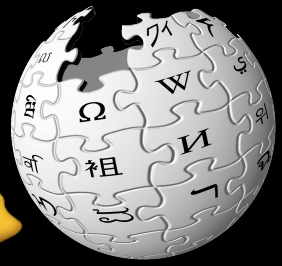
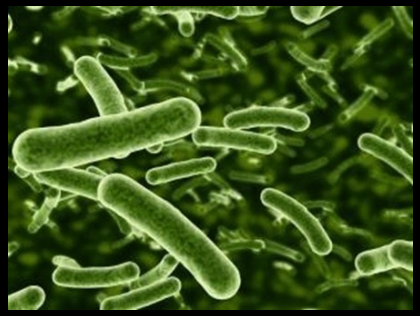
DEI
DEPARTAMENTO
DE ENGENHARIA INFORMÁTICA
TÉCNICO LISBOA

Public goods and collective risks

the origins of cooperation





Groups, kin, reciprocity, memory, reputations, moral systems, signals, gossip, networks, etc etc

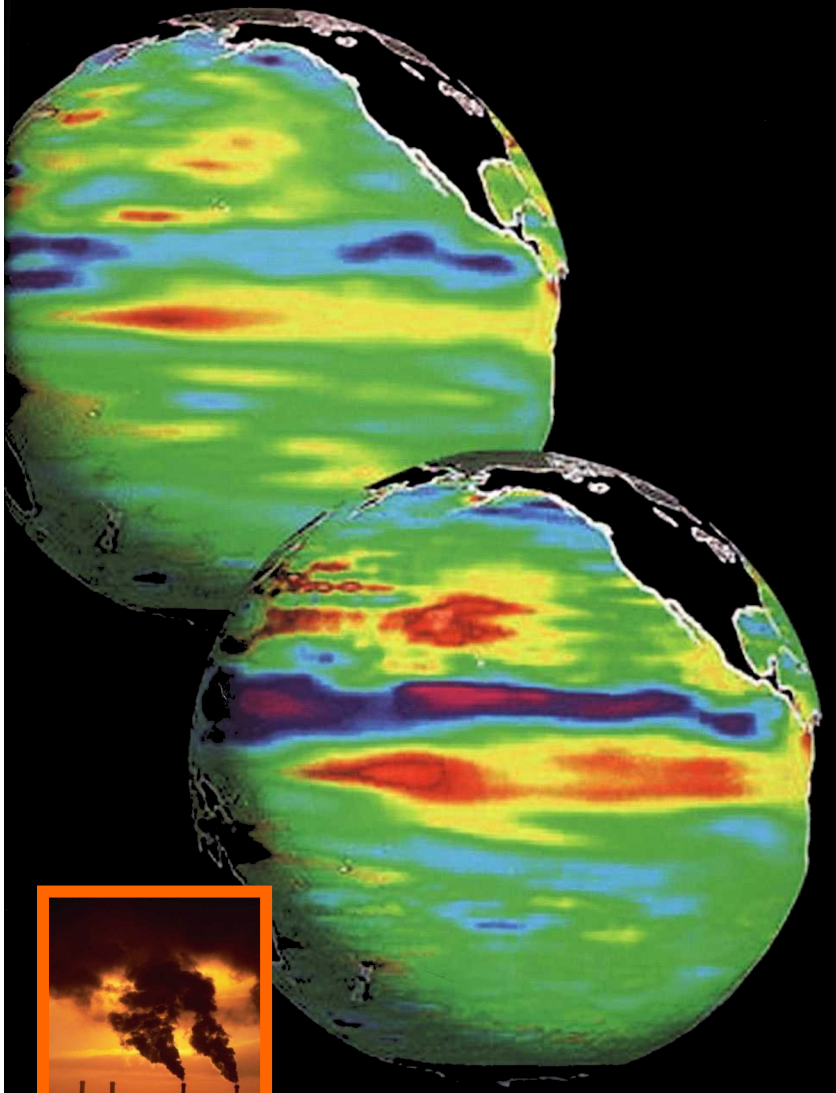
Are we missing a lot?

YES!!!

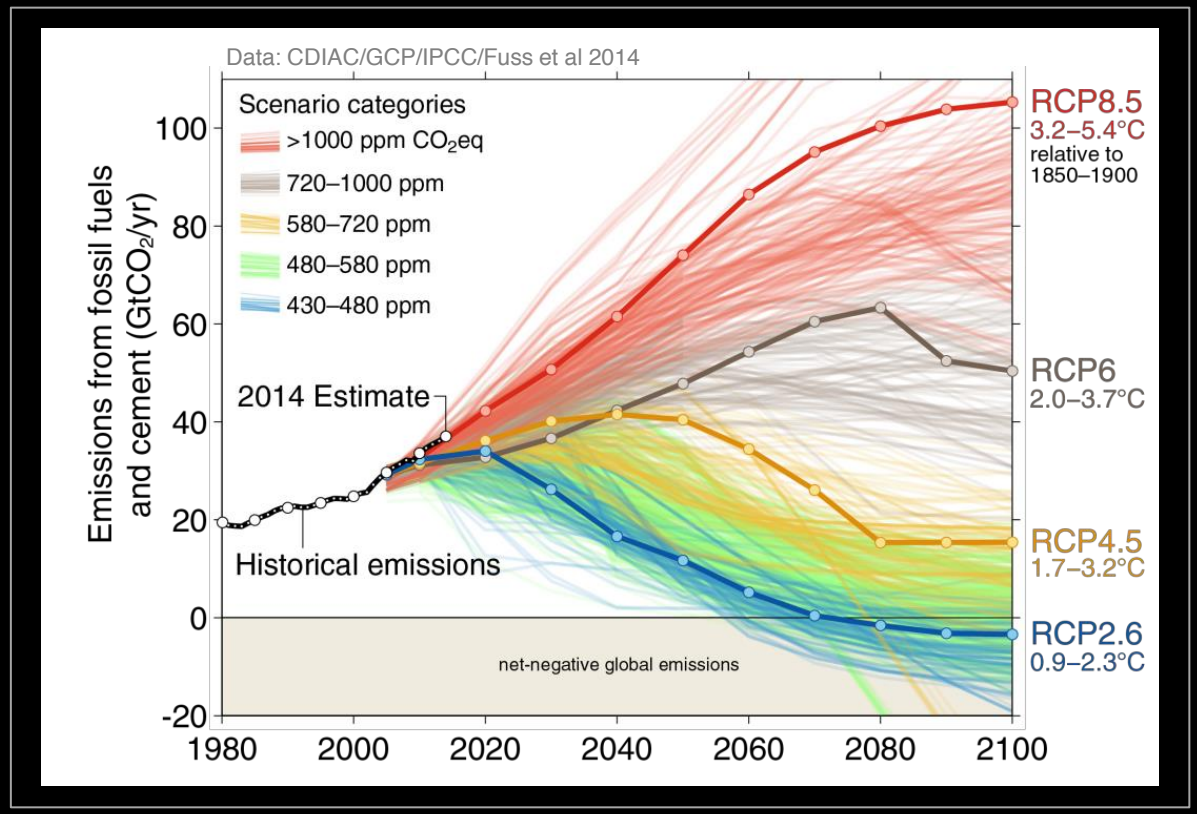
Missing pieces

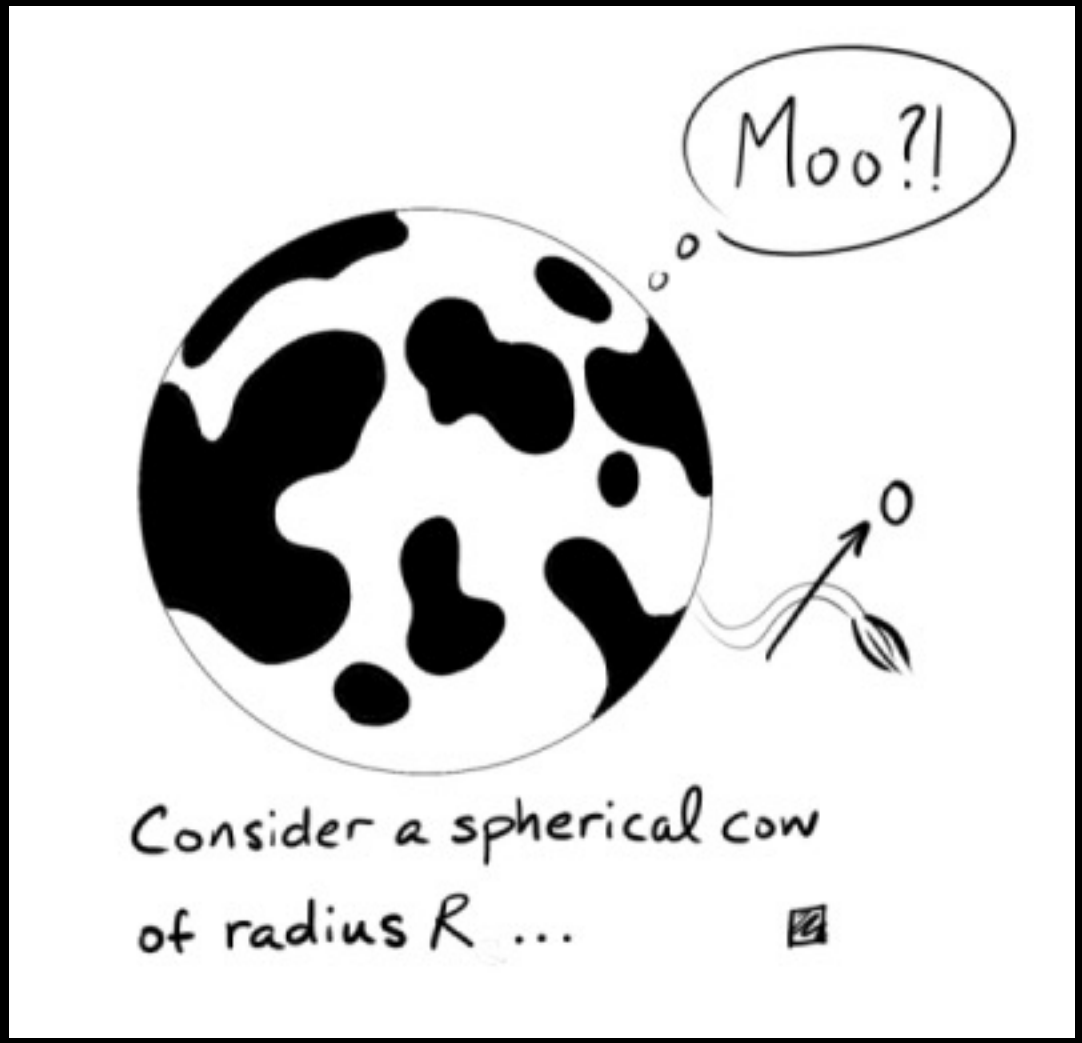
Often cooperation dilemmas involve many individuals simultaneously

The most important “cooperation game” we face...
and the one we can’t afford to lose



A sustainable planet requires… cooperation !



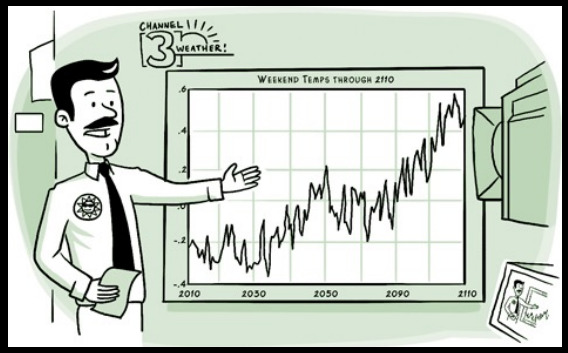

Moo?!
Consider a spherical cow
of radius R ...

overview

1

Behavioral experiments and population dynamics approaches to climate dilemmas

2

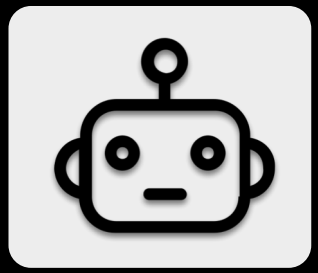


The role of uncertainty and scale

3

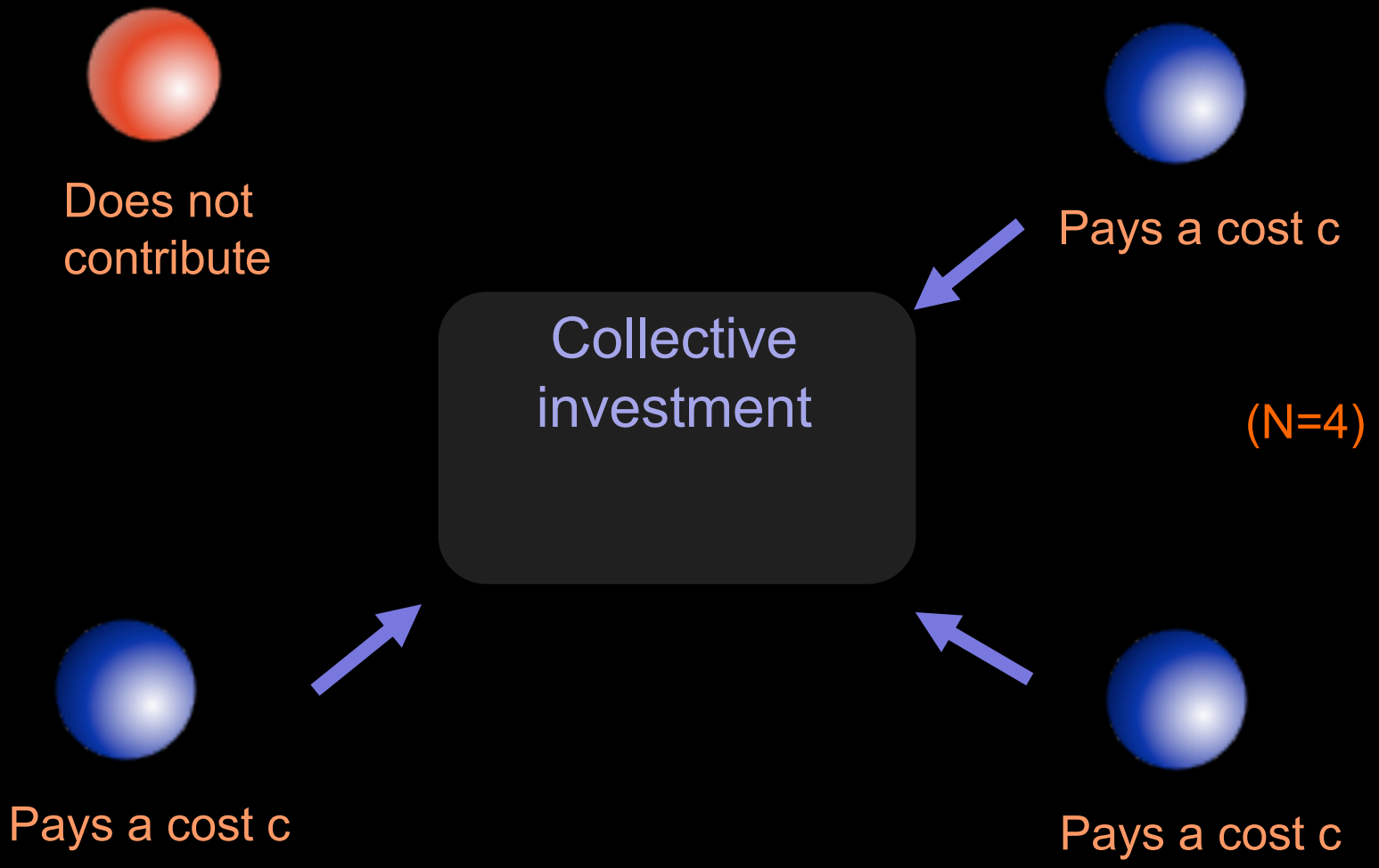
Self-organized institutions

4

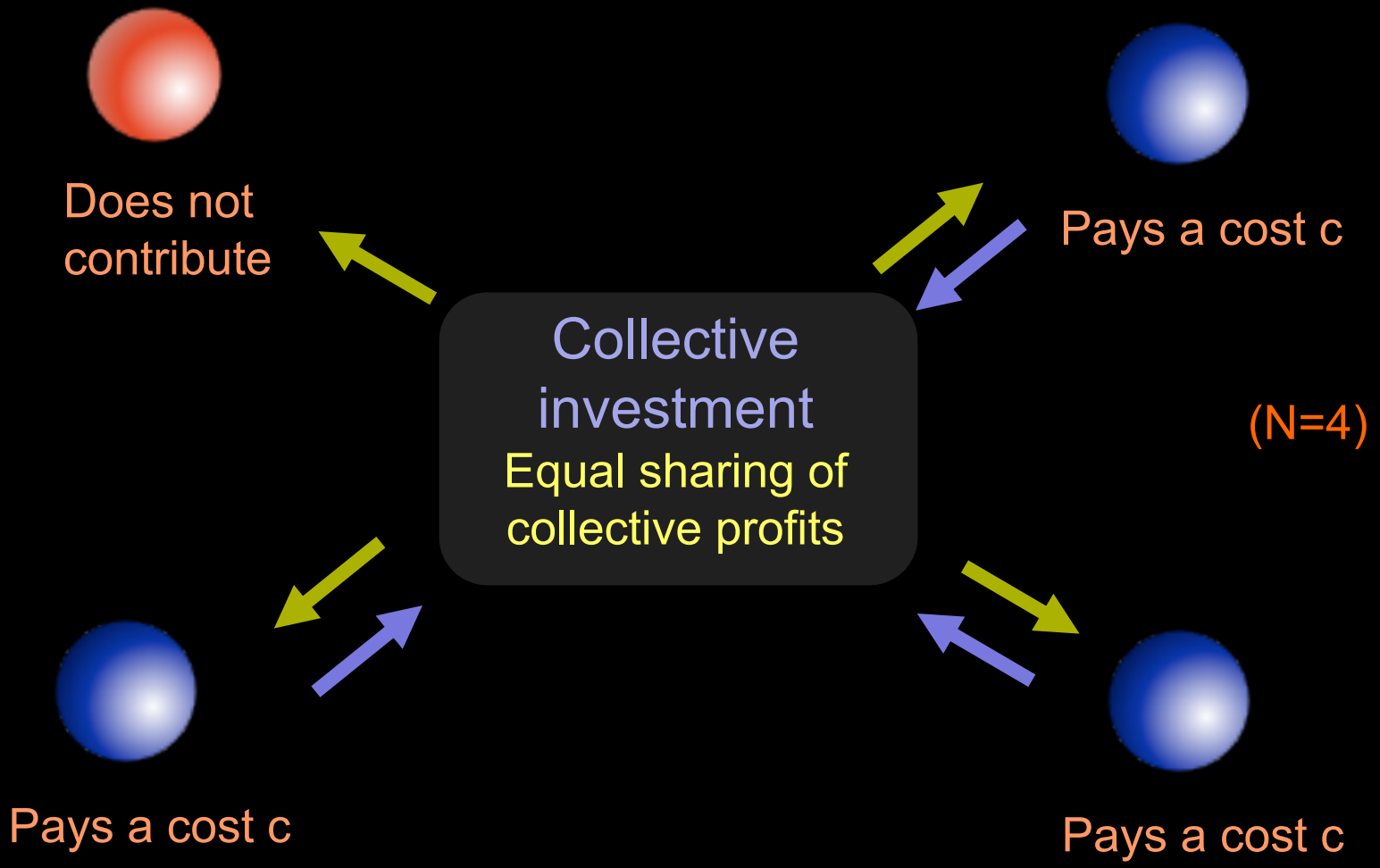


challenges ahead

Public goods games



Public goods games



The problem of discounting tomorrow

the challenge of minimizing the effects of climate change

It concerns everyone!
cooperation between individuals ?
cooperation between countries ?
cooperation between cities ?

the challenge of minimizing the effects of climate change

cooperation between ALL countries@once ?

& how about

multiple agreements involving few countries, states or regions?

