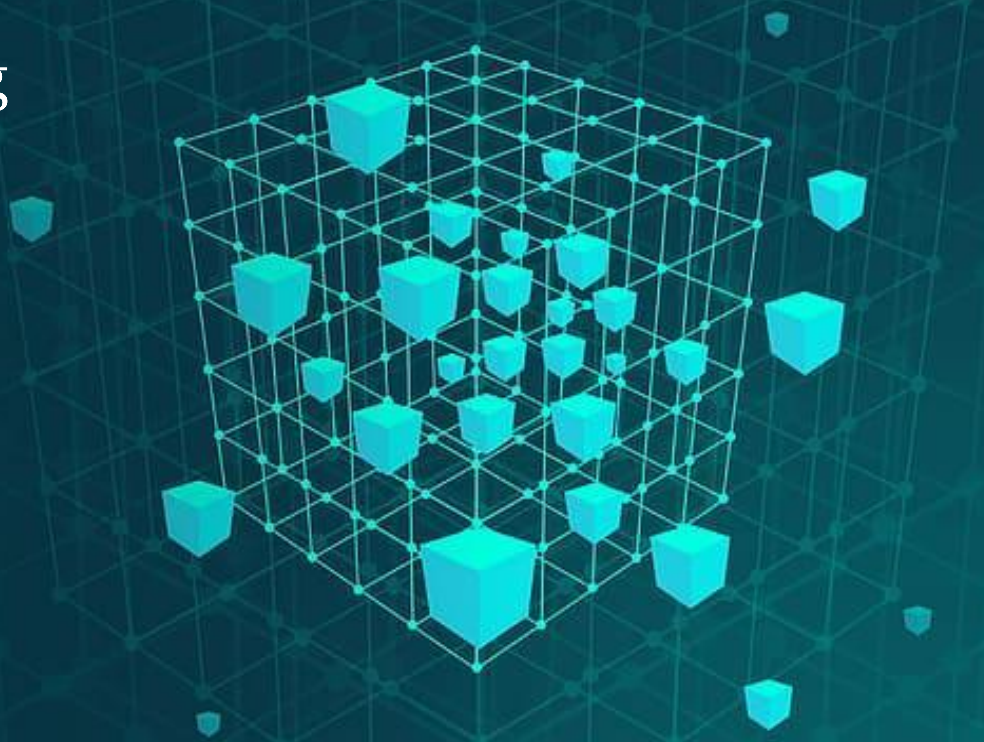
Reinforcement Learning Explains Conditional Cooperation and Its Moody Cousin

LENAERTS Tom
2025-2026



NATHAN François	(ULB)	514241
Mandel Eliot	(ULB)	493519
YILDIRIM Emirhan	(ULB)	459800

Table Of Content

I. Introduction
II. Environments
  II.i PDG
  II.ii PGG
III. Simulations
  III.i PDG
  III.ii PGG
IV. Additional Improvements
  IV.i Dynamical Aspiration
  IV.ii Free Riders
V. Conclusion

I. Introduction

EGT
(Evolutionary Game Theory)

RL
(Reinforcement Learning)

Conditional Cooperation (CC)

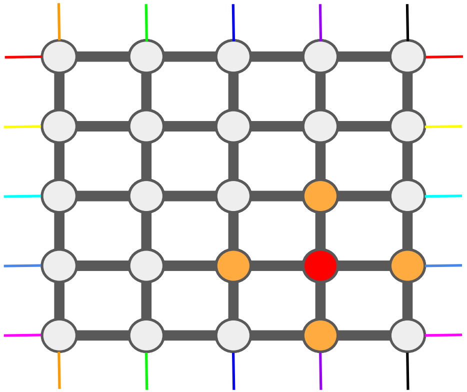
neigh C ⇒ agent C

Moody CC (MCC)

neigh D ⇒ independant agent

II.i PDG Environment

Prisoner Dilemma Game



neighbouring

	C	D
C	3	0
D	5	1

Payoff Matrix

1) Bush-Mosteller

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

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

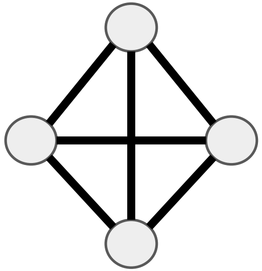
2) Misimplement the decision (ε)

$$\tilde{p}_t \equiv p_t(1 - \epsilon) + (1 - p_t)\epsilon$$

3) Payoff Reward

II.ii PGG Environment

<u>Public Good Game</u>



neighbouring

1) Bush-Mosteller

$$p_t = \begin{cases} p_{t-1} + (1 - p_{t-1})s_{t-1} & (a_{t-1} \geq X \text{ and } s_{t-1} \geq 0), \\ p_{t-1} + p_{t-1}s_{t-1} & (a_{t-1} \geq X \text{ and } s_{t-1} < 0), \\ p_{t-1} - p_{t-1}s_{t-1} & (a_{t-1} < X \text{ and } s_{t-1} \geq 0), \\ p_{t-1} - (1 - p_{t-1})s_{t-1} & (a_{t-1} < X \text{ and } s_{t-1} < 0). \end{cases}$$

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

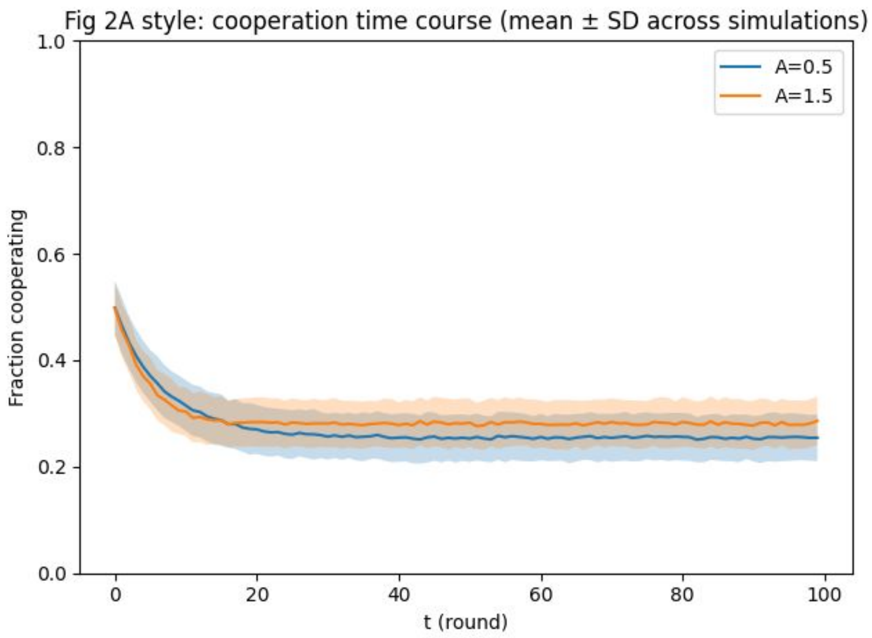
2) staked amount ε [0,1]

