2018 FutureICT2.0 Tallinn

53-2

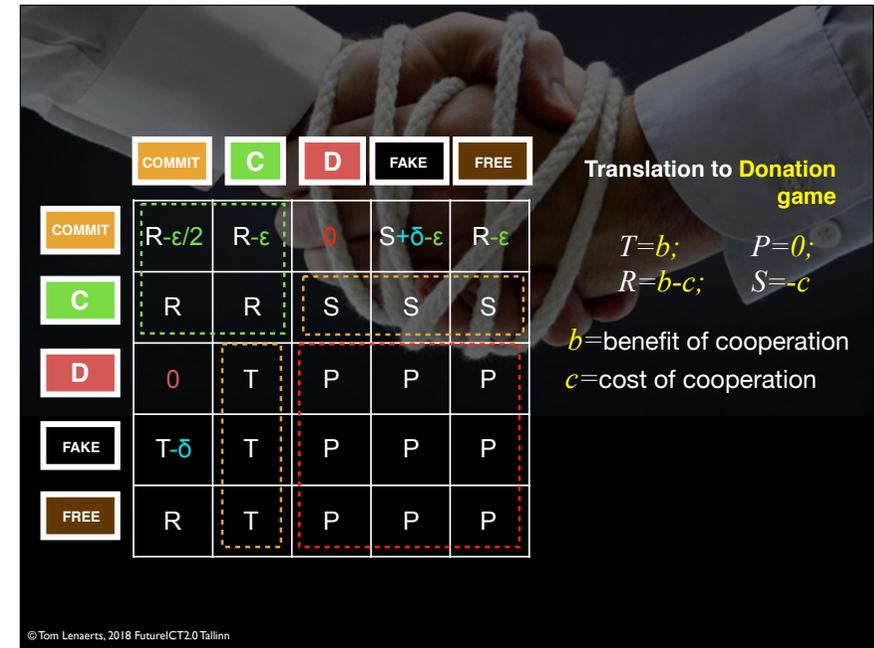
	COMMIT	C	D	FAKE	FREE
COMMIT	$R-\epsilon/2$	$R-\epsilon$	0	$S+\delta-\epsilon$	$R-\epsilon$
C	R	R	S	S	S
D	0	T	P	P	P
FAKE	$T-\delta$	T	P	P	P
FREE	R	T	P	P	P

© Tom Lenaerts, 2018 FutureICT2.0 Tallinn

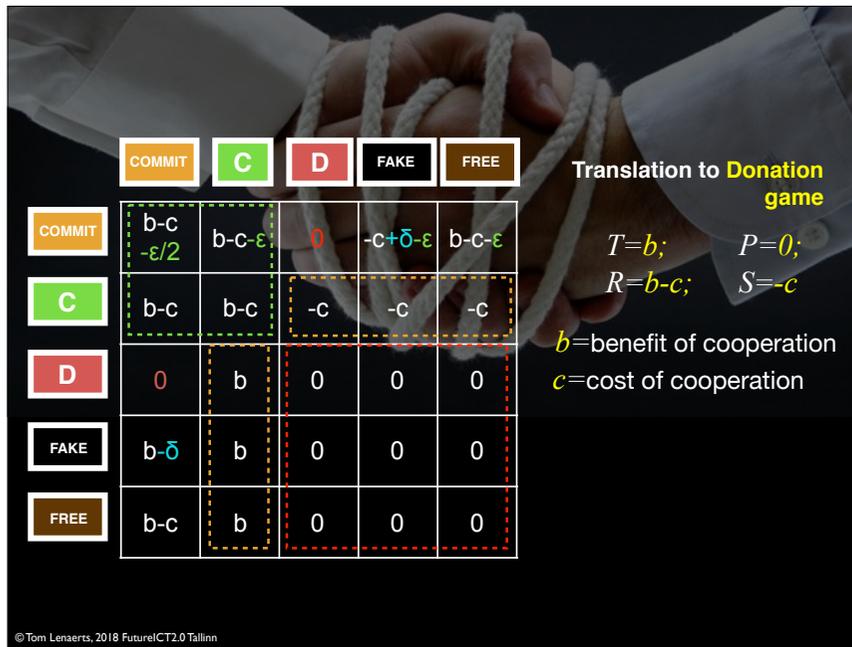
53-3



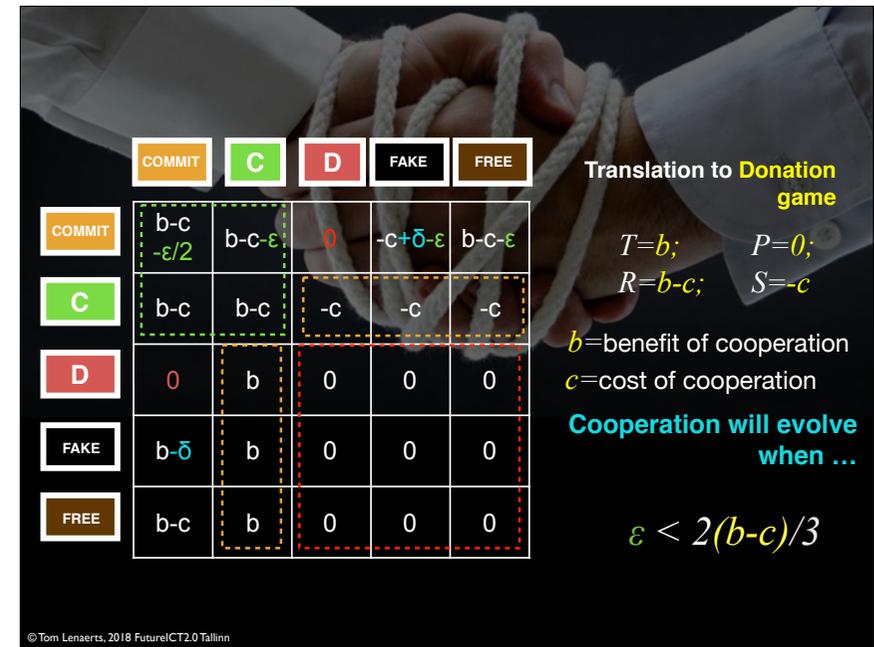
53-4



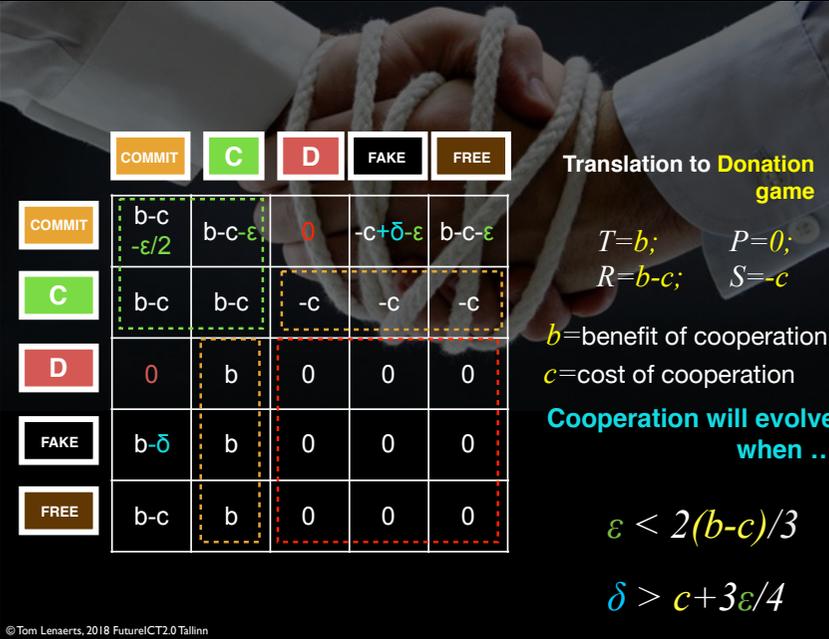
53-5



54-1



54-2



**Translation to Donation game**

	COMMIT	C	D	FAKE	FREE
COMMIT	$b-c$ $-\epsilon/2$	$b-c-\epsilon$	0	$-c+\delta-\epsilon$	$b-c-\epsilon$
C	$b-c$	$b-c$	-c	-c	-c
D	0	b	0	0	0
FAKE	$b-\delta$	b	0	0	0
FREE	$b-c$	b	0	0	0

