Reinforcement Learning Explains Conditional Cooperation and Its Moody Cousin

PloS Computational Biology 12, e1005034 (2016).

ERATO感謝祭season III 2016/08/09

江崎貴裕

共同研究者：堀田結孝・竹澤正哲・増田直紀

社会的ジレンマ

タダ乗り・抜け駆けなどの非協力的な行動が可能（合理的）なため
お互いに協力することが難しい状況

e.g. 軍備競争，投票行動，環境問題，　・・・

人間社会では、“意外に”協力が維持されている

なぜ協力が維持されうるのか

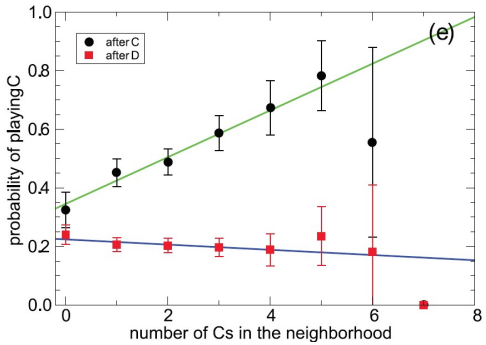
進化ゲーム理論的解釈の枠組み

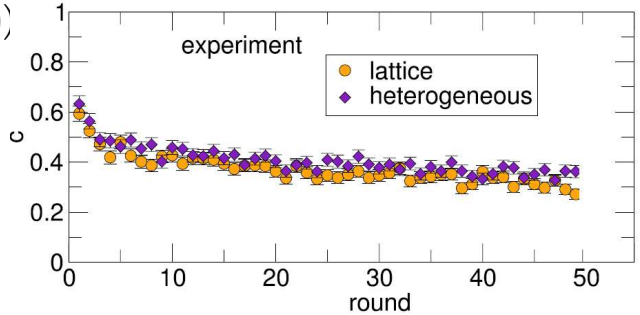
他人がやっているうまくいく戦略が分かると、それを真似する
淘汰の末、生き残った戦略が「協力的な戦略」だから協力がみられる

→互恵性が重要…
ネットワーク構造が重要…

多人数での行動実験

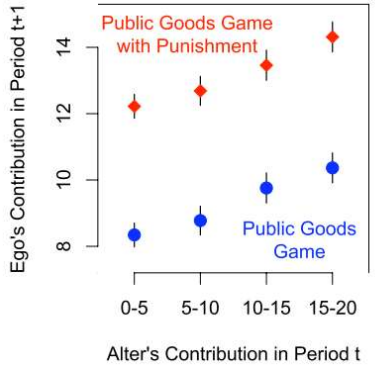
Grujic *et al.* PLOS ONE (2010)





Gracia-Lazaro *et al.* PNAS(2012)

Fowler *et al.* PNAS(2010)



なぜ協力が維持されうるのか

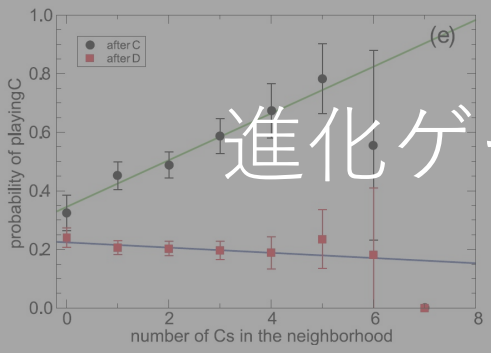
進化ゲーム理論的解釈の枠組み

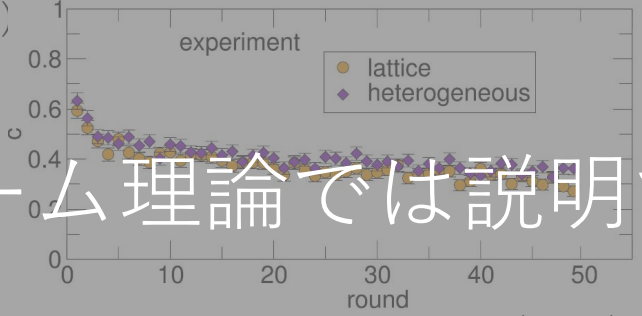
他人がやっているうまくいく戦略が分かると、それを真似する
淘汰の末、生き残った戦略が「協力的な戦略」だから協力がみられる

→互恵性が重要…
　ネットワーク構造が重要…

多人数での行動実験

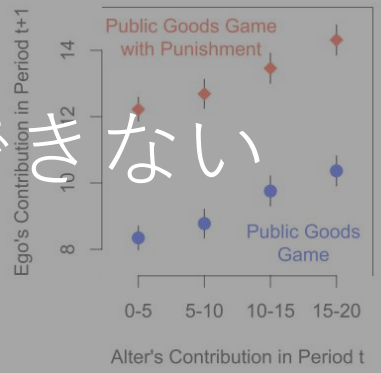
Grujic *et al.* PLOS ONE (2010)





Gracia-Lazaro *et al.* PNAS(2012)

Fowler *et al.* PNAS(2010)



進化ゲーム理論では説明できない

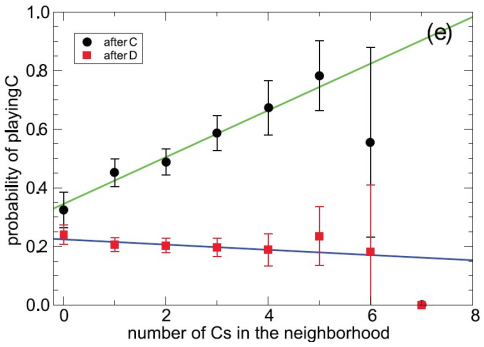
なぜ協力が維持されうるのか

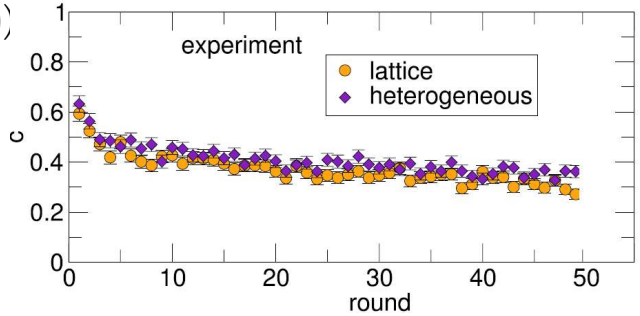
強化学習理論的解釈の枠組み

自分の行動の結果、うまくいったらその行動を強化する
うまくいかなかったら別の行動を試す。その結果として協力行動がみられる

多人数での行動実験

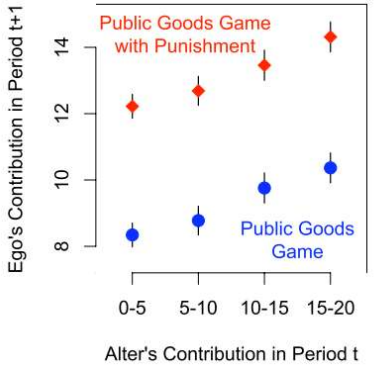
Grujic *et al.* PLOS ONE (2010)





Gracia-Lazaro *et al.* PNAS(2012)

Fowler *et al.* PNAS(2010)



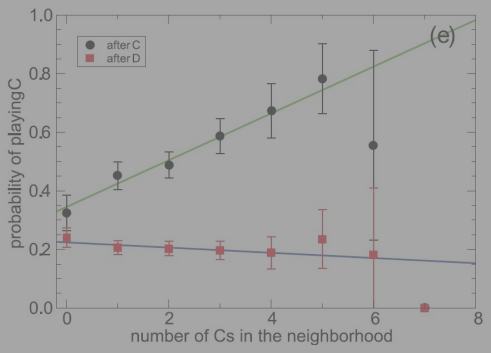
なぜ協力が維持されうるのか

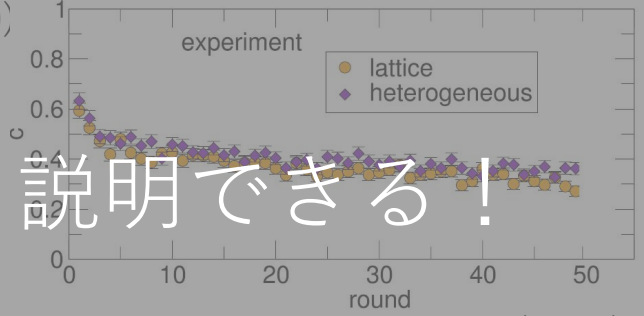
強化学習理論的解釈の枠組み

自分の行動の結果、うまくいったらその行動を強化する
うまくいかなかったら別の行動を試す。その結果として協力行動がみられる

多人数での行動実験

Grujic *et al.* PLOS ONE (2010)