$a_t = \eta(p_t, 0.2)$

3) Reward

$r = 1 - a_t + 0.4(a_t + \sum_{j=1}^{3} \tilde{a}_{j,t})$

III.i Simulations PDG



A = 0.5 / 1.5

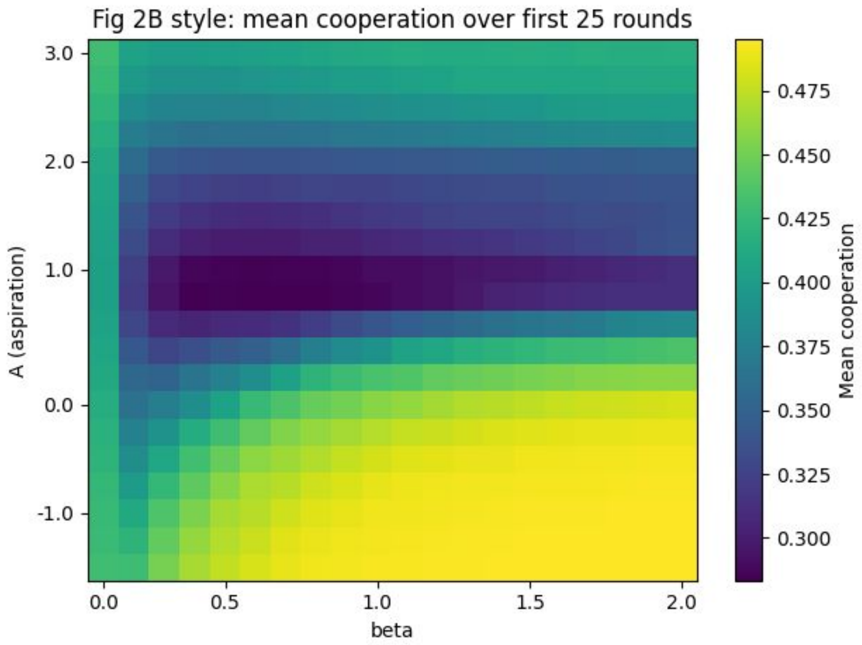
β = 0.2

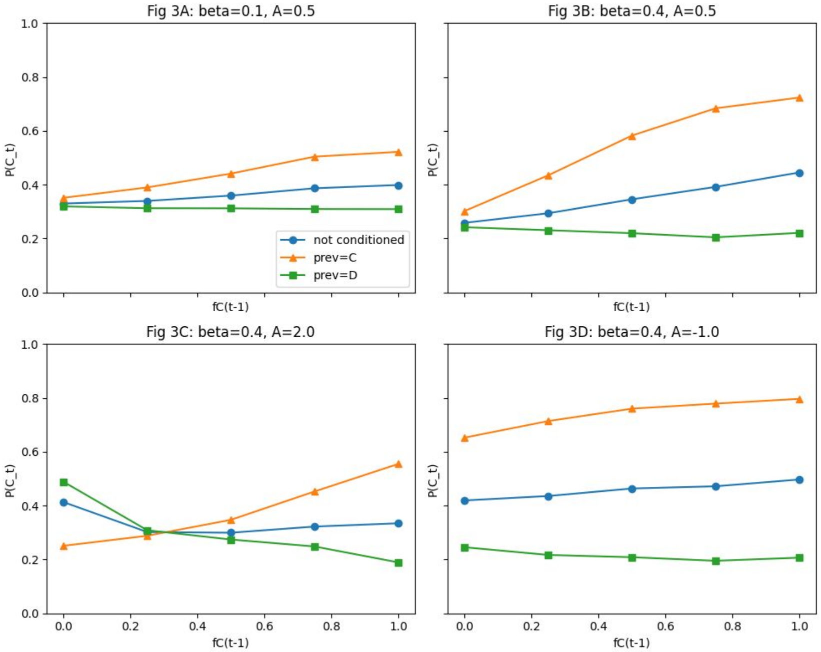
p_0 = 0.5

grid = 10 x 10

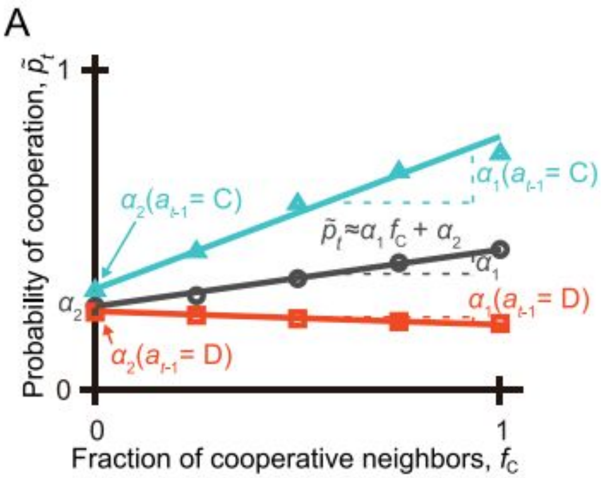
ε = 0.2

III.i Simulations PDG



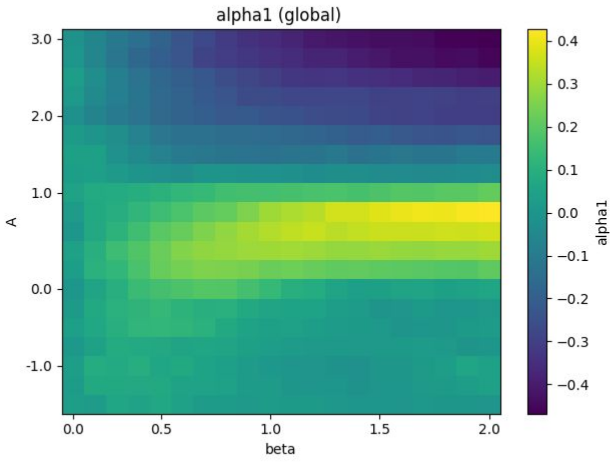


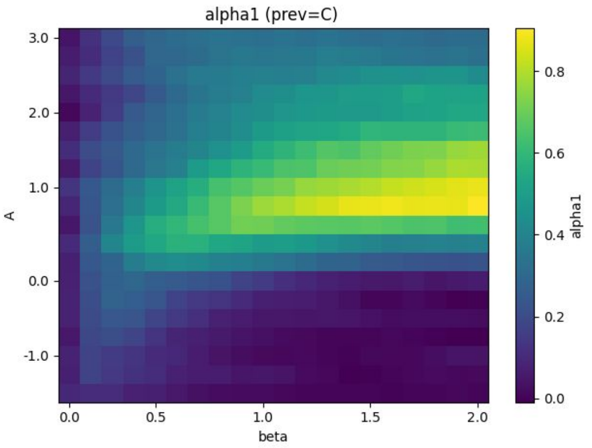