$T=b;$        $P=0;$   
 $R=b-c;$      $S=-c$

$b$ =benefit of cooperation  
 $c$ =cost of cooperation

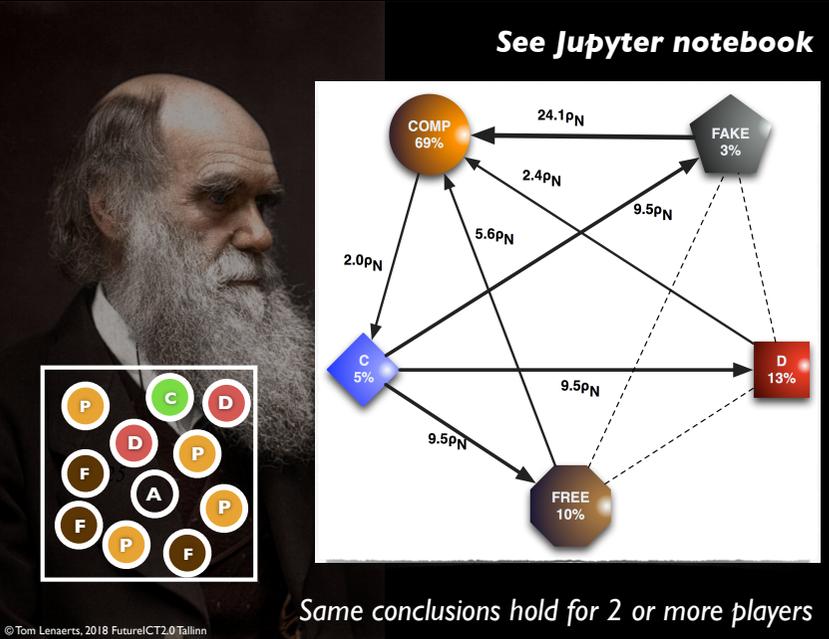
**Cooperation will evolve when ...**

$\epsilon < 2(b-c)/3$   
 $\delta > c+3\epsilon/4$

© Tom Lenaerts, 2018 FutureICT2.0 Tallinn

54-3

**See Jupyter notebook**



COMP 69%  
 FAKE 3%  
 C 5%  
 D 13%  
 FREE 10%

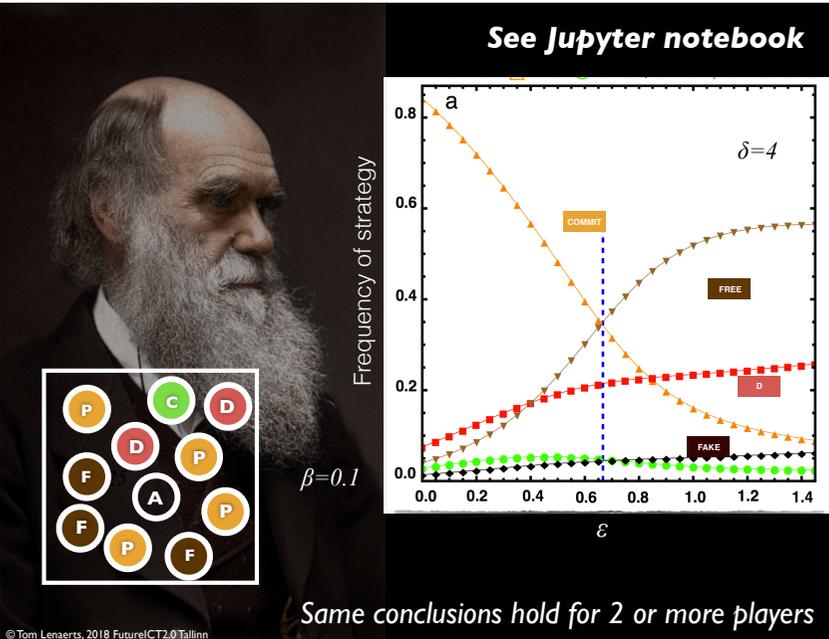
Interaction strengths:  
 COMP ↔ FAKE: 24.1pN  
 COMP ↔ C: 2.0pN  
 COMP ↔ D: 5.6pN  
 COMP ↔ FREE: 9.5pN  
 C ↔ D: 9.5pN  
 C ↔ FREE: 9.5pN  
 D ↔ FREE: 9.5pN  
 FAKE ↔ D: 9.5pN  
 FAKE ↔ FREE: 9.5pN

© Tom Lenaerts, 2018 FutureICT2.0 Tallinn

**Same conclusions hold for 2 or more players**

55

**See Jupyter notebook**



Frequency of strategy

$\delta=4$

$\beta=0.1$

COMMIT  
 FREE  
 D  
 FAKE

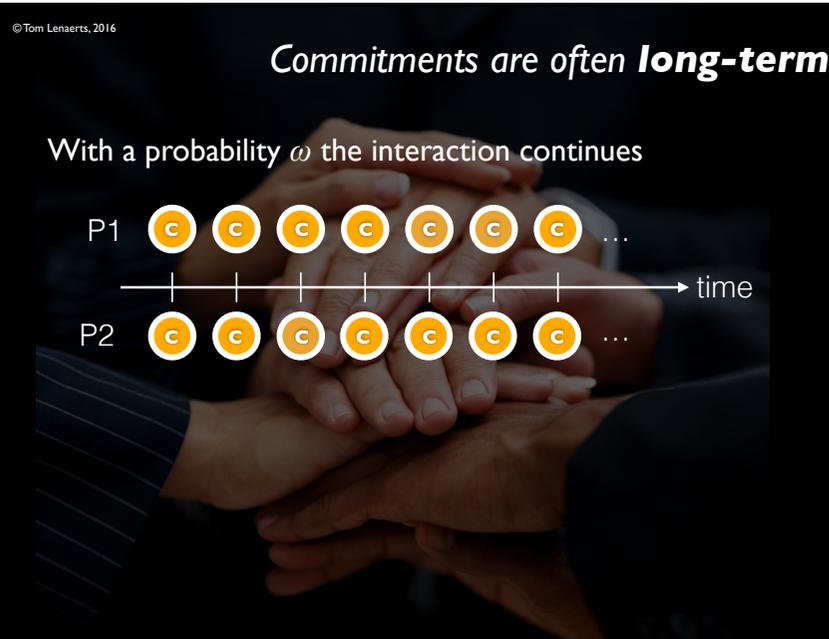
© Tom Lenaerts, 2018 FutureICT2.0 Tallinn

**Same conclusions hold for 2 or more players**

56

**Commitments are often long-term**

With a probability  $\omega$  the interaction continues



P1: C C C C C C ...  
 P2: C C C C C C ...

time

© Tom Lenaerts, 2016

57

© Tom Lenaerts, 2016

## Stochasticity, uncertainty and unpredictability

With a probability  $\omega$  the interaction continues

P1 C D C C C C D ...

P2 C C C D D C C ...

time

How to deal with mistakes ? (Occur with probability  $\alpha$ )

58-1

© Tom Lenaerts, 2016

## Stochasticity, uncertainty and unpredictability

With a probability  $\omega$  the interaction continues

P1 C D C C C C D ...

P2 C C C D D C C ...

time

How to deal with mistakes ? (Occur with probability  $\alpha$ )

**Should we collect the compensation or continue the agreement?**

58-2

© Tom Lenaerts, 2016

## Stochasticity, uncertainty and unpredictability

With a probability  $\omega$  the interaction continues

P1 C D C C C C D ...

P2 C C C D D C C ...

time

How to deal with mistakes ? (Occur with probability  $\alpha$ )

**Should we collect the compensation or continue the agreement?**

**Should one take revenge or apologise and forgive?**

58-3

© Tom Lenaerts, 2016

www.nature.com/scientificreports

# SCIENTIFIC REPORTS

**OPEN** Apology and forgiveness evolve to resolve failures in cooperative agreements

Received: 23 February 2015  
Accepted: 22 April 2015  
Published: xx xx xxx

Luis A. Martinez-Vaquero<sup>1,2</sup>, The Anh Han<sup>1</sup>, Luis Moniz Pereira<sup>1,2</sup> & T...