Collective-risk dilemma

[Milinski et al., PNAS 105 (2008) 2291]
[Santos & Pacheco, PNAS 2011]
[Vasconcelos, Santos, Pacheco, Nat. Climate Change 2013]
[Vasconcelos, Santos, Pacheco, Levin, PNAS 2014]
[Domingos, et al. eScience 2020]

6 players, 10 rounds
each player : 40 €
contribution in each round : 0 (selfish), 2 (fair) or 4 (altruistic)
cost for saving the planet : 120 €

if Σ contributions ≥ 120 €, planet is saved and each gets away with money left
if Σ contributions < 120 €, there is a collective disaster with a probability "r" :

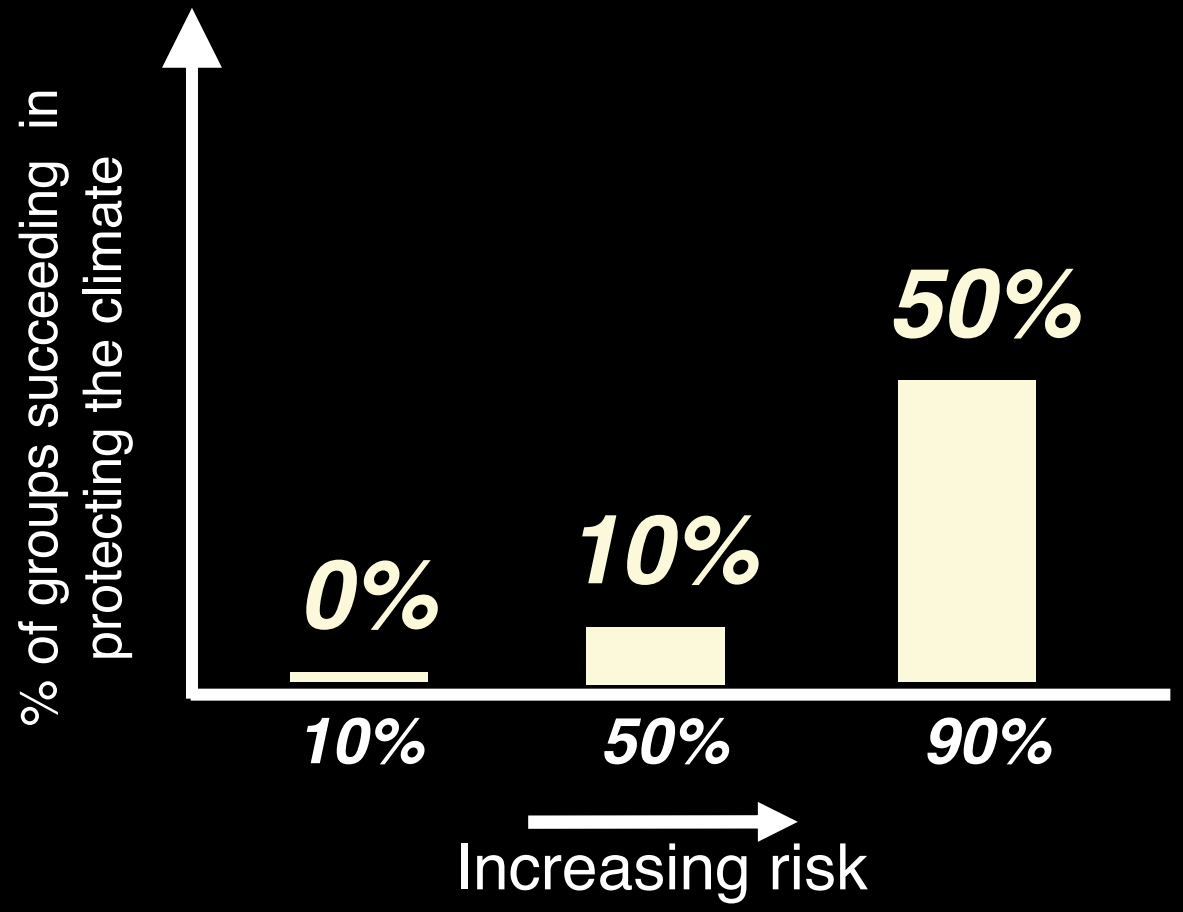
Increasing risk of disaster ↓

a) 10%
b) 50%
c) 90%

and all loose everything
…otherwise each gets away with the money left

Collective-risk dilemma

Dependence on the risk of collective failure:



[Milinski *et al.*, *PNAS* 195 (2008) 2291]

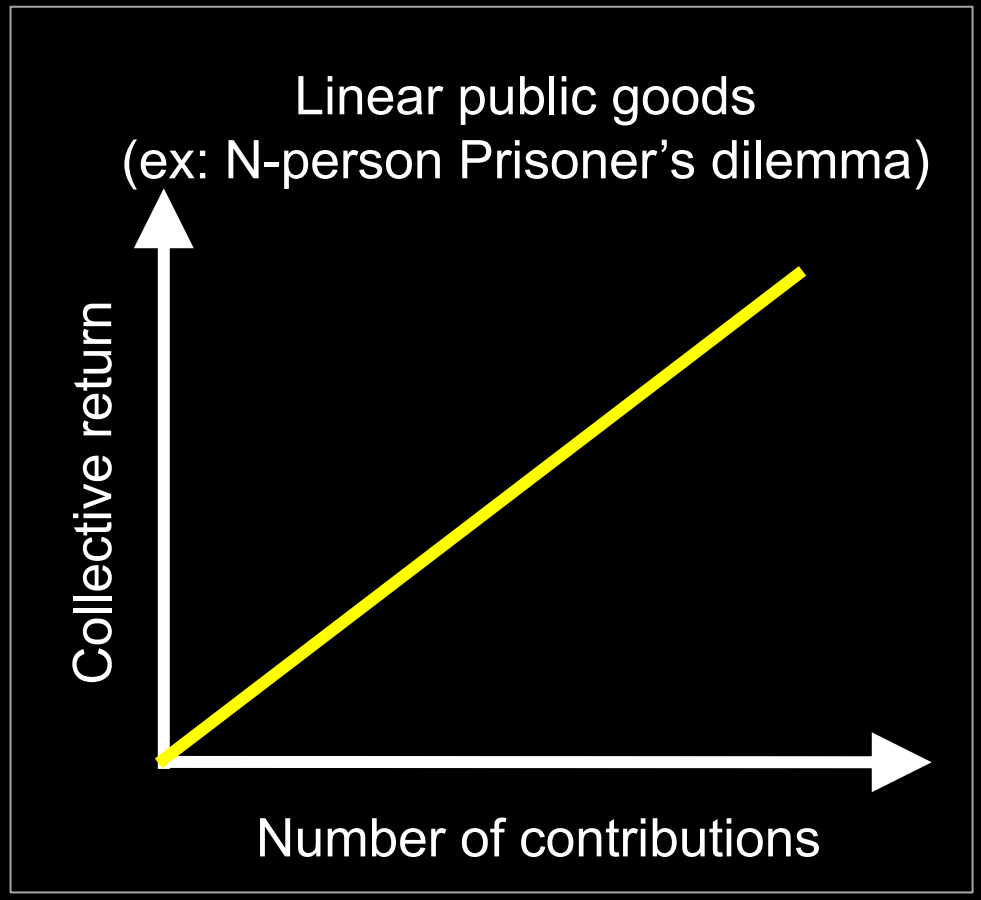
message from the game experiment on climate change

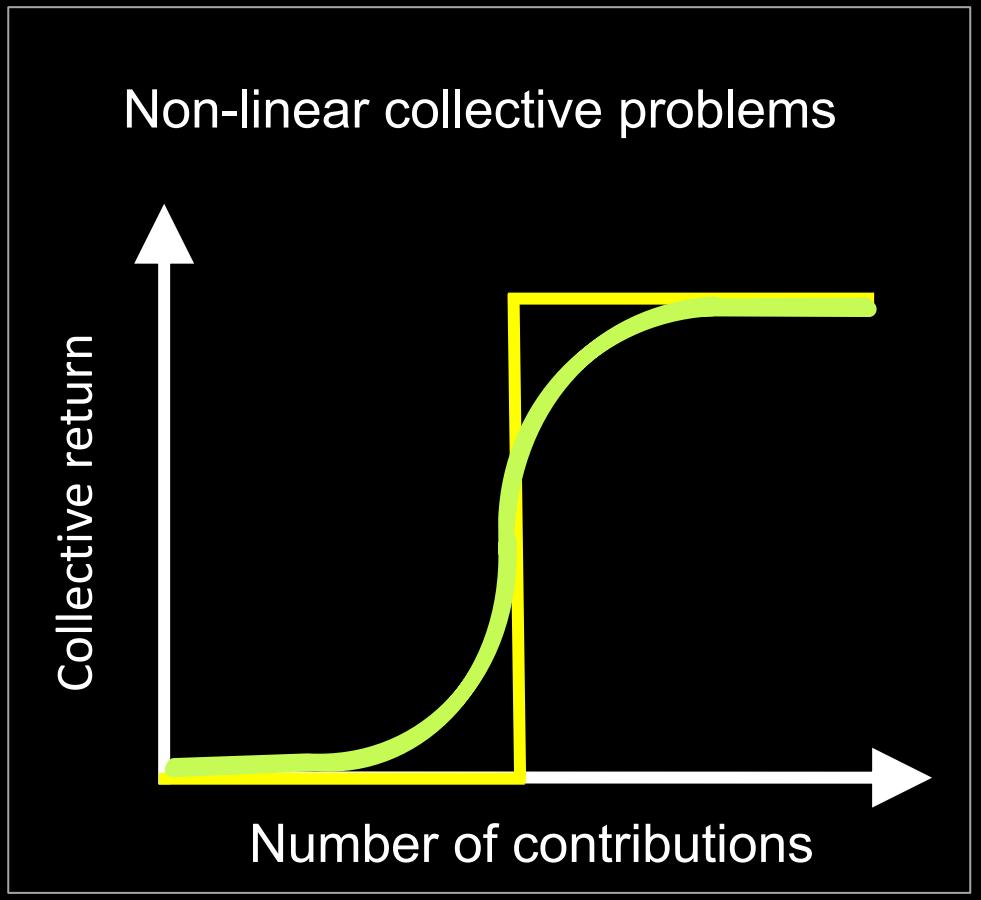
perceived risk of disaster ↗ cooperation ↗

➡ i) What's the nature of this dilemma?

ii) Can we model such behavioral dynamics?

Non-linear public goods



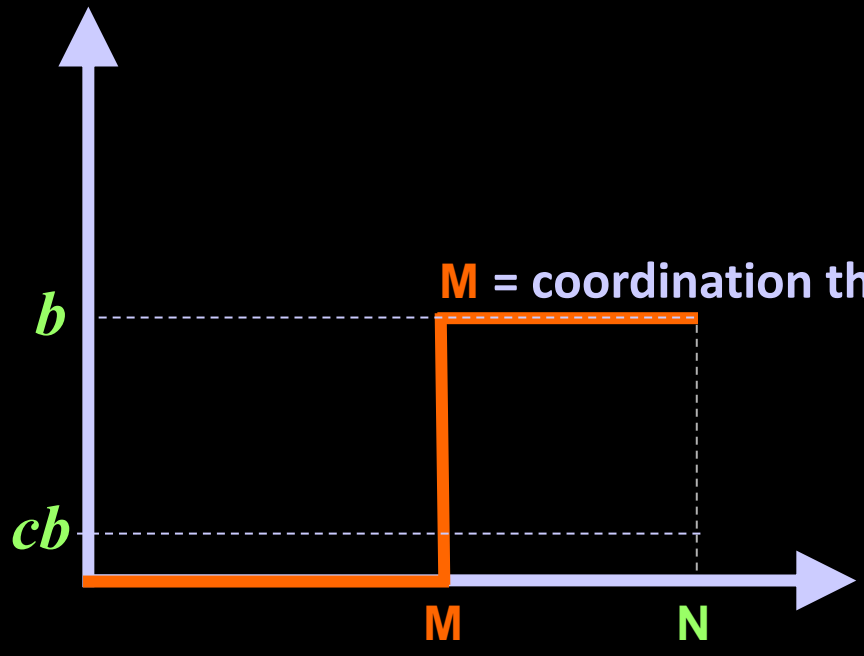


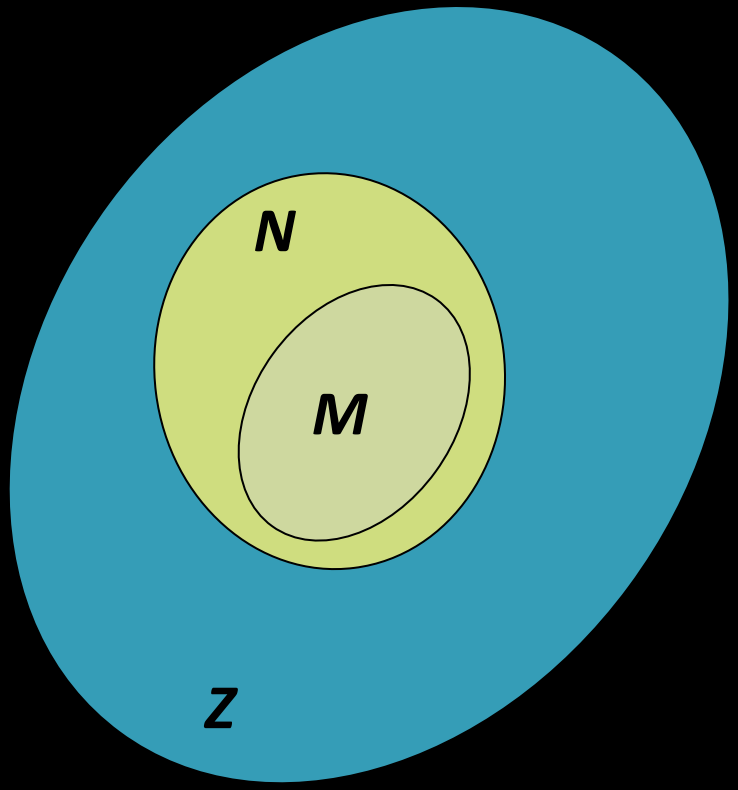
Souza, Pacheco, Santos (2009). Journal of Theoretical Biology,
Pacheco, Santos, Souza & Skyrms (2009) Proc. Royal Society B: Biol. Sci.
Santos, Pacheco (2011), Proc. Natl. Acad. Sci USA

N-person Coordination game . . . *with a risky twist*

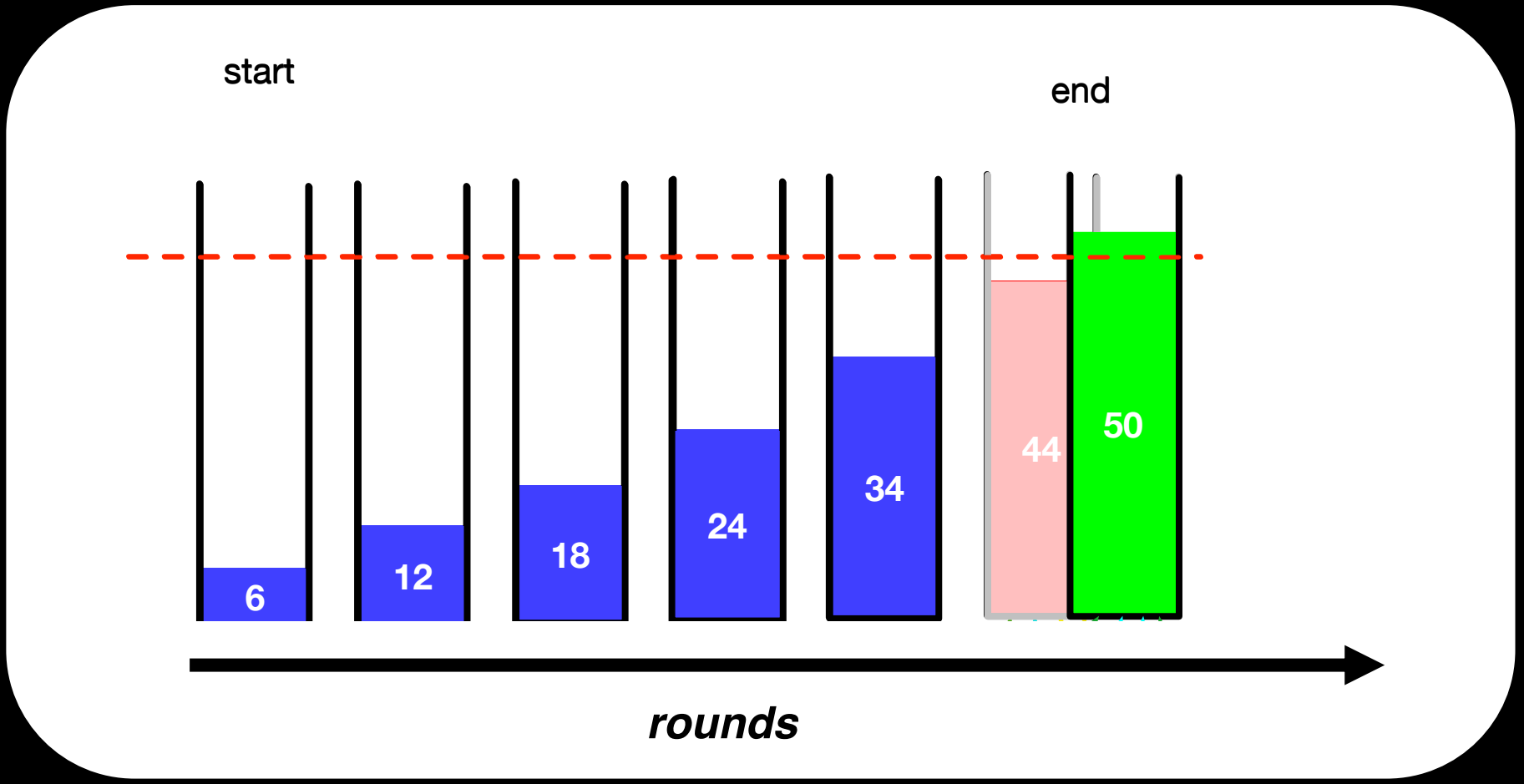
Cooperators **contribute a fraction** ***cb*** **(cost) of their initial endowment** ***b to the public good;***
Defectors **do not contribute.**

If Σ*cb* < *M* all loose everything with probability *r*
otherwise : everyone keeps all they have





N-person Coordination game . . . *with a risky twist*



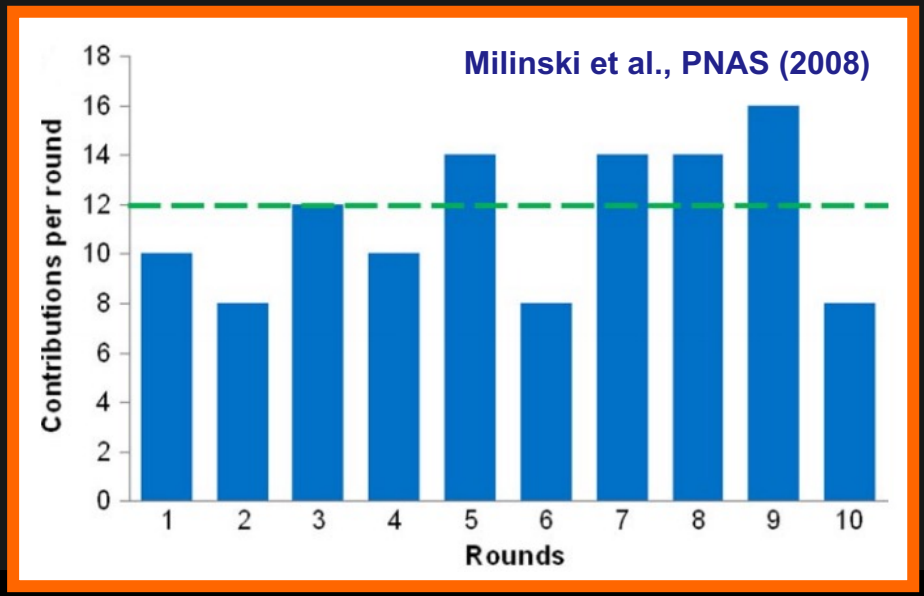
message from the game experiment on climate change

perceived risk of disaster ↗ cooperation ↗

i) What's the nature of this dilemma?

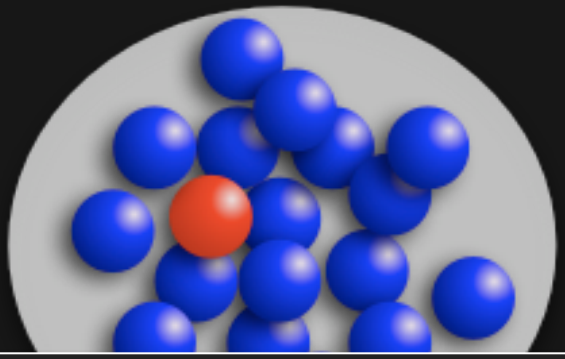
➡ ii) Can we model such behavioral dynamics?

rationality of players may not be an argument as individuals revise their strategy along the way

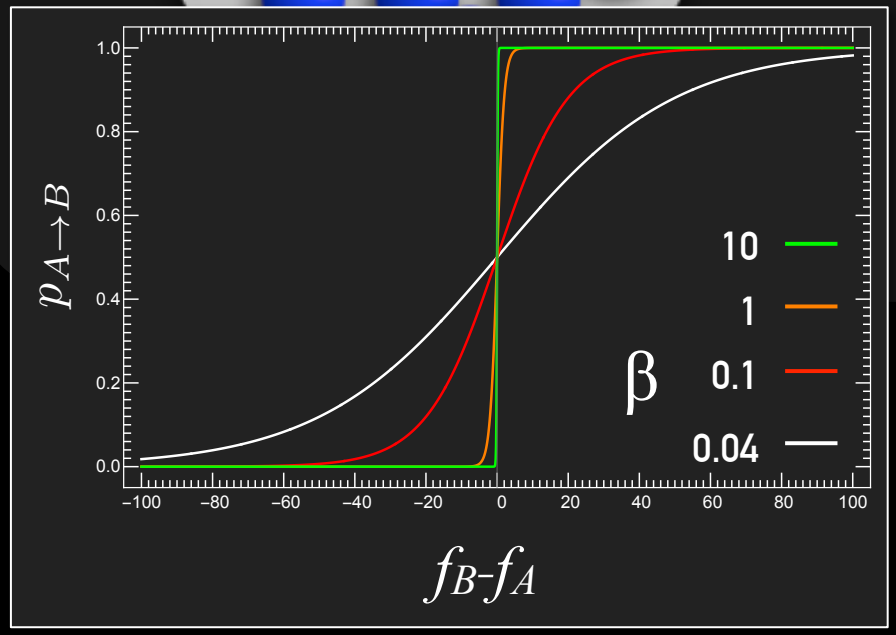


How do agents revise their strategies?

Imagine the simplest form of social learning:



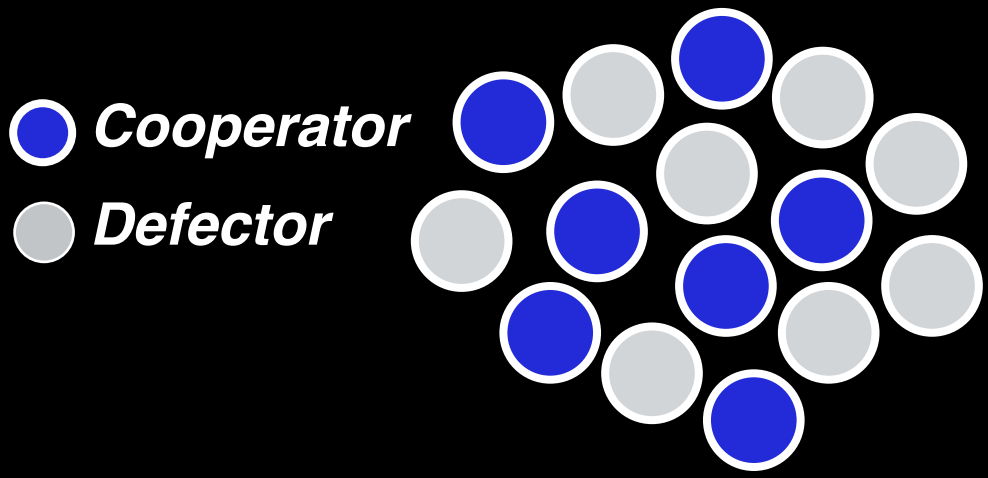
A imitates a random individual B with a probability that increases with the fitness difference.