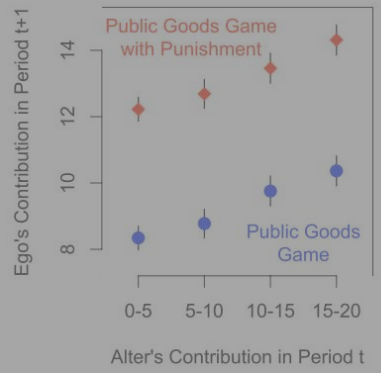




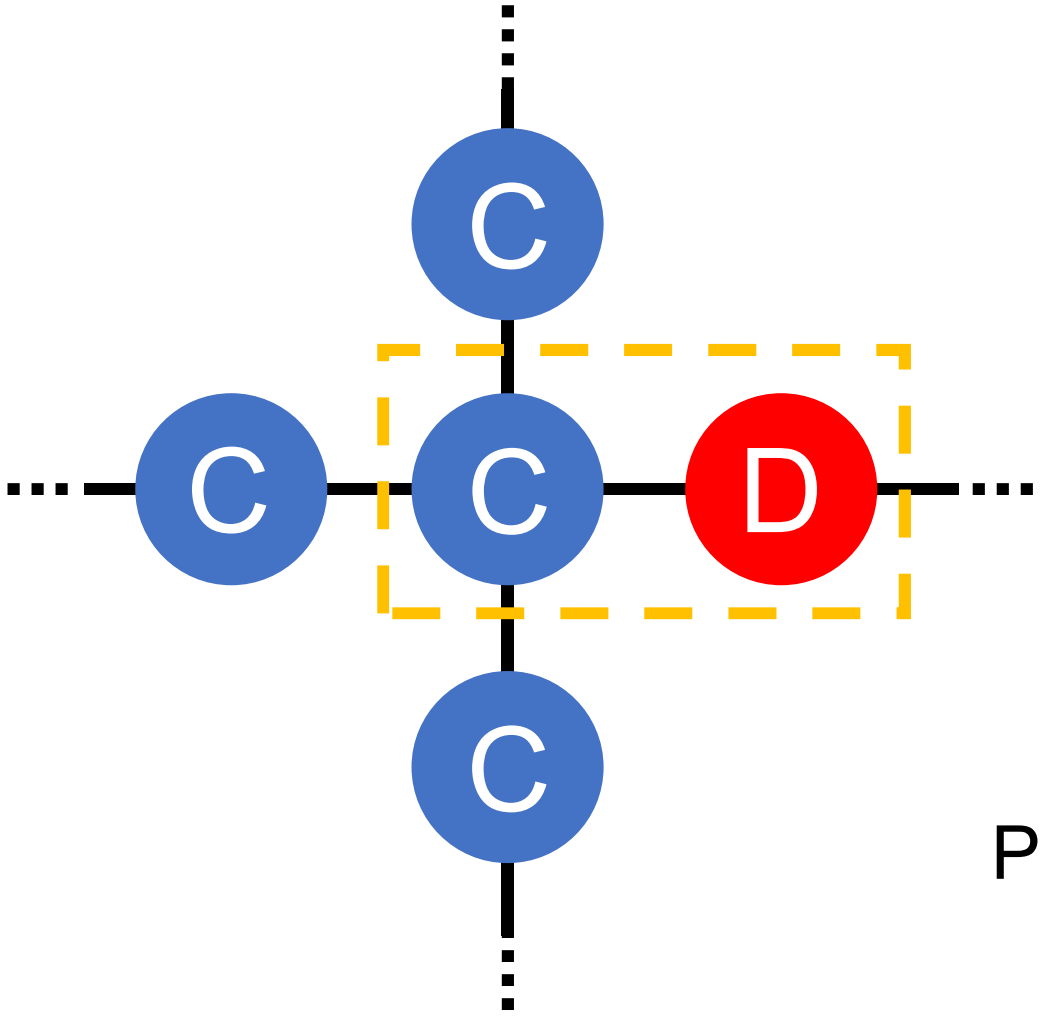
Gracia-Lazaro *et al.* PNAS(2012)

説明できる！

Fowler *et al.* PNAS(2010)



Prisoner's Dilemma Game on graphs



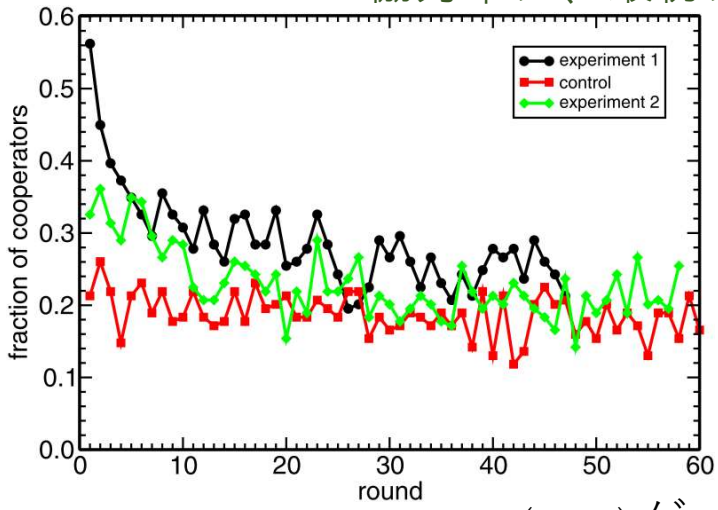
自分＼相手	C	D
C	3	0
D	5	1

$$\text{Payoff } r_t = \frac{3+3+3+0}{4}$$

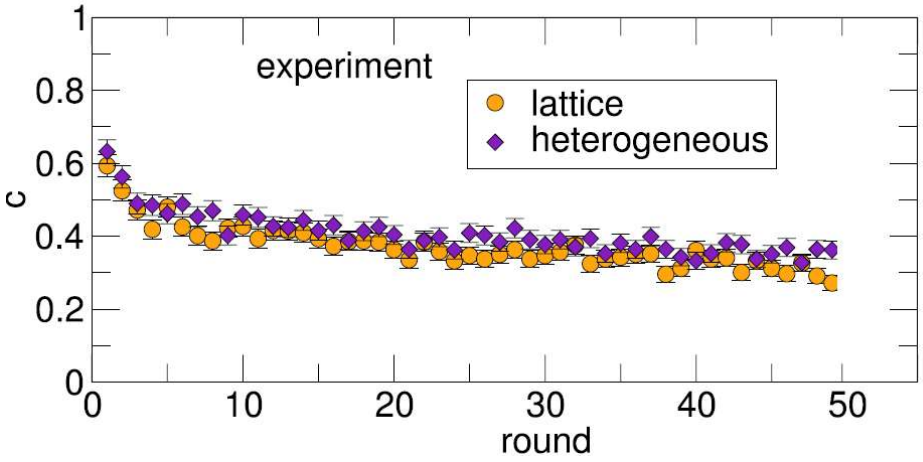
Experimental Findings

協力率は、最初は高くてだんだん下がってくる

協力率



Grujic *et al.* PLOS ONE (2010)



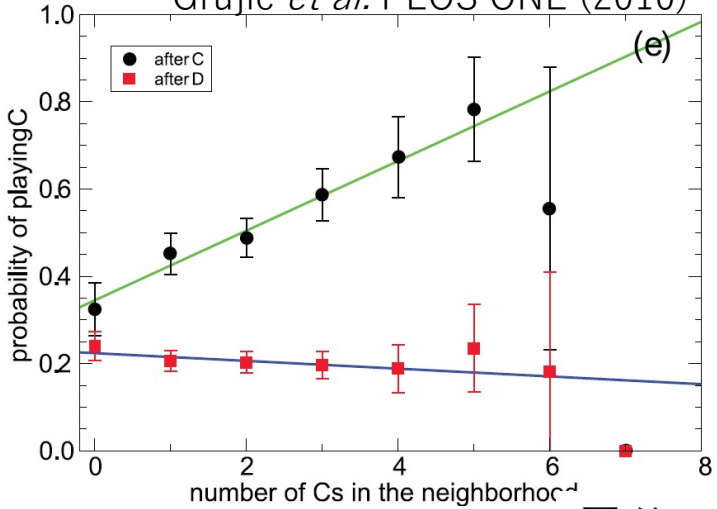
Gracia-Lazaro *et al.* PNAS(2012)

