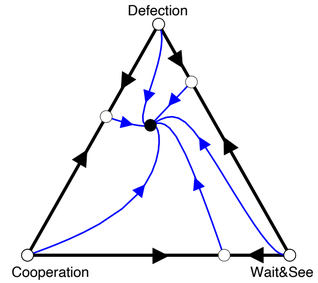
An overview

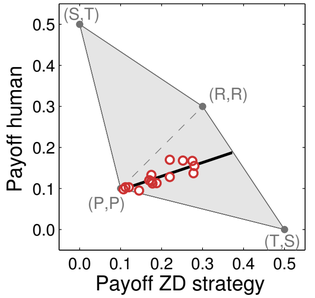
Yesterday's class (March 11, 2025)

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



Today's classes (March 12, 2025)

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



Tomorrow's class (March 13, 2025)

- Some current research: Reciprocity in complex environments

Evolution of cooperation: A Review

Editorial

Under Review

Most scientists regarded the new streamlined peer-review process as "twice an improvement."

Remark 3.1. Evolution of cooperation

- In theoretical biology, cooperation is often interpreted as a costly behavior that benefits someone else.
- These instances of cooperation are difficult to explain when we think that cooperation is properly modelled by a (one-shot) prisoner's dilemma.
- Instead, instances of cooperation may be driven by kin selection, group selection, direct reciprocity, additional incentives (reward and punishment), etc.
- This morning, we focussed on direct reciprocity; people are more cooperative when we interact repeatedly.
- However, people sometimes engage in interactions with others they hardly know. In that case, they often rely on reputations. One idea: people cooperate because of social norms.

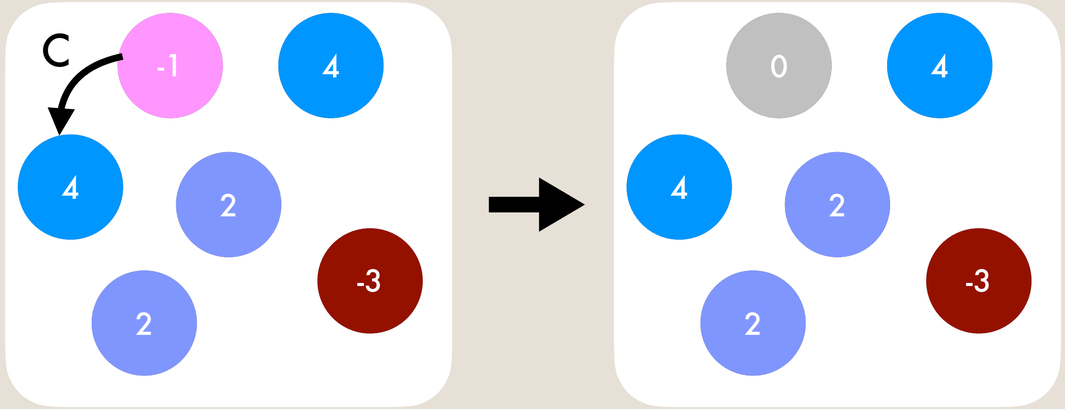
Remark 3.2. Social norms

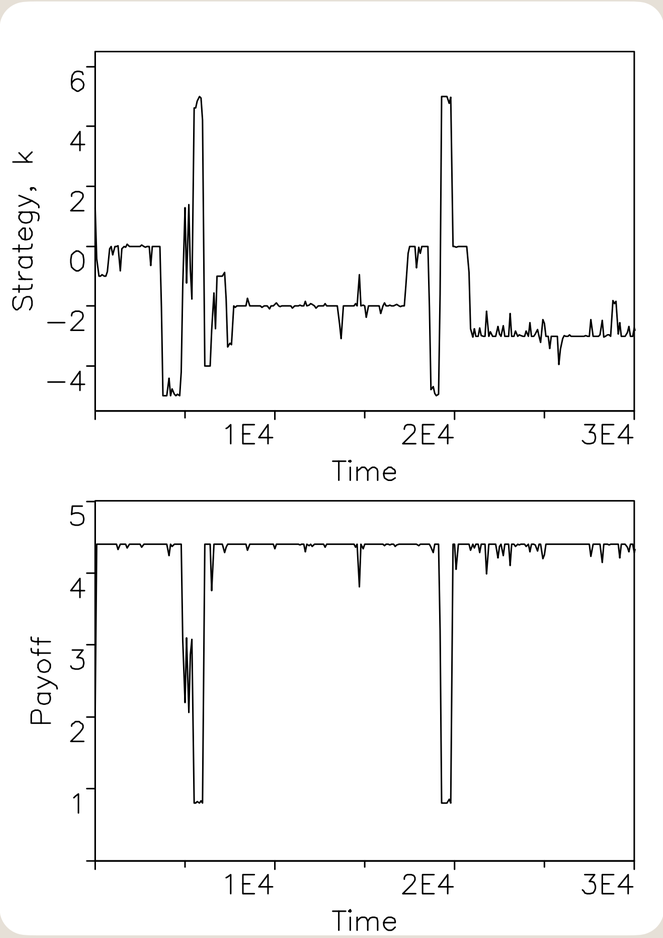
- There are many kinds of social norms out there.
- Such norms consist of rules that determine what people ought to do in certain situations, and what consequences misbehaviour should have.
- Social norms change in time. Beneficial norms, and norms that are easily enforced tend to stay. Others go.
  ⇒We can explore the evolution of norms with evolutionary game theory.
- Are there stable social norms that prevent people from defecting in the prisoner's dilemma?

Evolution of Indirect reciprocity: Image Scoring

Remark 3.3. Image scoring (Nowak & Sigmund 1998)

- Consider a population with n individuals
- In each time step, two individuals are randomly chosen. One individual is randomly determined to be the *donor*. The other individual is the *recipient*.
- Donors decide whether or not to pay a cost $c>0$ to deliver a benefit $b>c$ to the recipient.
- The donor's action affects the donor's image score (an integer). Every time the donor cooperates, the score goes up by one. Every time the donor defects, the score goes down by one. Scores are restricted to [-5,5].
- Here a donor's strategy takes the form of a threshold, that determines the minimum score $k \in [-5,6]$ that the co-player needs to have for the donor to cooperate.
- Unconditional defectors would use the maximum k. Unconditional cooperators would use the minimum k.
- What value of k would emerge if individuals update their strategies according to an evolutionary process?





