Playing Repeated **Coopetitive** Polymatrix Games with Small Manipulation Cost

Shivakumar Mahesh*, Nicholas Bishop, Le Cong Dinh, and **Long Tran-Thanh**

University of Warwick, UK

long.tran-thanh@warwick.ac.uk

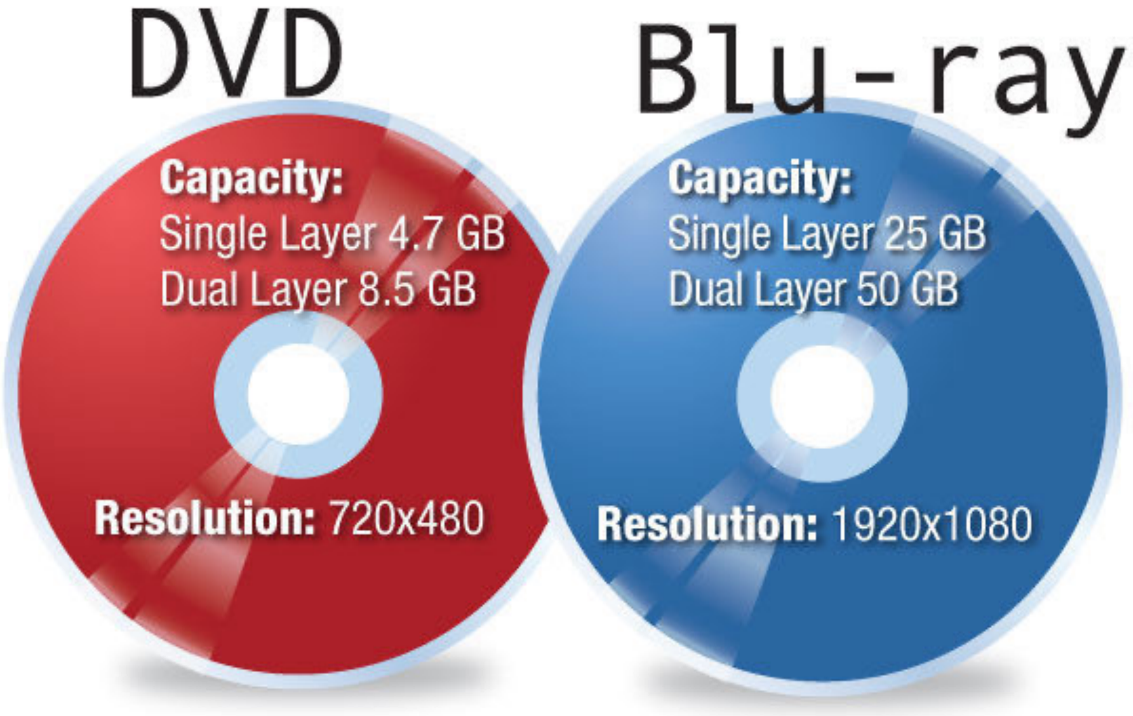
*work done when 1st author was an undergraduate student at Warwick

What’s Coopetitive Game?

- In order to win/perform well, one must cooperate with their opponents
- But they also need to know when to stop cooperating to become the winner/achieve their goal
- That is, they need to cooperate and compete at the same time (Nalebuff & Brandenburger, 1996)