ゲームの回数

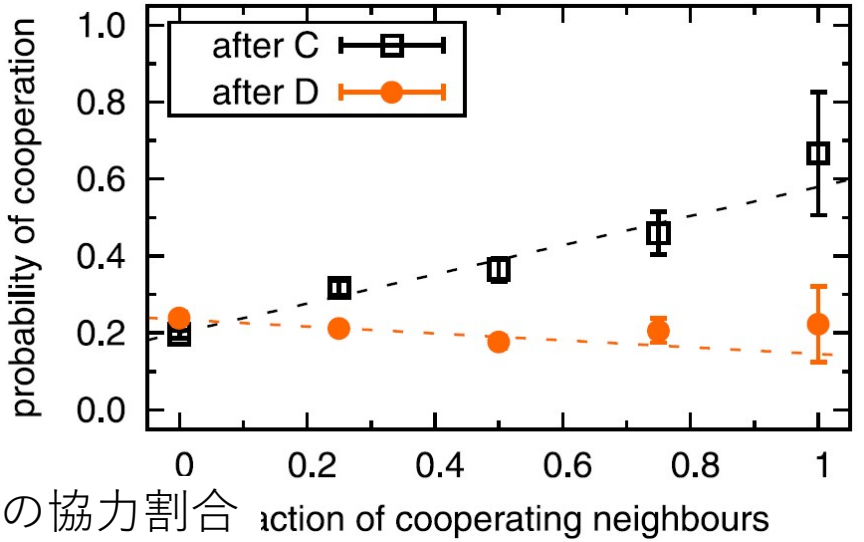
周りの人の協力割合に明らかに依存している

次回の協力率

Grujic *et al.* PLOS ONE (2010)



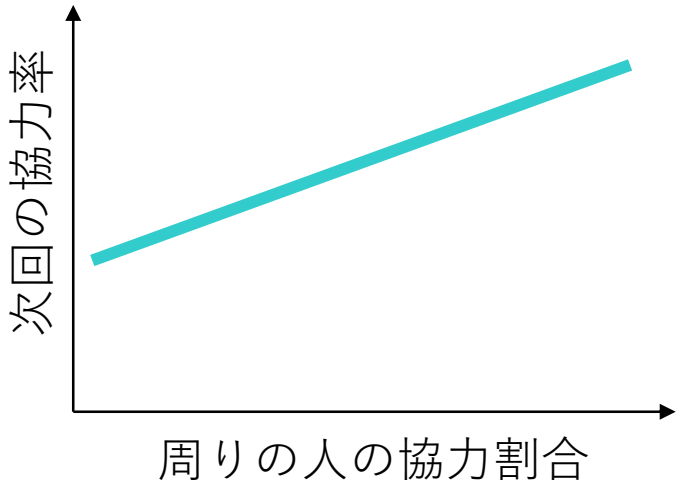
Grujic *et al.* Sci Rep (2014)



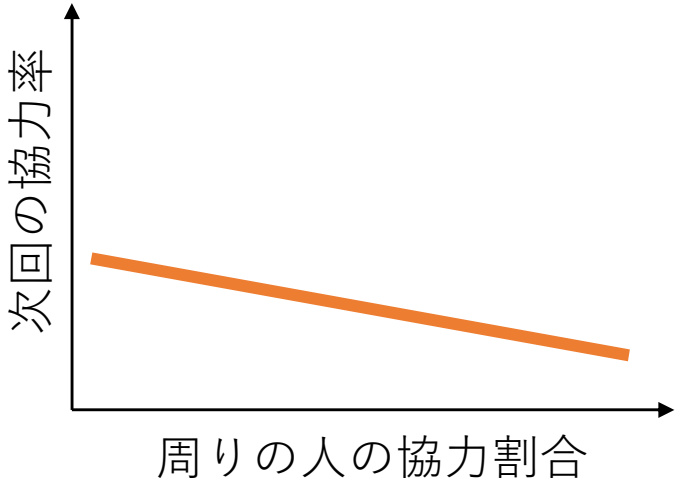
周りの人の協力割合

Experimental Findings

自分が協力した後

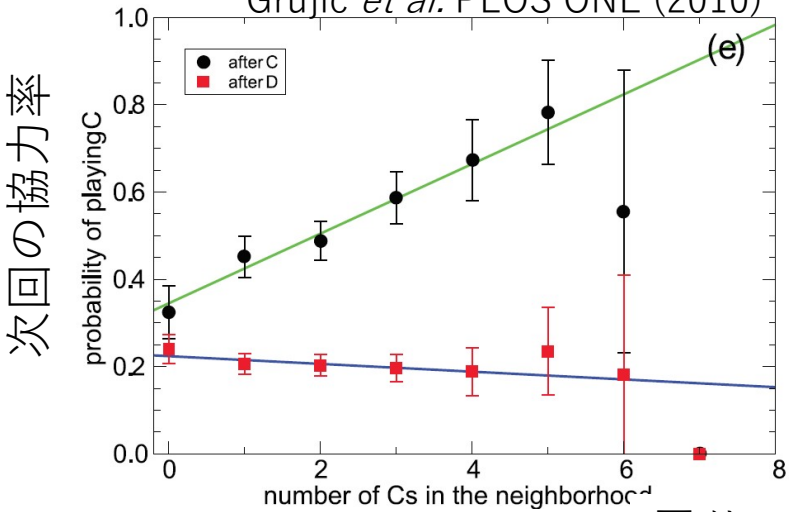


自分が裏切った後

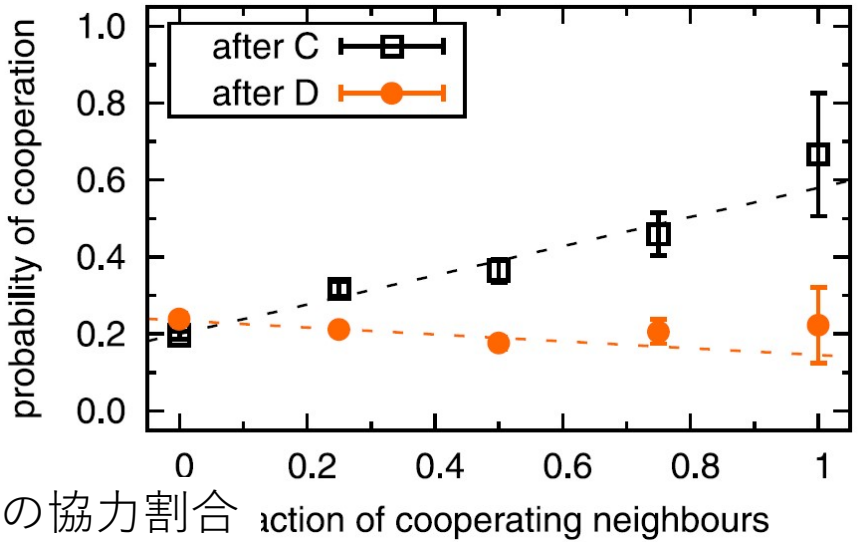


周りの人の協力割合に明らかに依存している

Grujic *et al.* PLOS ONE (2010)

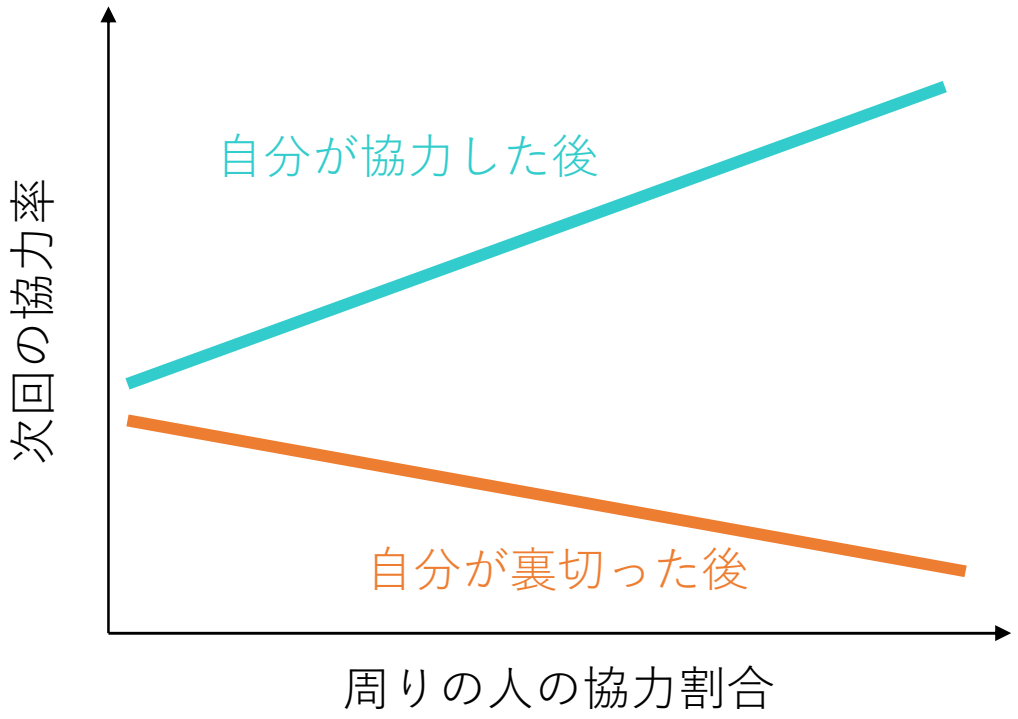


Grujic *et al.* Sci Rep (2014)