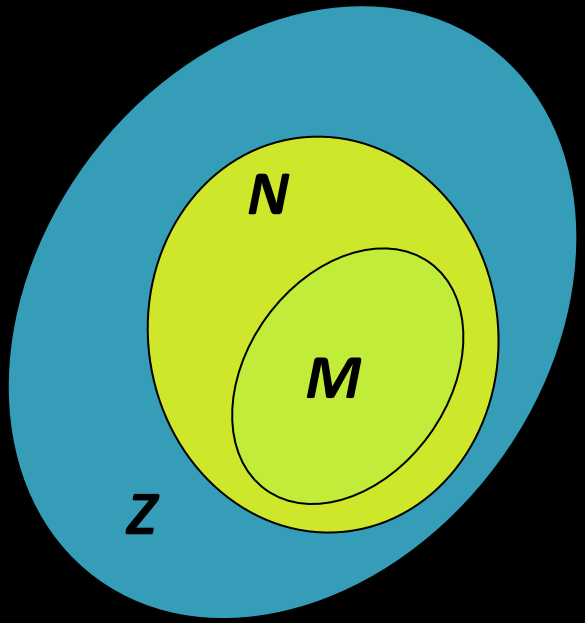


$$p = \left[1 + e^{-\beta(f_B - f_A)}\right]^{-1}$$

Additionally, with a “exploration” probability μ, agents may try a randomly chosen strategy

Method 1: Large-scale computer simulations

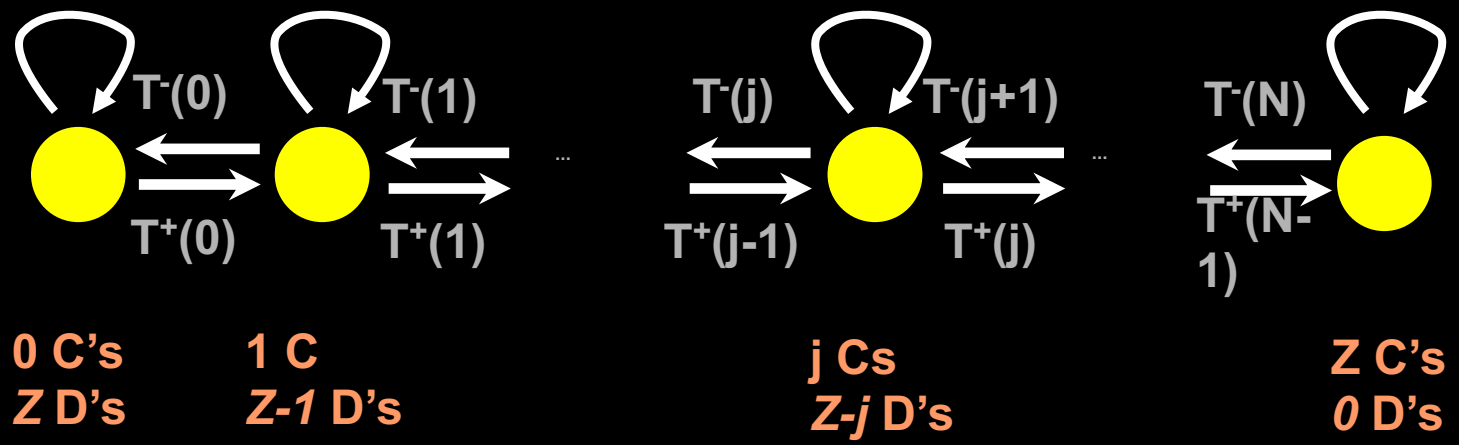


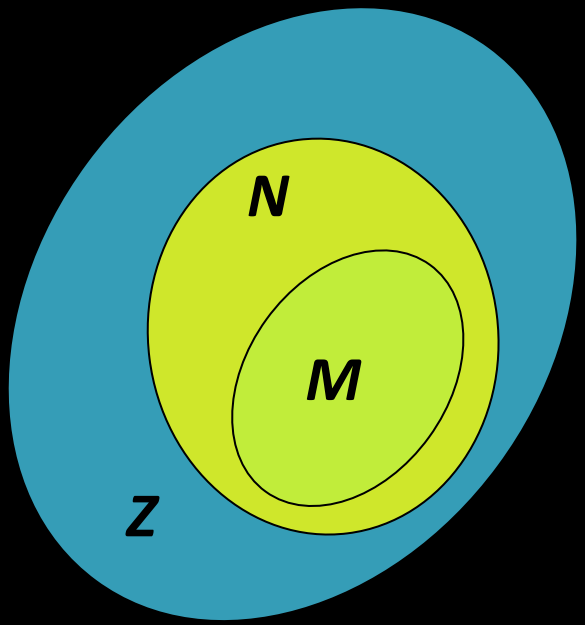


1) the *fitness* of each agent is given by the payoff received from randomly assigned groups of size N
2) Compute the *average* fraction of groups that surpass the threshold over many realizations

Method 2: A population-based Markov decision process

Z+1 possible states
(Z=population size, *j*=number of cooperators)



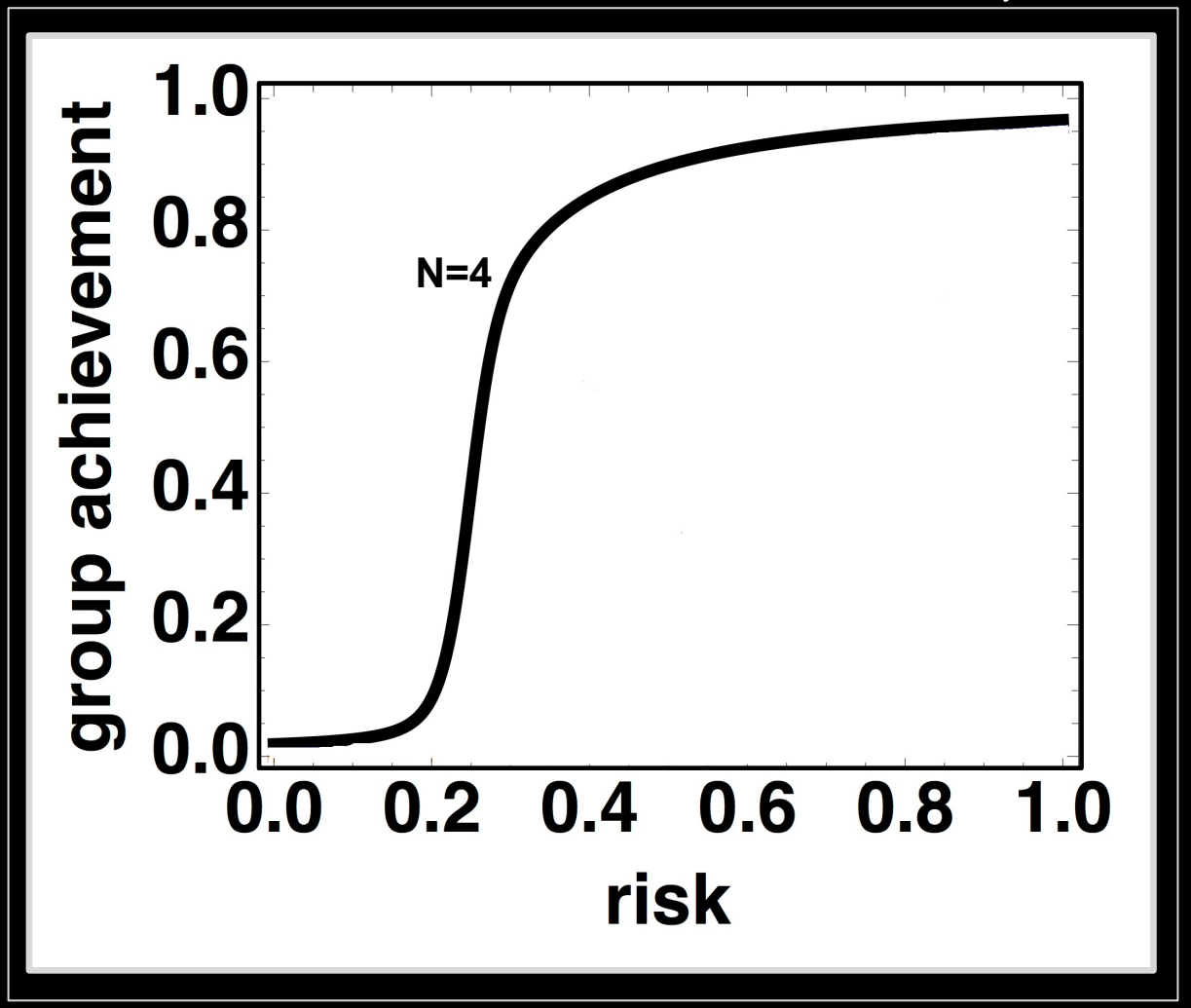


1) the *fitness* of each agent is given by the payoff received from randomly assigned groups of size N
2) Compute the *average* fraction of groups that surpass the threshold over many realizations

Average group achievement

Risk dependence | ex: Group Size (N) = 4

Analytical results



increasing the perception of risk opens a window of opportunity for cooperation

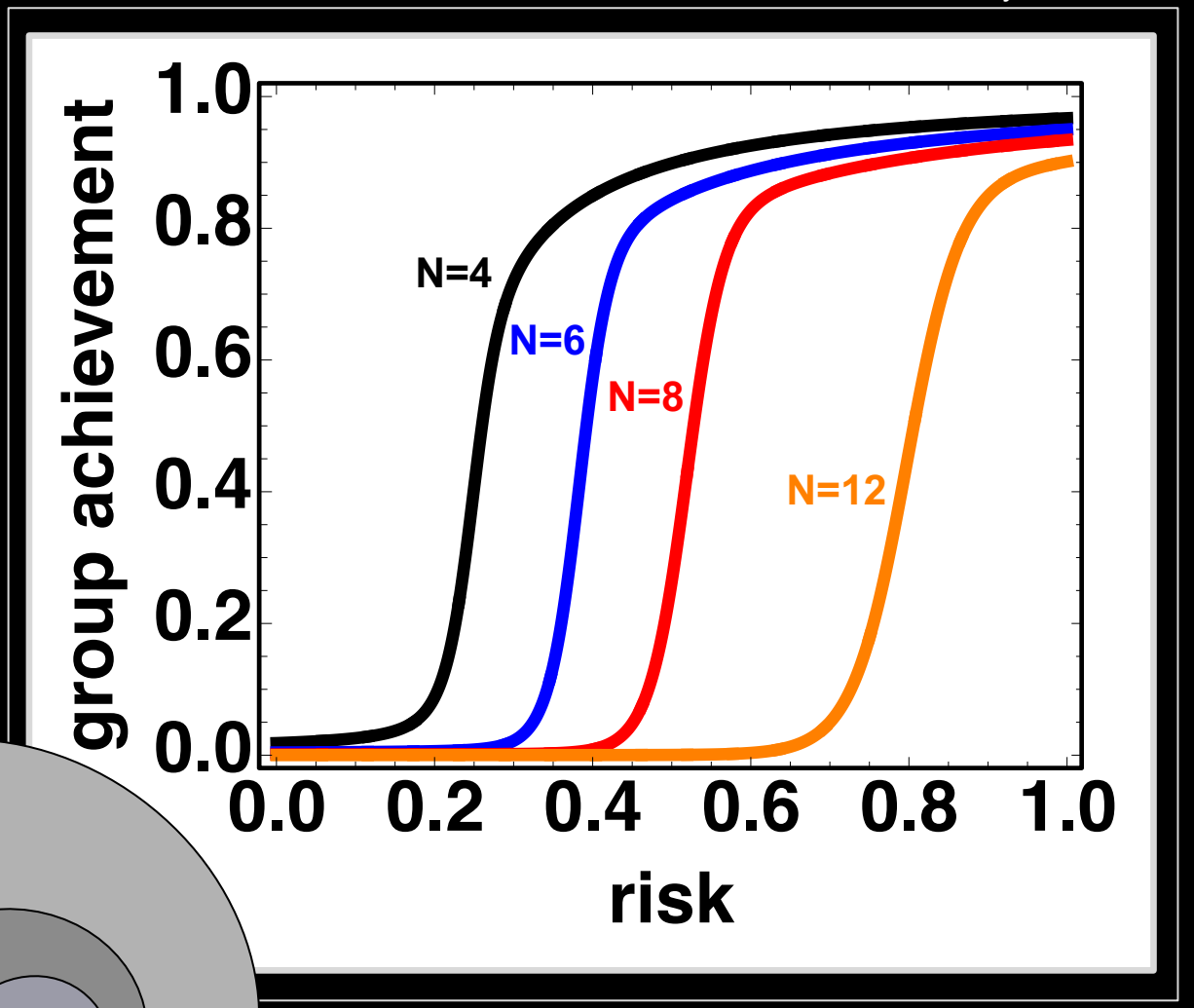
Santos, Pacheco, PNAS 2011
Vasconcelos, Santos, Pacheco, Levin PNAS 2014
Domingos, Santos, Lenaerts, 2021

$$Z = 100, \quad M = N/2$$

What's the best scale to reach an agreement?

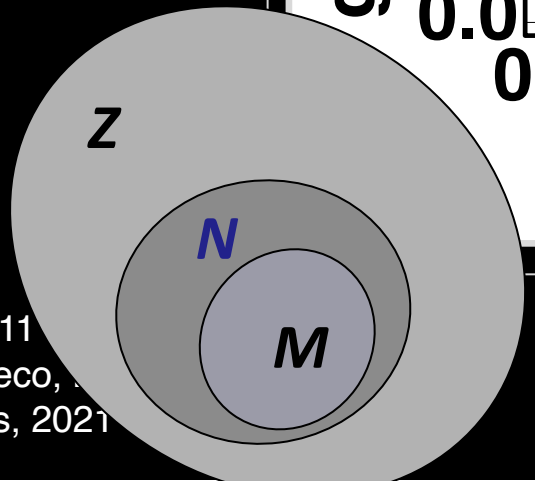
Group size dependence

Analytical results



As *the group size N* approaches *the population size Z* cooperation becomes less likely

Group size can change the nature of the dilemma



Santos, Pacheco, PNAS 2011
Vasconcelos, Santos, Pacheco,
Domingos, Santos, Lenaerts, 2021

$$Z = 100, \quad M = N/2$$

Side note: multi-agent reinforcement learning

Group size dependence

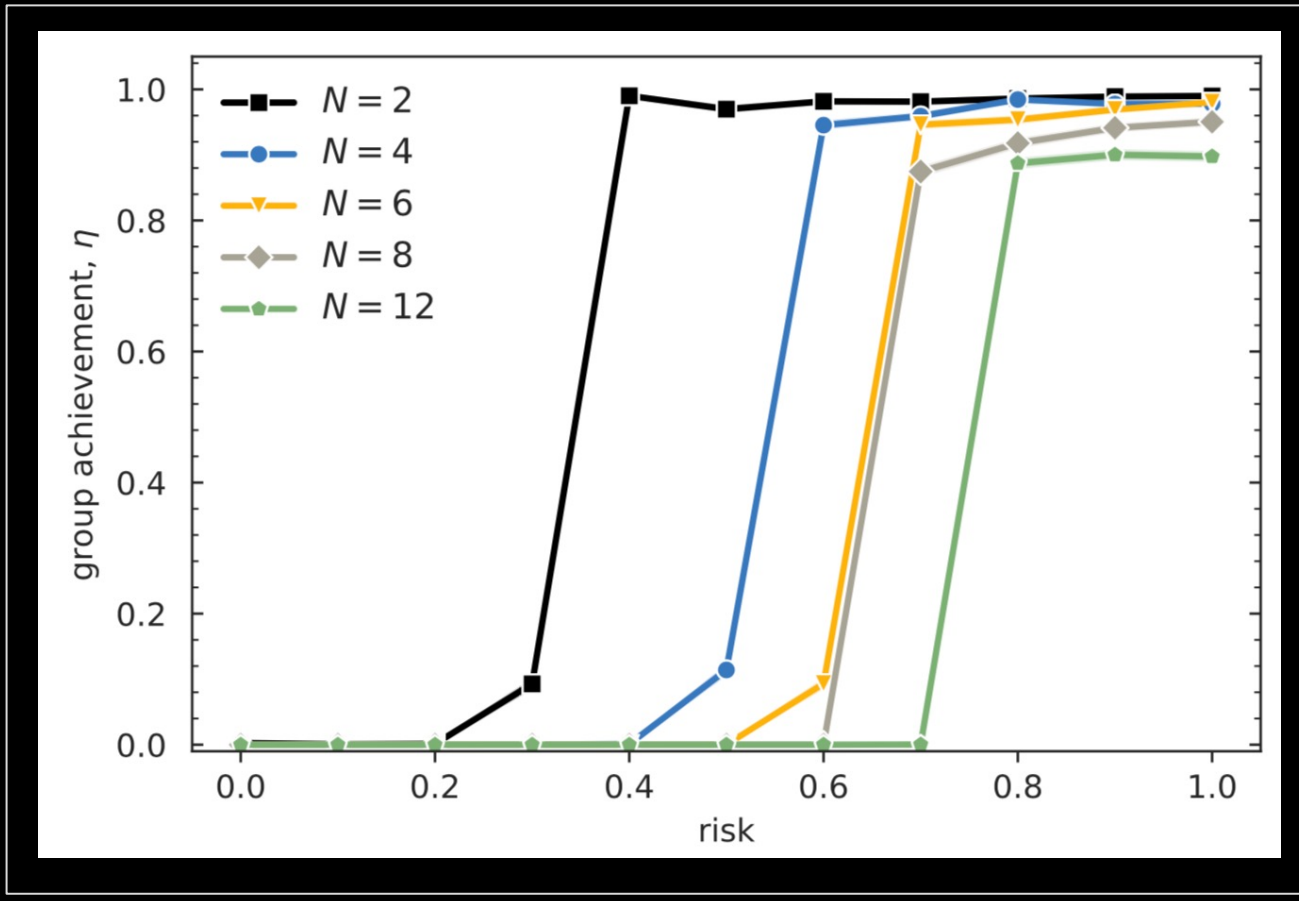
Q-learning

Similar results are obtained if we consider

Multi-agent reinforcement learning model with:

- Multiple rounds
- Decisions conditional to previous actions

Elias F. Domingos



Domingos, Grujić, Burguillo, Santos, Lenaerts. Modeling behavioral experiments on uncertainty and cooperation with population-based reinforcement learning, Simulation Modelling Practice and Theory (2021)

Population size = 50, number of rounds = 10, c/b= 0.5, α= 0.09,τ= 0.1

Maybe there's a better way…

one global agreement

a top-down answer,
to a global problem

Multiple, small-scale agreements

cooperation emerges from
decentralized initiatives

See also: Vasconcelos, Hannam, Levin, & Pacheco (2020), Sci. Rep.

Behavioral dynamics under polycentric governance

Large groups or low perception of risk

stable

unstable

Santos, Pacheco, PNAS 2011

Behavioral dynamics under polycentric governance

Multiple small-scale agreements & high perception of risk

stable

unstable

Santos, Pacheco, PNAS 2011

A multi-agent perspective on a well-known idea

Evolution and a self-organization analysis suggests that
a bottom-up or polycentric approach leads to better cooperative standards

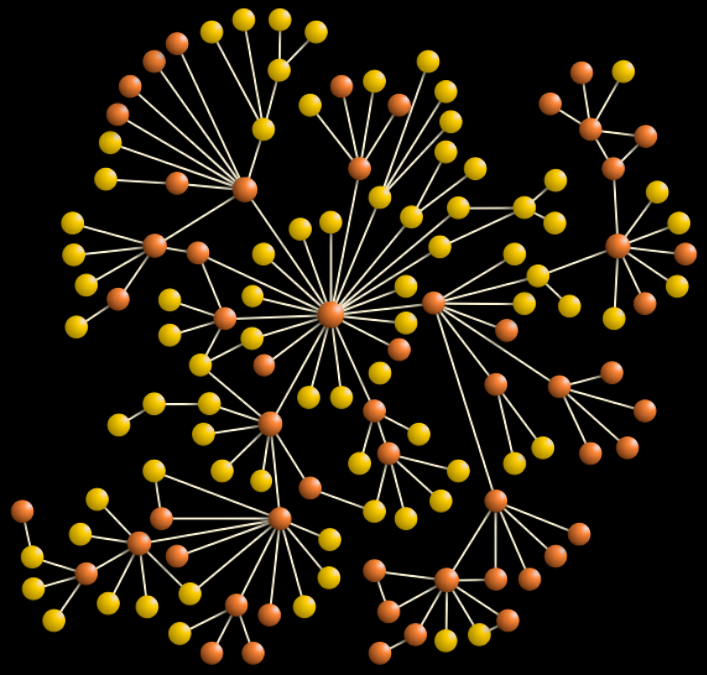
"Think globally, act locally"

"Simply waiting for resolution of these issues at a global level, without trying out policies at multiple scales because they lack a global scale, is not a reasonable stance."

Elinor Ostrom, *A Polycentric Approach for Coping with Climate Change*, 2009

how to optimize polycentrism in practice ?

how should we assign players to particular groups?



Networks of overlapping agreements

how to define these networks ?
some ideas . . .

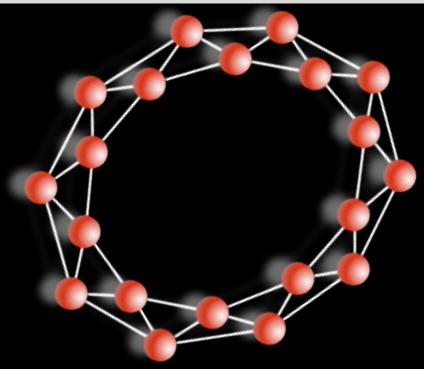
maybe networks could be defined based on groups of countries bound by common interests

- ***Regional interests on Alternative forms of energy***
- ***similar means of managing CO_2 emmisions***
- ***joint interest in local commons***
- ***etc . . .***

or groups of regions bound by common interests
(ex: California in USA, Catalonia in Spain, Bavaria in Germany, etc.)