https://cruciformstuff.com/2023/07/30/betrayal/

Example 1: Blue-Ray vs. DVD



https://fr.tipard.com/resource/blu-ray-vs-dvd.html

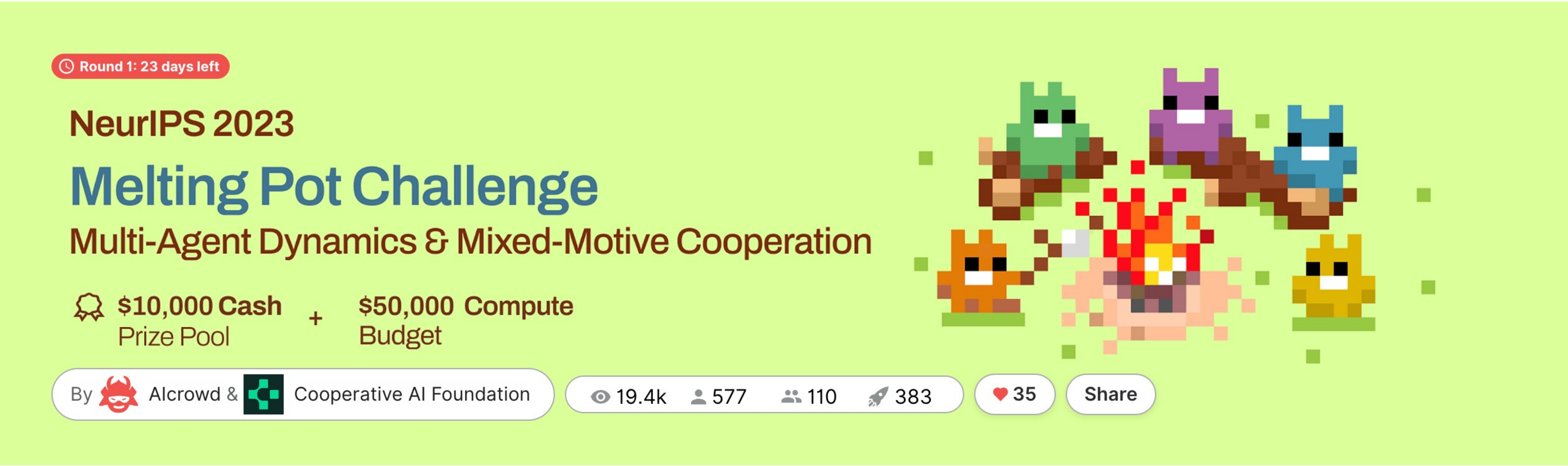
Example 2: Tour de France

https://www.ef.fr/blog/language/les-principaux-termes-de-cyclisme-connaitre-pour-regarder-le-tour-de-france/

Recent Interests from the AI Community

Google Deepmind + Cooperative AI Foundation's Melting Pot Challenge (hosted at NeurIPS 2023)

https://www.aicrowd.com/challenges/meltingpot-challenge-2023



Research Questions

In AI, we consider a multi-agent sequential decision-making version of coopetitive games:

- Who to cooperate with?
- How to signal/incentivise others to collaborate
- When to switch side?

This Paper’s Focus

- Aim: Proof of Concept
- Simplified setting
- 3 players
- Repeated games
- Polymatrix games
- Signaling: payoff manipulation

Payoff Manipulation Explained

- In our setting no explicit communication between agents is allowed (e.g., no negotiation/bargaining theory)

- Instead, we allow one agent to modify another agent's payoff by:

  - Sacrificing from their own payoffs (e.g., gift, bribery, etc) -> increasing the other's payoff

  - Enforce some penalties -> decreasing opponent's payoff

  - Examples: multiplayer video games, nature, etc.

Problem Formulation

- 3 players: P1, P2, P3 (we are P1) – repeated game (each round they play the same game)
- Polymatrix game:
  - Game can be decomposed to sum or pairwise 2-player games
  - Payoff = sum of pairwise payoffs defined by pairwise payoff matrices $A^{(i,j)}$
- Payoff manipulation: P1 can modify $A^{(2,1)}$ and $A^{(3,1)}$
- Payoff of P1:

$$x^T A^{(1,2)} y + x^T A^{(1,3)} z - ||M^{(2,1)} - A^{(2,1)}||_\infty - ||M^{(3,1)} - A^{(3,1)}||_\infty$$

- Payoff of P2 & P3:

$$y^T M^{(2,1)} x + y^T A^{(2,3)} z$$

$$z^T M^{(3,1)} x + z^T A^{(3,2)} y$$

Winning Policies

Objective: P1 will have higher total/average payoff than P2 and P3

Idea: We are interested in a certain type of behaviour (policy) that can lead to winning the game

- Suppose P1 plays i* action for all the rounds
- Suppose P2 has a **strictly dominant** strategy j* against i*, similarly P3 has a **strictly dominant** strategy k* against i*
- Also, suppose $u_1(i^*, j^*, k^*) > \max\left\{u_2(i^*, j^*, k^*), u_3(i^*, j^*, k^*)\right\}$
- Then by consistently playing i*, P1 would eventually *win the game*