Experimental Findings

Moody Conditional Cooperation (MCC)



なぜ人間がこのような行動をとるのか、ちゃんとした説明がない

×進化ゲーム理論

Reinforcement Learning (BM model)

前回行動

次回協力確率

$$p_t = \begin{cases} p_{t-1} + (1 - p_{t-1})s_{t-1} & (a_{t-1} = \mathrm{C}, s_{t-1} \geq 0) \\ p_{t-1} + p_{t-1}s_{t-1} & (a_{t-1} = \mathrm{C}, s_{t-1} < 0) \\ p_{t-1} - p_{t-1}s_{t-1} & (a_{t-1} = \mathrm{D}, s_{t-1} \geq 0) \\ p_{t-1} - (1 - p_{t-1})s_{t-1} & (a_{t-1} = \mathrm{D}, s_{t-1} < 0) \end{cases}$$

満足	協力増やす	
不満	協力減らす	
満足	裏切り増やす	
不満	裏切り減らす	

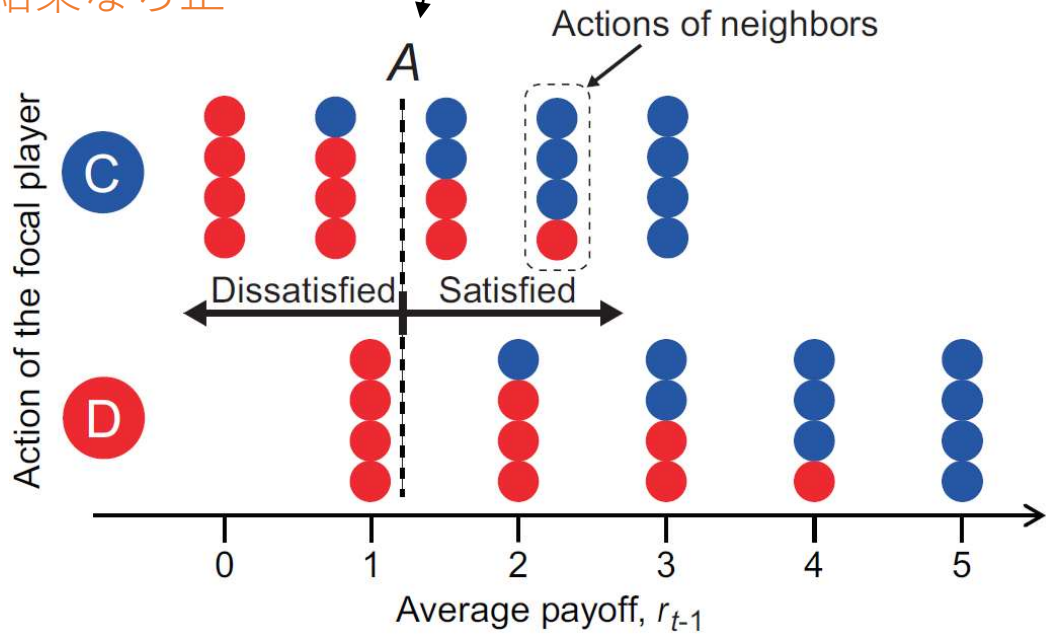
$$s_{t-1} = \tanh\left[\beta(r_{t-1} - A)\right]$$

満足な結果なら正

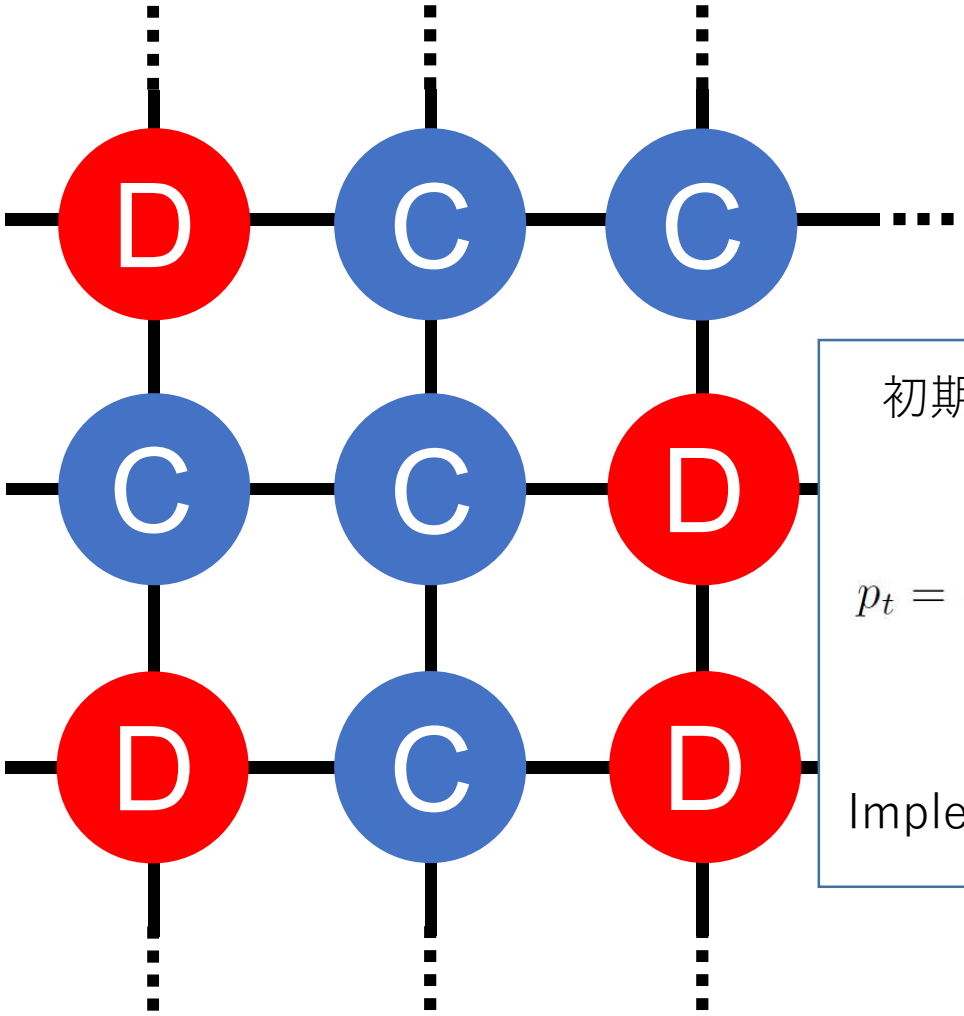
期待するpayoff

自分が協力した場合

自分が裏切った場合



Numerical Simulations



10 × 10 square lattice

自分＼相手	C	D
C	3	0
D	5	1

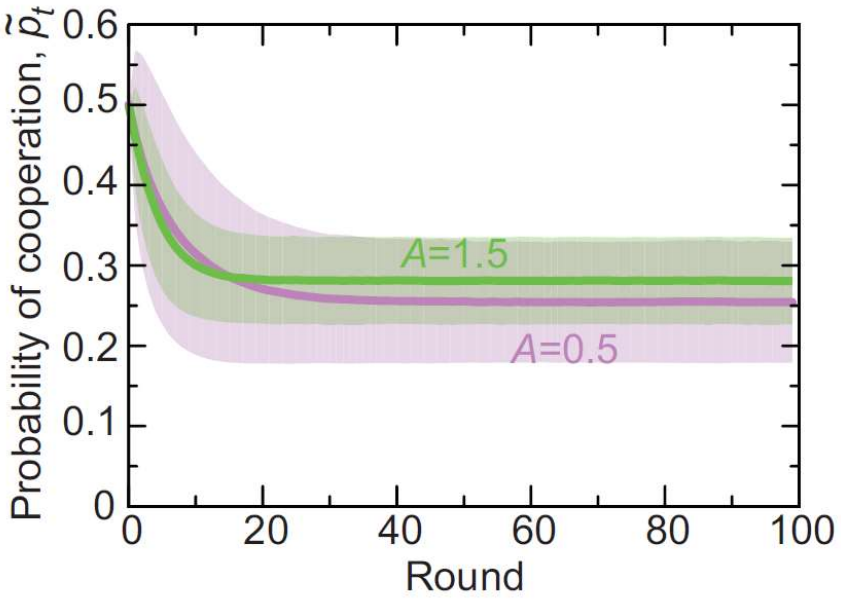
初期条件： $p_0 = 0.5$

$$p_t = \begin{cases} p_{t-1} + (1 - p_{t-1})s_{t-1} & (a_{t-1} = \mathrm{C}, s_{t-1} \geq 0) \\ p_{t-1} + p_{t-1}s_{t-1} & (a_{t-1} = \mathrm{C}, s_{t-1} < 0) \\ p_{t-1} - p_{t-1}s_{t-1} & (a_{t-1} = \mathrm{D}, s_{t-1} \geq 0) \\ p_{t-1} - (1 - p_{t-1})s_{t-1} & (a_{t-1} = \mathrm{D}, s_{t-1} < 0) \end{cases}$$