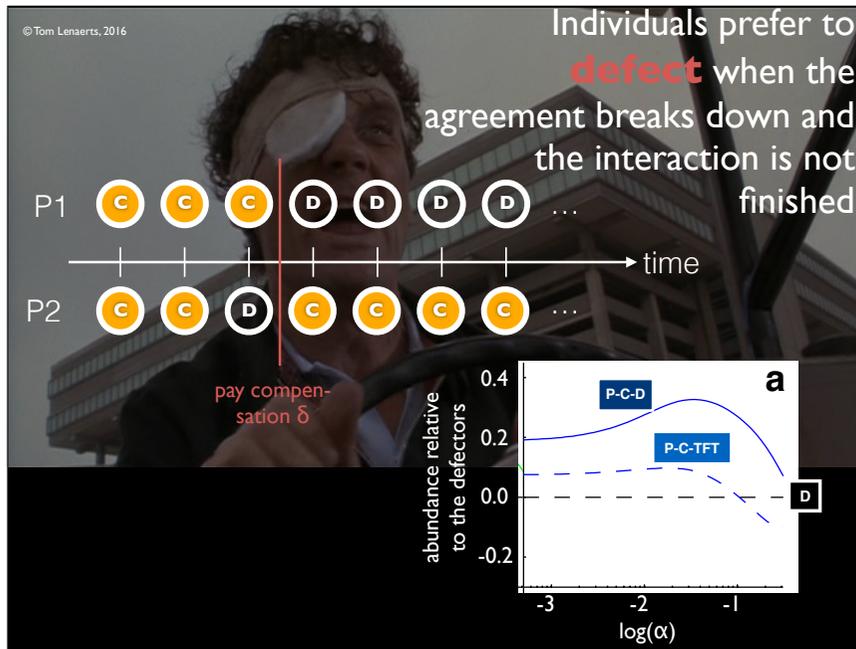
Making agreements on how to behave has been shown to be an evolution one-shot social dilemmas. However, in many situations agreements aim to mutually beneficial interactions. Our analytical and numerical results reveal which conditions revenge, apology and forgiveness can evolve and deal w agreements in the context of the iterated Prisoners Dilemma. We show th fails, participants prefer to take revenge by defecting in the subsisting en costly apology and forgiveness reveals that, even when mistakes are frequ levels of cooperation. In short, even when to err is human, revenge, apolo evolutionarily viable strategies which play an important role in inducing o dilemmas.

MICHAEL E. McCULLOUGH

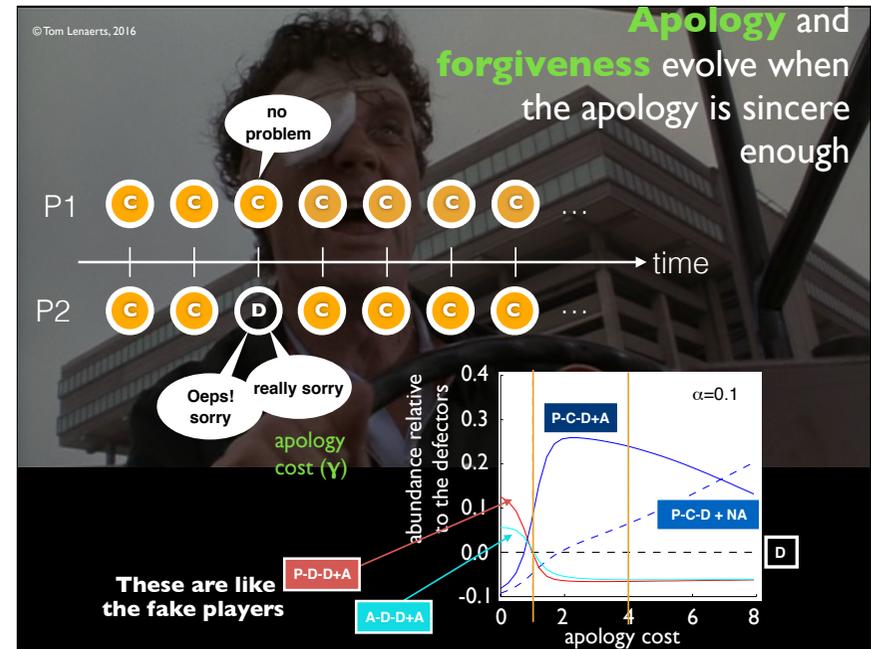
## BEYOND REVENGE

THE EVOLUTION OF THE FORGIVENESS INSTINCT

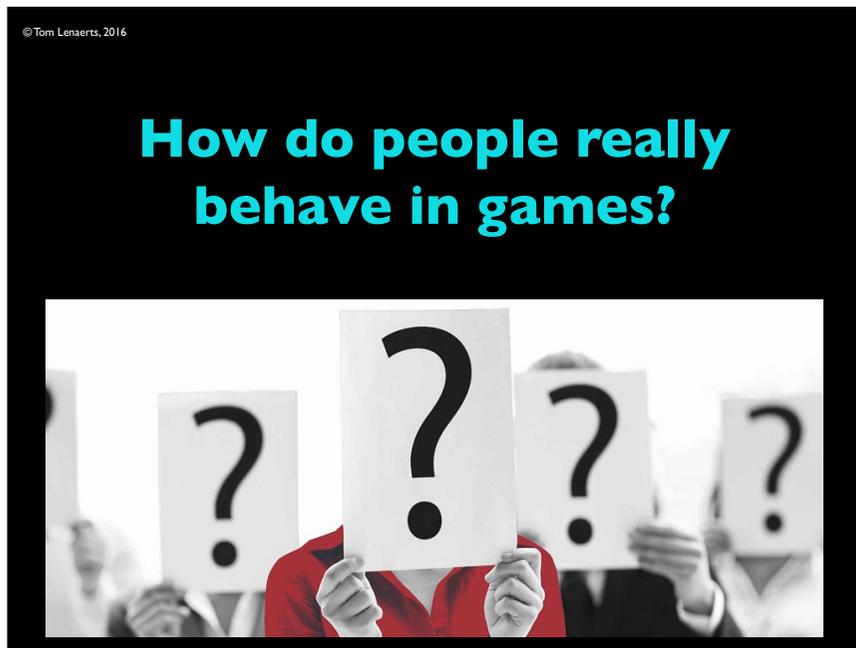
59



60



61



62

© Tom Lenaerts, 2015

## Inference from experiments

Vrije Universiteit Brussel

UNIVERSITÉ LIBRE DE BRUXELLES

ONLINE RECRUITMENT SYSTEM FOR ECONOMIC EXPERIMENTS

Main Sign up Calendar Rules Privacy policy FAQ

WELCOME TO BEEL  
Brussels Experimental Economics Lab

associated with scientific questions typical from the Vrije Universiteit Brussel and their experiments.

anonymously interact with other participants during their experimental gains, where it ends on the decisions made by oneself at live other tasks than making decisions during this is explicitly mentioned in an invitation.

63

Tom Lenaerts - Omins 2019

How to arrive at a legally binding universal agreement to avoid a global 2° temperature increase?

64

Tom Lenaerts - Omins 2019

### Current Climate Pledges Aren't Enough to Stop Severe Warming

Existing pledges under the Paris Agreement won't curb warming to two degrees Celsius above preindustrial levels, a new UN report warns.

BY MICHAEL GRESHKO 3 MINUTE READ

PUBLISHED OCTOBER 31, 2017

The world is not doing enough to curb its collective carbon emissions, a [new UN report](#) warns.

In an audit of the [global Paris Agreement](#) released Tuesday, the UN Environment Programme finds that if action to combat climate change is limited to just current pledges, the Earth will get at least three degrees Celsius (5.4 degrees Fahrenheit) warmer by 2100 relative to preindustrial levels.

This amount of warming would vastly exceed the Paris Agreement's goal, which is to limit global warming by the end of the century to two degrees

Voluntary pledges require **trust** and **cooperative** participants

SCIENTIFIC REPORTS

OPEN Good Agreements Make Good Friends

Non-costly, non-enforced agreements lead to defection

SUBJECT ABFAJ The Anh Han<sup>1,2</sup>, Luis Morán Pereira<sup>1</sup>, Francisco C. Santos<sup>1,2</sup> & Tom Lenaerts<sup>1,2</sup>

65

Tom Lenaerts - Omins 2019

## How to study decision-making in the climate change issue.

66-1

Tom Lenaerts - Omins 2019

## How to study decision-making in the climate change issue.

How to model this problem?