III.i Simulations PDG

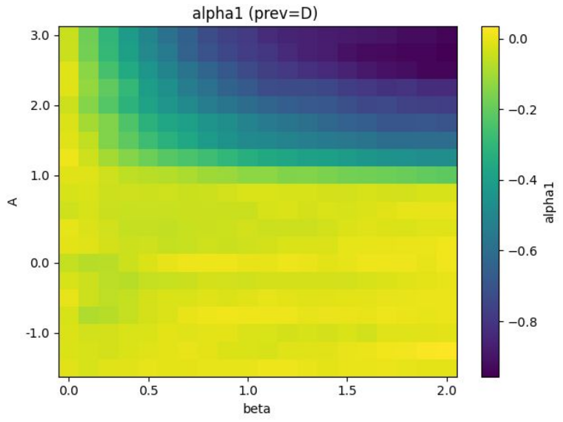


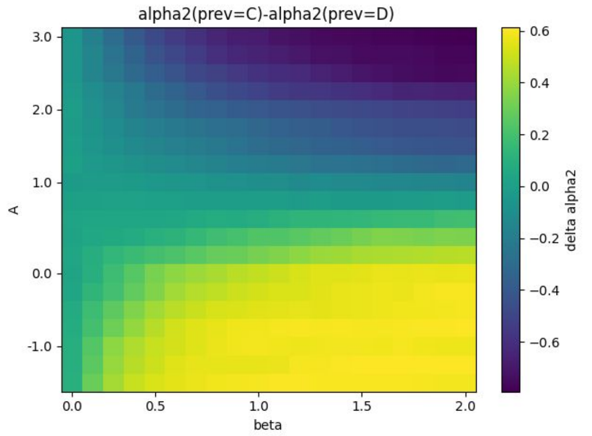
$A < \sim 1$

$\beta > 1$









III.ii Simulations PGG

A = 0.9

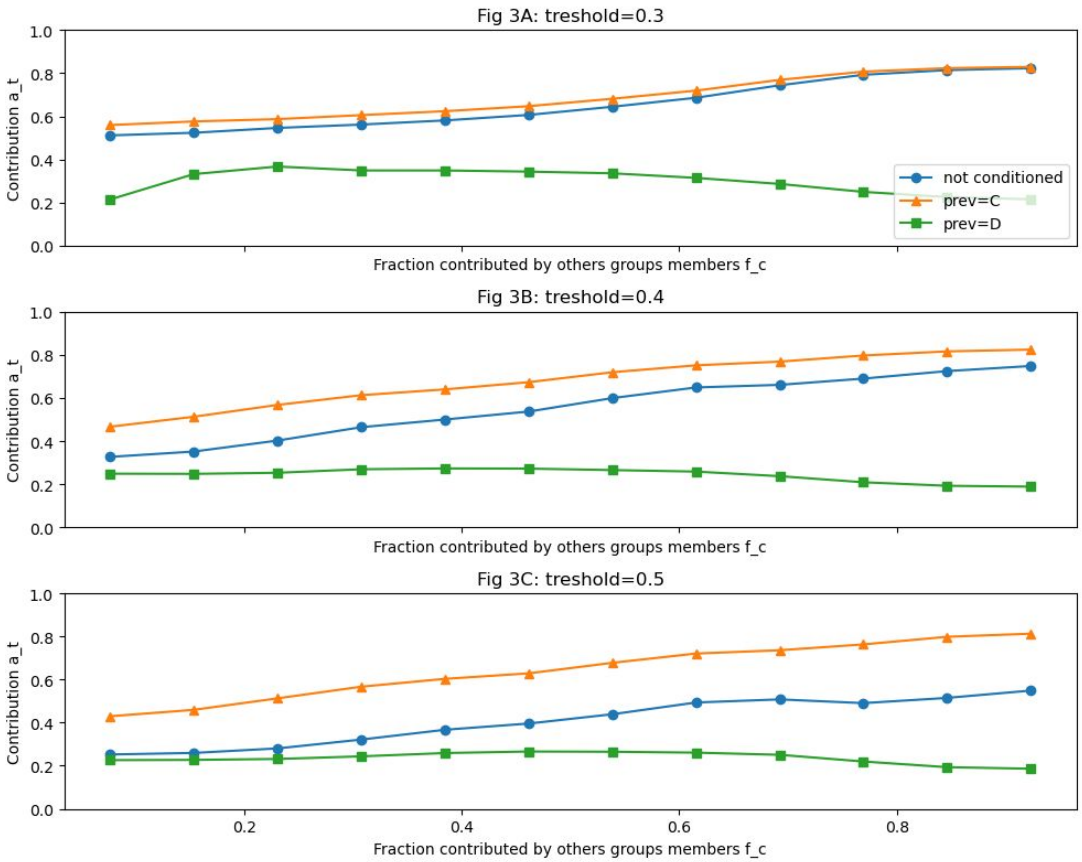
β = 0.4

p_0 = 0.5

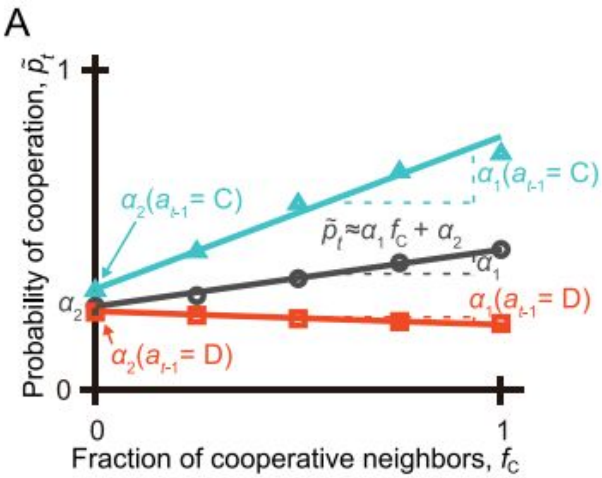
number player = 4

ε = 0.2

Treshold = 0.3 / 0.4 / 0.5

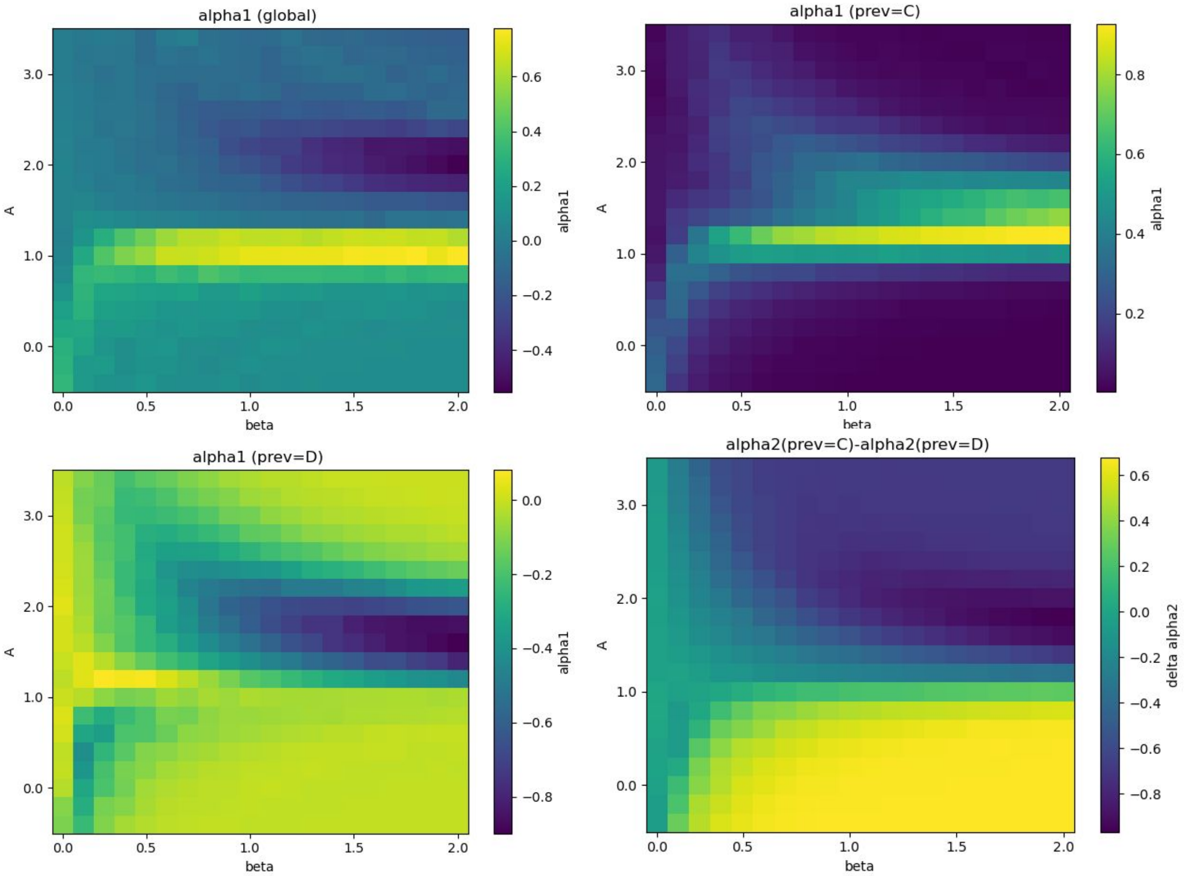


III.ii Simulations PGG



$0 < A < 1$