Implementation error： 確率 ε で逆の行動をする

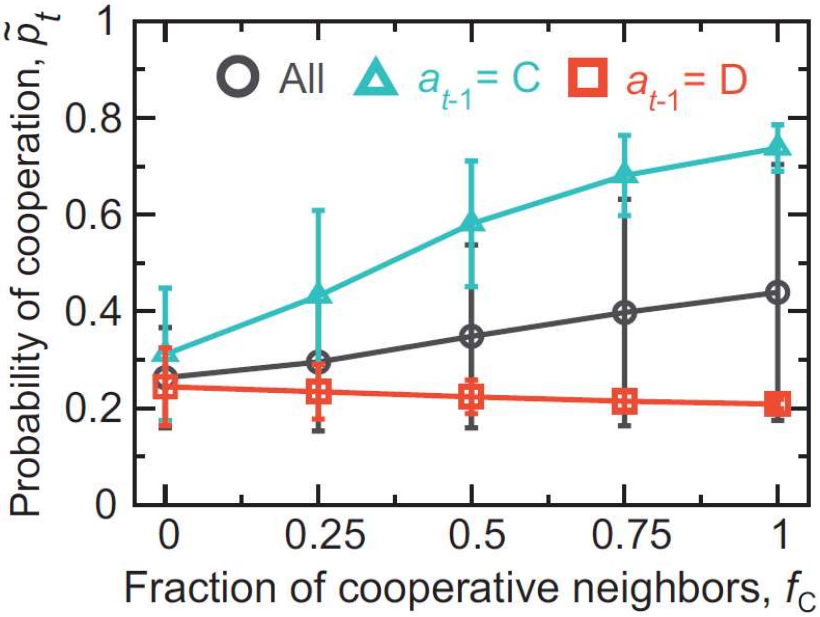
Results

協力率



✔実験結果と概ね整合的な振る舞い

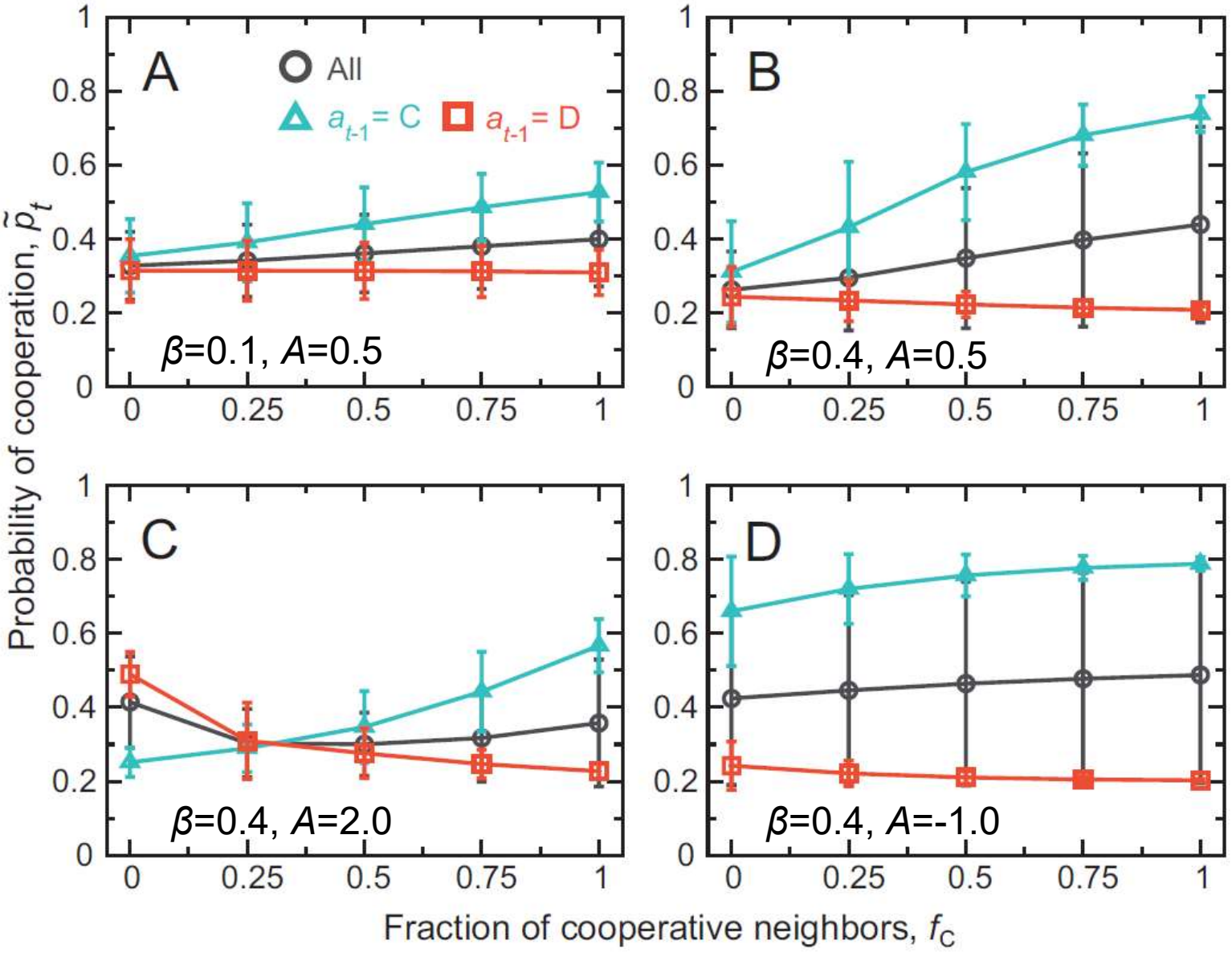
次回の協力率



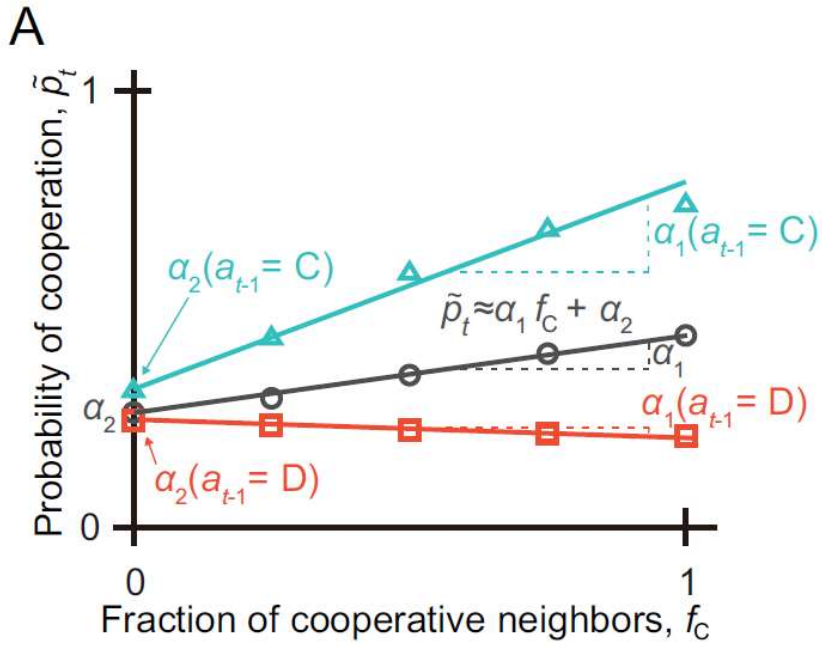
✔実験結果と概ね整合的な振る舞い
（β=0.4, A=0.5）

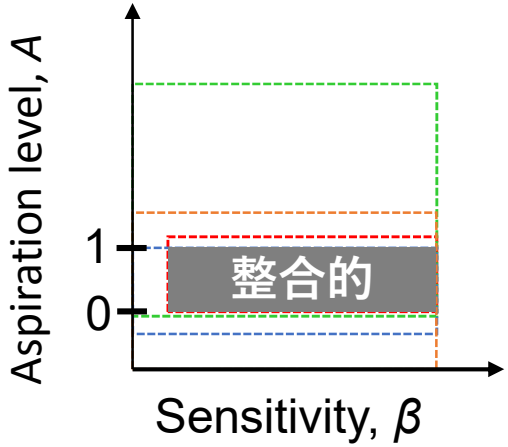
Results

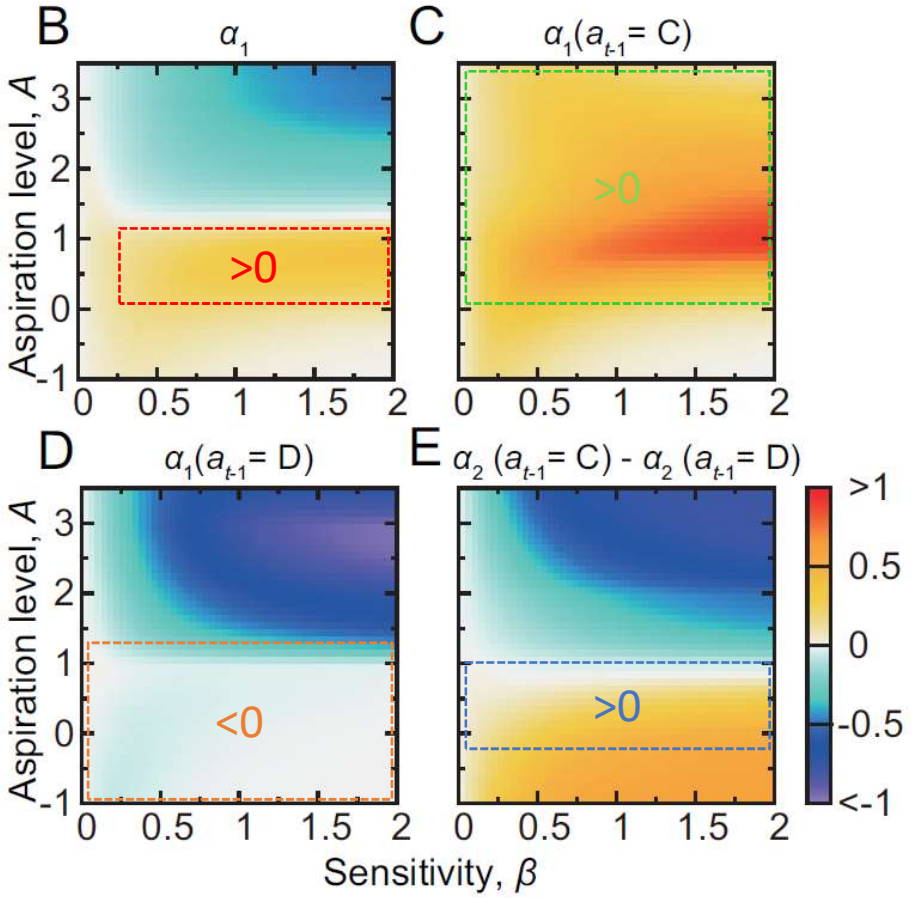
パラメータの値によっては実験結果と矛盾する



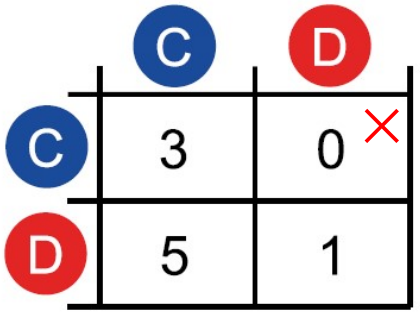
Results

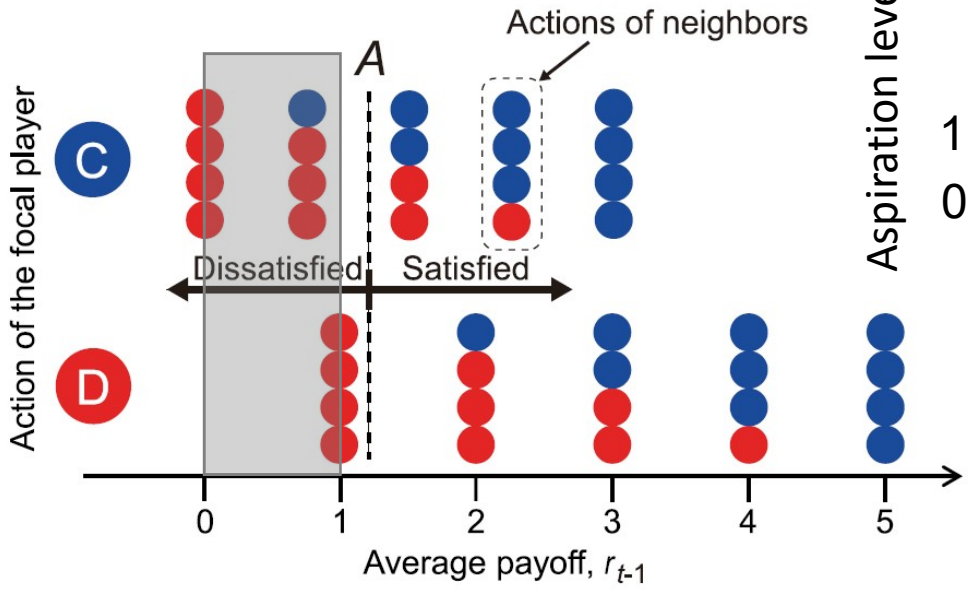

A
Probability of cooperation, $\tilde{p}_t$
1
0
$\alpha_2(a_{t-1}=\mathrm{C})$