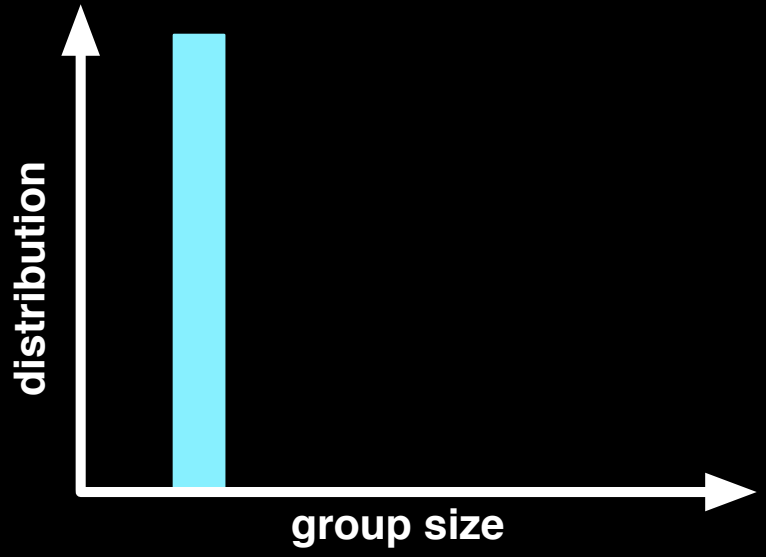
Introducing diversity

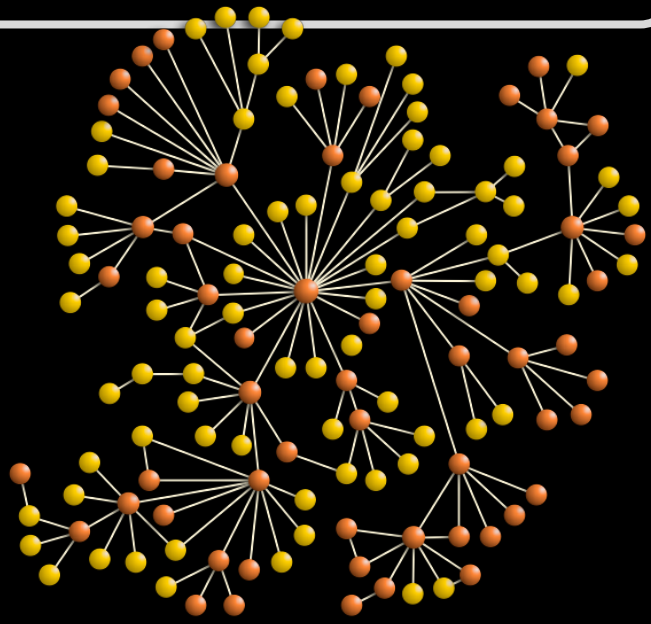
graph heterogeneity leads to diversity in the number and size of the collective dilemmas each agent participates



Homogeneous

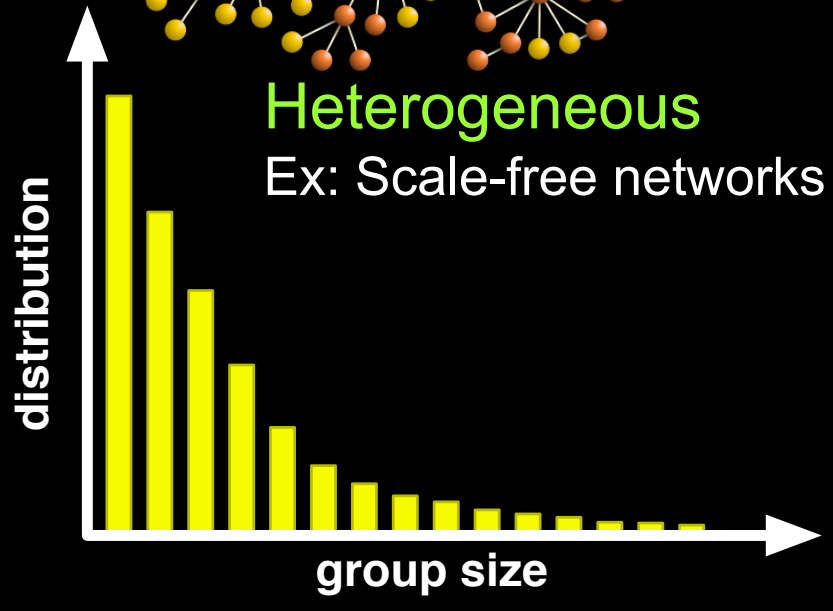
Ex: regular networks, well-mixed populations





Heterogeneous

Ex: Scale-free networks



Social-political networks and the role of heterogeneity

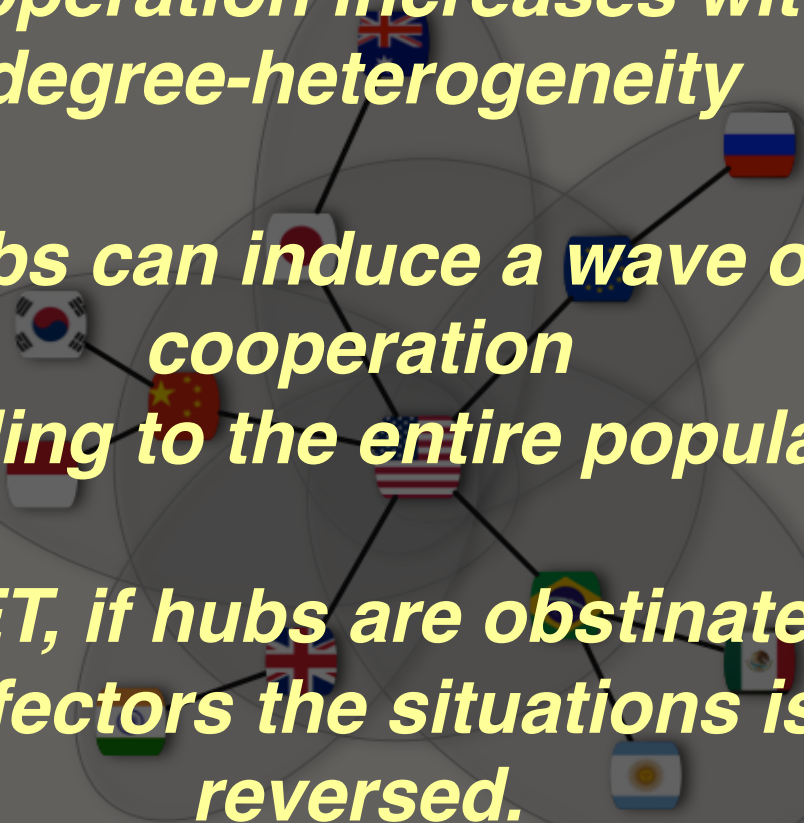
Collective-risk dilemmas in structured populations

each neighborhood defines a game/group.

Cooperation increases with degree-heterogeneity

Hubs can induce a wave of cooperation spreading to the entire population

YET, if hubs are obstinate defectors the situations is reversed.

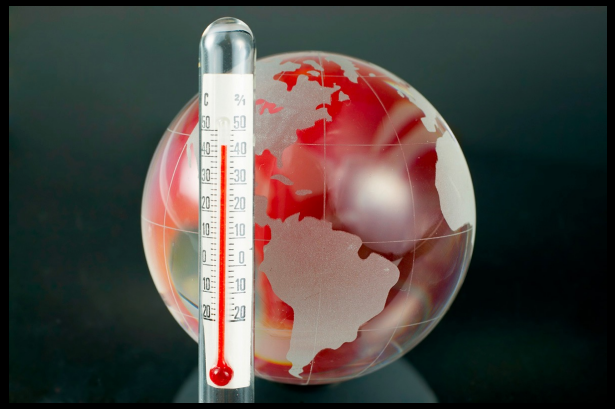


[FC Santos, MD Santos, JM Pacheco, Nature 454, 213-216 (2008)]

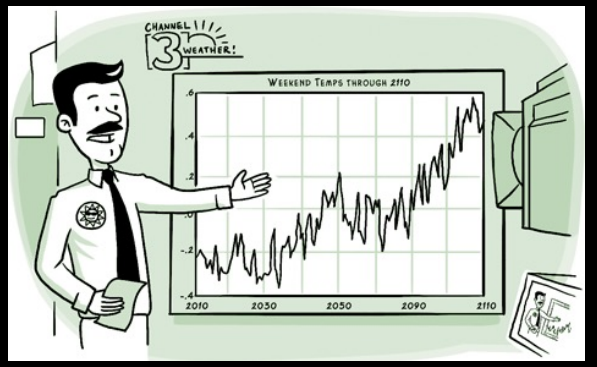
Can we improve cooperation even further?

the challenge of minimizing the effects of climate change

failure of global summits has been attributed (among many other issues) to:



low perception of risk



uncertainty (goals & timing)

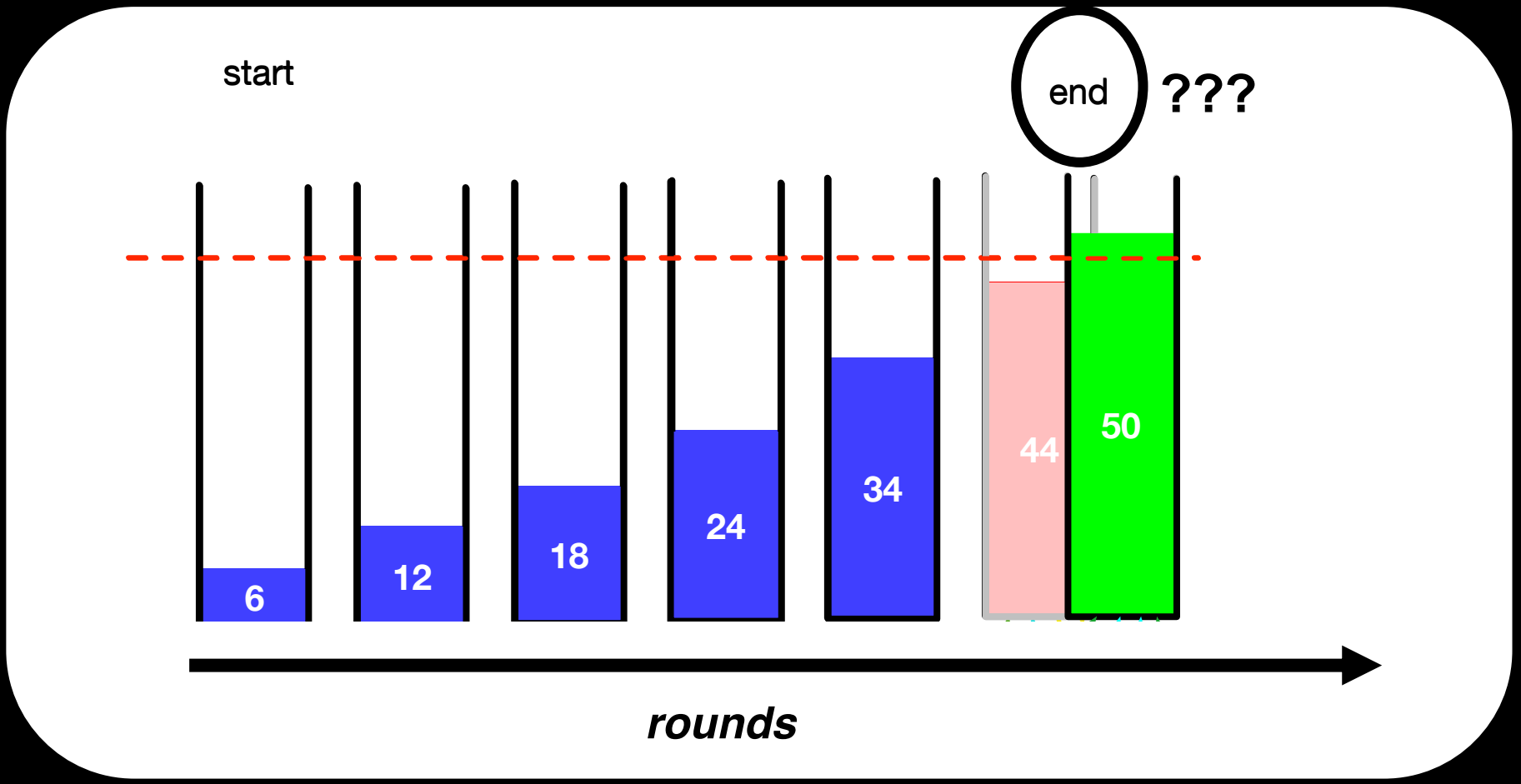
conflicting policies between rich and poor parties

Domingos-Fernandez et al., iScience 2020
Vasconcelos et al, Phys Life Reviews 2015

no institution to monitor, reward or sanction

Decision-making under timing uncertainty

We are unsure when it will be too late, or when a disaster might happen…



Decision-making under timing uncertainty

A total of 246 participants.

Experiment 1 (No uncertainty)

- ***F***
- ***D***

Expe

- ***M***
- ***Pr***

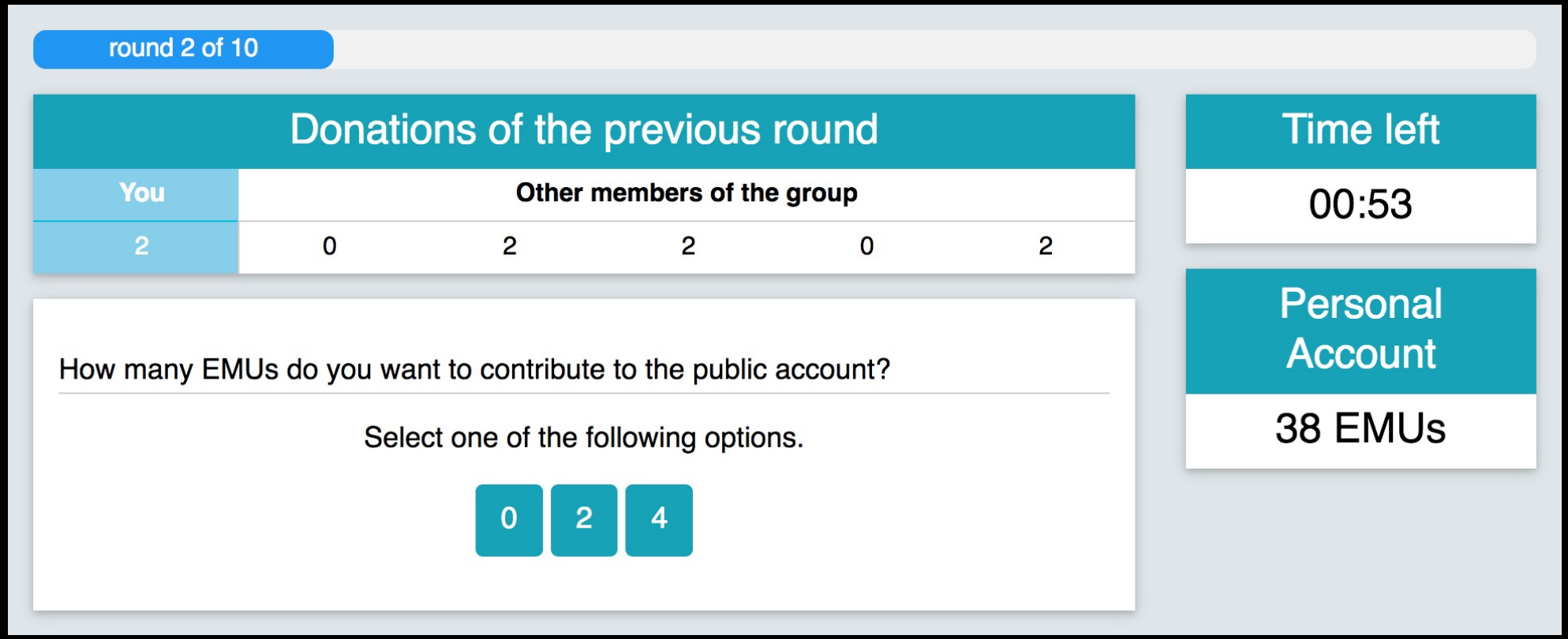
Expe

- ***M***
- ***Pr***

Elias F. Domingos

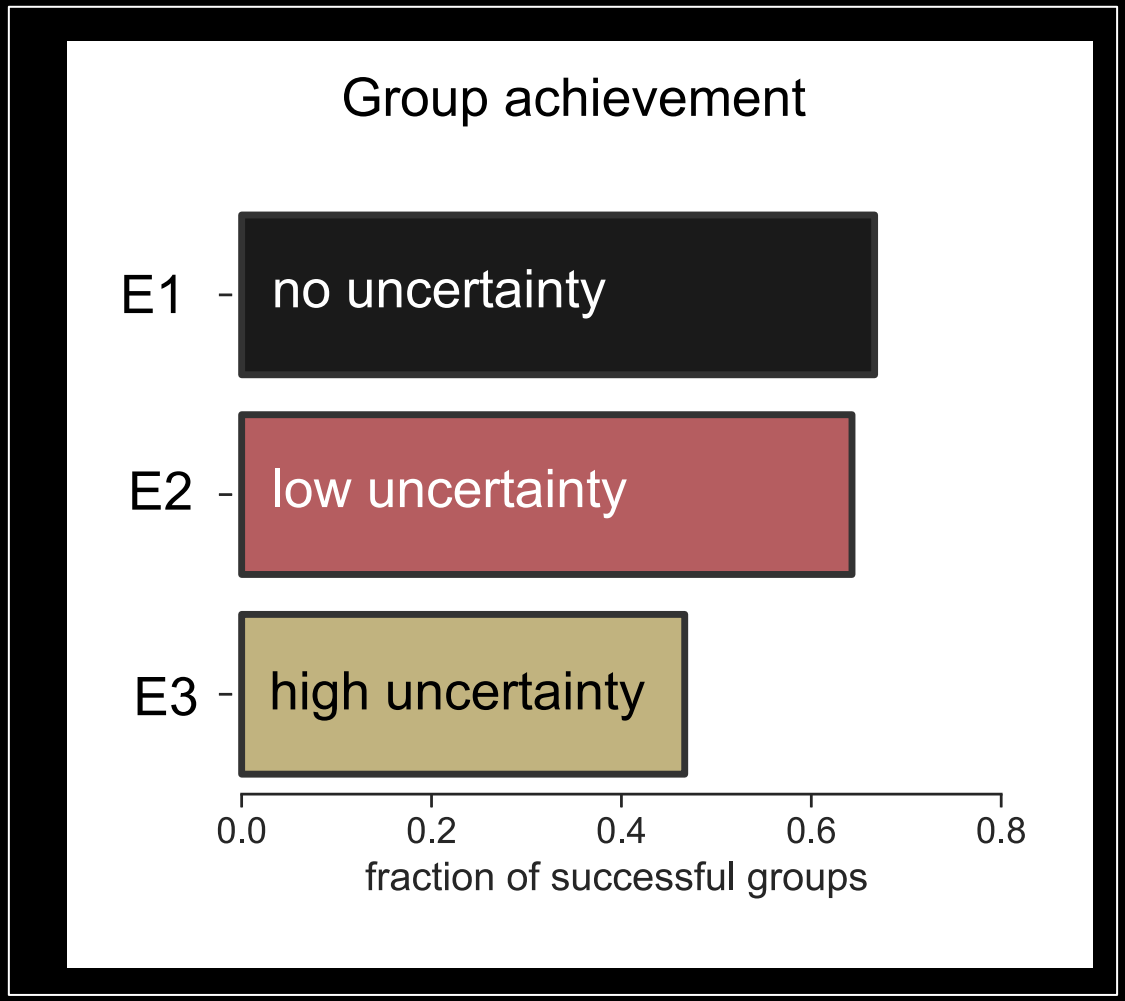
[In all cases: average number of rounds = 10]

Decision-making under timing uncertainty



Timing uncertainty: experimental results

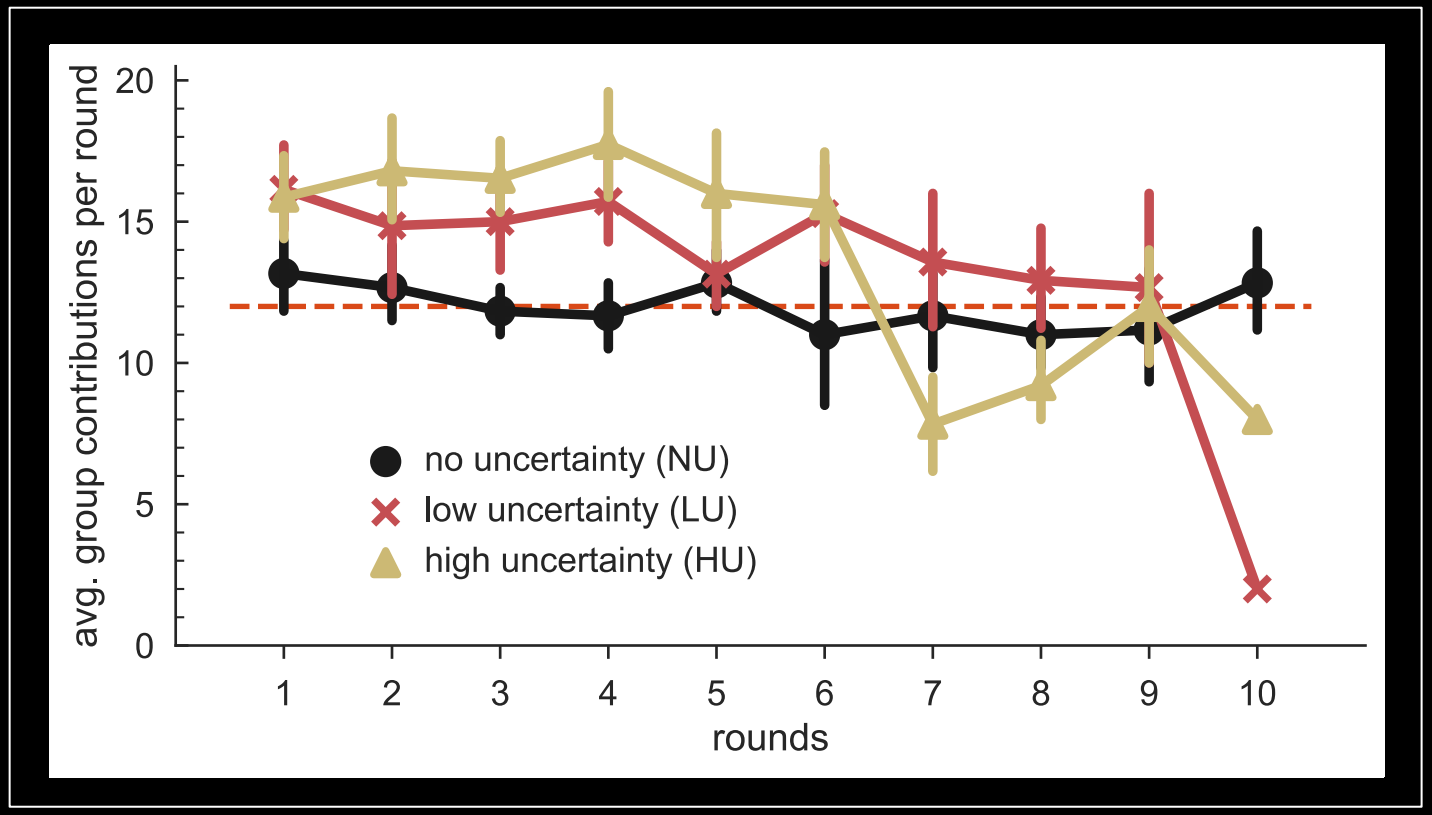
Timing uncertainty reduces collective action



E. F. Domingos et al. iScience (2020)

Timing uncertainty: experimental results

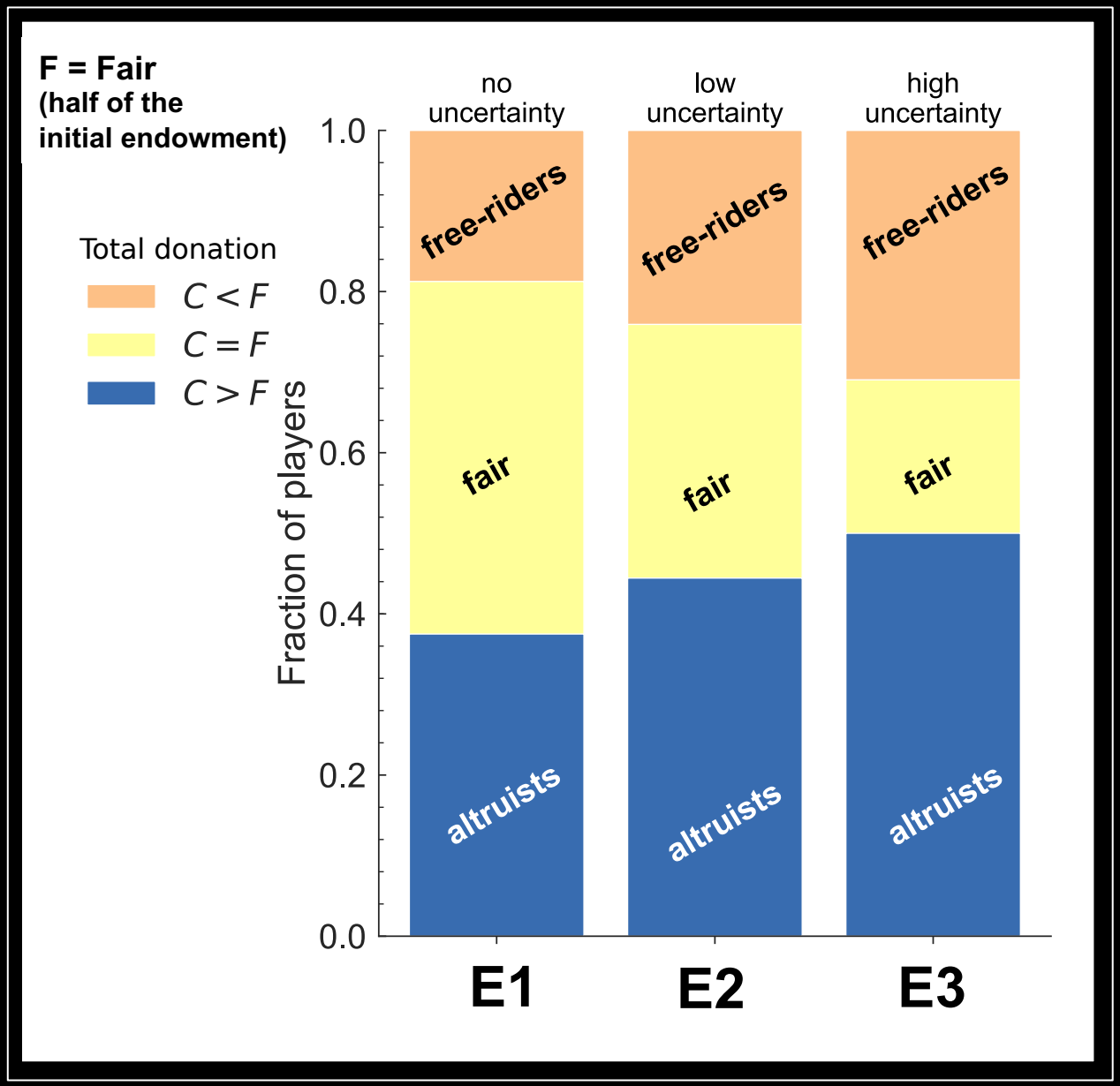
Players invest earlier under uncertainty…but not all!