$\beta > 0.5$

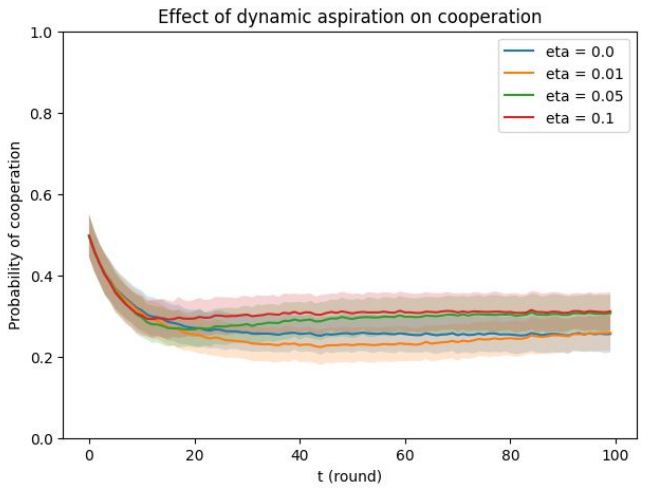


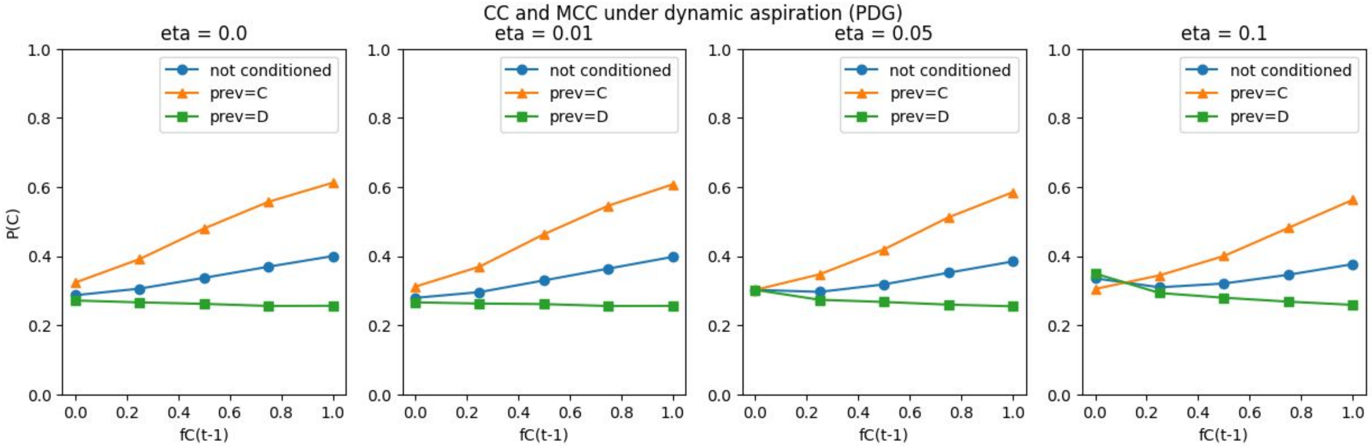
IV.i Dynamical Aspiration

PDG

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

$$A_{t+1} = (1-\eta)A_t + \eta r_t$$



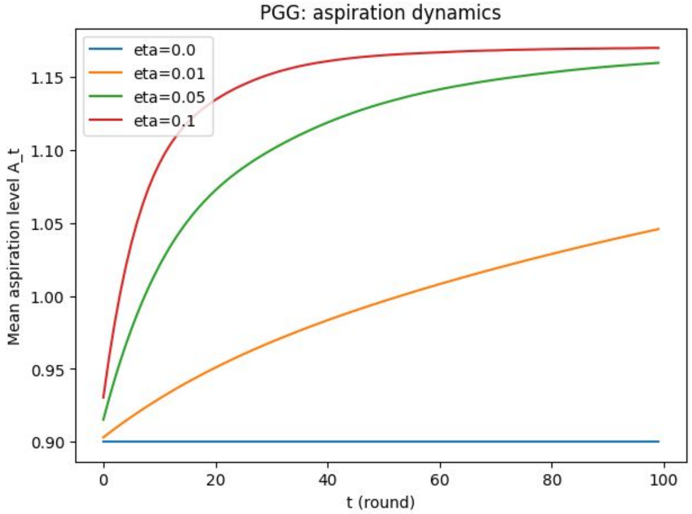


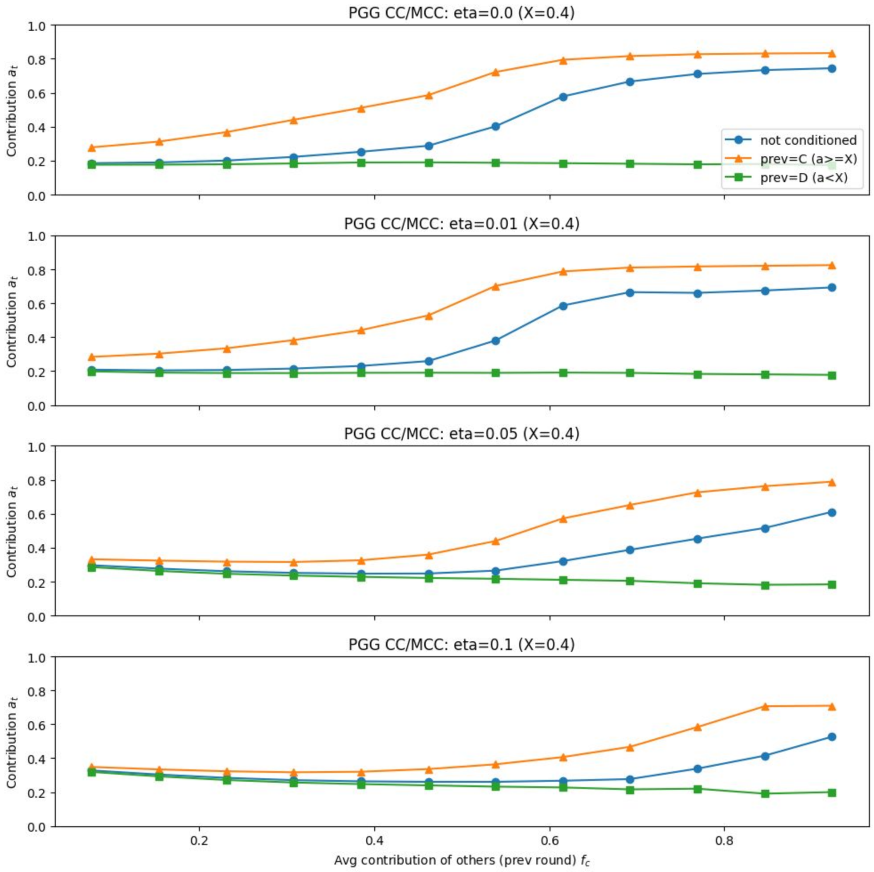
IV.i Dynamical Aspiration

PGG

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