Evolution of Indirect reciprocity: Image Scoring, revisited

How should we define goodness?—reputation dynamics in indirect reciprocity

Hisashi Ohtsuki*, Yoh Iwasa

Remark 3.4. Instability of Image Scoring

- Leimar & Hammerstein (2001): Image scoring cannot be a stable strategy. Even if the co-player has a bad score, individuals may wish to cooperate in order to maintain their own reputation.
- Major insight: To maintain cooperation, a population's reputation system should differentiate between *justified* and *unjustified defections*.
  That is, when assessing someone's action, one should not only take into account *what that person did*, but also to *whom*.

Evolution of cooperation through indirect reciprocity

Olof Leimar[1*] and Peter Hammerstein[2]

Remark 3.5. Searching for stable cooperative social norms

- Ohtsuki & Iwasa (2004) considered a model in which individuals can either have a *good* or a *bad* reputation (i.e., there are only two possible scores).
- However, they allow for more complex social norms. These norms consist of two components (α, β).
- The first component is called the social norm's assessment rule. It determines how observers update a donor's reputation:

$$\alpha = (\alpha_{gCg}, \alpha_{gCb}, \alpha_{bCg}, \alpha_{bCb}, \alpha_{gDg}, \alpha_{gDb}, \alpha_{bDg}, \alpha_{bDb}) \in \{0,1\}^8$$

- The second component is called the action rule. It determines how donors make their decision whether or not to cooperate with a given recipient

$$\beta = (\beta_{gg}, \beta_{gb}, \beta_{bg}, \beta_{bb}) \in \{0,1\}^4$$

- In principle, this gives rise to $2^{12} = 4{,}096$ possible social norms to consider.

Evolution of indirect reciprocity: Ohtsuki & Iwasa (2004)

How should we define goodness?—reputation dynamics in indirect reciprocity

Hisashi Ohtsuki*, Yoh Iwasa

Remark 3.6. Examples of third-order norms

- Unconditional cooperators (ALLD)

$$(\alpha_{gCg}, \alpha_{gCb}, \alpha_{bCg}, \alpha_{bCb}, \alpha_{gDg}, \alpha_{gDb}, \alpha_{bDg}, \alpha_{bDb}) = (0,0,0,0,0,0,0,0)$$
$$(\beta_{gg}, \beta_{gb}, \beta_{bg}, \beta_{bb}) = (0,0,0,0)$$

- Unconditional cooperators (ALLC)

$$(\alpha_{gCg}, \alpha_{gCb}, \alpha_{bCg}, \alpha_{bCb}, \alpha_{gDg}, \alpha_{gDb}, \alpha_{bDg}, \alpha_{bDb}) = (1,1,1,1,1,1,1,1)$$
$$(\beta_{gg}, \beta_{gb}, \beta_{bg}, \beta_{bb}) = (1,1,1,1)$$

- Simple Scoring (SCO)

$$(\alpha_{gCg}, \alpha_{gCb}, \alpha_{bCg}, \alpha_{bCb}, \alpha_{gDg}, \alpha_{gDb}, \alpha_{bDg}, \alpha_{bDb}) = (1,1,1,1,0,0,0,0)$$
$$(\beta_{gg}, \beta_{gb}, \beta_{bg}, \beta_{bb}) = (1,0,1,0)$$

Remark 3.7. Searching for stable cooperative social norms

Question: Among all these pairs (α, β), can we identify all social norms with the following two properties:

- If the whole population adopts it, then everybody always cooperates in the long run
- The norm is self-enforcing (no population member can gain a higher payoff by deviating from the social norm).

If these norms exist, how do they look like?

Evolution of indirect reciprocity: Ohtsuki & Iwasa (2004)

How should we define goodness?—reputation dynamics in indirect reciprocity

Hisashi Ohtsuki*, Yoh Iwasa

Remark 3.8. The "*Leading eight*"

- Using analytical methods and numerical computations, Ohtsuki and Iwasa identified eight social norms that satisfy both properties.

- What are their shared characteristics?

  - Cooperating with a good recipient should always yield a good reputation.

  - Defecting against a good recipient should always yield a bad reputation.

  - A good donor who defects against a bad recipient should keep his/her good reputation ("justified punishment")

- The norms disagree on how one should assess
  - Good donors who cooperate with bad recipients
  - Bad donors who cooperate with bad recipients
  - Bad donors who defect with bad recipients

Assessment rule	* L1	Consistent Standing L2	Simple Standing L3	* L4	* L5	Stern Judging L6	Staying L7	Judging L8
Good cooperates with Good	g	g	g	g	g	g	g	g
Good cooperates with Bad	g	b	g	g	b	b	g	b
Bad cooperates with Good	g	g	g	g	g	g	g	g
Bad cooperates with Bad	g	g	g	b	g	b	b	b
Good defects against Good	b	b	b	b	b	b	b	b
Good defects against Bad	g	g	g	g	g	g	g	g
Bad defects against Good	b	b	b	b	b	b	b	b
Bad defects against Bad	b	b	g	g	g	g	b	b

Action rule	L1	L2	L3	L4	L5	L6	L7	L8
Good meets Good	C	C	C	C	C	C	C	C
Good meets Bad	D	D	D	D	D	D	D	D
Bad meets Good	C	C	C	C	C	C	C	C
Bad meets Bad	C	C	D	D	D	D	D	D

Evolution of indirect reciprocity: The leading eight

Remark 3.9. Stern Judging

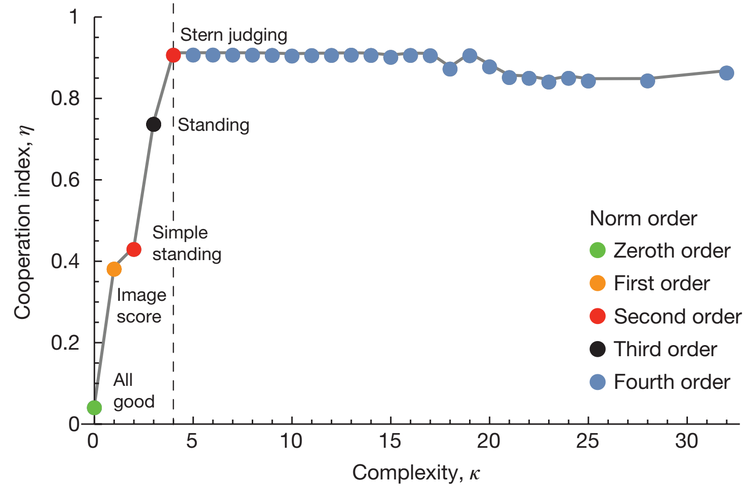
- One important rule among the leading eight is called “Stern Judging” (L6):
  The only behaviors that should yield a good reputation are cooperating with good people and defecting with bad people.