$\alpha_1(a_{t-1}=\mathrm{C})$
$\tilde{p}_t \approx \alpha_1 f_C + \alpha_2$
α_1
α_2
$\alpha_1(a_{t-1}=\mathrm{D})$
$\alpha_2(a_{t-1}=\mathrm{D})$
0
1
Fraction of cooperative neighbors, f_C

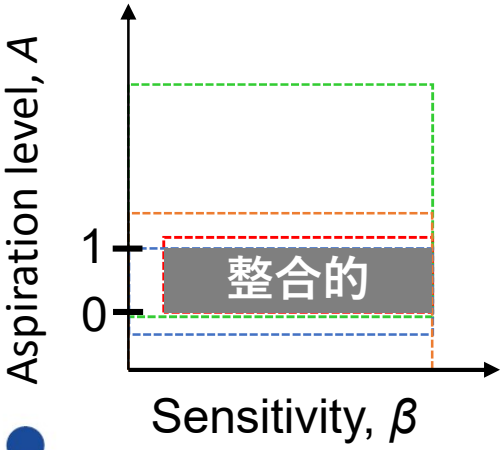

Aspiration level, A
1
0
整合的
Sensitivity, β

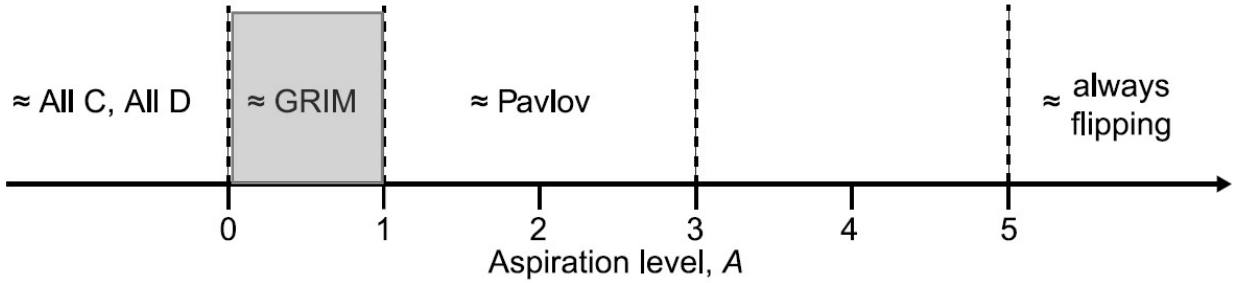

B
α_1
C
$\alpha_1(a_{t-1}=\mathrm{C})$
>0
>0
D
$\alpha_1(a_{t-1}=\mathrm{D})$
E
$\alpha_2(a_{t-1}=\mathrm{C}) - \alpha_2(a_{t-1}=\mathrm{D})$
<0
>0
Aspiration level, A
3
2
1
0
-1
0
0.5
1
1.5
2
>1
0.5
0
-0.5
<-1
Sensitivity, β