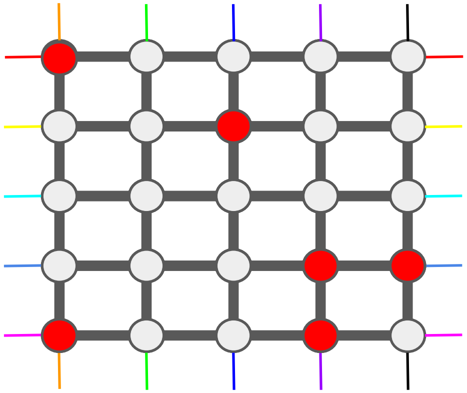
$$A_{t+1} = (1-\eta)A_t + \eta r_t$$



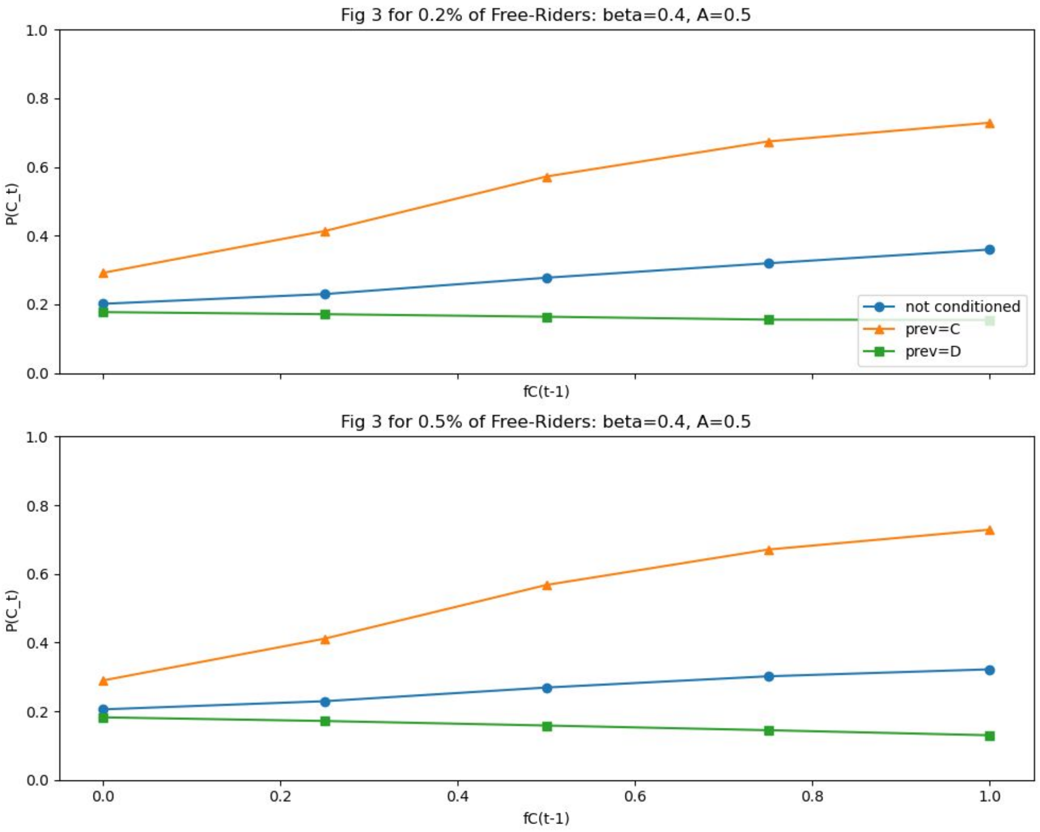


IV.ii Free Rider

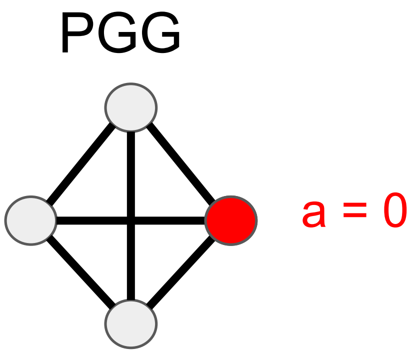
PDG

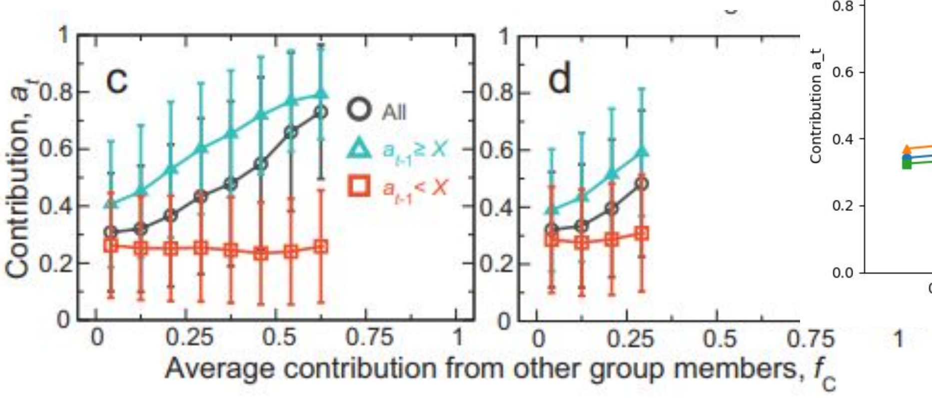


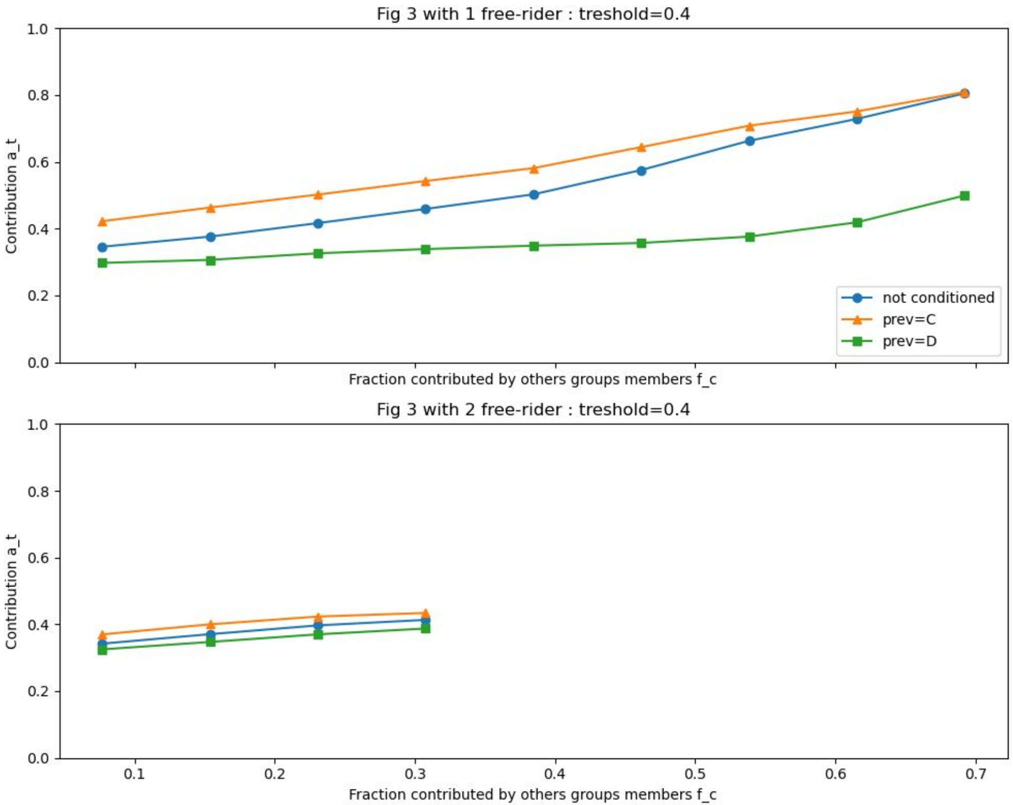
a = 0



IV.ii Free Rider







V. Conclusion

- Reconstruction of paper indicating that CC and MCC are behaviors naturally observable in dynamic learning