66-2

# How to study decision-making in the climate change issue.

How to model this problem?

How do humans make decisions?

66-3

# How to study decision-making in the climate change issue.

How to model this problem?

How do humans make decisions?

What features influence this decision?

66-4

## 1 Uncertainty encourages group reciprocity and polarization in a high-risk climate change problem

3

4 Elias Fernández Domingos<sup>1,2,6,\*</sup>, Jelena Grujić<sup>1,2,\*</sup>, Juan C. Burguillos<sup>4</sup>, Georg Kirchsteiger<sup>3</sup>, Francisco C. Santos<sup>4,5</sup>, and Tom Lenaerts<sup>1,2,2</sup>

6

7 <sup>1</sup>AI lab, Computer Science Department, Vrije Universiteit Brussel, Pleinlaan 9, 3<sup>rd</sup> floor, 1050 Brussels, Belgium

8

9 <sup>2</sup>MLG, Département d'Informatique, Université Libre de Bruxelles, Boulevard du Triomphe, CP 212, 1050 Brussels, Belgium

10

11 <sup>3</sup>ECARES, Université Libre de Bruxelles, Av. Roosevelt 42, CP 114, 1050 Brussels, Belgium

12

13 <sup>4</sup>INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, IST-Taguspark, 2744-016 Porto Salvo, Portugal

14

15 <sup>5</sup>ATP-group, 2744-016 Porto Salvo, Portugal.

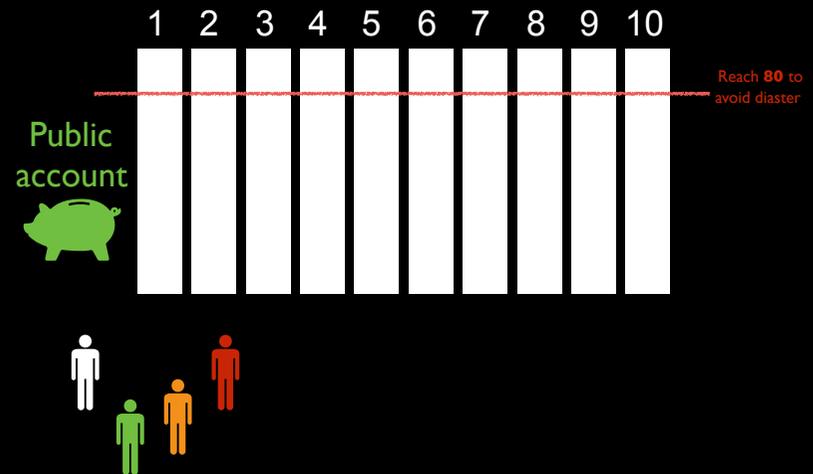
16

16 <sup>6</sup>Department of Telematic Engineering, University of Vigo, 36310 Vigo, Spain