E. F. Domingos et al. iScience (2020)

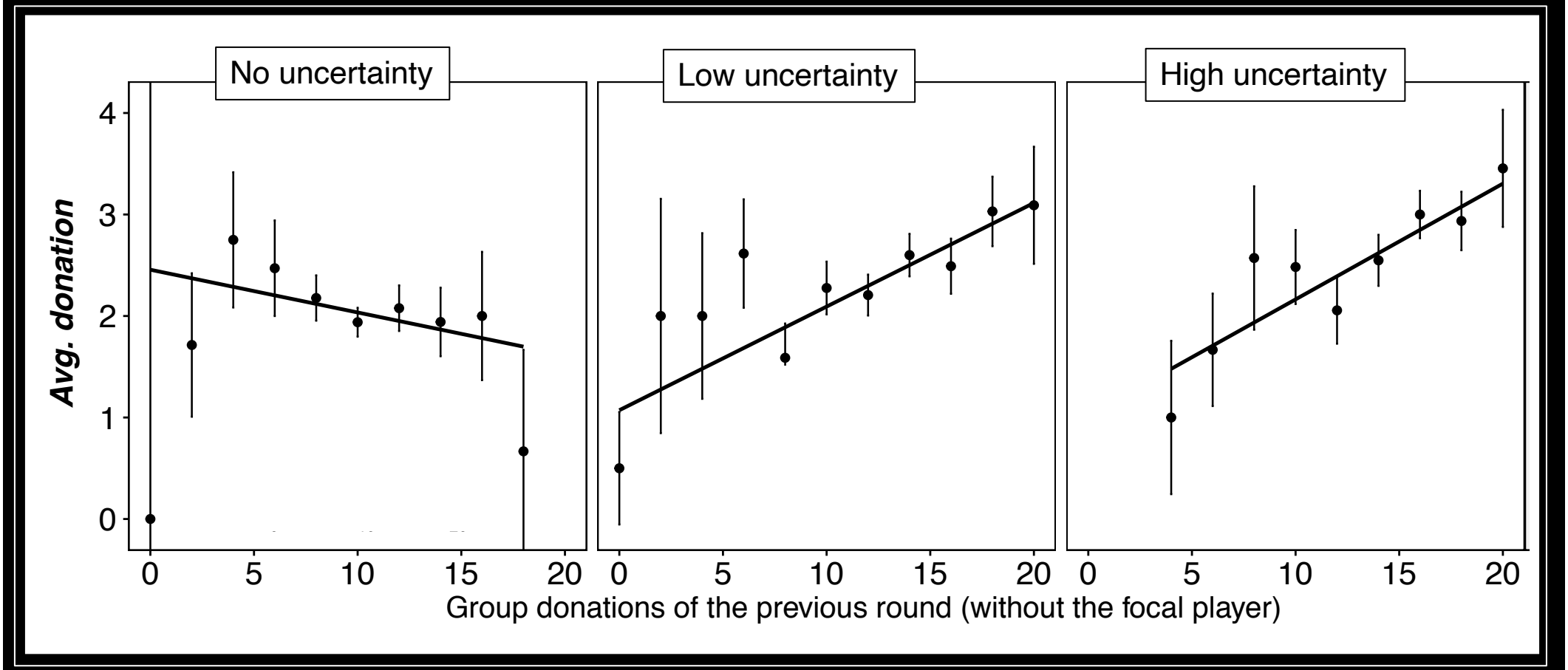
Timing uncertainty: experimental results

Uncertainty leads to polarization ☹



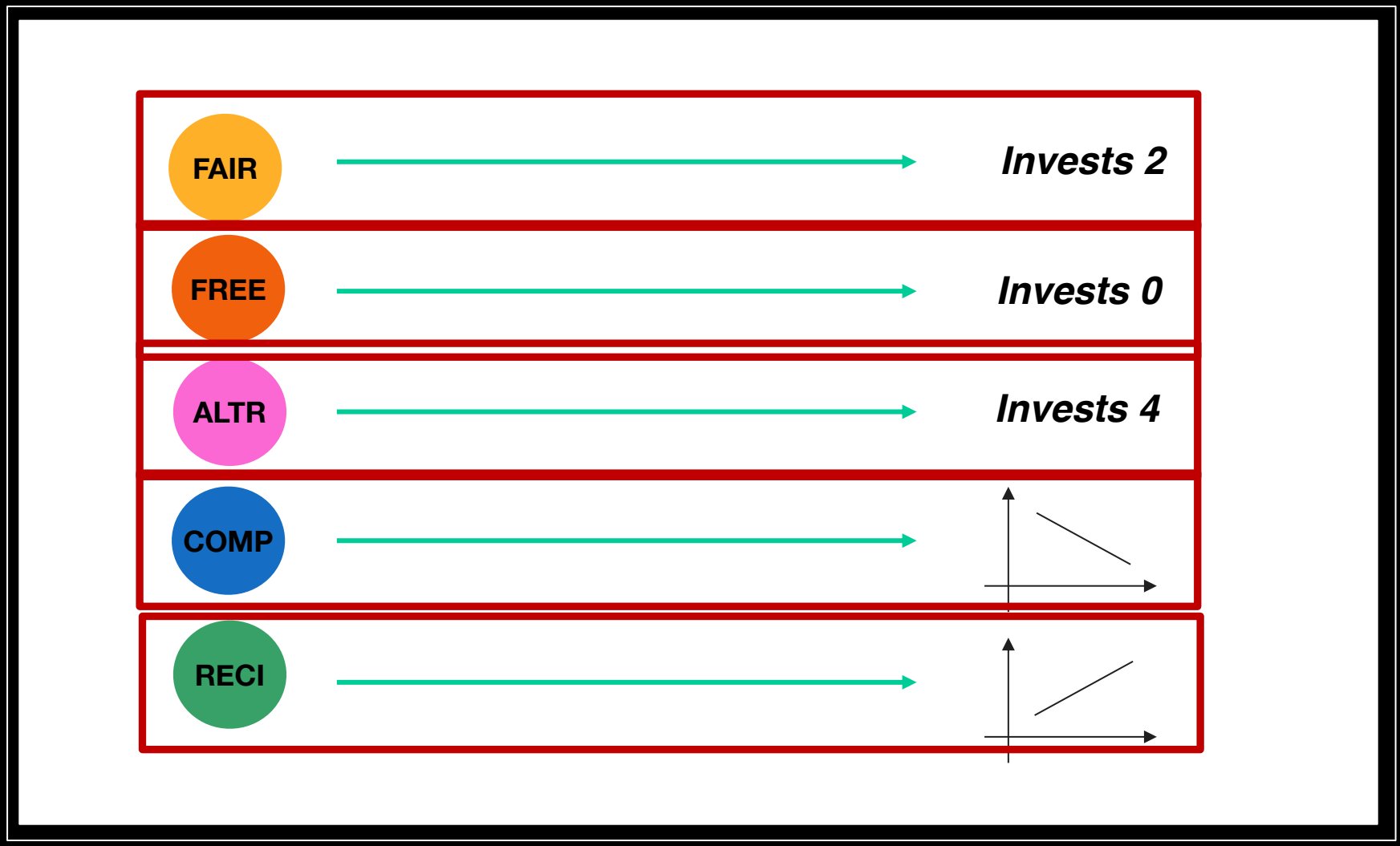
Timing uncertainty: experimental results

…and foster reciprocal strategies (as Tit-for-Tat)



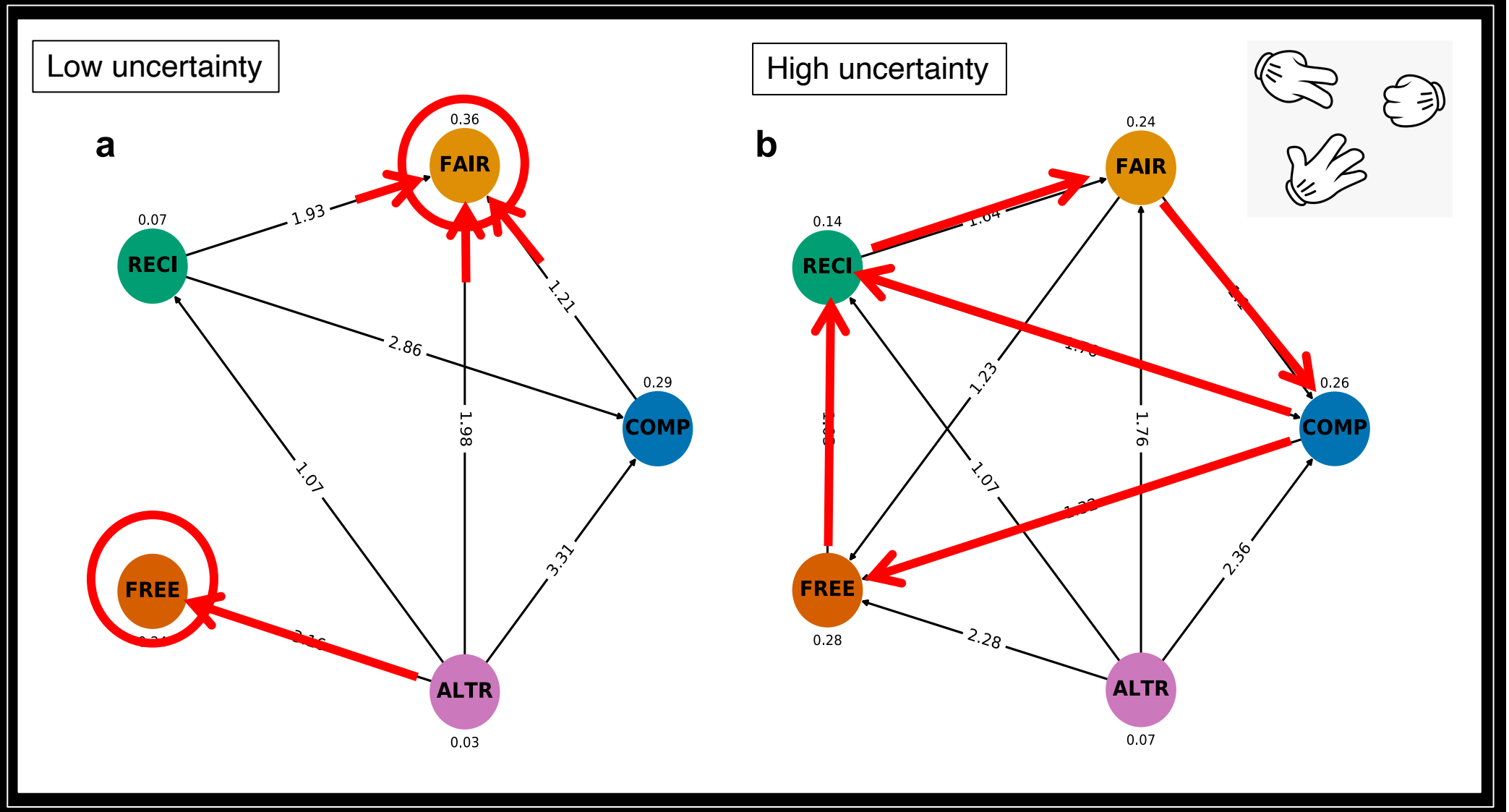
Timing uncertainty: Understanding the results

Clustering of strategies:



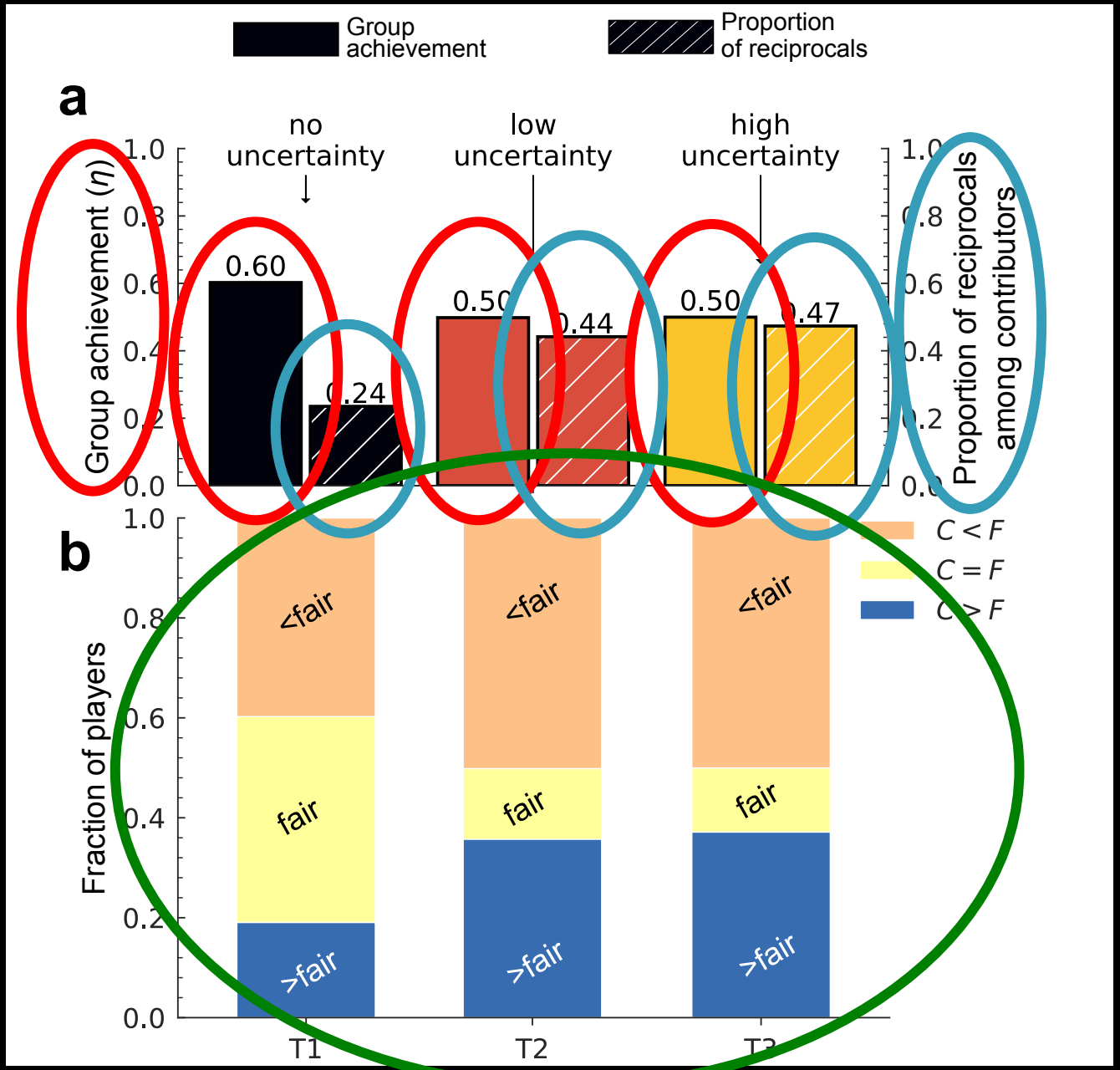
Uncertainty creates a new ecology of behaviors

Ex: 5 strategies



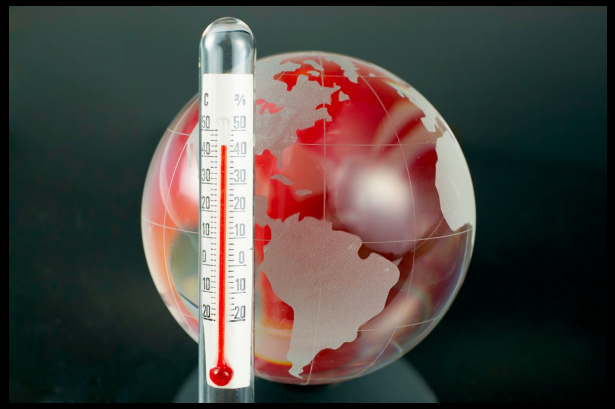
(Analytical results under the small-mutation approximation)

Timing uncertainty: theoretical results



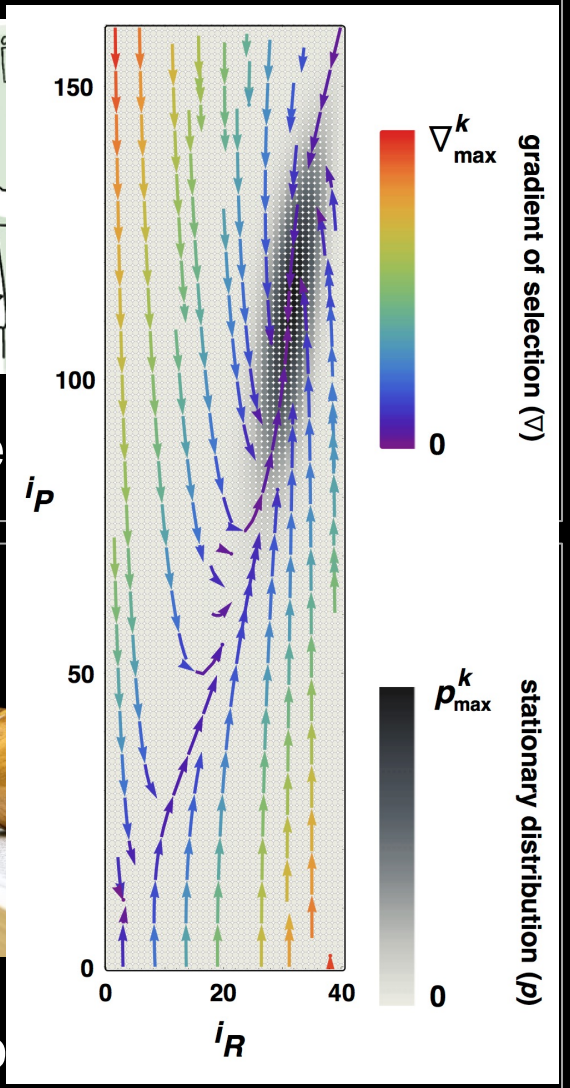
the challenge of minimizing the effects of climate change

failure of global summits has been attributed (among many other issues) to:



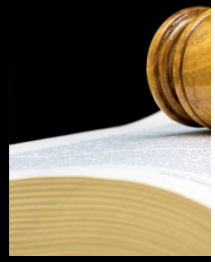
low perception of risk

unce



Vasconcelos, Santos, Pacheco, Levin, PNAS 2014

conflicting policies between rich and poor parties

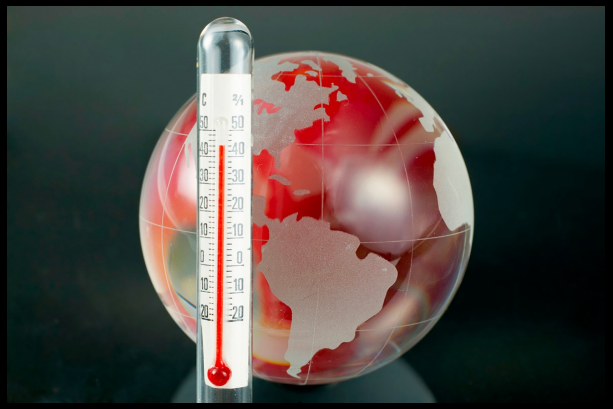


no

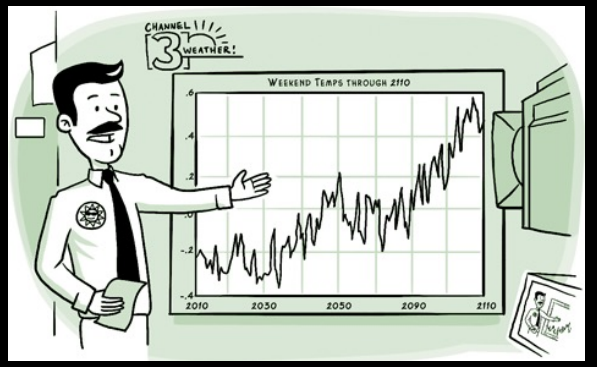
monitor and sanction

the challenge of minimizing the effects of climate change

failure of global summits has been attributed (among many other issues) to:



low perception of risk



uncertainty (goals & timing)

conflicting policies between rich and poor parties

no institution to monitor, reward or sanction

the challenge of minimizing the effects of climate change

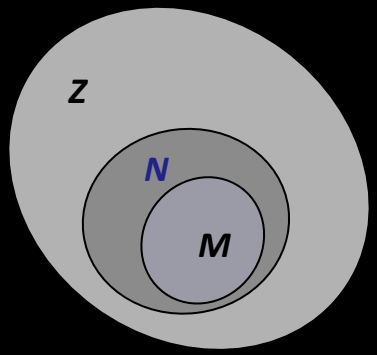
1. What's the the impact of policing?
2. Can these institutions be stable?
   Institutions may be costly, leading to a second-order free-riding problem
3. What should be the scale of such institutions?
   Should we have a single global institution or multiple (local) institutions?
4. What type of institution is more likely to emerge?