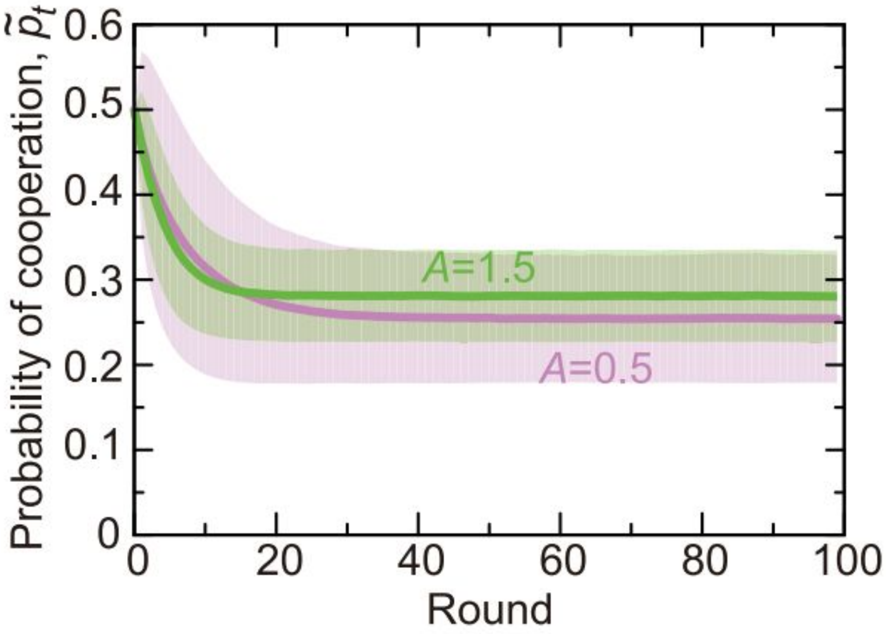
- The parameters do not have a significant impact (convergence speed changes but behavior remains observable).

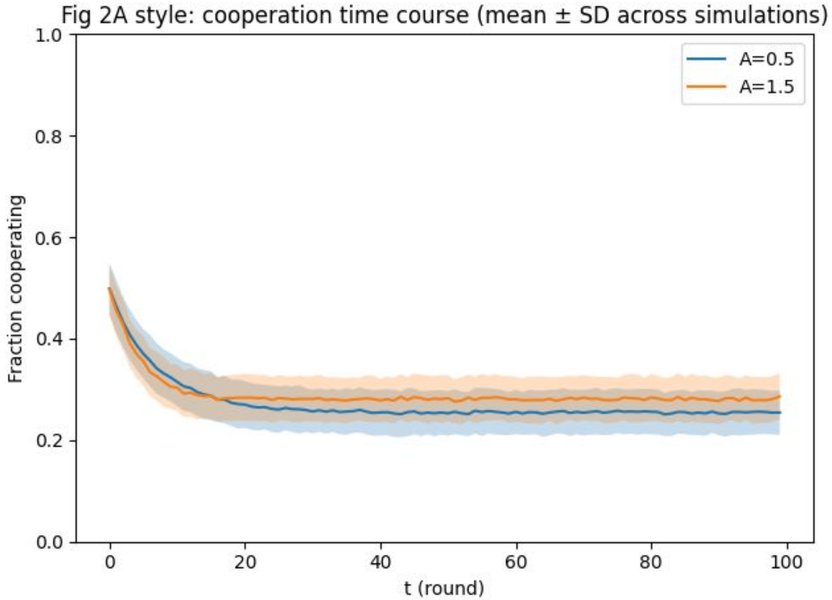
- It is easier to force cooperation than defection.

- Study interaction with neighbors (more neighbors for the PDG or no interaction with all neighbors for the PGG)

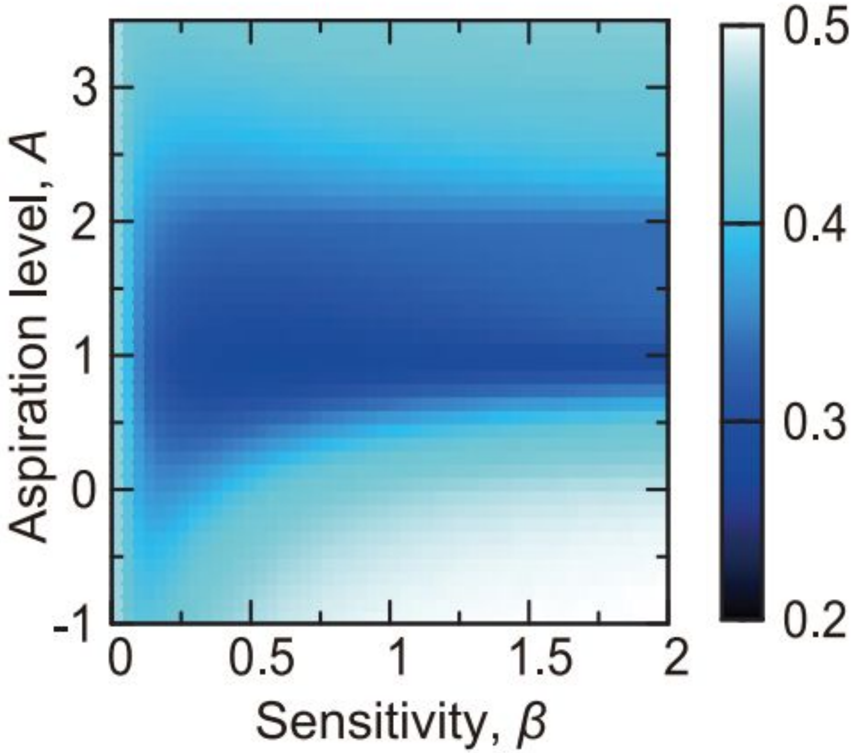
- Use heterogeneous parameters (one variable for each agent) and for other social dilemmas.