- Among all norms that guarantee full cooperation, this is the one of the lowest complexity (Santos et al 2018).

- Evaluations consistent with this norm can be found even in toddlers (as young as five months old); toddlers do not only show a preference for individuals who helped others, but also for individuals who harmed those who hindered others (Hamlin et al 2011).

Social norm complexity and past reputations in the evolution of cooperation

Fernando P. Santos[1,2], Francisco C. Santos[1,2] & Jorge M. Pacheco[2,3,4]



How infants and toddlers react to antisocial others

J. Kiley Hamlin[a,1], Karen Wynn[b], Paul Bloom[b], and Neha Mahajan[b]

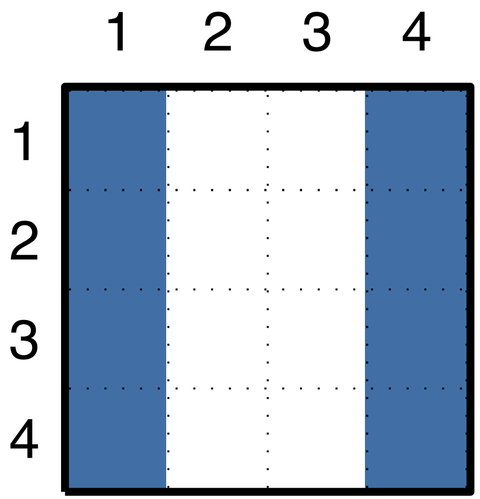
Infants prefer a nasty moose if it punishes an unhelpful elephant

Evolution of indirect reciprocity: The impact of noise

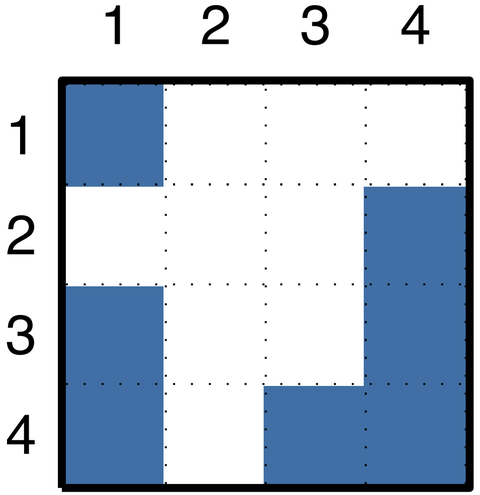
Remark 3.10. Indirect reciprocity with noise

- One strong assumption in the model on the leading-eight: All relevant information is public and there are no perception errors.
- One way to represent these agreements is to consider the image matrix $M(t) = \left(m_{ij}(t)\right)$, with $m_{ij} = 1$ if and only if player i considers j to be good.
- Under public information and no noise these image matrices only depend on the column index j.
- When there is private information, or some individuals misinterpret a donor's action, there can be disagreements.

Question: Assume there are a few initial disagreements between the members of a population. Over time, do these disagreements disappear or do they proliferate?
And how does this depend on the population's social norm?



Under public information



Under private information

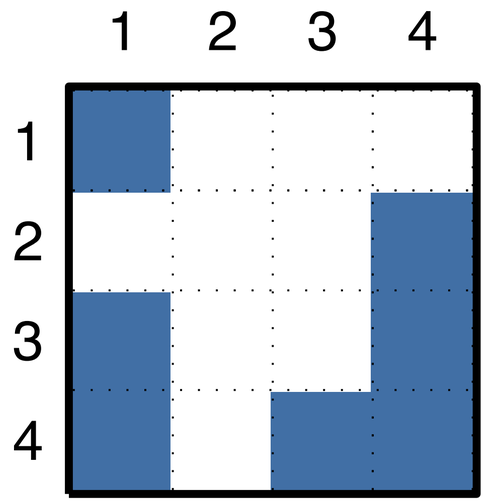
Evolution of indirect reciprocity: The impact of noise

Indirect reciprocity with private, noisy, and incomplete information

Christian Hilbe[a,1], Laura Schmid[a], Josef Tkadlec[a], Krishnendu Chatterjee[a], and Martin A. Nowak[b,c,d]

Remark 3.10. Indirect reciprocity with noise (continued)

- Consider a population in which players assign reputations based on private and noisy information
- Moreover, suppose people in the population follow different social norms; one third uses ALLC, one third uses ALLD, and one third uses some leading-eight social norm.
- Assume initially everyone considers everyone as good, and donors and recipients are randomly chosen as before. However, now some population members may not observe a certain interaction. Even if they observe it, there may be a small probability that the donor's action is misinterpreted.
- How do the population's image matrices evolve in time?