Issue: such situation does not always exist ☹

Solution: create such solution via (minimal) payoff matrix manipulation!!! ☺

Existence of Dominant Solvable Games

Goal: Design a game via (optimally) manipulating $M^{(2,1)}$ and $M^{(3,1)}$ such that P2 has a **strictly dominant** strategy j* against i*, similarly P3 has a **strictly dominant** strategy k* against i* (for some i* action of P1)

Result 1: such dominant solvable game exists for any original 3-player polymatrix games

Even more, if we fix i*, j*, and k* in advance -> there exists a dominant solvable game for (i*,j*,k*)

Issue 1: How to achieve $u_1(i^*, j^*, k^*) > \max\left\{u_2(i^*, j^*, k^*), u_3(i^*, j^*, k^*)\right\}$

Issue 2: What happens if P2 and P3 are learning agents?

Consistent Agents

Definition 1. *(Consistent Agent) Suppose that for an agent there exists an action a^* that is the unique best response for her for every round of the game. Suppose that within T rounds of the game, the number of rounds the agent plays action a^* is T^*. If $\mathbb{P}\left(\lim_{T\to\infty} \frac{T^*}{T} = 1\right) = 1$ then the agent is 'consistent'.*

Consistent agent:

- There is a same fixed best action for that agent in **every round**
- Event: the fraction of number of times the agent plays this best action tends to 1
- Probability of this event = 1

Persistent Agents

Definition 4. *(Persistent Agent) Suppose that the action k^* is the best action in hindsight for player 3 eventually, with probability 1. That is,*

$$\mathbb{P}\Big(\boldsymbol{e}_{k^*} = \arg\max_{\boldsymbol{z}\in\Delta_l} U_3(\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{z})_{t=1}^{T} \text{ eventually}\Big) = 1$$

Let T^* denote the number of rounds within T rounds, that player 3 plays action k^*. If $\mathbb{P}\Big(\lim_{T\to\infty}\frac{T^*}{T} = 1\Big) = 1$ then player 3 is 'persistent'.

Persistent agent:

- There is a same fixed best action for that agent **from some round** τ (i.e., eventually)
- Event: the fraction of number of times the agent plays this best action tends to 1
- Probability of this event = 1

Main Results

Winning dominance solvable policies:

- Each action of P1= $(a_t^1, M_t^{(2,1)}, M_t^{(3,1)})$
- Makes P1 is the winner of the resulting dominant solvable game

Theorem 1: If P2 and P3 are consistent agents then there exists a winning dominance solvable policy for P1

Theorem 2: If P2 is consistent and P3 is persistent, then there exists a winning dominance solvable policy for P1

Theorem 3: These winning dominance solvable policies, if exist, can be calculated in polynomial running time

Additional Objectives

- Winning by largest margin
- Winning by lowest inefficiency ratio
- Maximising the egalitarian social welfare

Winning by Largest Margin

Margin of P1:

$$\min \left\{ \mathbb{E}\left[U_1(\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{z}_t)_{t=1}^{\infty} - U_2(\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{z}_t)_{t=1}^{\infty} \right], \mathbb{E}\left[U_1(\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{z}_t)_{t=1}^{\infty} - U_3(\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{z}_t)_{t=1}^{\infty} \right]$$

- How much better the (expected) average payoff of P1 is compared to the others'

Theorem 6: *If winning dominance solvable policies exist*, then there exists an algorithm that can find the **largest margin dominance solvable policy**, with running time that is polynomial in the number of actions of the players.

Winning by Lowest Inefficiency Ratio

Inefficiency ratio: the ratio between the **cost for modifying the payoff matrices** and the **expected increase in long run payoffs** from the worst-case payoff.

$$\frac{\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{(i,j)\in P} \|A_t^{(i,j)} - A_0^{(i,j)}\|_\infty}{\mathbb{E}\Big[\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \Big(\mathbf{x}_t^T A_t^{(1,2)} \mathbf{y}_t + \mathbf{x}_t^T A_t^{(1,3)} \mathbf{z}_t\Big)\Big] - K}$$

where $K = \min_{i,j,k} \left(A^{(1,2)}(i,j) + A^{(1,3)}(j,k)\right)$ *is the minimum revenue for player 1.*

Theorem: *If winning dominance solvable policies exist*, then there exists an algorithm that can find the **winning dominance solvable policy with the lowest inefficiency ratio**, with running time that is polynomial in the number of actions of the players.