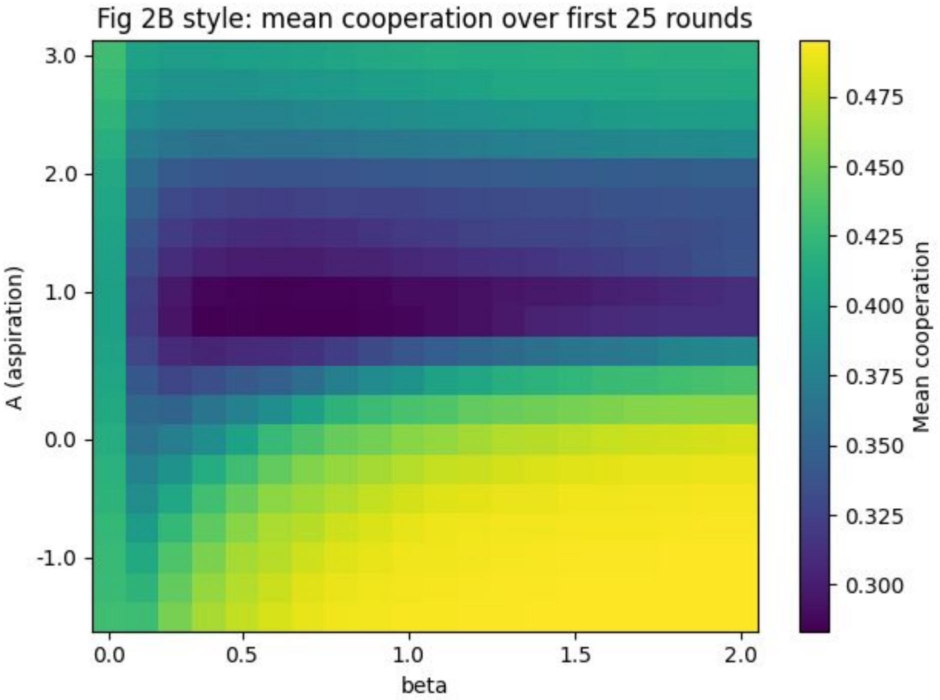
Appendix : Fig 2A



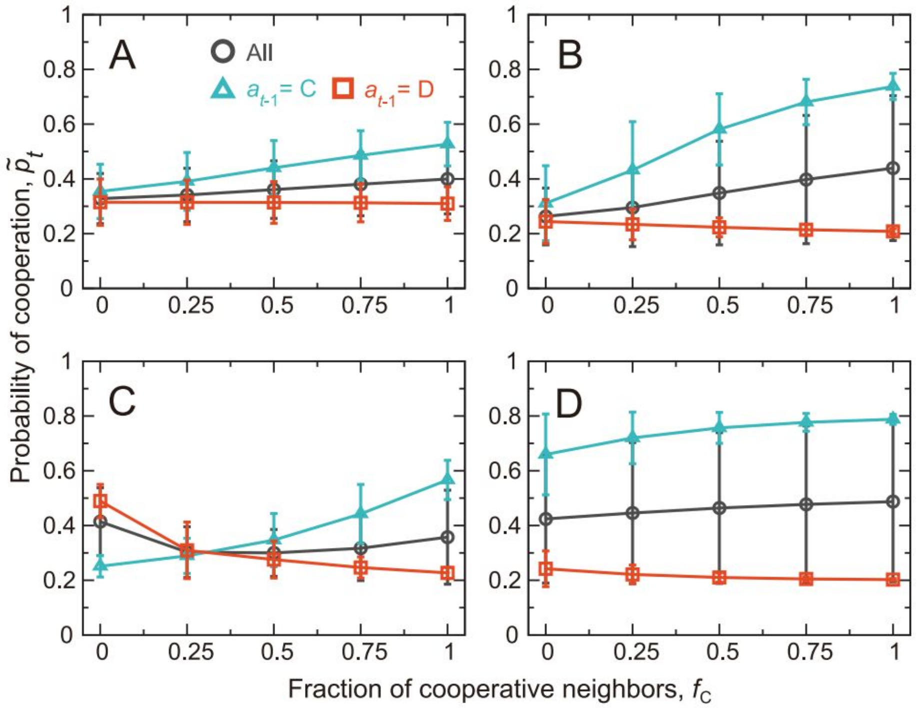


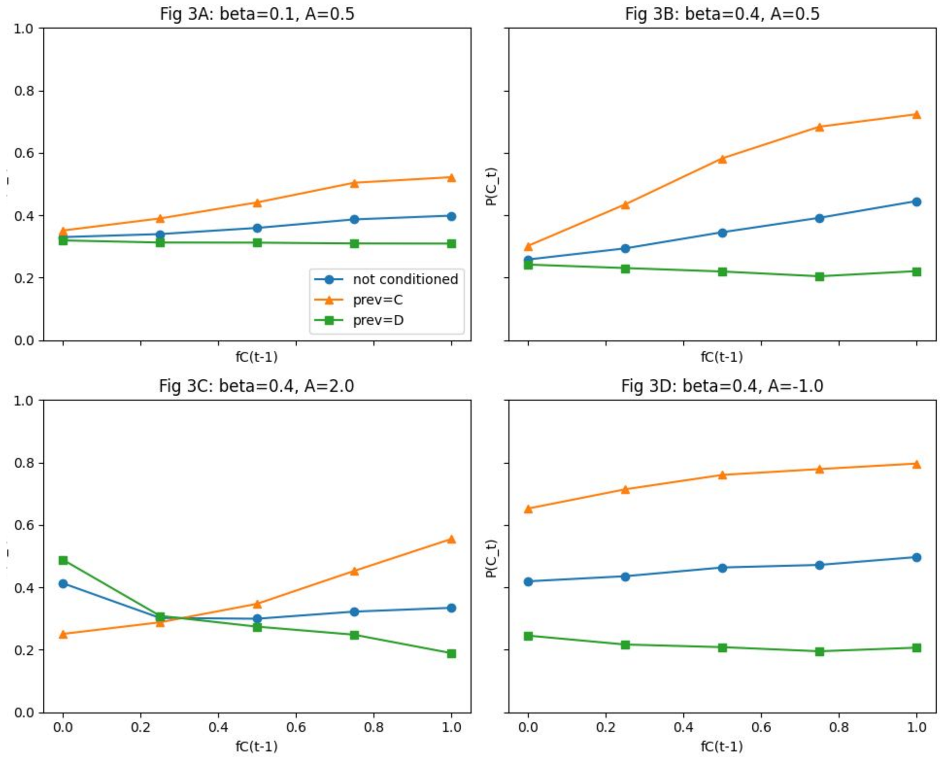
Appendix : Fig 2B



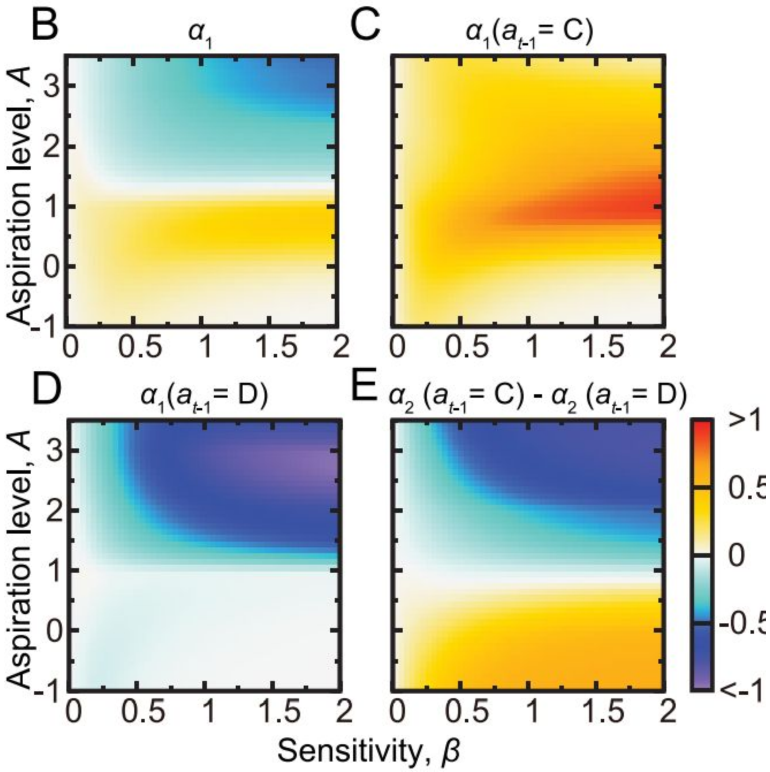