Under private information

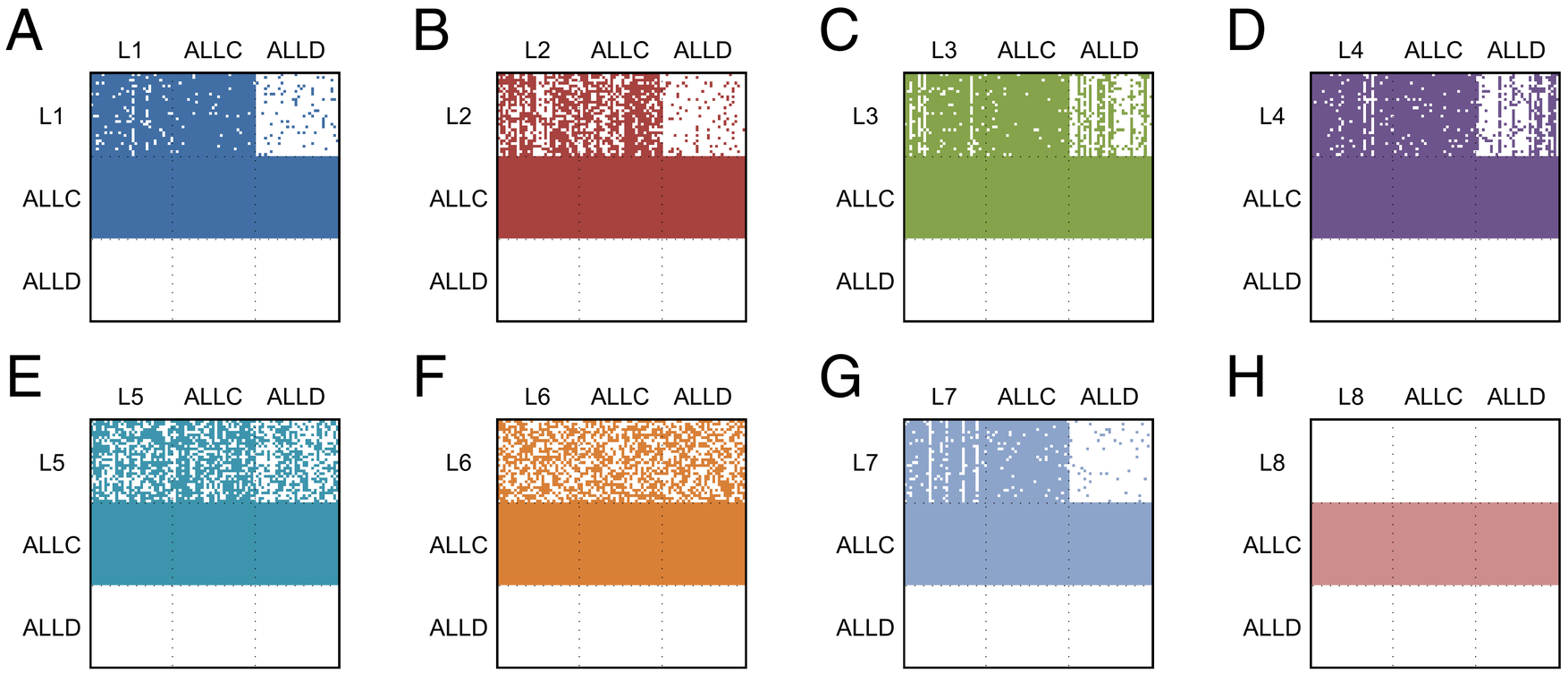
Evolution of indirect reciprocity: The impact of noise

Indirect reciprocity with private, noisy, and incomplete information

Christian Hilbe[a,1], Laura Schmid[a], Josef Tkadlec[a], Krishnendu Chatterjee[a], and Martin A. Nowak[b,c,d]

Reputation dynamics with private information

We assume the population consists in equal parts of a leading-eight strategy, of ALLC, and ALLD.
Snapshot after 10^6 interactions:



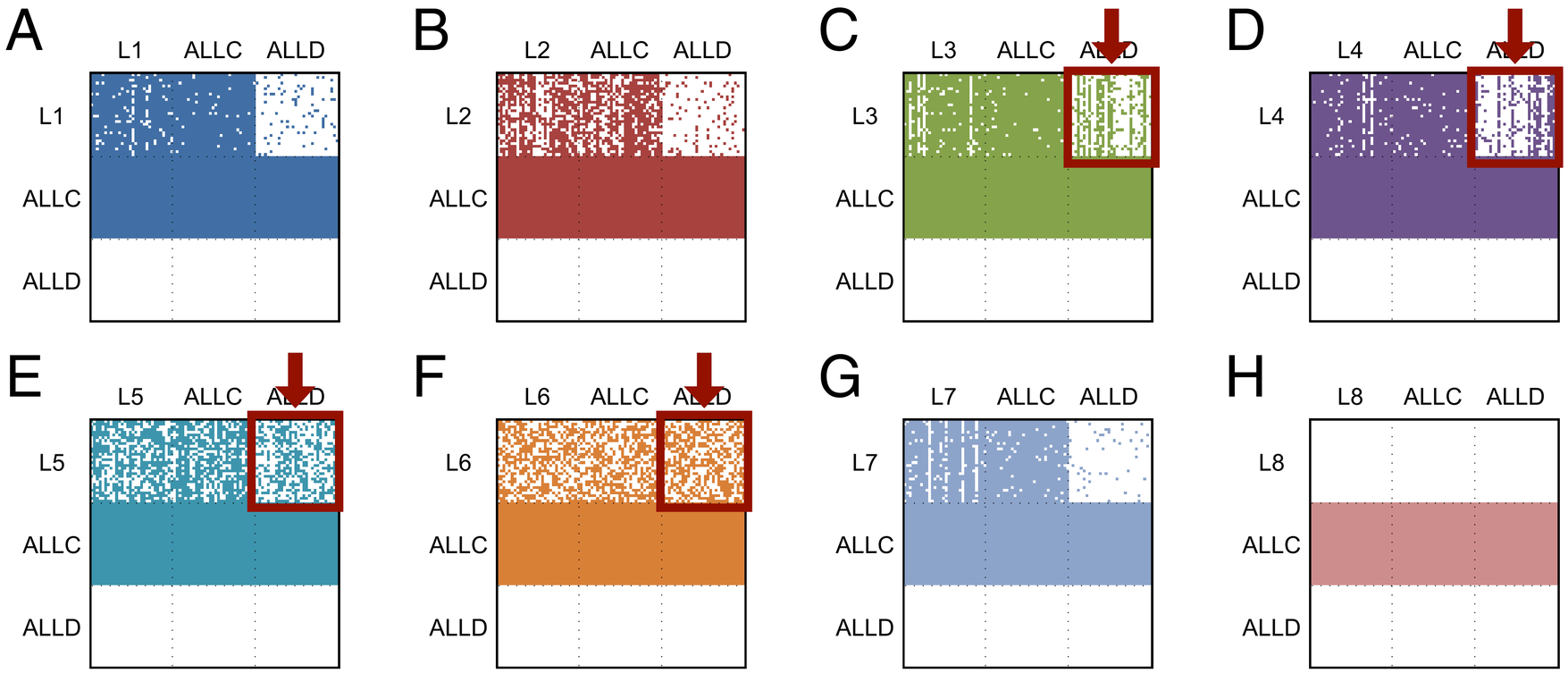
Evolution of indirect reciprocity: The impact of noise