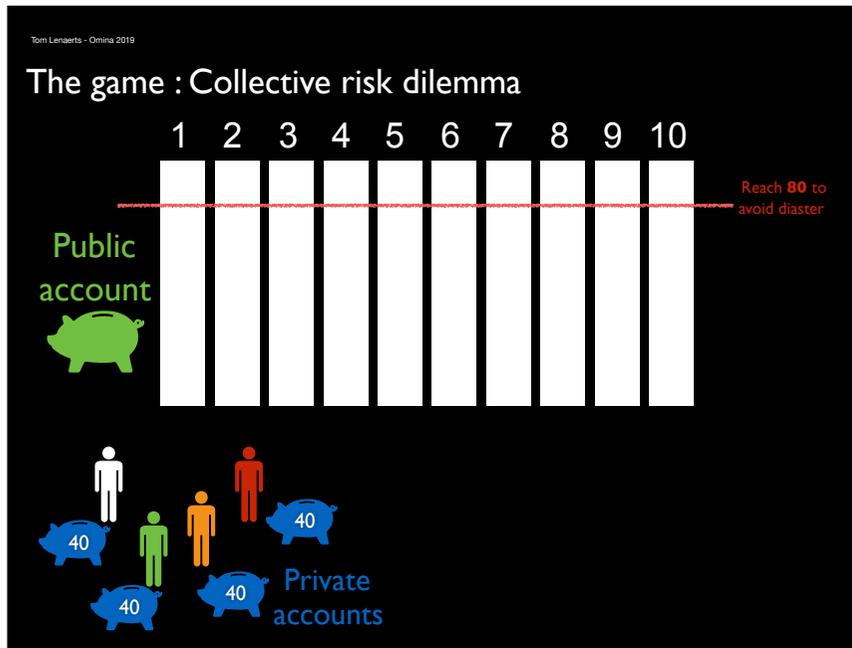
PRELIMINARY

67

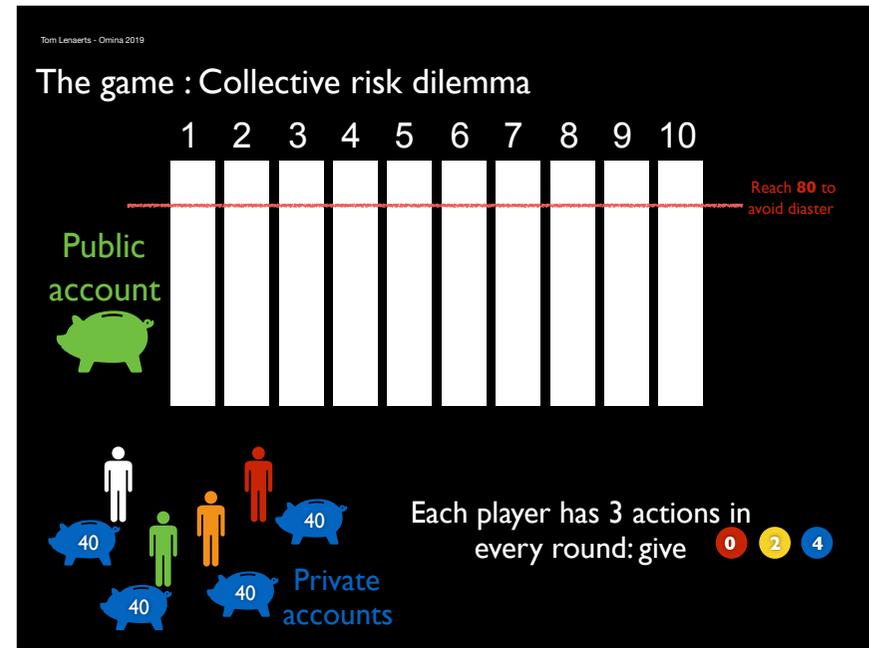
## The game : Collective risk dilemma



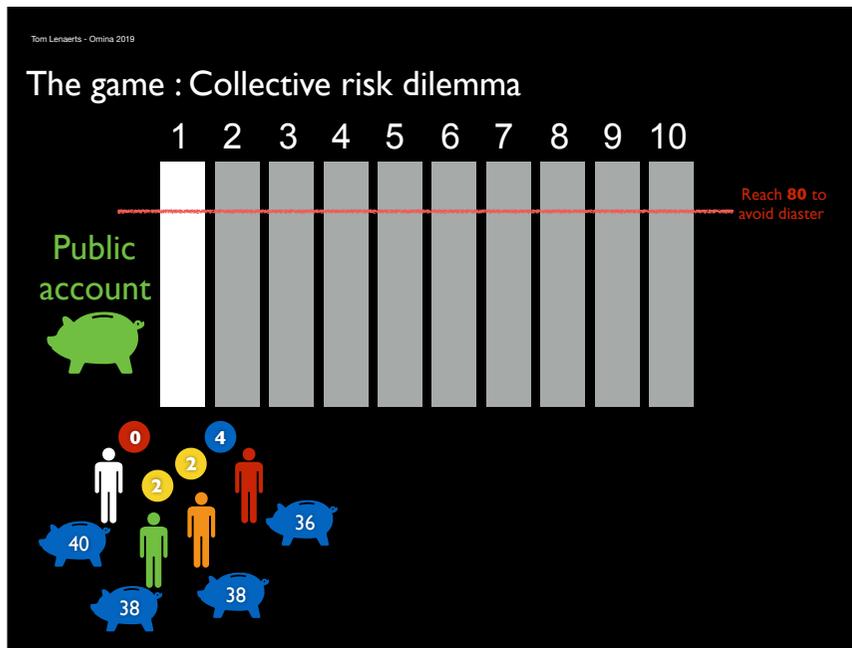
68-1



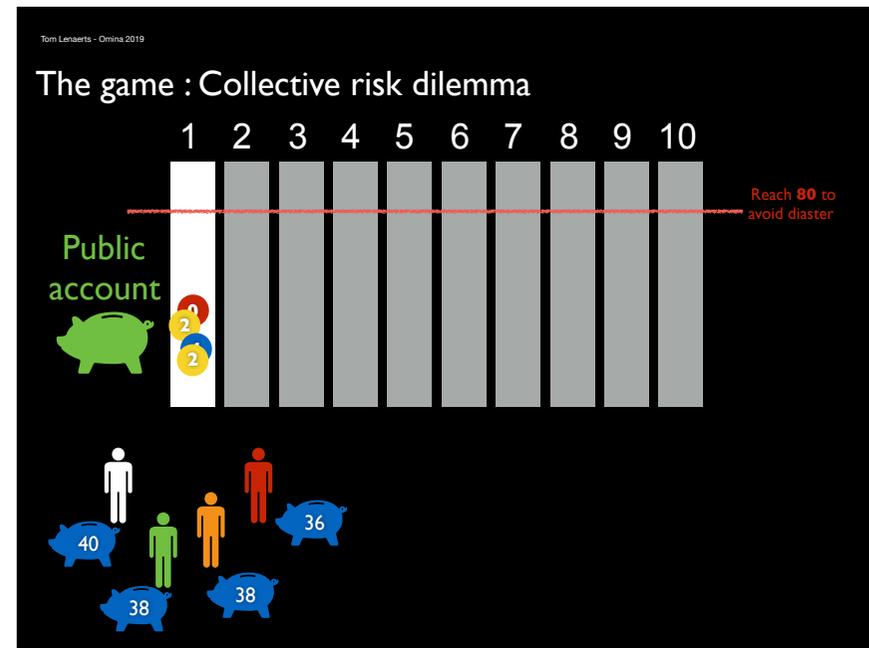
68-2



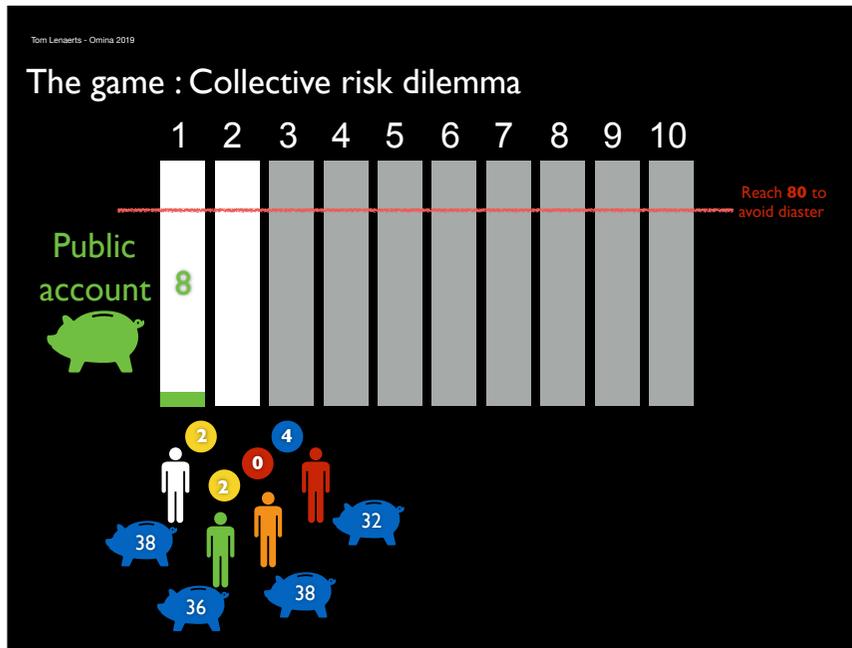
68-3



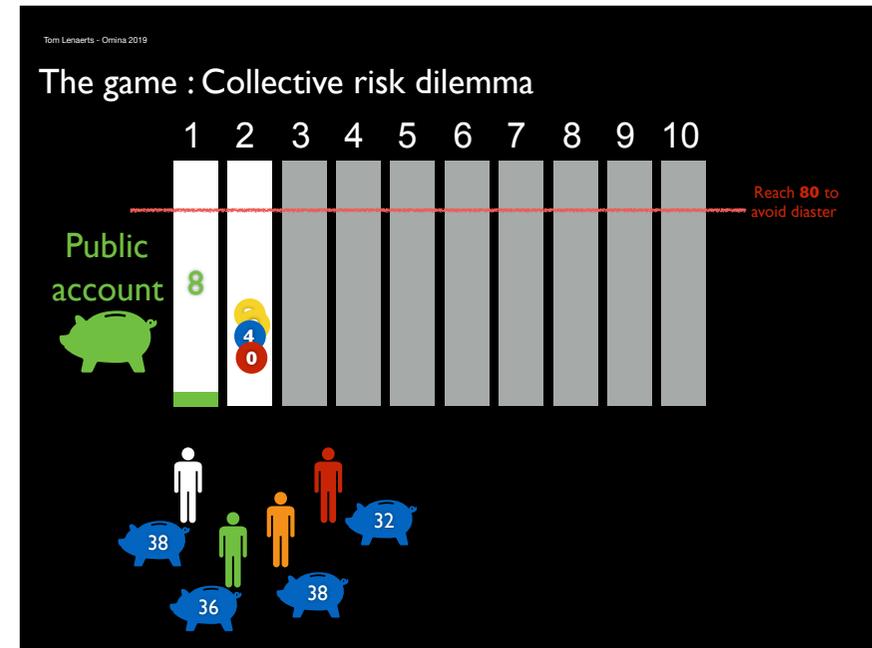
69-1



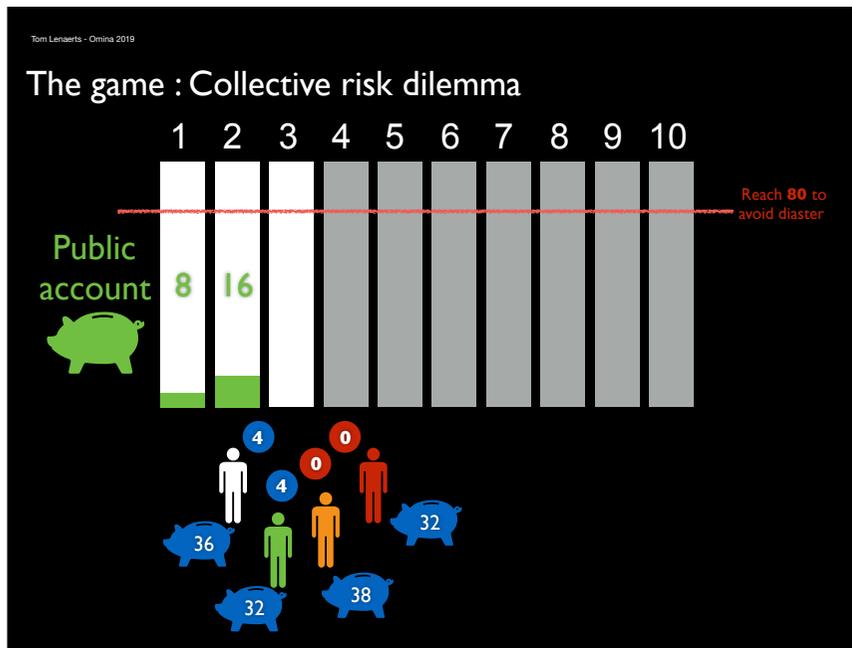
69-2



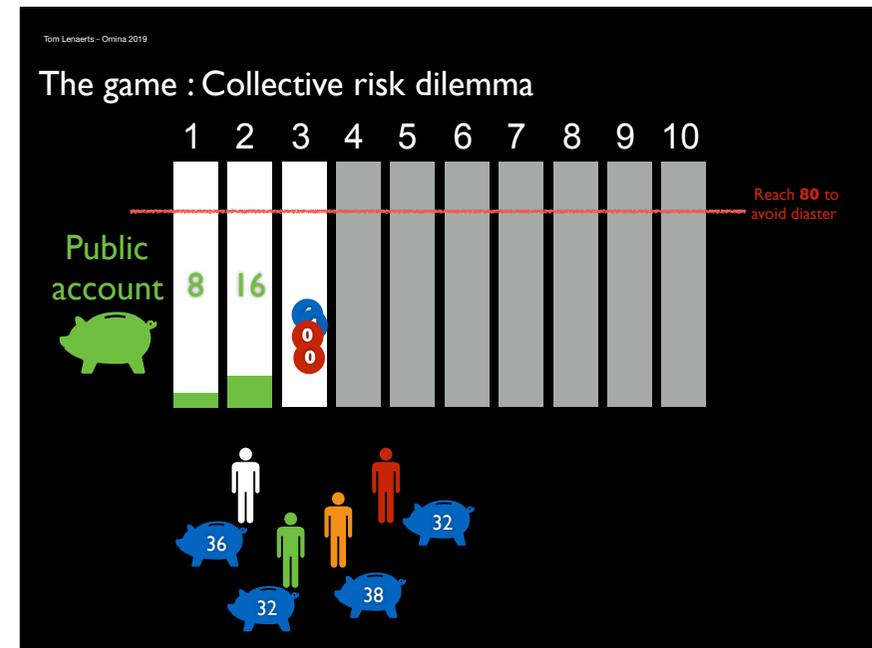
70-1



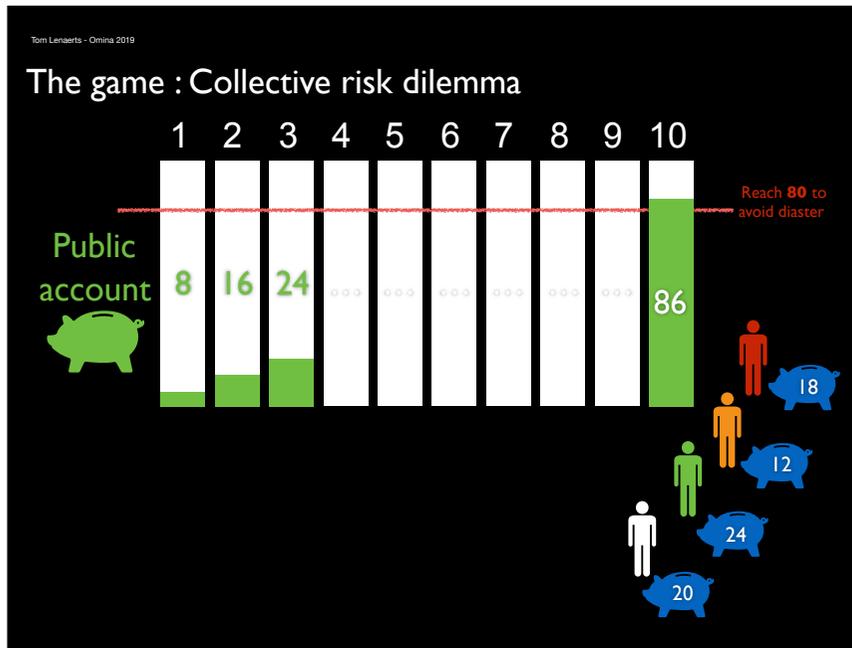
70-2



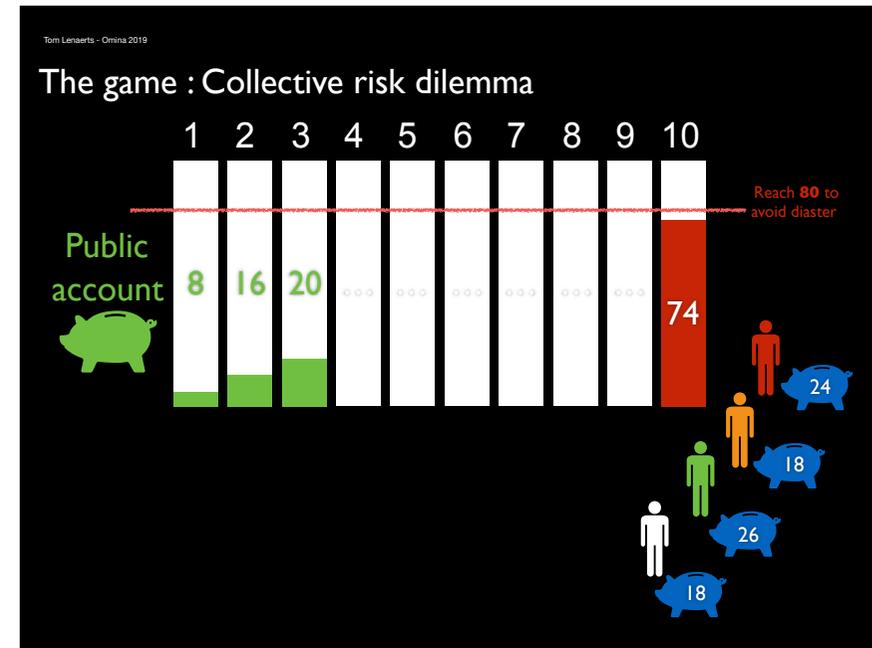
71-1



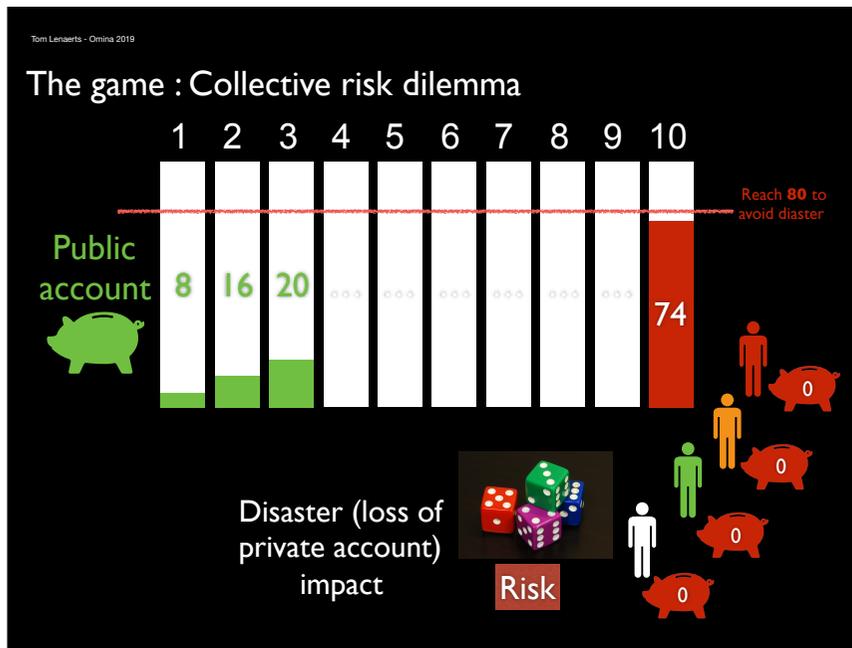
71-2



72



73-1



73-2

Tom Lenaerts - Omina 2019

### Uncertainties

74-1

# Uncertainties



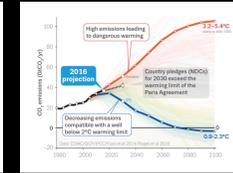
impact

74-2

# Uncertainties



impact



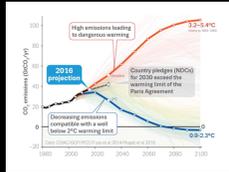
threshold

74-3

# Uncertainties



impact



threshold



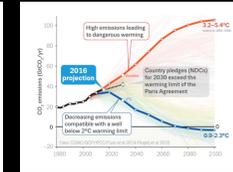
timing

74-4