Interpretations





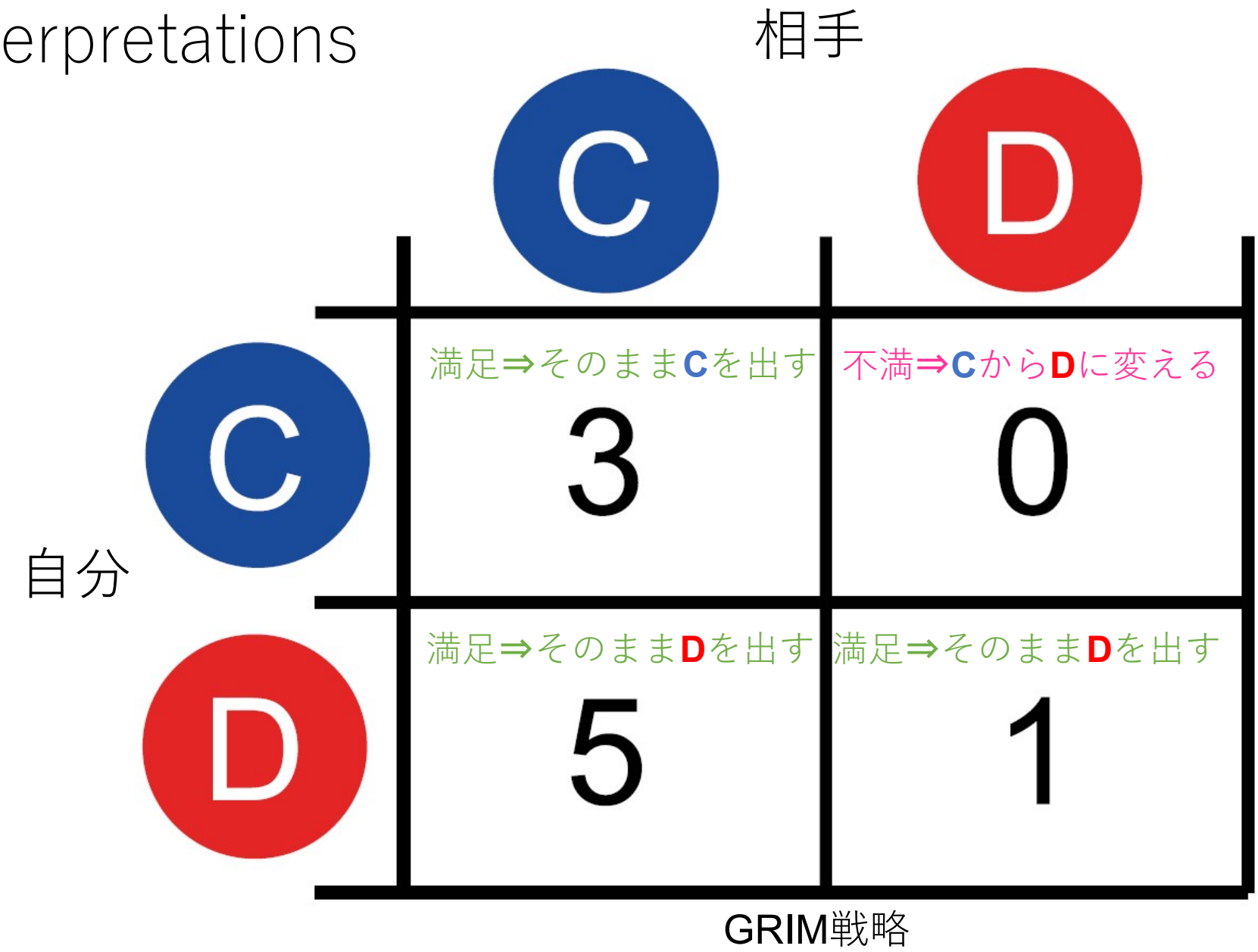




$\beta \to +\infty$、二人ゲームに対応する古典的な戦略

前回の結果に満足したら確実に同じ手を出す
不満だったら、確実に逆の手を出す

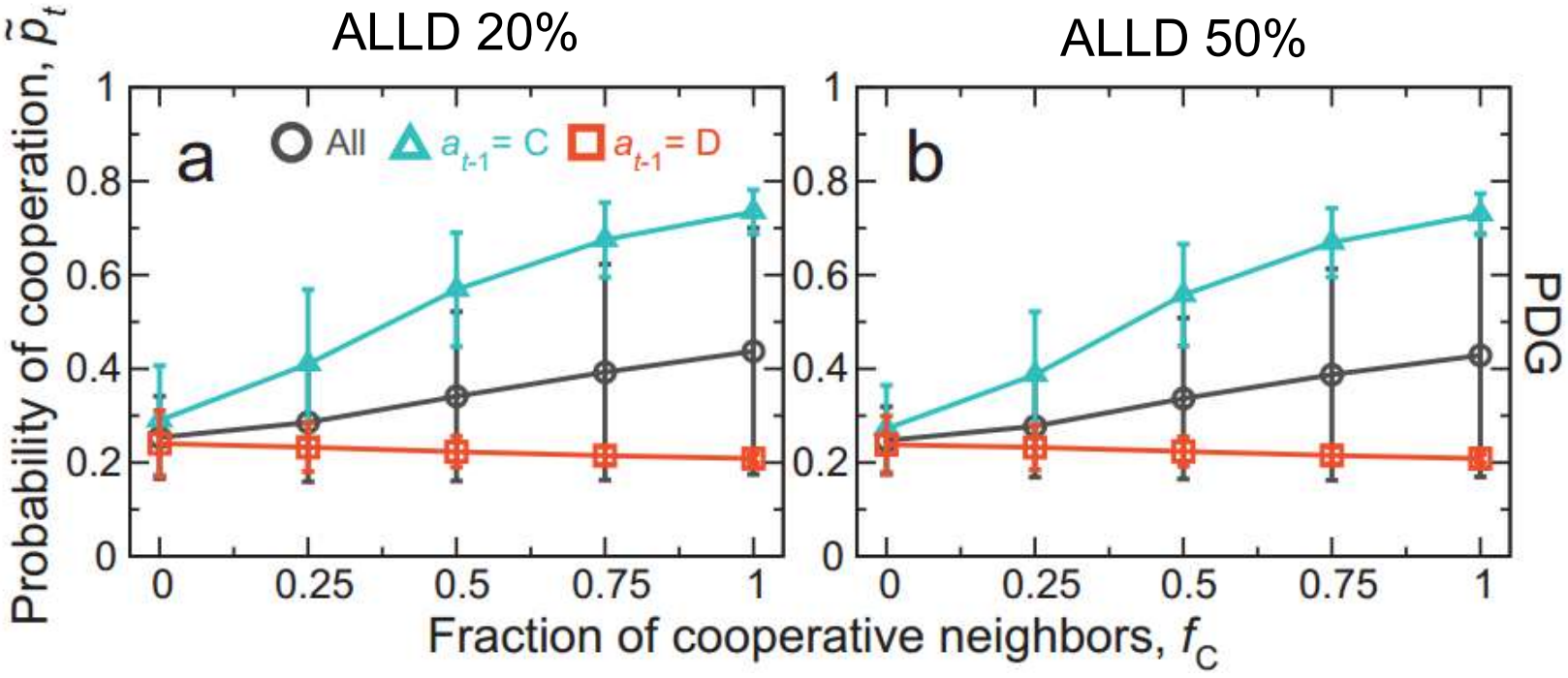
Interpretations



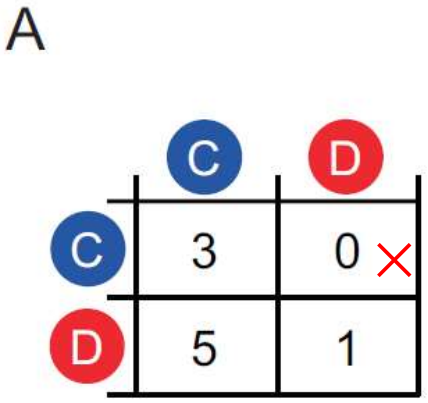
ALLD戦略（Dしか出してこない人）に対して強い戦略として知られる

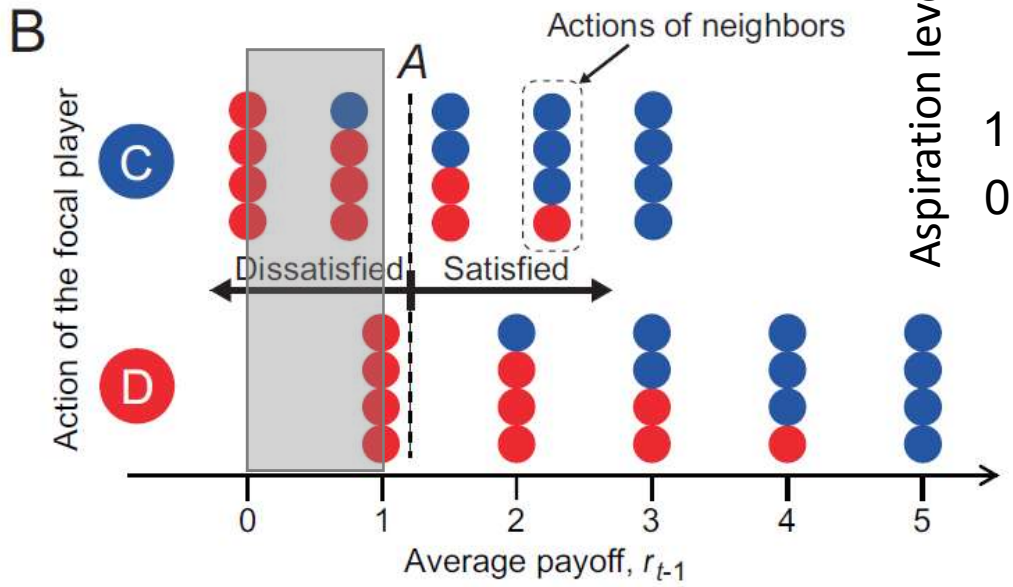
Robustness test

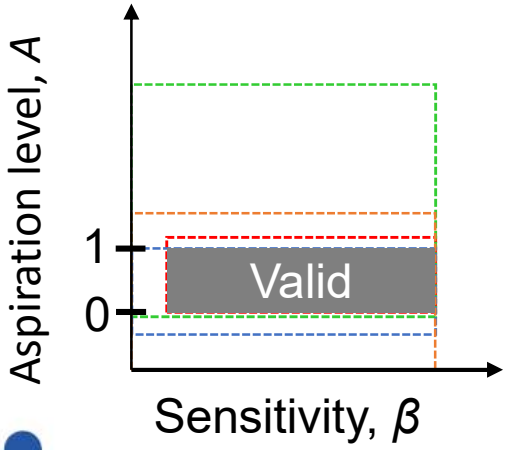