Indirect reciprocity with private, noisy, and incomplete information

Christian Hilbe[a,1], Laura Schmid[a], Josef Tkadlec[a], Krishnendu Chatterjee[a], and Martin A. Nowak[b,c,d]

Reputation dynamics with private information

We assume the population consists in equal parts of a leading-eight strategy, of ALLC, and ALLD.
Snapshot after 10^6 interactions:



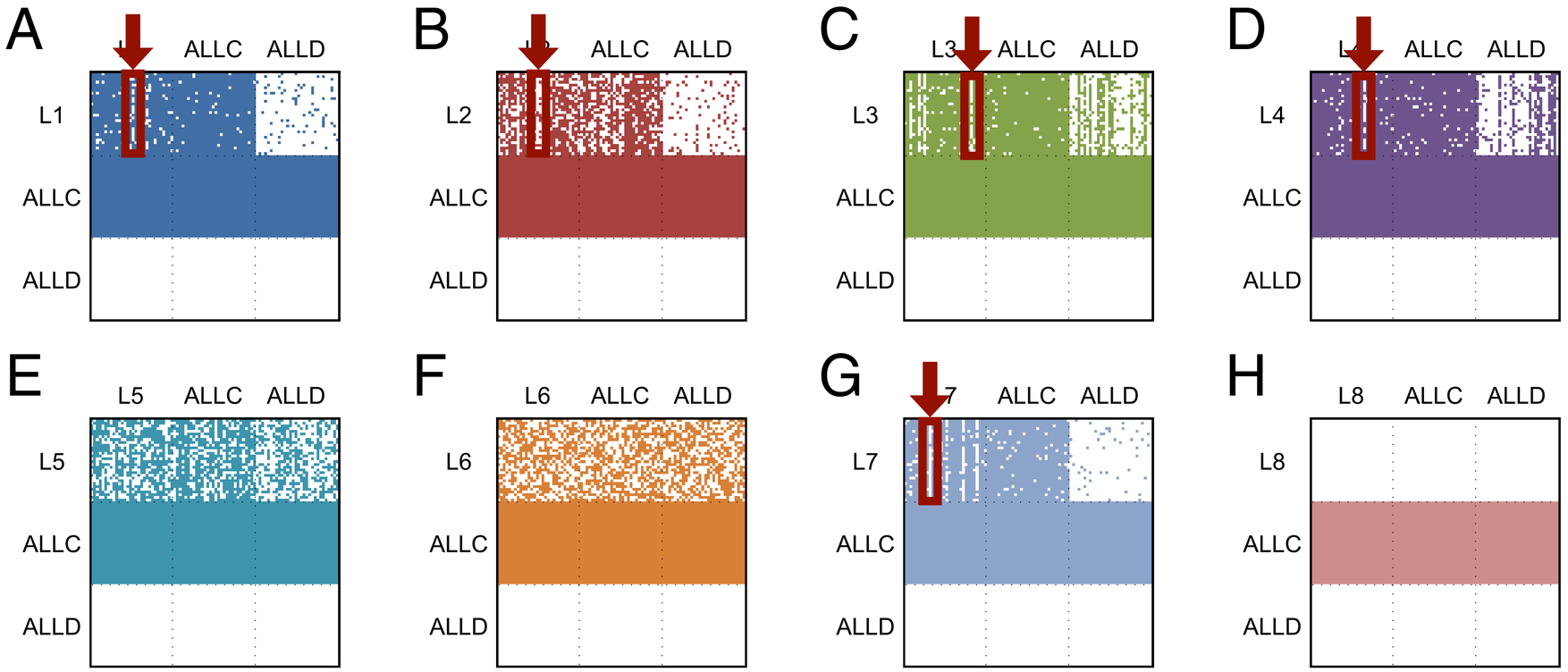
Evolution of indirect reciprocity: The impact of noise

Indirect reciprocity with private, noisy, and incomplete information

Christian Hilbe[a,1], Laura Schmid[a], Josef Tkadlec[a], Krishnendu Chatterjee[a], and Martin A. Nowak[b,c,d]

Reputation dynamics with private information

We assume the population consists in equal parts of a leading-eight strategy, of ALLC, and ALLD.
Snapshot after 10^6 interactions:



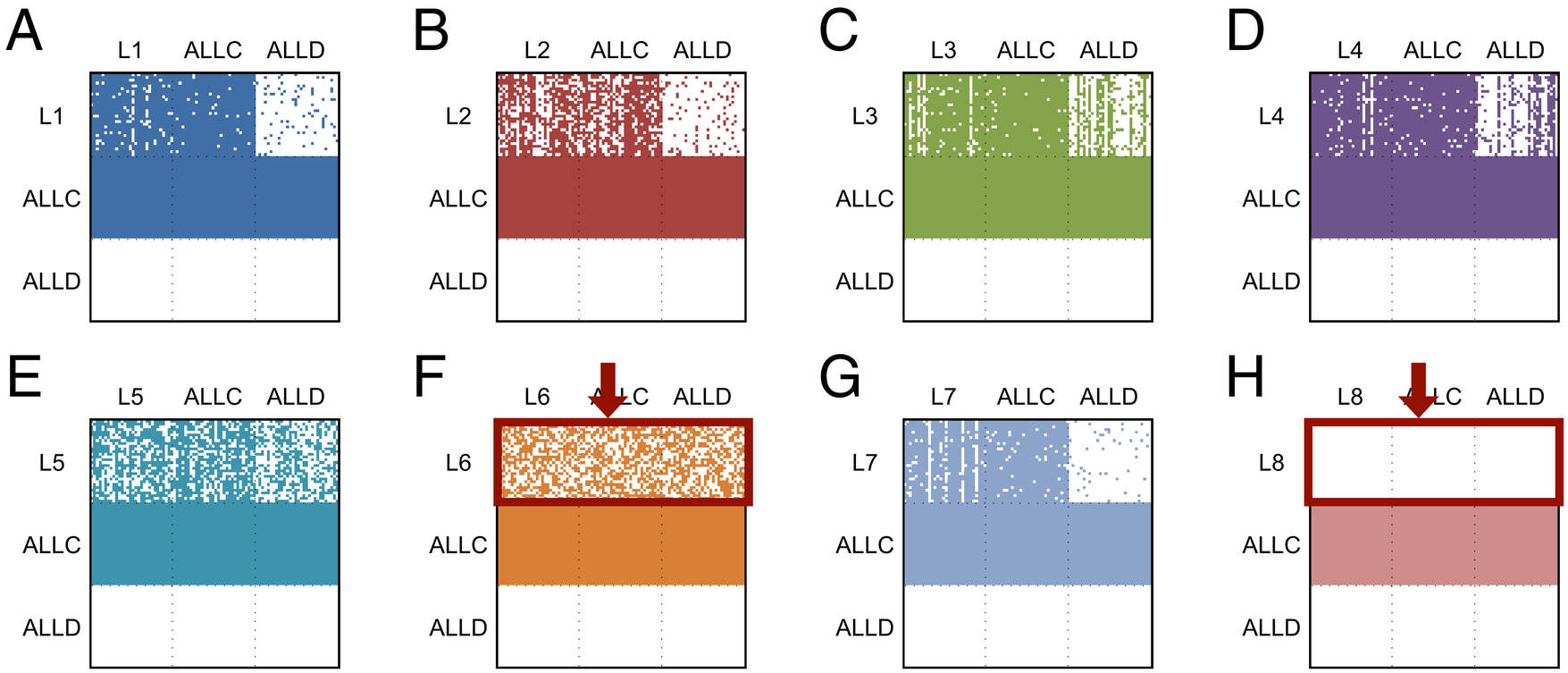
Evolution of indirect reciprocity: The impact of noise

Indirect reciprocity with private, noisy, and incomplete information

Christian Hilbe[a,1], Laura Schmid[a], Josef Tkadlec[a], Krishnendu Chatterjee[a], and Martin A. Nowak[b,c,d]

Reputation dynamics with private information

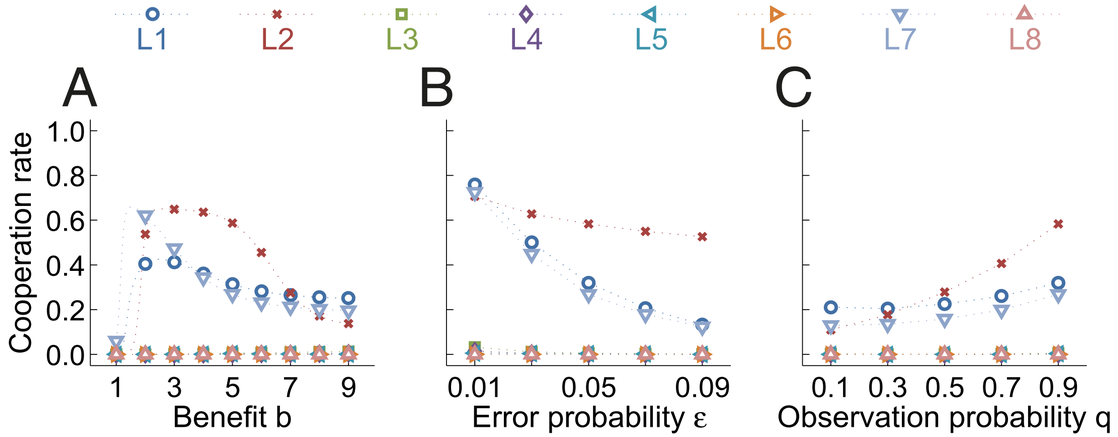
We assume the population consists in equal parts of a leading-eight strategy, of ALLC, and ALLD.
Snapshot after 10^6 interactions:



Evolution of indirect reciprocity: The impact of noise

Remark 3.10. Indirect reciprocity with noise (continued)

- Overall, noise seems to make it harder for the leading-eight to be stable.
- Once there are errors, players may think of each other as bad, although they apply exactly the same norm
- These disagreements can spread
- Some norms fail to maintain cooperation altogether.



Possible solutions

Empathy

Evolution of empathetic moral evaluation

Arunas L Radzvilavicius[1]*, Alexander J Stewart[2], Joshua B Plotkin[1]*

eLife, 2019

Institutions

Adherence to public institutions that foster cooperation

Arunas L. Radzvilavicius[1,2], Taylor A. Kessinger[1,2] & Joshua B. Plotkin[1]✉

Nature Communications, 2021

Gossip

Explaining the evolution of gossip

Xinyue Pan[a,1], Vincent Hsiao[b], Dana S. Nau[b,c], and Michele J. Gelfand[d,e,1]

PNAS, 2024

Opinion synchronization

Indirect reciprocity under opinion synchronization

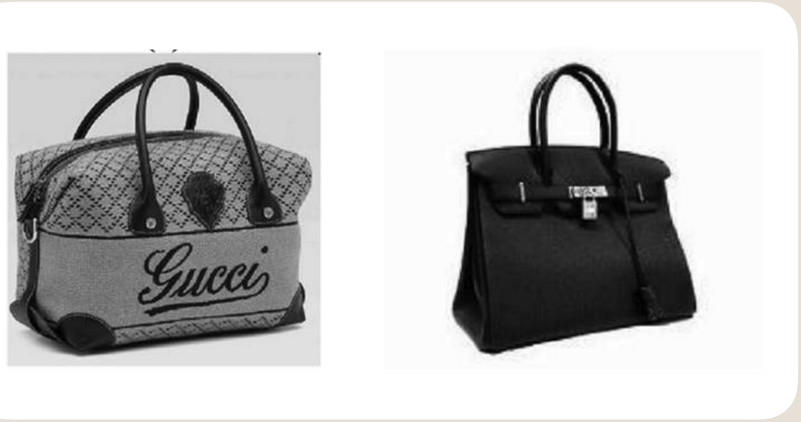
Yohsuke Murase[a,1] and Christian Hilbe[b]