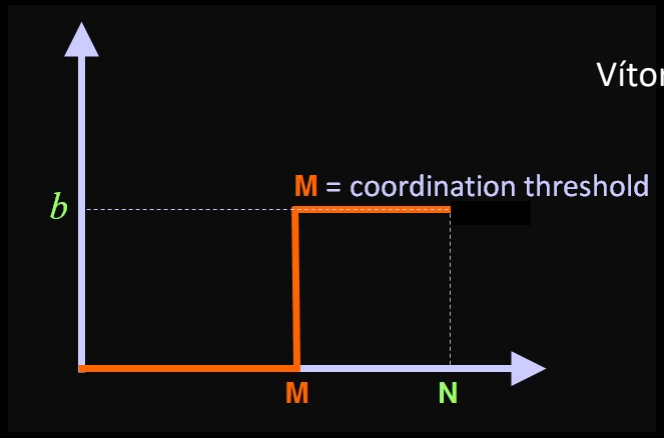
Couto, Pacheco, Santos, J Theor Biol 2020
A. Góis, FP Santos, JM Pacheco, FC Santos, Sci. Rep. 2019
Vasconcelos, Santos, Pacheco, Nature Climate Change 2013

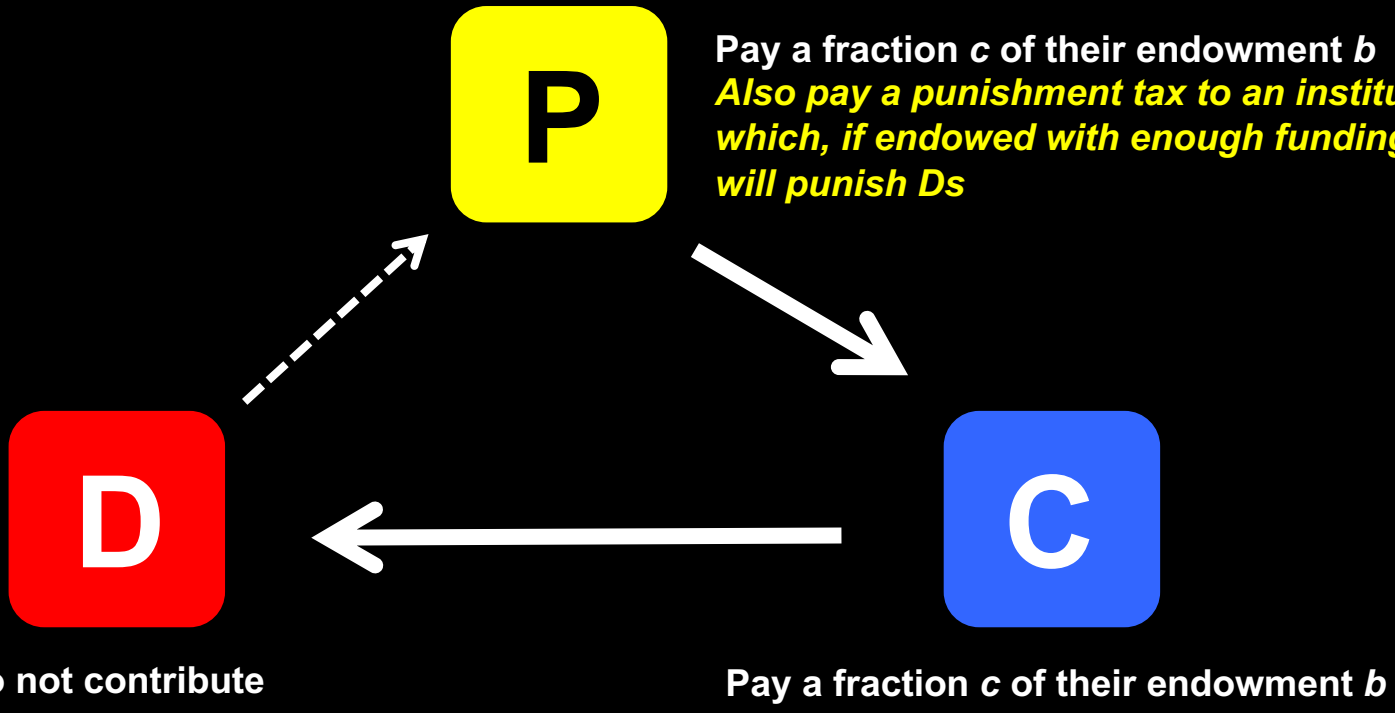
A second-order free-riding problem

Vítor V. Vasconcelos

3 strategies & 2 dilemmas:
- ***Cooperators (Cs)***,
- ***Defectors (Ds)*** **and**
- ***Punishers (Ps)***



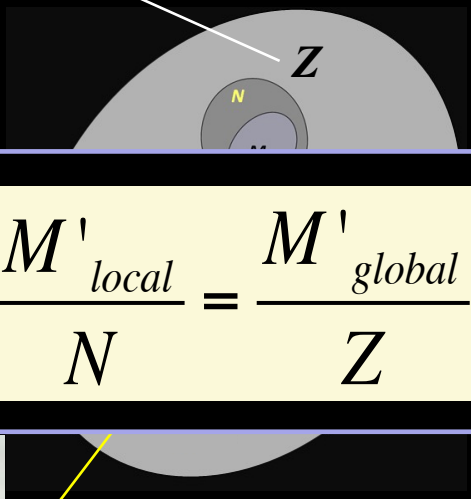




Vasconcelos, et al. Nature Climate Change 2013

global versus local institutions

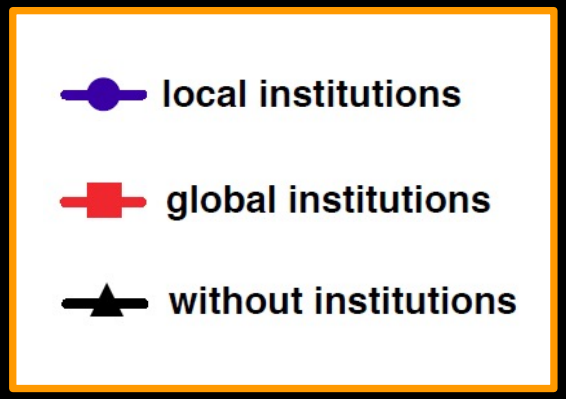
one *global* institution



$$\frac{M'_{local}}{N} = \frac{M'_{global}}{Z}$$

multiple *local* institutions

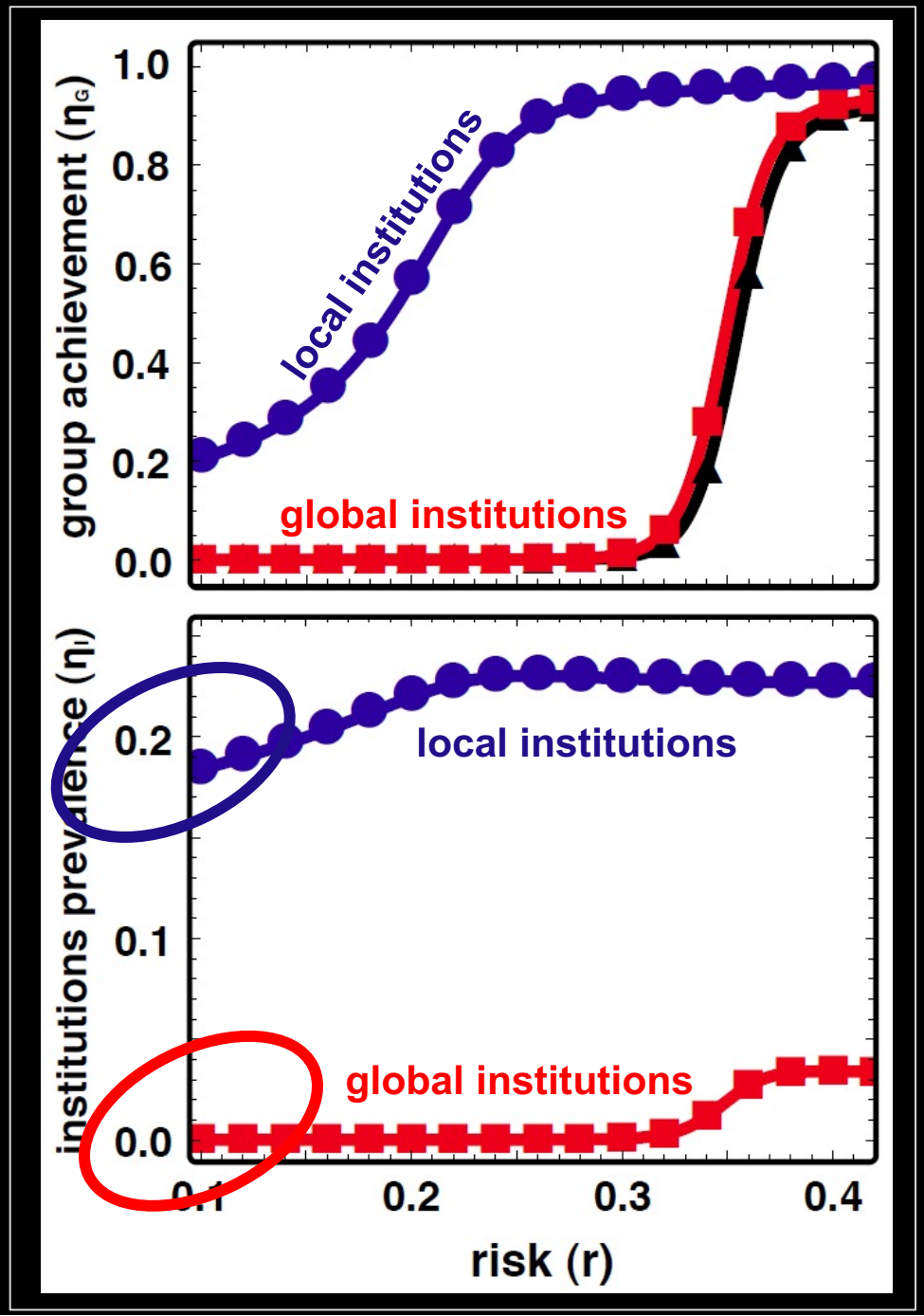
global versus local sanctioning institutions



local institutions play a crucial role when perception of risk is small;

global institutions (alone) provide marginal improvements in promoting cooperation

global institutions demand the coordinated efforts of many individuals, and cannot be maintained for long periods.



Z=100, M/N=0.75, $M'_{local}/N=M'_{global}/Z=0.25$, $\mu=1/Z$

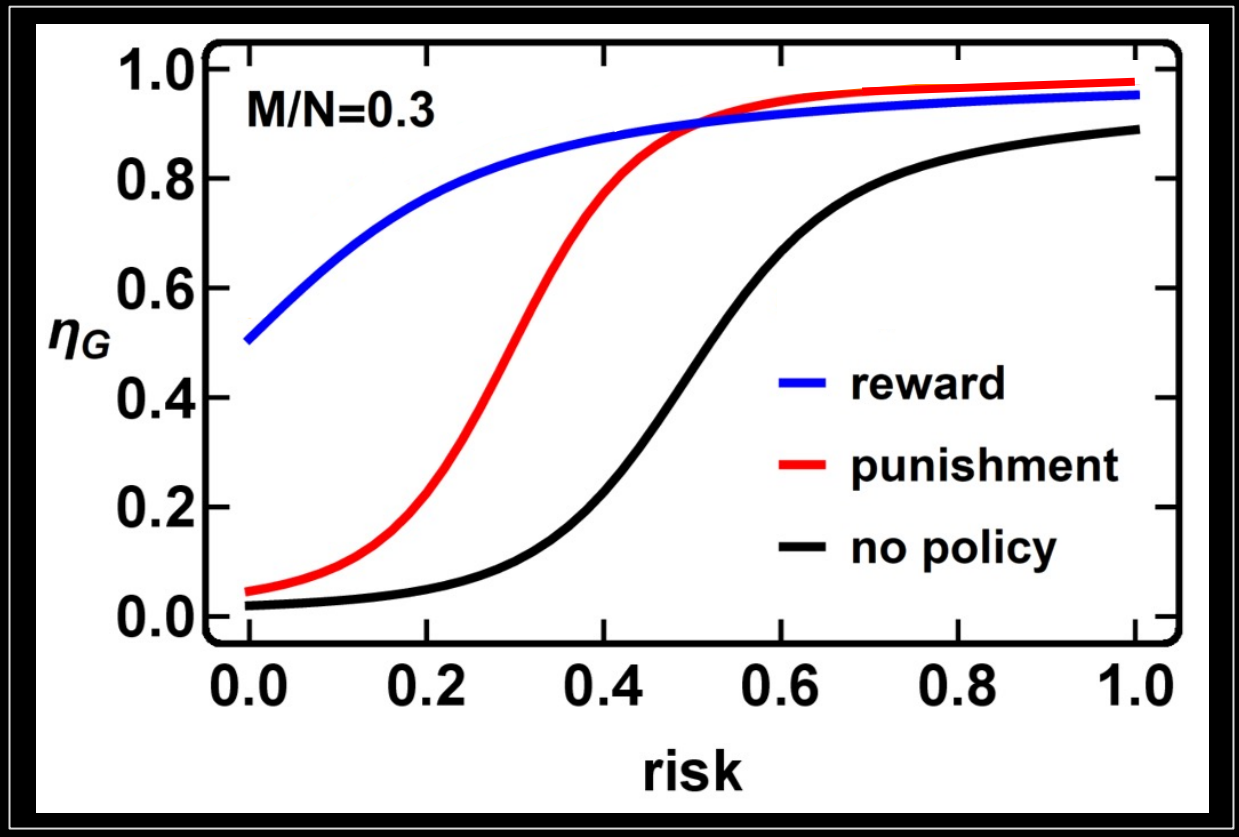
To a large extent, we fail at creating self-sustained or stable institutions ☹

How can we overcome the second-order free-riding problem?

Institutional rewards vs sanctions in collective-risk dilemmas

Institutional rewards are crucial to foster cooperation when the risk is low;
Sanctions are instrumental to maintain cooperation.

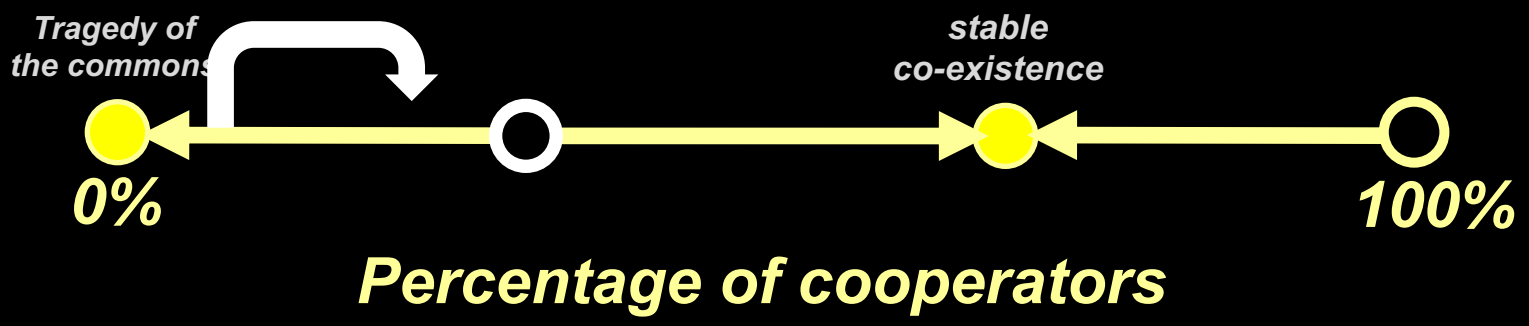
António Góis



A Góis, et al. Reward and Punishment in Climate Dilemmas Scientific Reports 2019
[& IST-MEIC MSc thesis, 2018]

Reward and punishment in climate dilemmas

***Institutional rewards** are crucial when cooperation is low;*
Sanctions are instrumental to maintain cooperation.



stable

unstable

A Góis, et al. Reward and Punishment in Climate Dilemmas Scientific Reports 2019
[& IST-MEIC MSc thesis, 2018]

Reward and punishment in climate dilemmas

Institutional rewards are crucial when cooperation is low;

Sanctions are instrumental to maintain cooperation.

stable

unstable

A Góis, et al. Reward and Punishment in Climate Dilemmas Scientific Reports 2019
[& IST-MEIC MSc thesis, 2018]

Ostrom's Design Principles

1. The resource has clearly defined boundaries
2. Use and provision of the resource is adapted to local conditions
3. Rules and decisions are made through collective-choice arrangements that allow most resource users to participate
4. Rules are enforced through effective monitoring by monitors accountable to the users
5. Violation of community rules are punished with graduated sanctions
6. Conflicts and issues are resolved through low cost and accessible conflict resolution mechanisms
7. The right of resource users to self-govern is recognized by higher level authorities
8. In the case of larger common-pool resources: rules are organized and enforced through multiple layers of nested enterprises

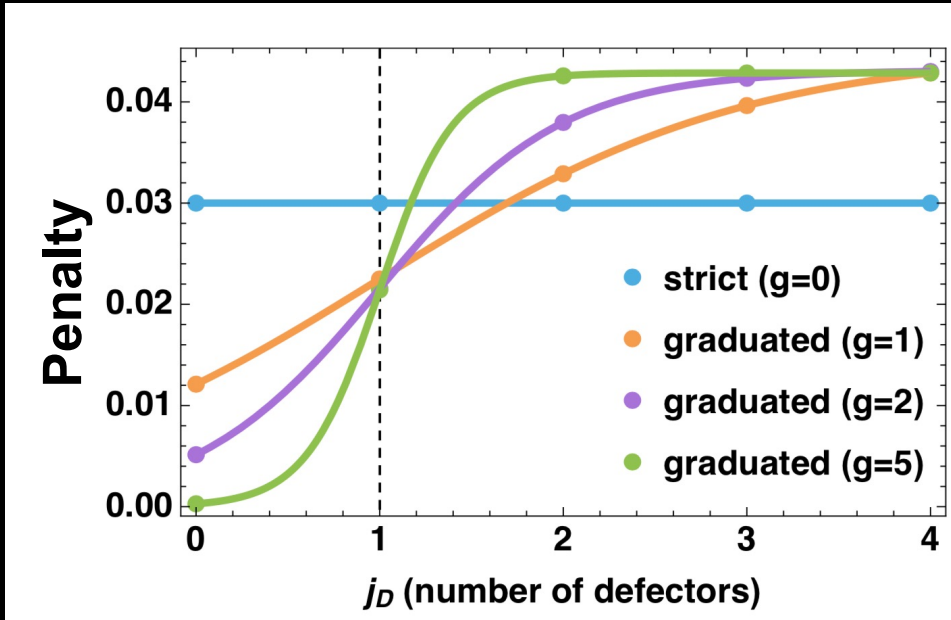
Let's get back to Ostrom's design principles

8. In the case of larger common-pool resources, rules are organized and enforced through multiple layers of nested enterprises

Ostrom's Design Principles

1. The resource has clearly defined boundaries
2. Use and provision of the resource is adapted to local conditions
3. Rules and decisions are made through collective-choice arrangements that allow most resource users to participate
4. Rules are enforced through effective monitoring by monitors accountable to the users
5. Violation of community rules are punished with graduated sanctions
6. Conflicts and issues are resolved through low cost and accessible conflict resolution mechanisms
7. The right of resource users to self-govern is recognized by higher level authorities
8. In the case of larger common-pool resources: rules are organized and enforced through multiple layers of nested enterprises

Let's get back to Ostrom's design principles



5. Sanctions whose severity is gradually adjusted to the harm caused by the defector; In this case, the costs of an institution may also be graduated

