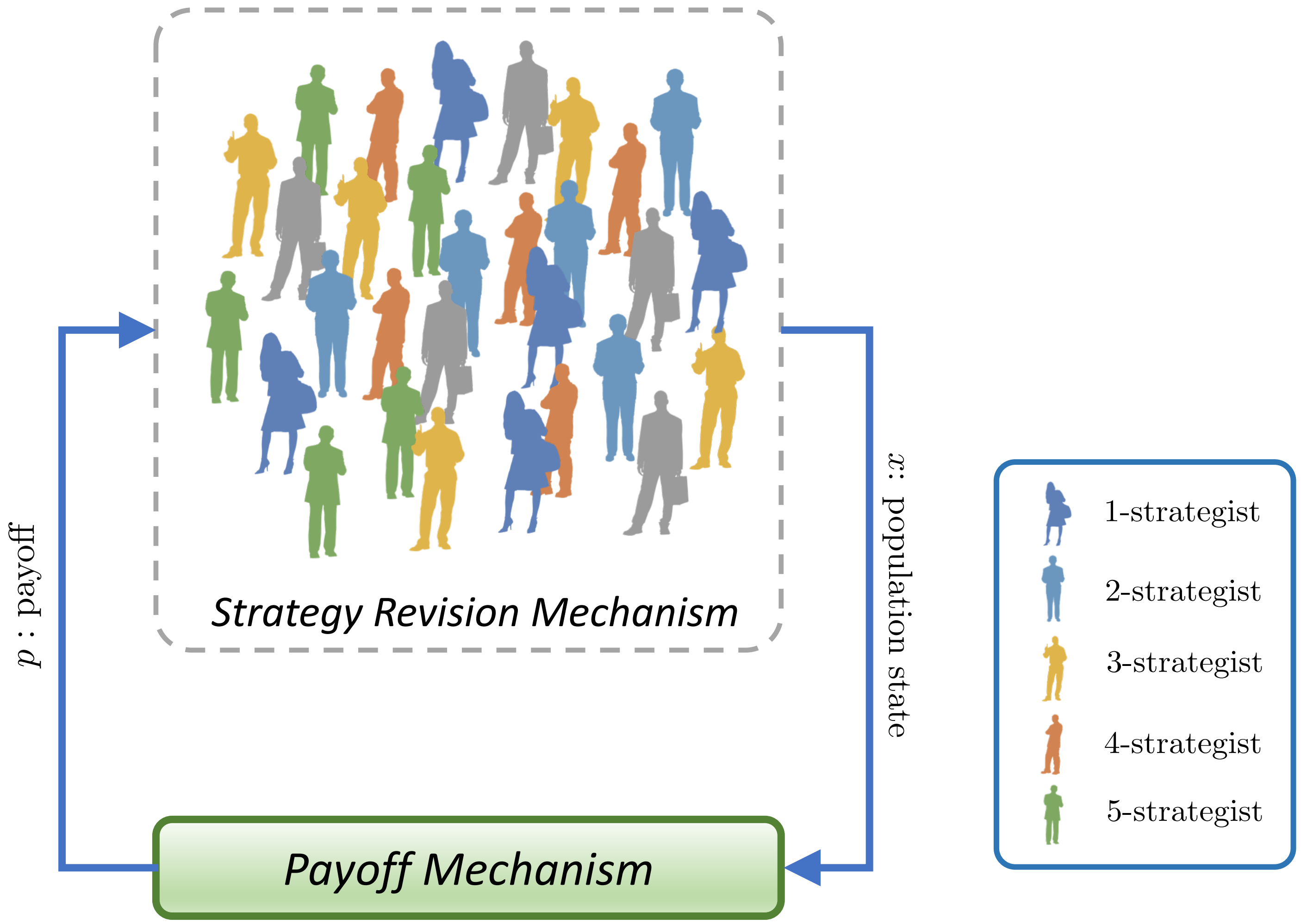
Tutorial Session on
"Payoff Dynamics and Higher-Order Learning in Population Games"