ALLD（裏切りしかしない人）が一定数いても振る舞いは変わらない
（＝実験結果と整合的）



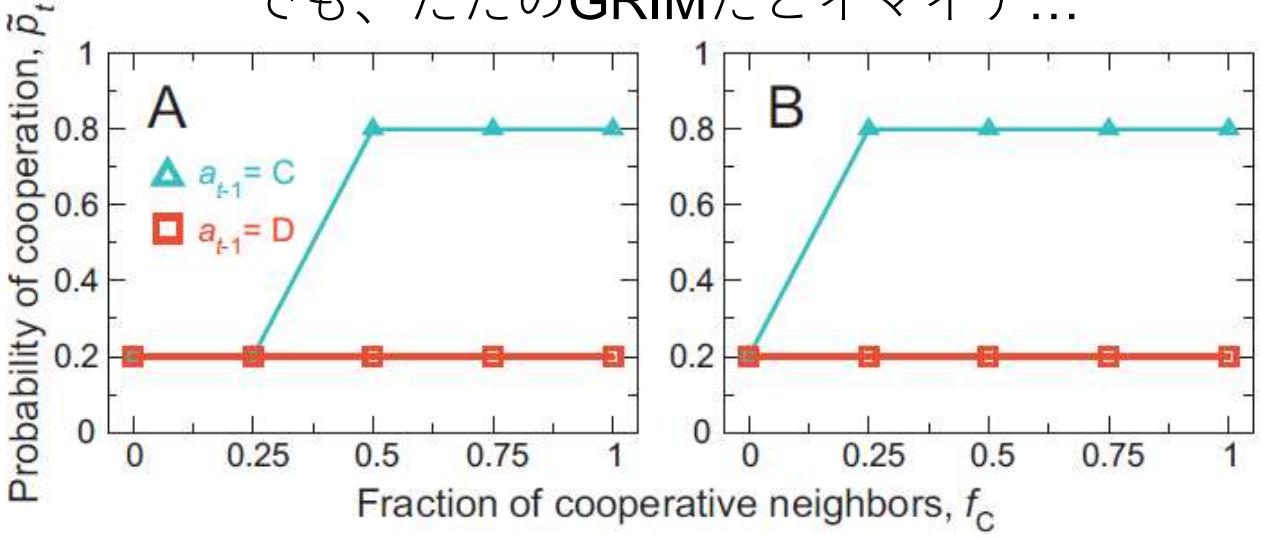
Interpretations



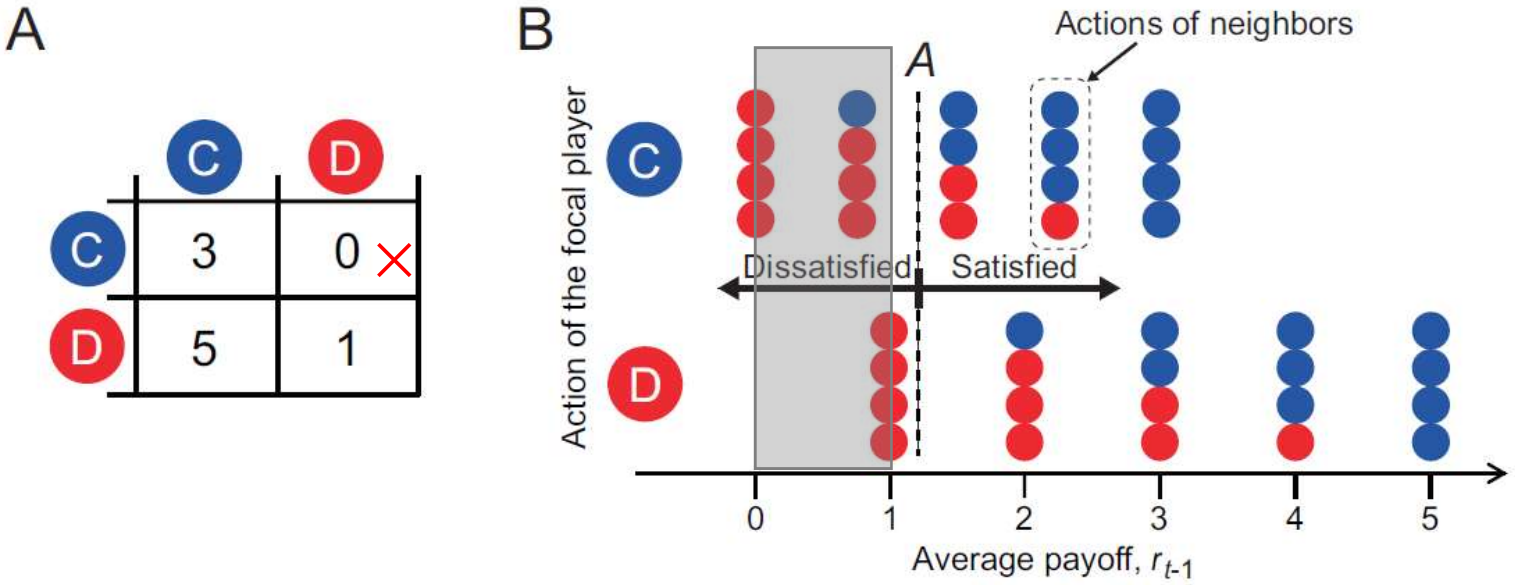




でも、ただのGRIMだとイマイチ…



Summary



・実験で見られる人間の振る舞い(CC&MCC)を強化学習で説明できた
→進化ゲーム理論の限界（？）
　他の状況についても強化学習のほうがよく記述できるかもしれない

・そのときのパラメータの値はGRIM的な振る舞いに対応する
←GRIMは裏切りしかしてこない相手(=free rider)に対して強い戦略

ご清聴ありがとうございました