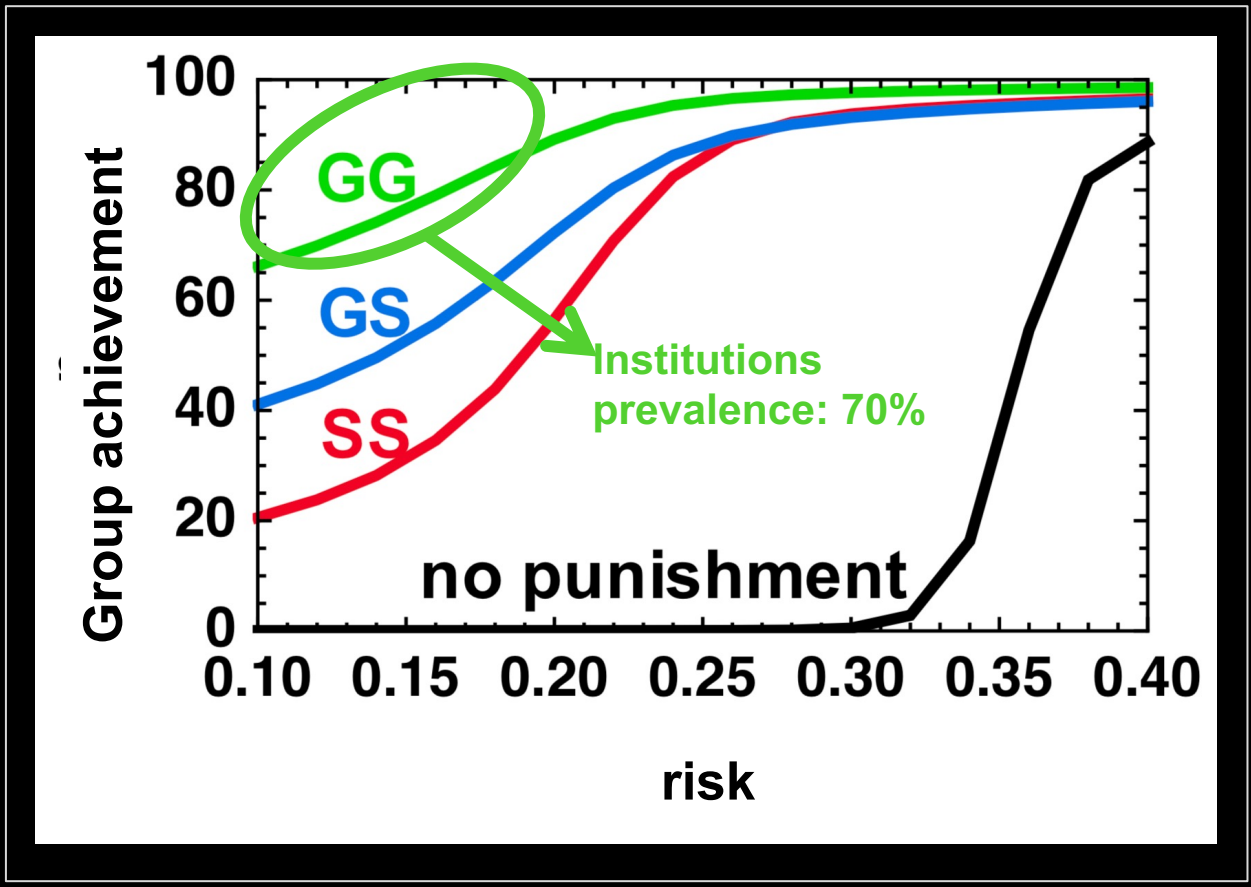
Couto, et al., J Theor Biol. 2020 [& IST MSc-MEFT thesis 2019]

$$\pi_x \rightarrow \pi_x(j_D) \propto \frac{1}{1+e^{-g[j_D-(N-n_{PG})]}}$$

Marta G. Couto

Graduated sanctions allow for the evolution of cooperation and stable institutions

GG: graduated sanctions and graduated costs
GS: graduated sanctions and strict costs
SS: strict sanctions and costs



Marta G. Couto

MG Couto, JM Pacheco, FC Santos, Governance of Risky Public Goods under Graduated Punishment, Journal of Theoretical Biology (2020)

Again, in all cases, higher success is obtained with decentralized initiatives

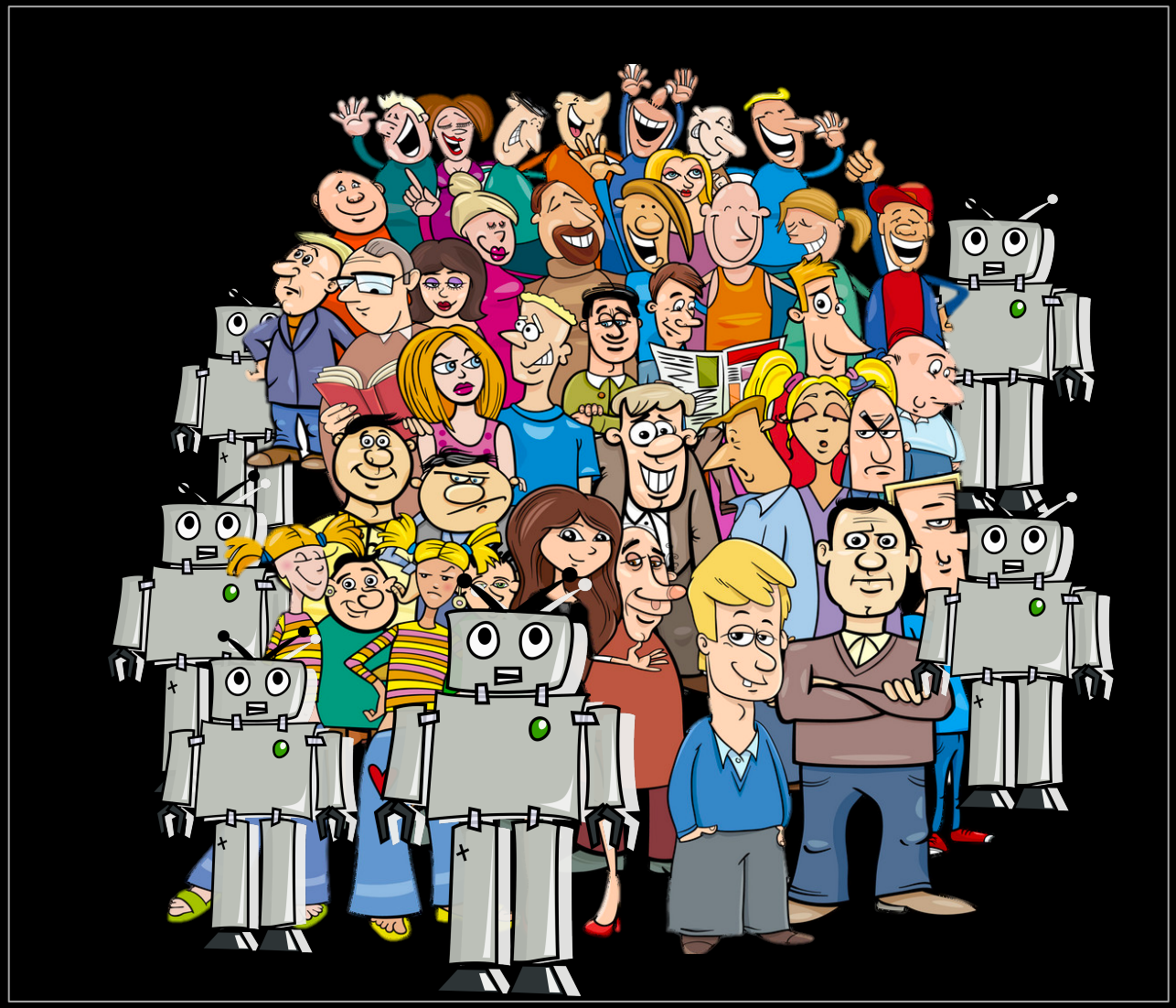
one global institution

a top-down answer,
to a global problem

Multiple, small-scale institutions

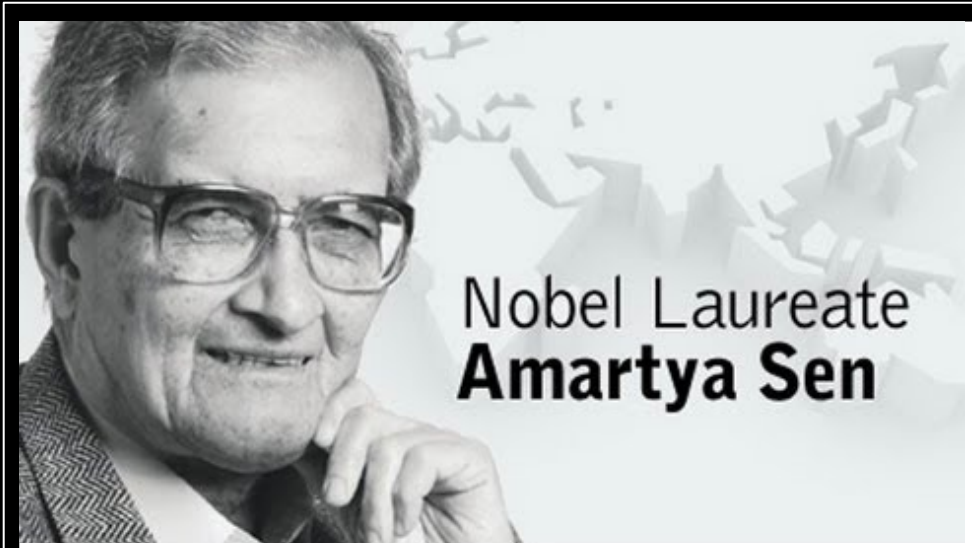
cooperation emerges from decentralized initiatives

What's next?
From social to pro-social AI

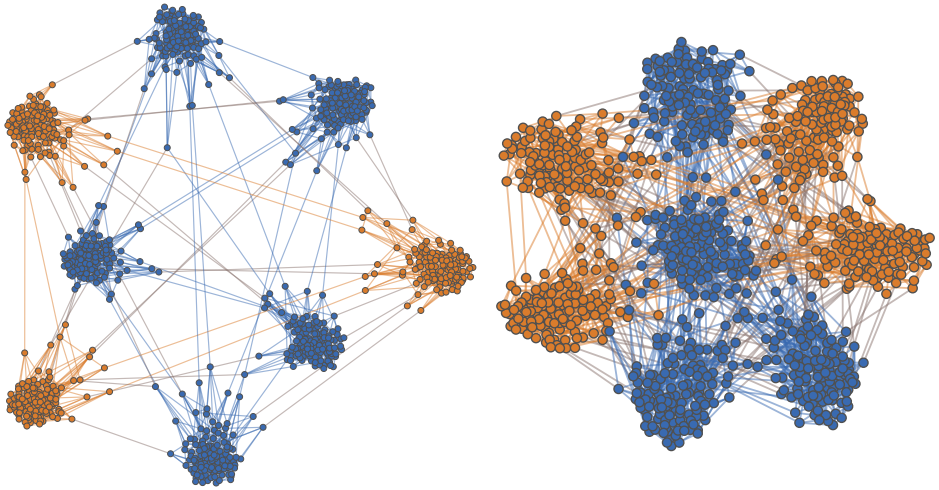


FP Santos, Pacheco, Paiva, FC Santos, Evolution of fairness in hybrid populations of humans and agents, AAAI 2019
A Paiva, FP Santos, & FC Santos, Engineering pro-sociality through autonomous agents, AAAI 2018

Challenges in “pro-social computing”



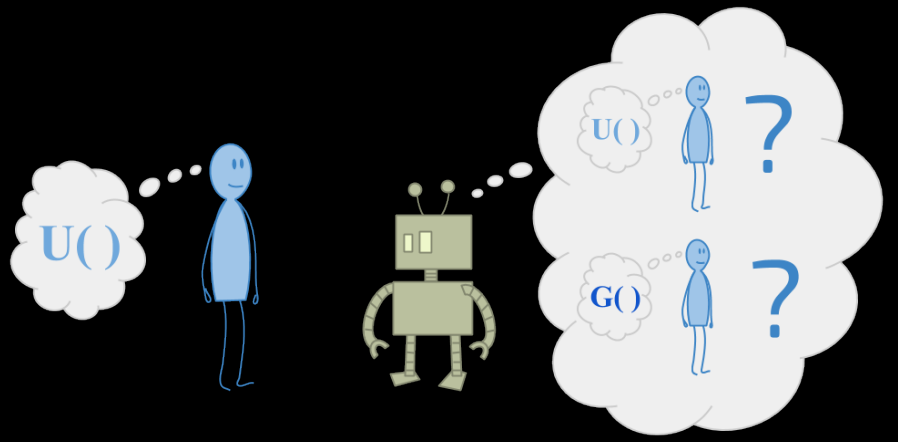
“a society or an economy can be Pareto-optimal and still be perfectly disgusting”



Example: Efficient social matching algorithms in online networks may steer polarization and intolerance

(FP Santos, FC Santos, JM Pacheco, SA Levin, AAMAS, 2021)

Challenges in “pro-social computing”

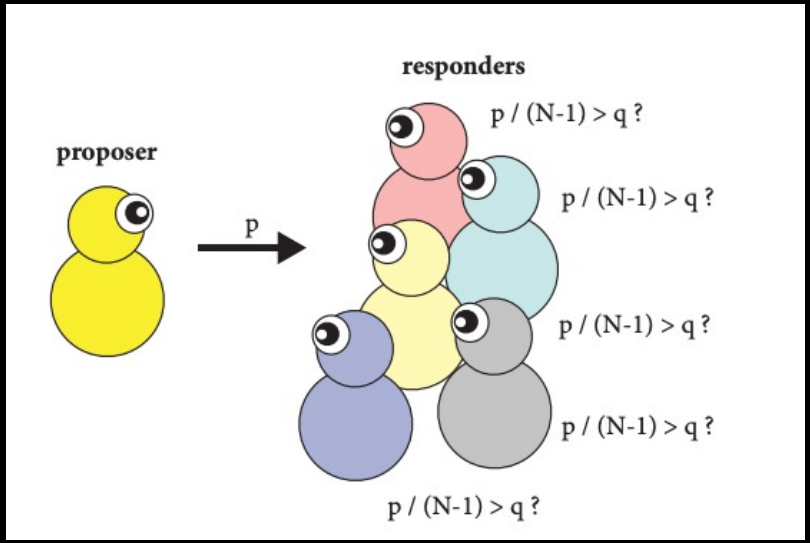


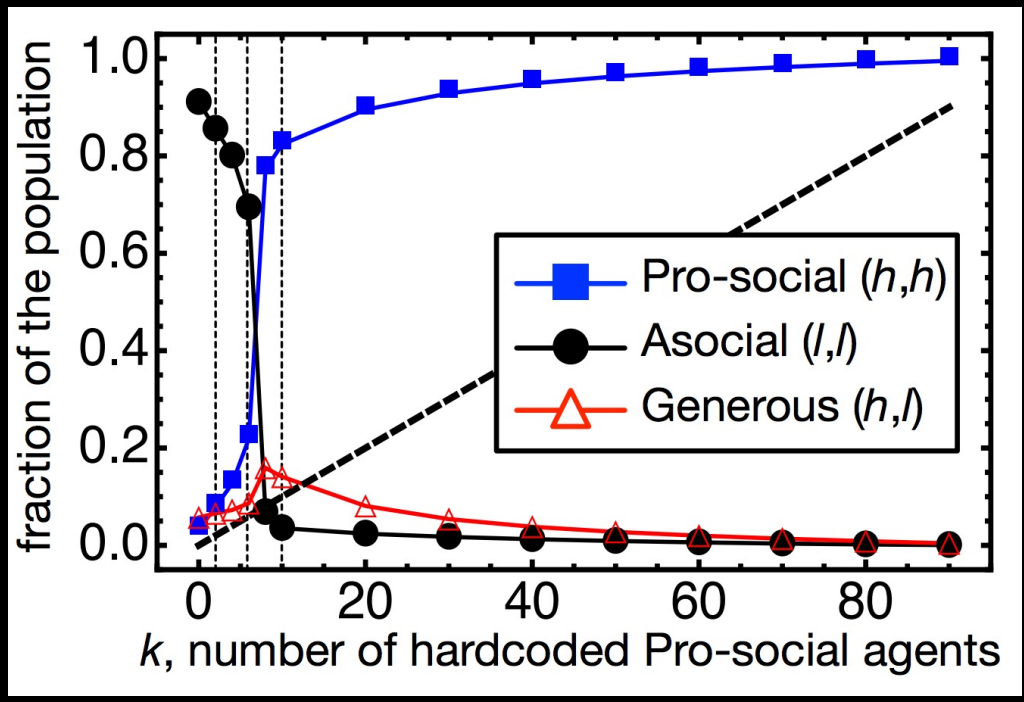
1) Which properties should be added to artificial agents to foster cooperation among humans (and machines)?
2) Do humans cooperate more in the presence of artificial entities? How do we delegate our choices?

FP Santos, Pacheco, Paiva, FC Santos, Evolution of fairness in hybrid populations of humans and agents, AAAI 2019
A Paiva, FP Santos, & FC Santos, Engineering pro-sociality through autonomous agents, AAAI 2018

Example: fairness in populations of humans and agents

Example: fairness in populations of humans and agents



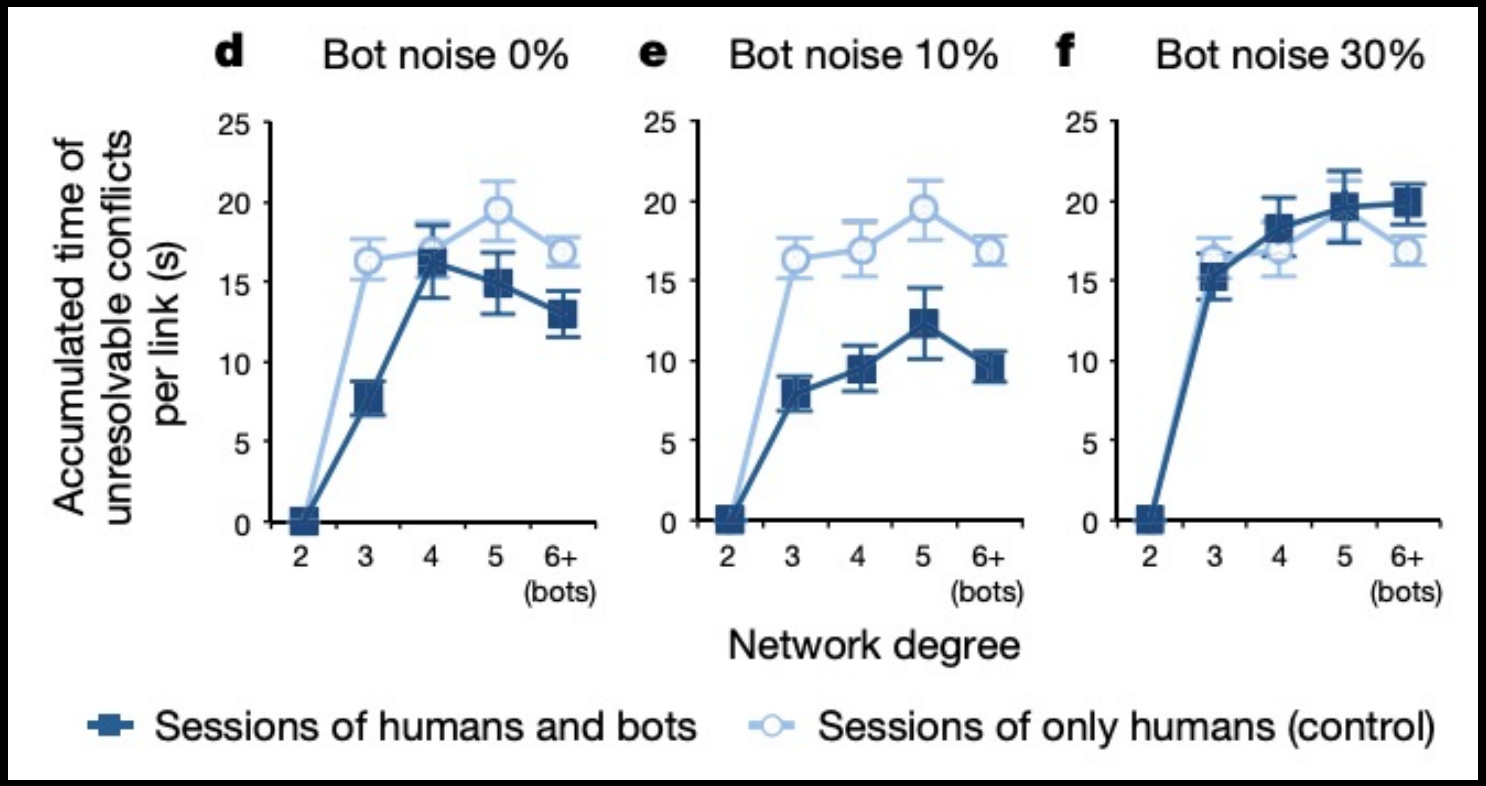


Including a small minority of hardcoded Pro-social agents favors the evolutionary stability of fair states, even for *soft* group decision rules.

FP Santos, Pacheco, Paiva, FC Santos, Evolution of fairness in hybrid populations of humans and agents, AAAI 2019

Example: The role of noisy agents

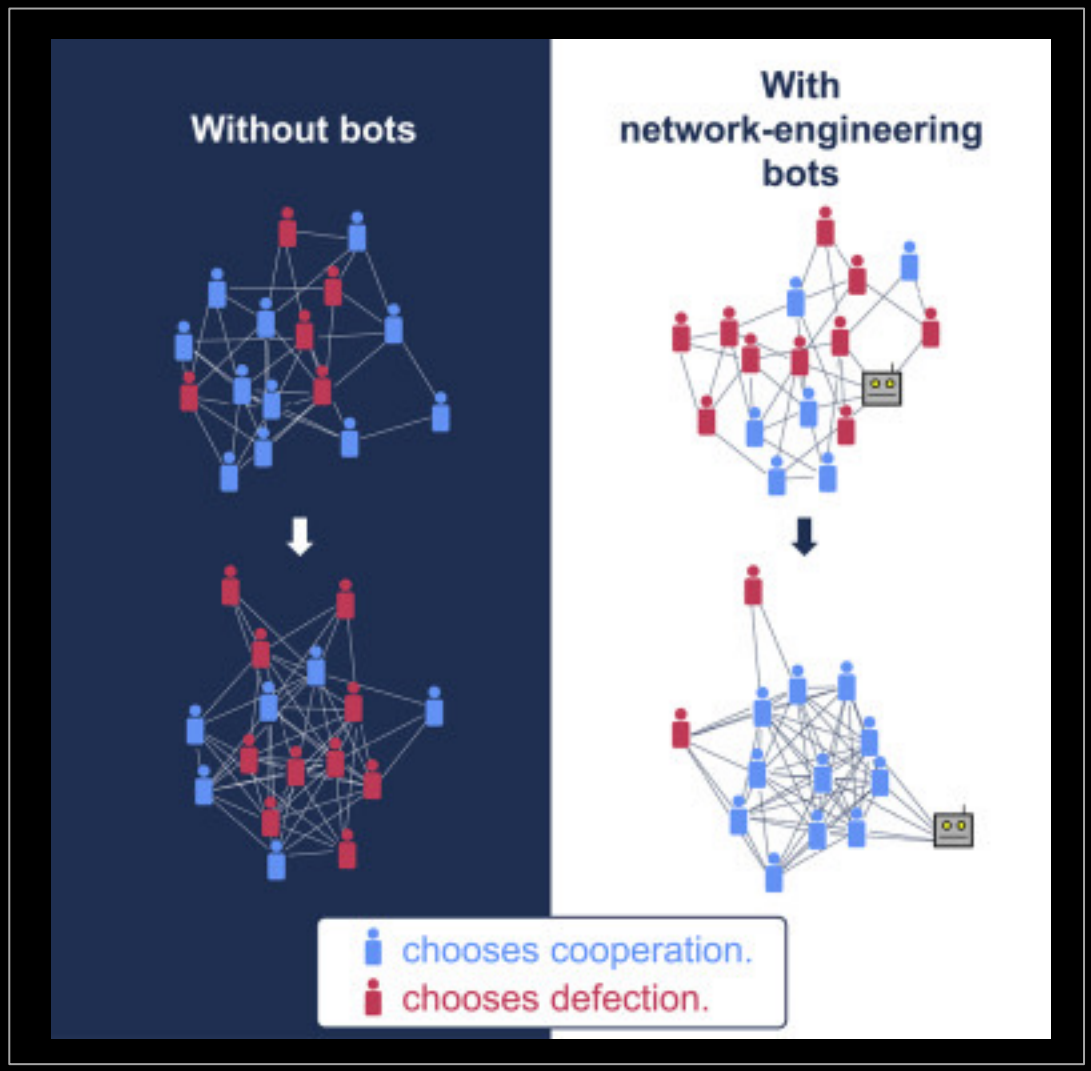
Colour coordination game:
The collective goal is for every node to have a colour different than all of its neighbour nodes

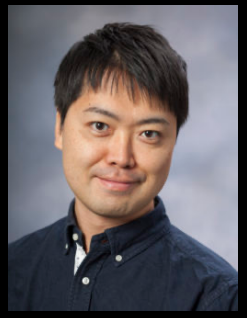


Bots acting with small levels of random noise and placed in central locations meaningfully improve the collective performance of human groups, accelerating the median solution time by 55.6%.