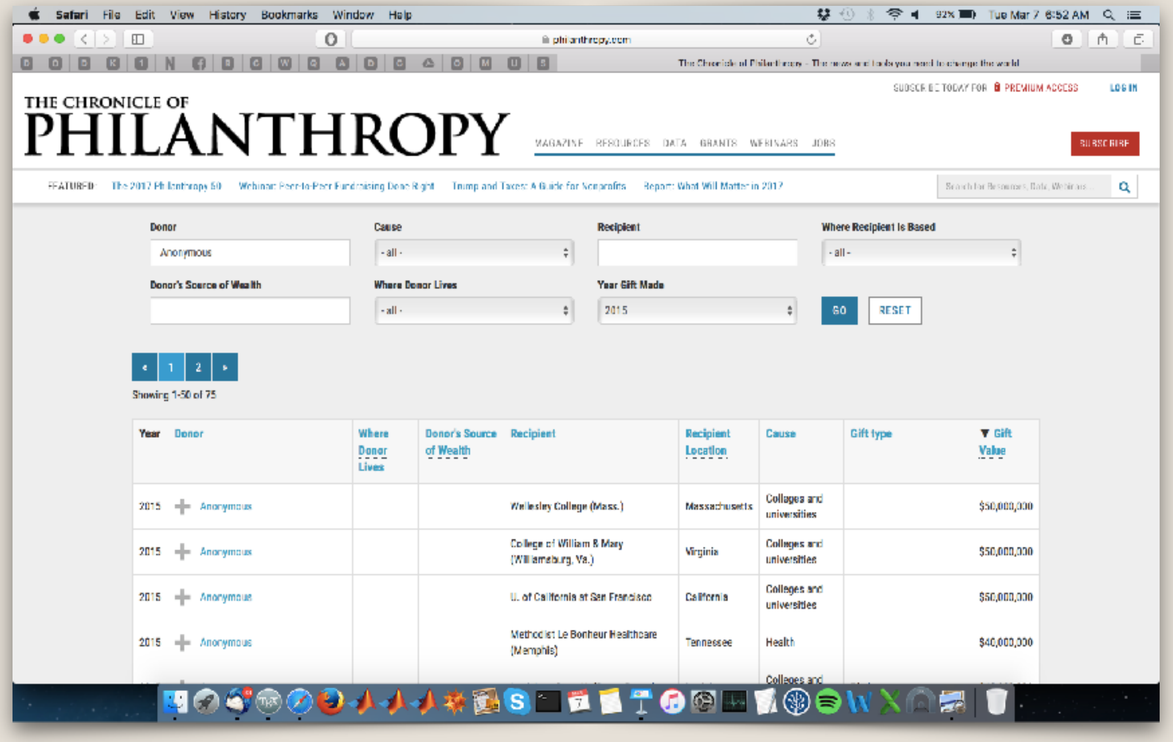
PNAS, 2024

Evolution of social norms: Humility

Remark 3.11. Why being humble?

- Interesting observation: Indirect reciprocity is all about cooperating publicly, such that other people would learn about it.
- Similar points can be made with other types of models: For example, if you want to indicate your wealth, you are incentivised to engage in "conspicuous consumption"
- At the same time, you often observe social norms that suggest people should be humble; they should not brag about their achievements, nor about their good deeds.
- Examples: Anonymous donations, avoiding to brag about your latest high-impact publication, inconspicuous consumption
- Why pay a cost to do something good, or purchase something valuable, and then hiding the cost?

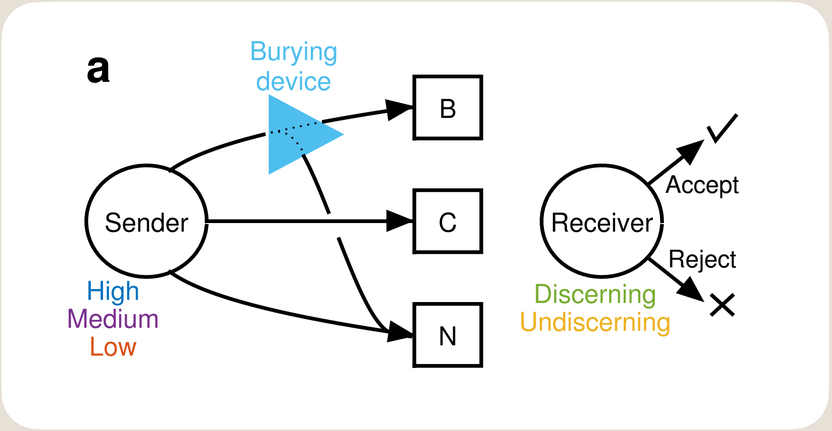


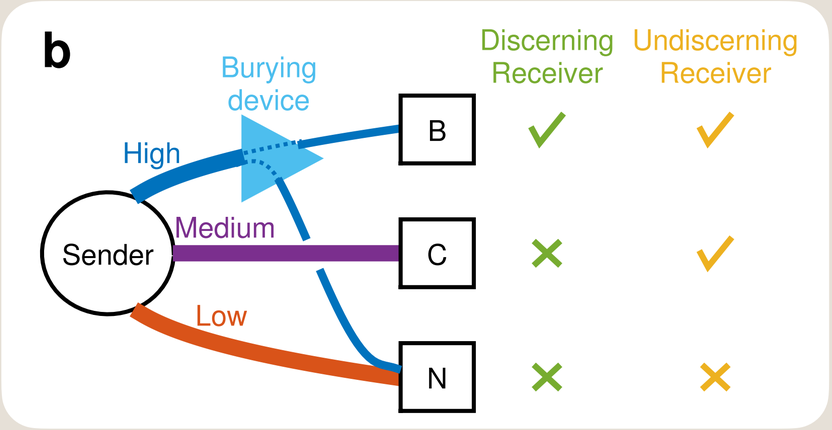


Evolution of social norms: Humility

Remark 3.12. The signal-burying game

- Consider an asymmetric game between two players, a sender and a receiver.
- There are three possible types of senders, $i \in$ {high, medium, low}, and there are two types of receivers $j \in$ {discerning, undiscerning}
- Senders can choose whether or not to pay a cost c>0 for a good signal. If they do, they can decide whether to send a clear signal, or whether to bury it.
- Clear signals become publicly known. Buried signals only become known with some probability r_θ. But if they become known, receivers also learn that the sender buried the signal.
- Depending on the signal they observe, receivers choose whether or not to accept that sender. If accepted, the payoff is a_{ij} for the sender, and b_{ij} for the receiver.
- Is there an equilibrium in which some senders bury?