# Uncertainties



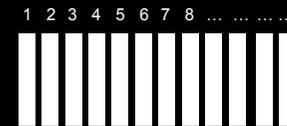
impact



threshold



timing



Game ends with probability  $\omega$



74-5

Tom Lenaerts - Omina 2019




**No uncertainty**

- 6 participants
- Risk 90%
- Threshold 120
- Endowment 40
- Rounds 10**

75-1

Tom Lenaerts - Omina 2019




**No uncertainty**

- 6 participants
- Risk 90%
- Threshold 120
- Endowment 40
- Rounds 10**

**Low uncertainty**

- Same **BUT**
- $\omega = 1/3$
- min. 8 rounds
- Avg. rounds 10**

75-2

Tom Lenaerts - Omina 2019




**No uncertainty**

- 6 participants
- Risk 90%
- Threshold 120
- Endowment 40
- Rounds 10**

**Low uncertainty**

- Same **BUT**
- $\omega = 1/3$
- min. 8 rounds
- Avg. rounds 10**

**High uncertainty**

- Same **BUT**
- $\omega = 1/5$
- min. 6 rounds
- Avg. rounds 10**

75-3

Tom Lenaerts - Omina 2019

round 2 of 10

Donations of the previous round					
You	Other members of the group				
2	0	2	2	0	2