Maximising Egalitarian Social Welfare

Egalitarian social welfare: The lowest payoff among the players'

Definition 9. *The Egalitarian Social Welfare of a strategy profile* $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ *is defined to be*

$$\mathcal{S}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) := \min \left\{ U_1(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), U_2(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), U_3(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \right\}$$

Theorem: There exists an algorithm that can find the dominance solvable policy that **maximizes egalitarian social welfare** with running time that is polynomial in the number of actions of the players.

Application 1: 3-Player Iterated Prisoner's Dilemma

Action space = {C, D}

$$A_0^{(i,j)} = \begin{bmatrix} 3 & 0 \\ 5 & 1 \end{bmatrix} \text{ if } i < j \text{ and } A_0^{(i,j)} = \begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix} \text{ if } i > j$$

For P1, a winning strategy would be always playing D (and both P2 and P3 also defect all the time)

- But this one has 0 margin as well
- Can we design a better policy with positive margin, and **incentivises cooperation**?

We show that for $0 < \varepsilon \leq \frac{7}{6}$ we set $\hat{A} = \begin{bmatrix} 3 & 5 \\ 3/2 + \epsilon & -1/2 \end{bmatrix}$

P1 plays D and manipulates opponents' payoff matrices to $\hat{A}$

Theorem: system will converge to (D,C,C) and P1 wins with large (positive) margin

Application 2: Social Distancing Game

Inspired by Zinkevic's Lemonade Stand Game

Winning the game:

Theorem 1: P1 can win the game with **negligible manipulation cost**

Egalitarian social welfare:

Theorem 2: P1 plays position 12 and use $\hat{A}$ and $\tilde{A}$ to manipulate the payoff of P2 and P3, then the **egalitarian social welfare is maximised**

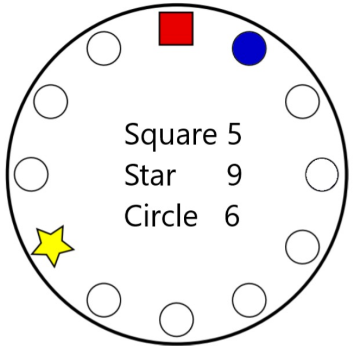


Fig. 1. Example Social Distancing Game

$$\hat{A}(k,l) = \begin{cases} d(k,l) & \text{if } k \neq 12 \\ d(k,l) - 1 - 2\epsilon & \text{if } k = 12 \text{ and } l \neq 5 \\ d(k,l) + 1 - \epsilon & \text{if } k = 12 \text{ and } l = 5 \end{cases}$$

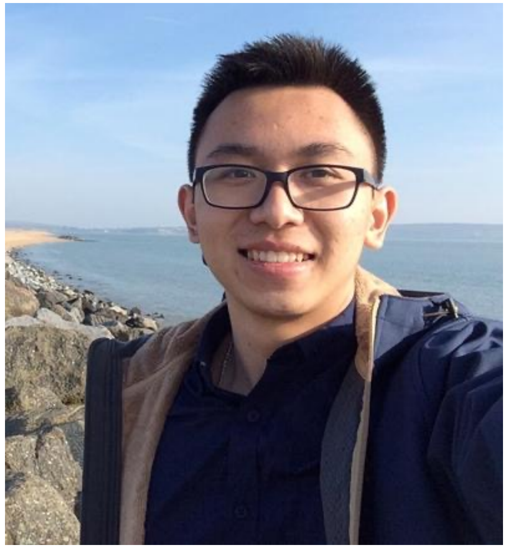
$$\tilde{A}(k,l) = \begin{cases} d(k,l) & \text{if } k \neq 12 \\ d(k,l) - 1 + \epsilon & \text{if } k = 12 \text{ and } l \neq 7 \\ d(k,l) + 1 - \epsilon & \text{if } k = 12 \text{ and } l = 7 \end{cases}$$

Arxiv Version of the Paper

- Arxiv link: https://arxiv.org/abs/2110.13532
- More analysis:
  - batch policies,
  - dominant solvability types,
  - Numerical results
- More applications:
  - Electric cars vs. petrol cars
  - Battle of buddies
- Full proofs

Shiva Mahesh

Nick Bishop



Le Cong Dinh

Many thanks for your Attention!