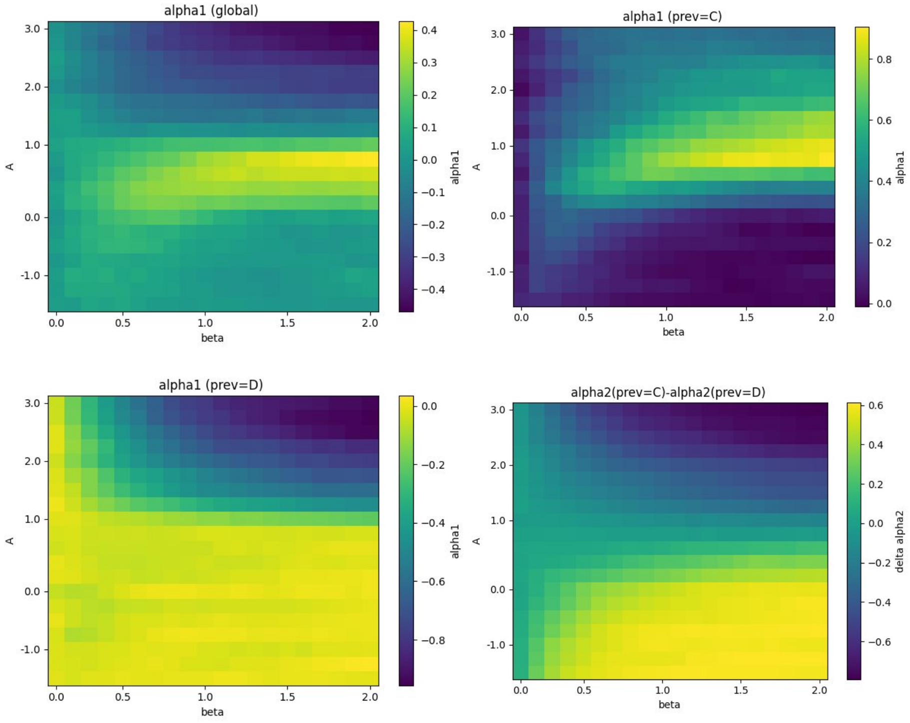
Appendix : Fig 3



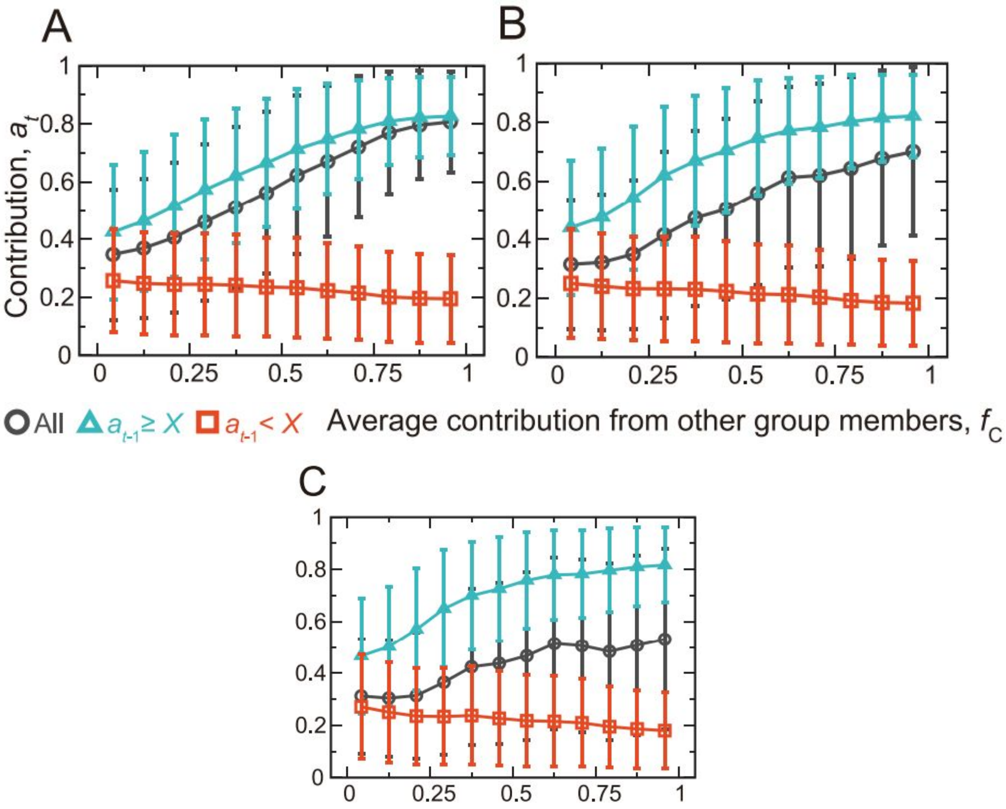


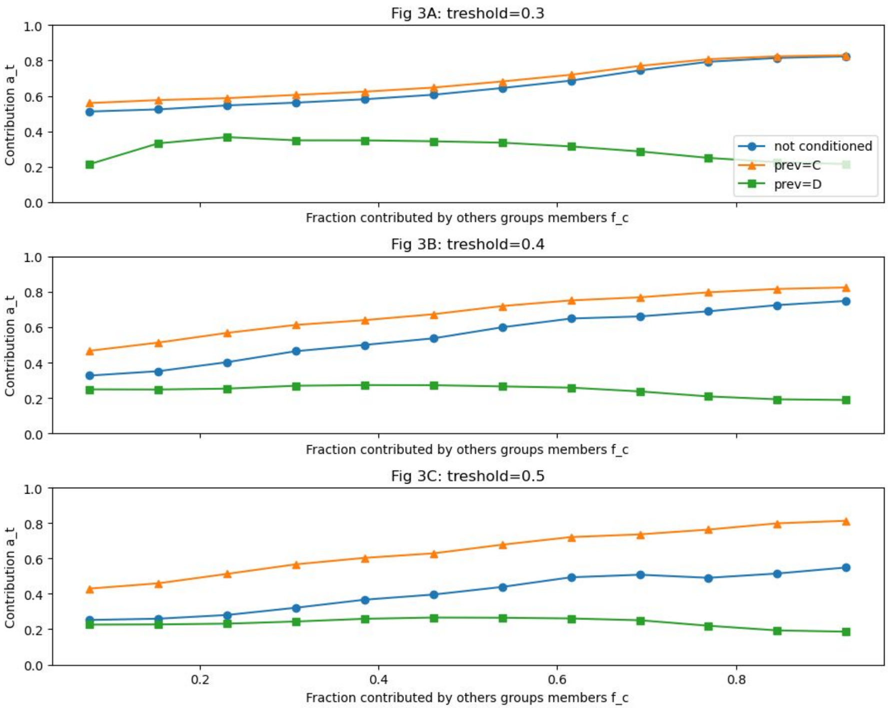
Appendix : Fig 4





Appendix : Fig 5 A-B-C





Appendix : Fig 5 D-E-F-G