Time left: 00:53

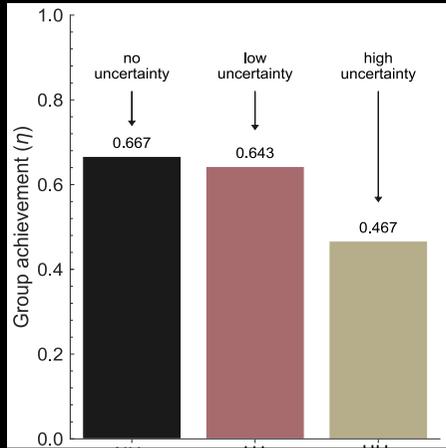
Personal Account: 38 EMUs

How many EMUs do you want to contribute to the public account?

Select one of the following options.

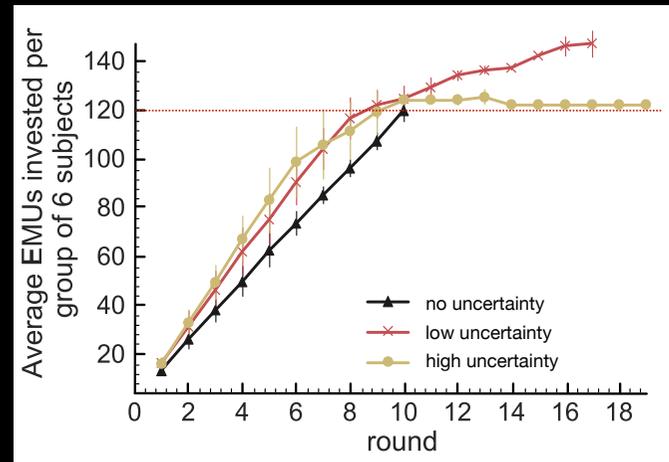
0 2 4

76



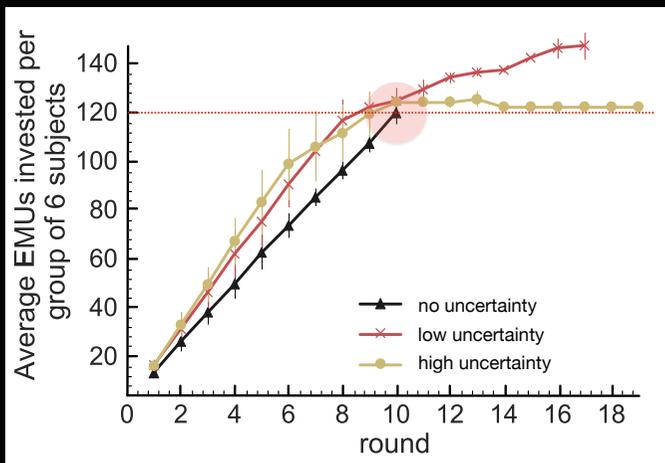
No significant differences in group success

77



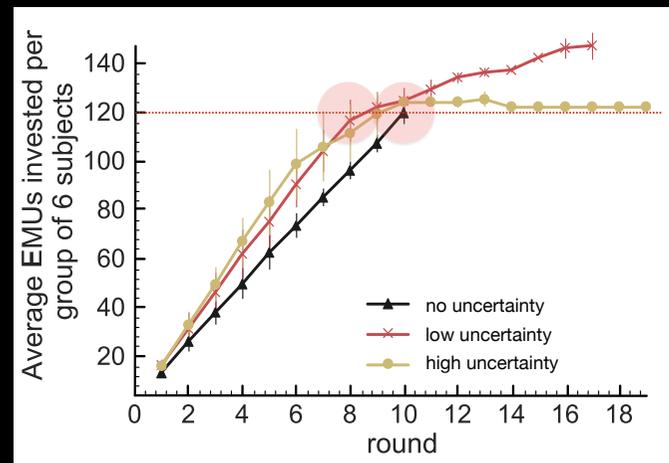
Players invest early in when there is timing uncertainty

78-1



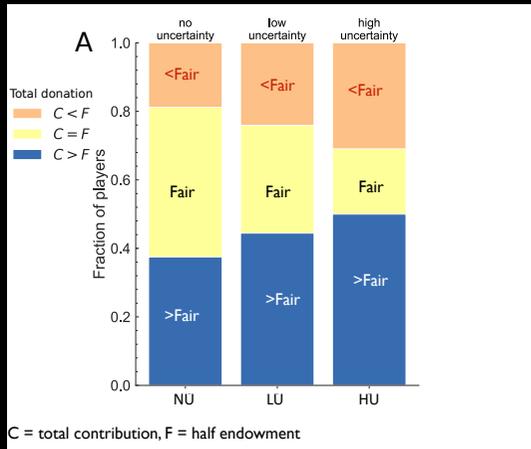
Players invest early in when there is timing uncertainty

78-2



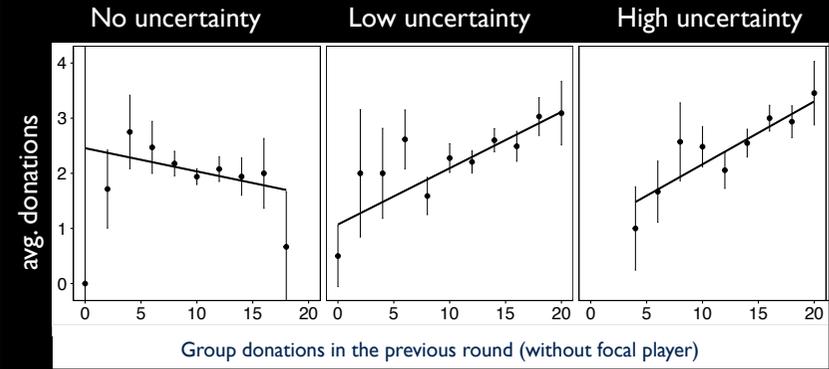
Players invest early in when there is timing uncertainty

78-3



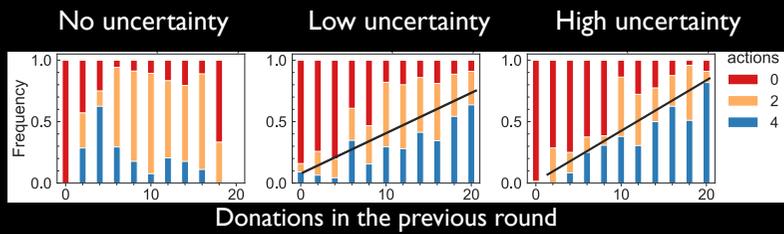
Participants become more polarised

79



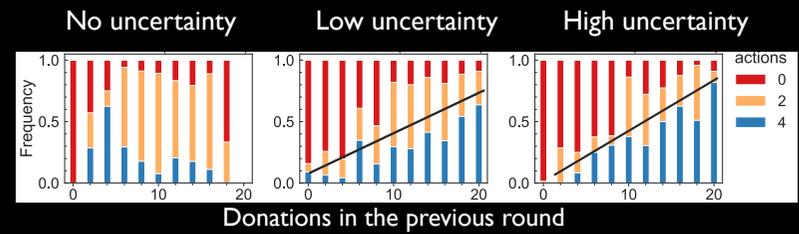
Group reciprocal strategies are needed to avoid disaster

80



What other behaviours ?