Buried signals equilibrium

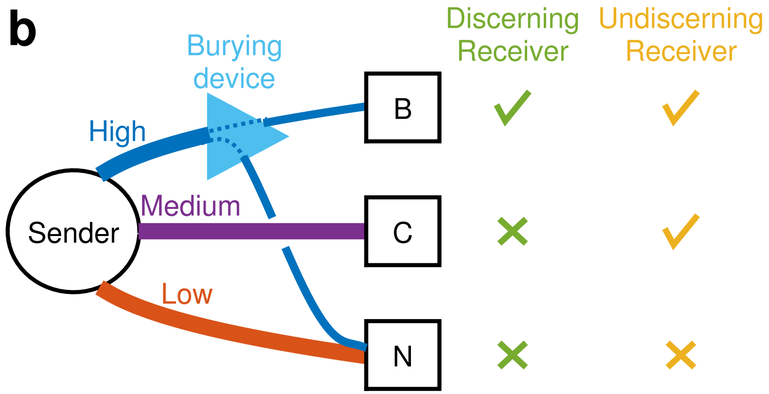
Evolution of social norms: Humility

Remark 3.13. Interpretation

- Why make anonymous donations?
  Makes a lot of sense if you generally do not care about the opinions of the general public, but you do care about the opinion of close family members, who are likely to learn about your anonymous donation anyways.

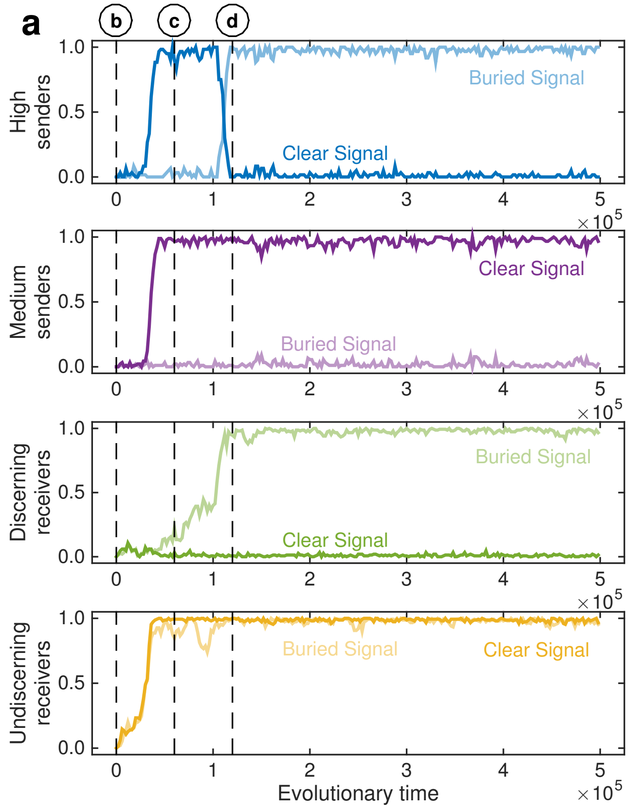
- Why being subtle in your fashion choices?
  Wearing a brand handbag may signal wealth, but it also signals that you really want everyone to know it, instead of only those people who are sophisticated enough to know the subtle signals of expense.

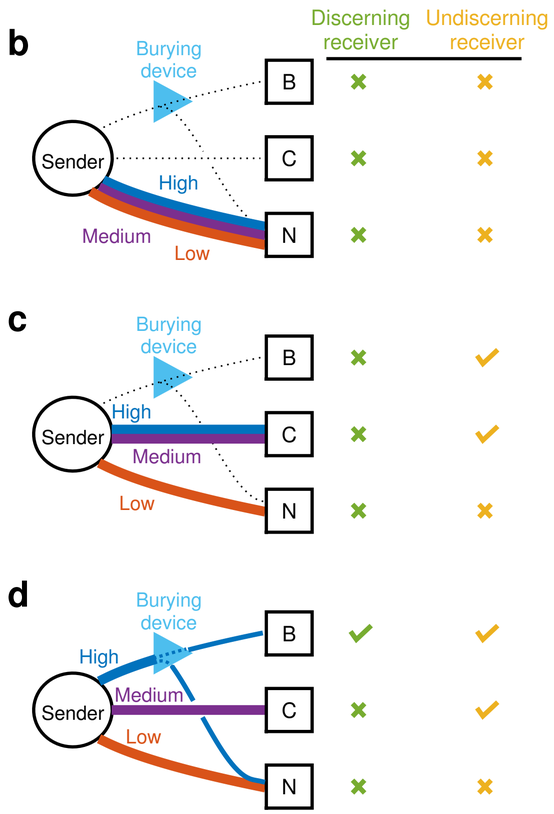
- Burying such signals is a great way of showing that you are only interested in a particular group of receivers.



Buried signals equilibrium

Evolution of social norms: Humility





Evolution of the Buried signals equilibrium

Summary

Some things you should have learned today:

1. Models of indirect reciprocity can explain why cooperation might be the result of social norms. They lead to interesting reputation dynamics!

2. However, when reputations are noisy, and information is imperfect, cooperation can be surprisingly fragile. There are models out there that use this observation to explain the evolution of empathy, and gossip, etc.

3. Social norms can also be rather nuanced; as an example, we discussed why people might value modesty. Also such norms can be explained with game theory (in this case: with a signaling model).