Hirokazu Shirado and Nicholas A. Christakis, Locally noisy autonomous agents improve global human coordination in network experiments. Nature, Vol. 545, pp. 370-374, 2017.

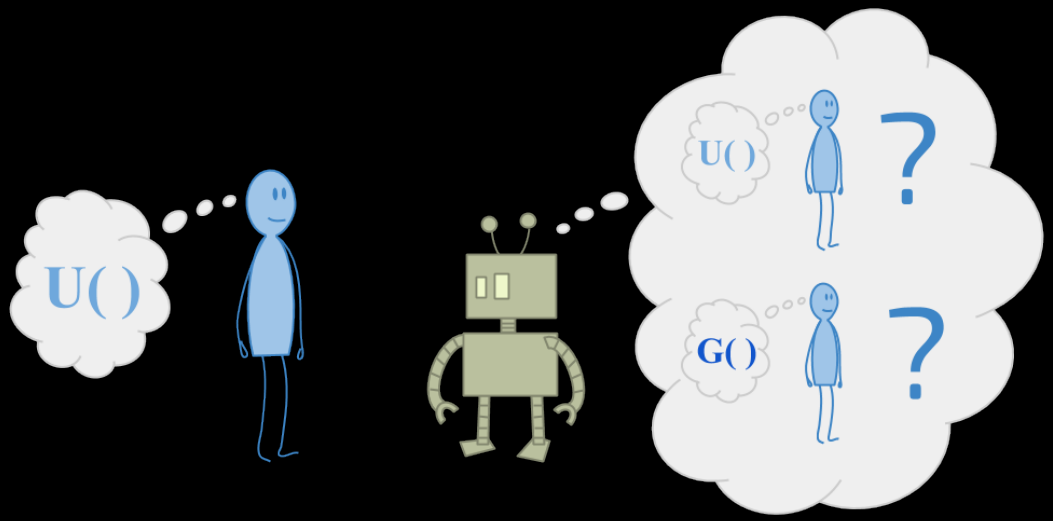
Example: Network engineering using autonomous agents





Hirokazu Shirado and Nicholas A. Christakis, Network engineering using autonomous agents increases cooperation in human groups. iScience, Vol. 23, No. 9, 101438, 2020.

Challenges in “pro-social computing”



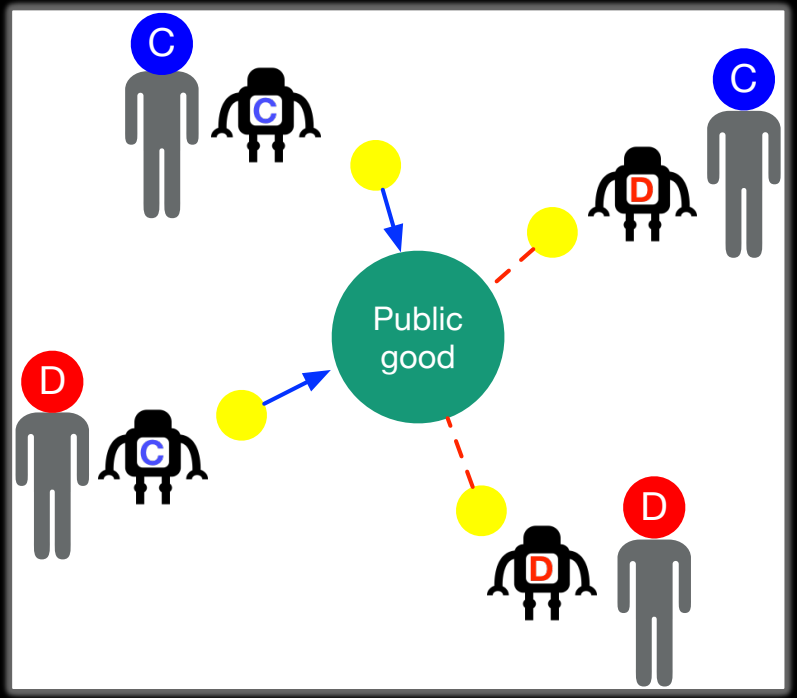
1) Which properties should be added to artificial agents to foster cooperation among humans (and machines)?
2) Do humans cooperate more in the presence of artificial entities? How do we delegate our choices?

FP Santos, Pacheco, Paiva, FC Santos, Evolution of fairness in hybrid populations of humans and agents, AAAI 2019
A Paiva, FP Santos, & FC Santos, Engineering pro-sociality through autonomous agents, AAAI 2018

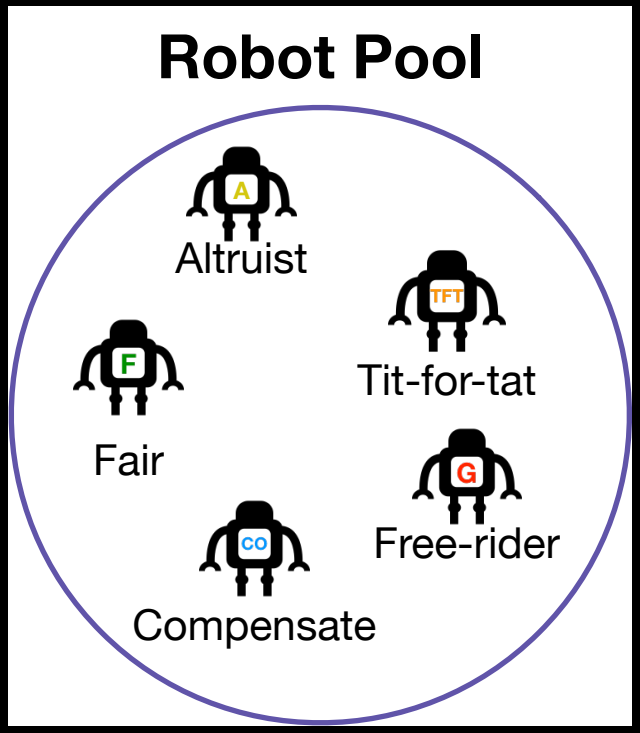
Another example: Climate action through proxies

Delegation

Does delegating to an artificial agent affect cooperation?



What type of agents will humans select?



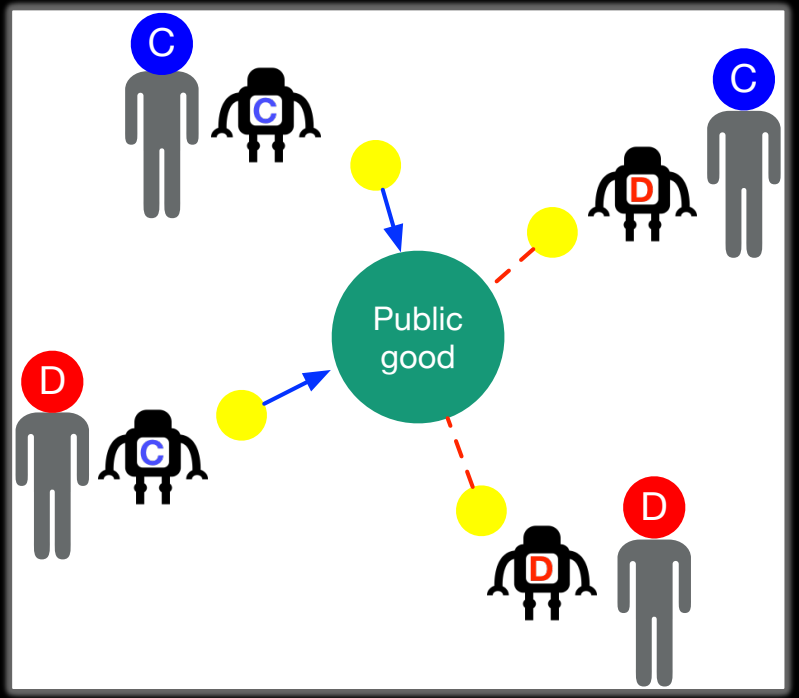
Elias F. Domingos

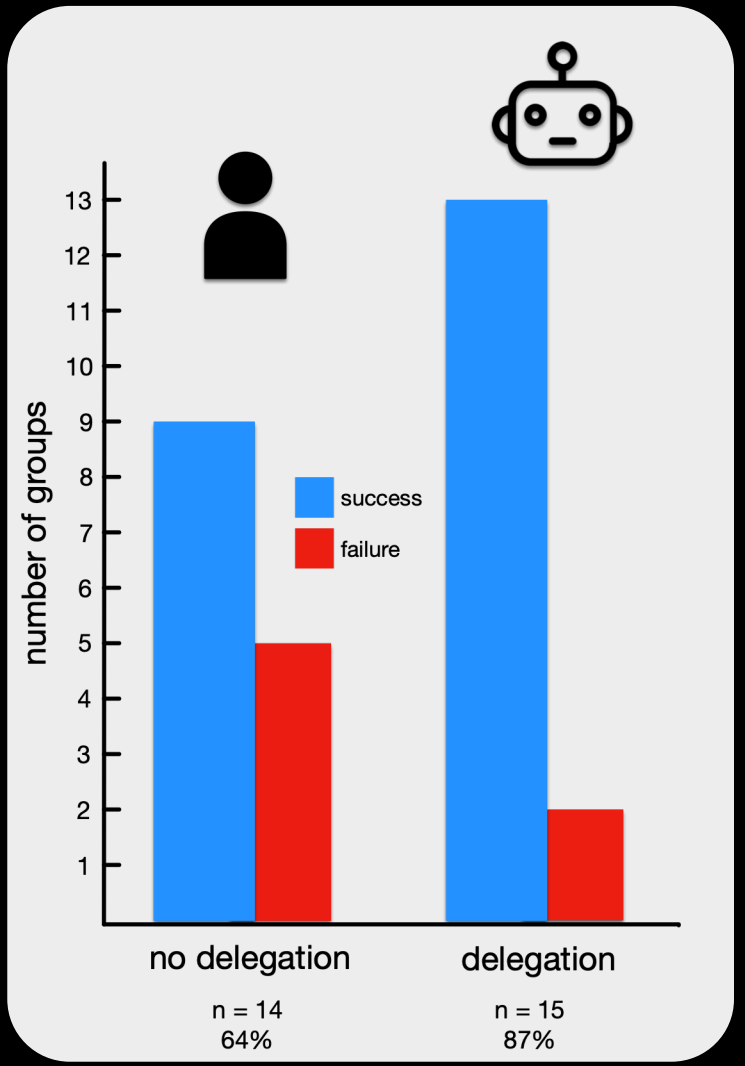
E. F. Domingos, FC Santos, T. Lenaerts et al., arXiv:2103.07710 (2021)
[see also de Melo, Marsella & Gratch, PNAS 2019]

Another example: Climate action through proxies

Delegation

Does delegating to an artificial agent affect cooperation? YES!!





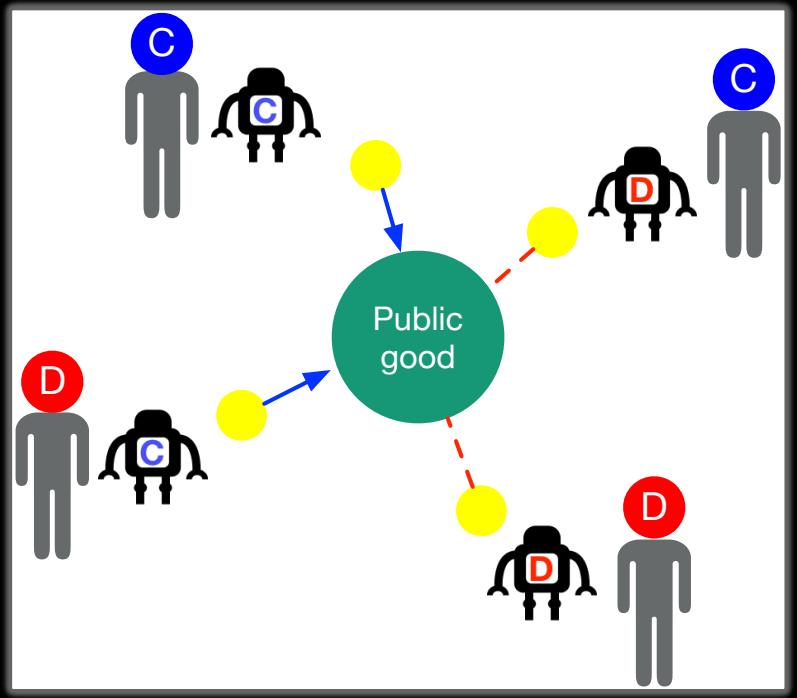
Elias F. Domingos

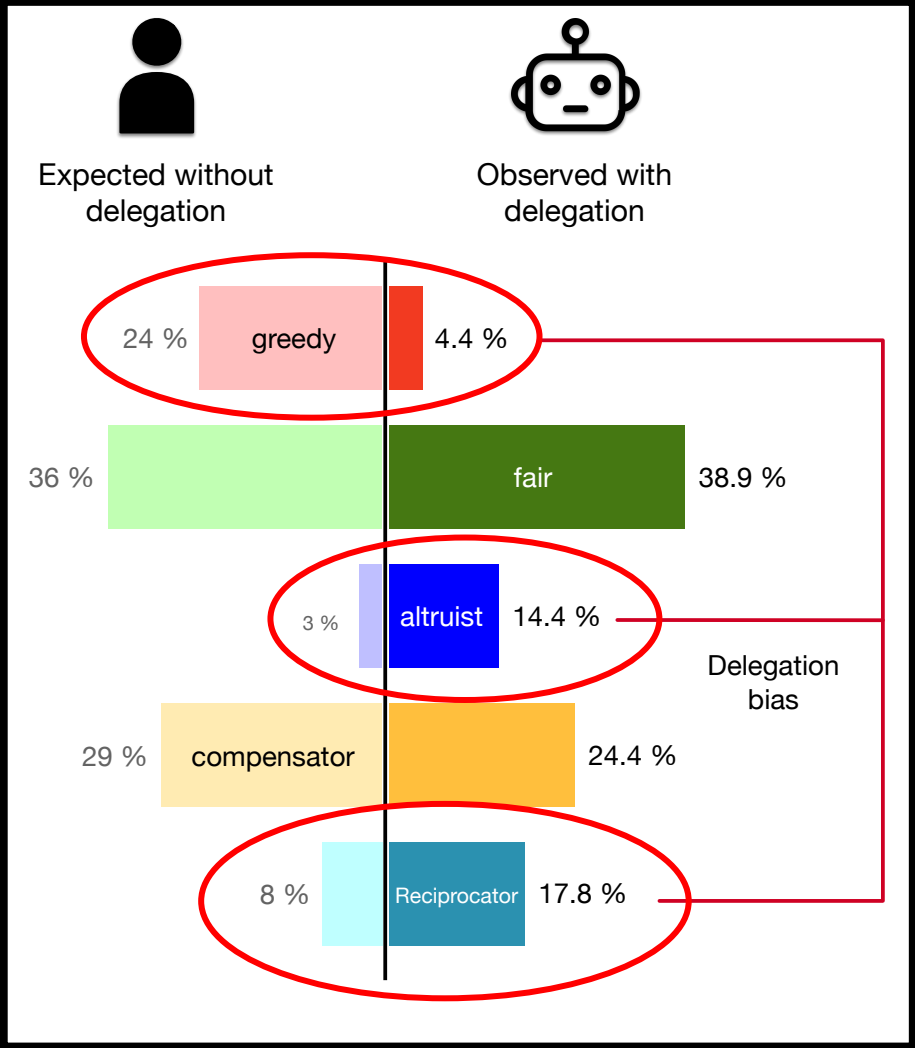
E. F. Domingos, FC Santos, T. Lenaerts et al., arXiv:2103.07710 (2021)
[see also de Melo, Marsella & Gratch, PNAS 2019]

Another example: Climate action through proxies

Delegation

Does delegating to an artificial agent affect cooperation? YES!!





E. F. Domingos, FC Santos, T. Lenaerts et al., arXiv:2103.07710 (2021)
[see also de Melo, Marsella & Gratch, PNAS 2019]

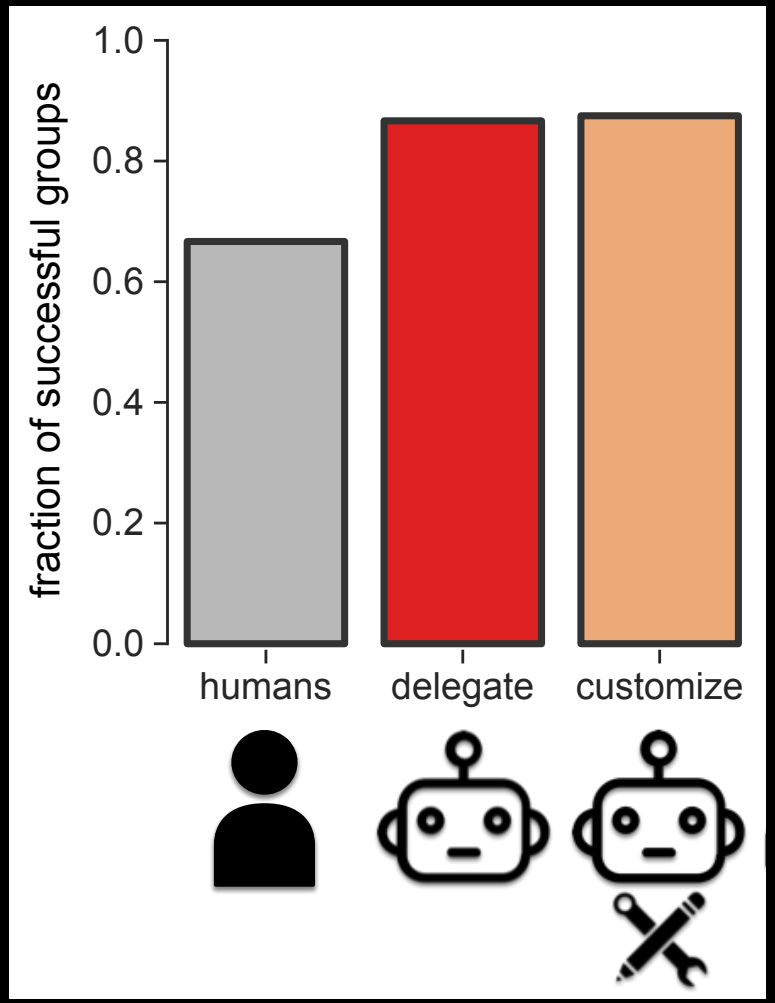
Another example: Climate action through proxies

Delegation

Does ***customization*** affect cooperation?

Build your own agent

For Round=1:10
If Round==1: give a_0
Else
If $a_{-1}(t-1) > T_p$: give a_a
Else If $a_{-1}(t-1) > T_p$: give a_b
Else: give a_c



E. F. Domingos, FC Santos, T. Lenaerts et al., arXiv:2103.07710 (2021)
[see also de Melo, Marsella & Gratch, PNAS 2019]

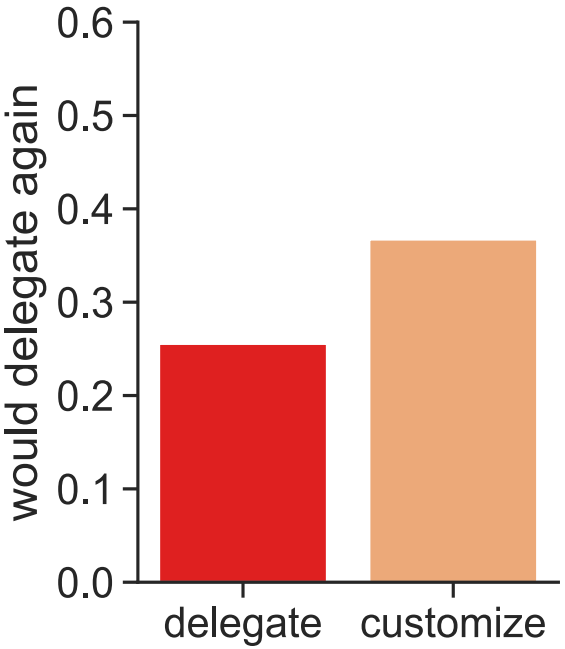
Another example: Climate action through proxies

Delegation

Does ***customization*** affect ***trust***?

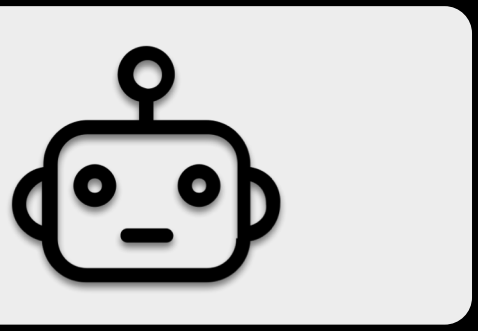
Build your own agent

```
For Round=1:10
    If Round==1: give a0
    Else
        If a-1(t-1) > Tp: give aa
        Else If a-1(t-1) > Tp: give ab
        Else: give ac
```



E. F. Domingos, FC Santos, T. Lenaerts et al., arXiv:2103.07710 (2021)
[see also de Melo, Marsella & Gratch, PNAS 2019]

It’s important to keep in mind where the problem is... and why we want to solve it



Super pro-social robot?

It's important to keep in mind where the problem is... and why we want to solve it

@ Pogo by Walt Kelly

Next Thursday :

First day of our NetSci Workshop