81-1



Uncertainty induces reciprocal behaviour

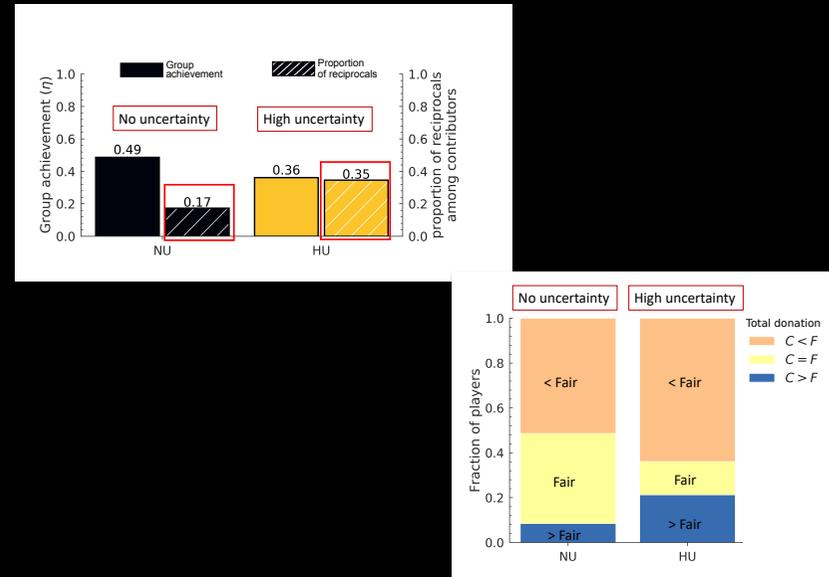
What other behaviours ?

81-2

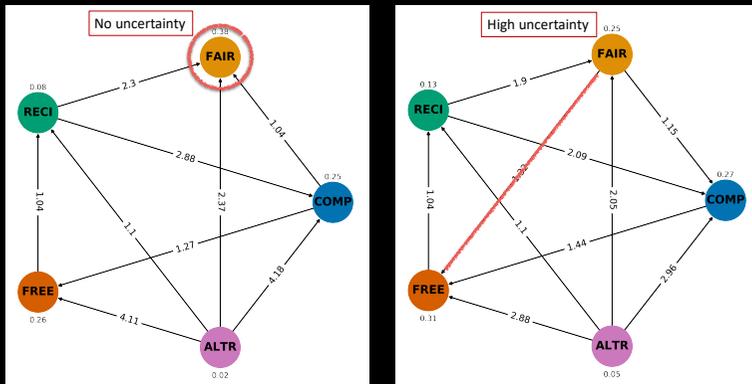
- F** Fair players, give 2
- D** Defecting players (**free riders**), give 0
- A** Altruistic players, give 4
- C** Compensators players, give 0 when prior investment was  $\geq 10$ , otherwise 4
- R** Reciprocators players, give 4 when prior investment was  $\geq 10$ , otherwise 0

\*All end when threshold is reached

82



83



84

**Uncertainty influences behaviour**

**Timing uncertainty**

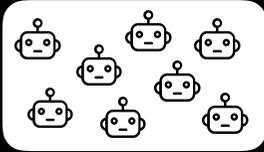
Leads to **polarisation**

Success requires **reciprocal mechanisms**

**Investing early** promotes coordination among peers

**EGT models explain experiments**

85



# Will AI agents do better?

What happens when we delegate the responsibility to an agent of our choosing

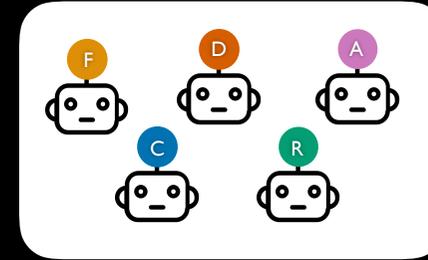
Do we like to delegate our choices to agents?

What agents will be selected?

Will the outcome improve?

86

# Delegation game

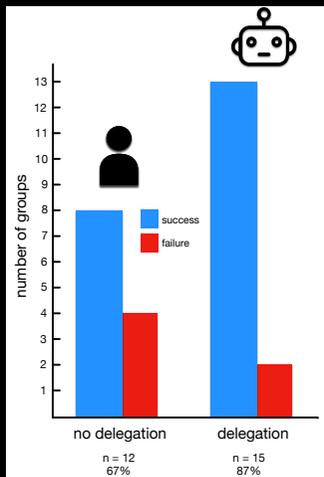


Which agent should play the CRD\* for you?

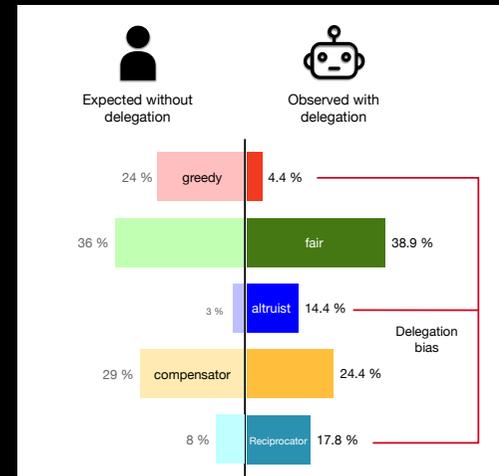


\*without uncertainty

87

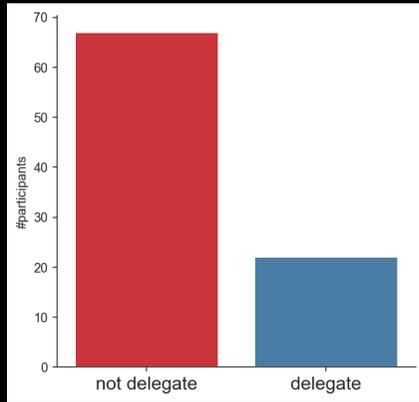


88



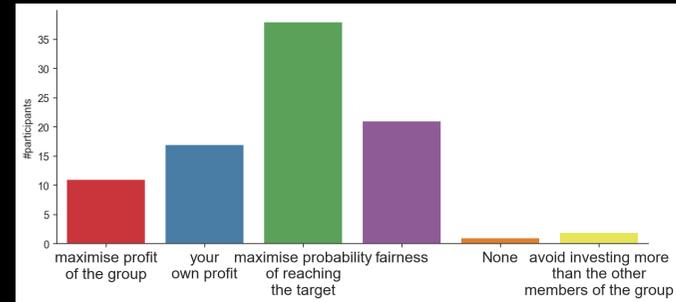
People's choices are heterogeneous

89



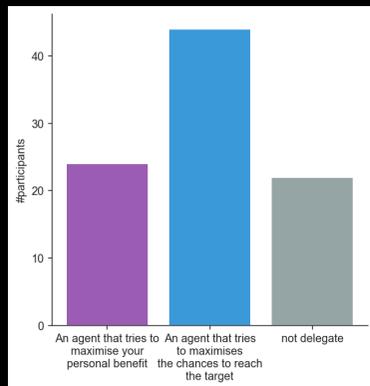
They would not delegate again

90



Most wanted to reach the target

91



Would select an agent that maximises the chance to reach the target

92



Delegation to agents increases success, but **humans prefer to act for themselves**

All understood that the goal is to reach the target to guarantee a benefit

93



Delegation forces to **think about future rewards**

**Fear of betrayal in the game disappears** as  
behaviour is fixed

**Social norms are no longer used** when  
delegating to an agent

Alimone, J. A., & Houser, D. (2012). What you don't know won't hurt you: a laboratory analysis of betrayal aversion. *Experimental Economics*, 15(4), 571-588.  
Ensthaler, L., Huck, S., & Leutgeb, J. (2019). Games played through agents in the laboratory—a test of Prat & Rustichini's model. *Games and Economic Behavior*.  
Prat, A., & Rustichini, A. (2003). Games played through agents. *Econometrica*, 71(4), 989-1026.  
Tom Lenaerts - Omnia 2019

94



95