From Population Games to Payoff Dynamics Models: A Passivity-Based Approach

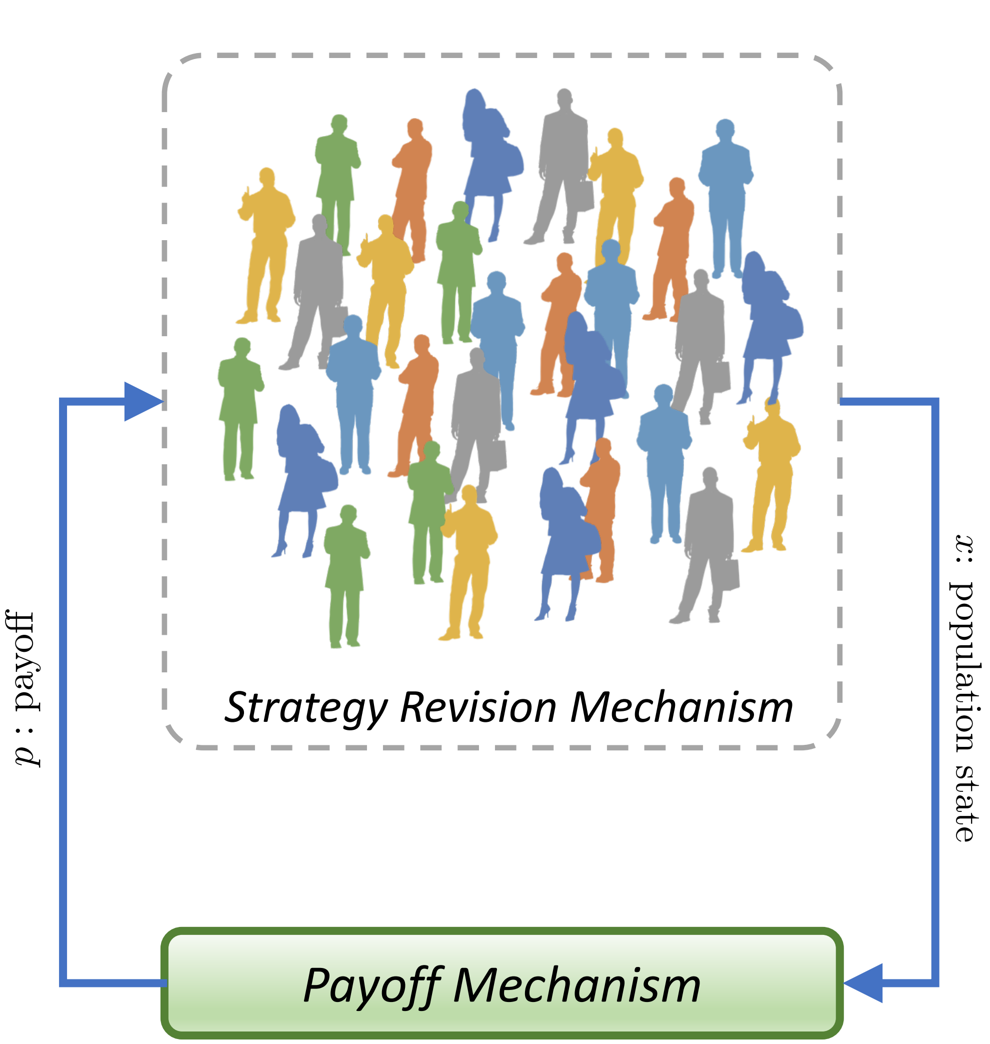
Dec. 13, 2019

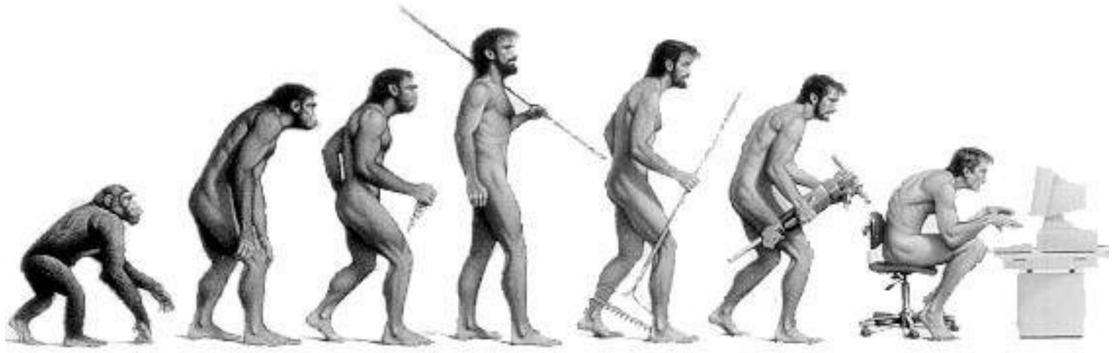
Shinkyu Park[♮], Nuno Martins[♯], and Jeff S. Shamma[♭]
[♮]Princeton University
[♯]University of Maryland
[♭]King Abdullah University of Science and Technology (KAUST)

Decision Making in Large Populations



Decision Making in Large Populations

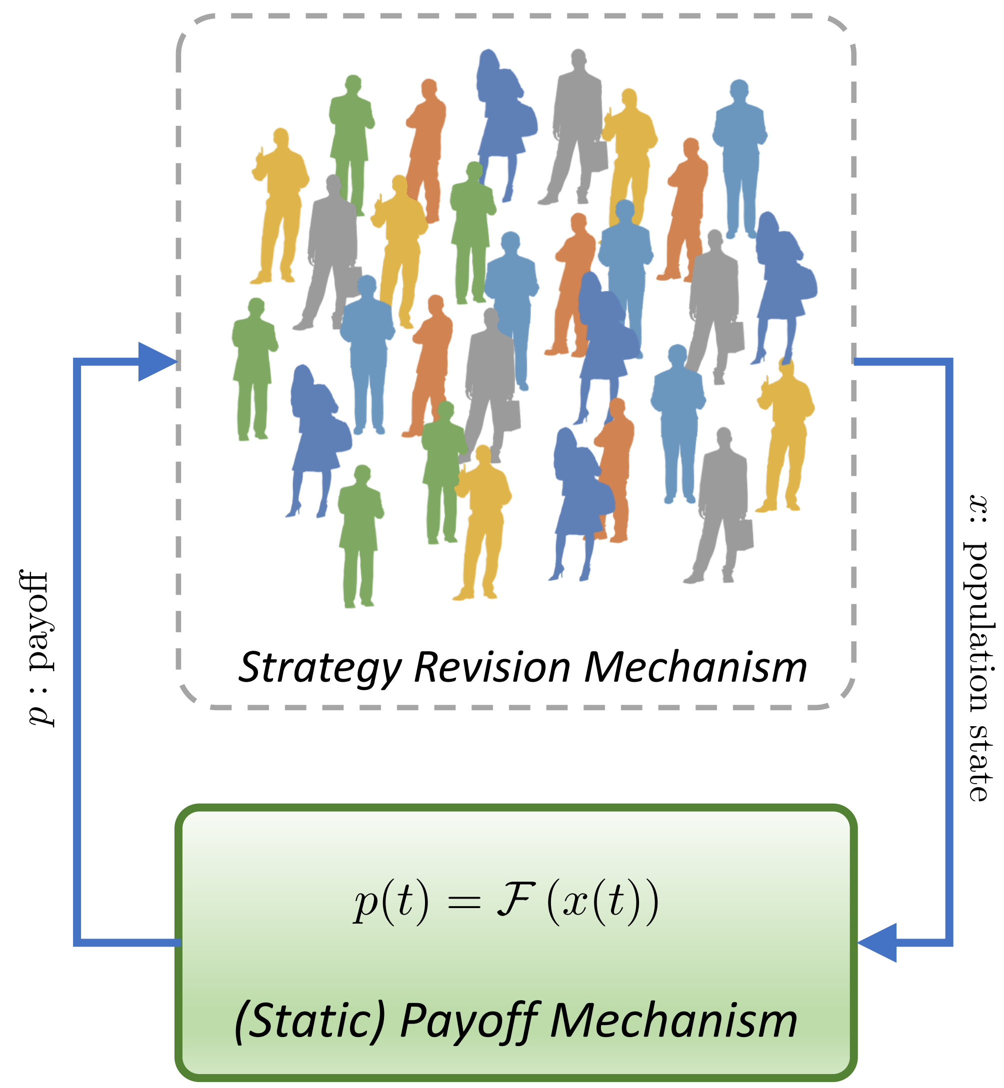


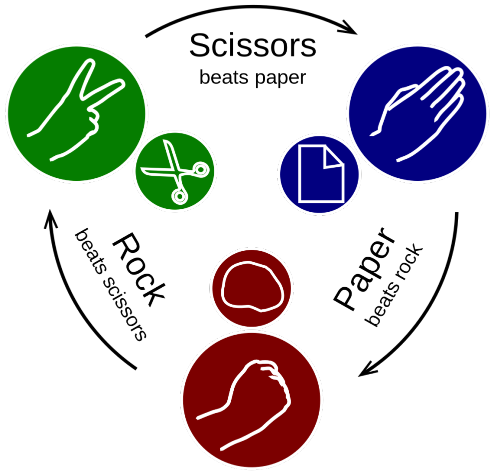


Evolutionary Biology

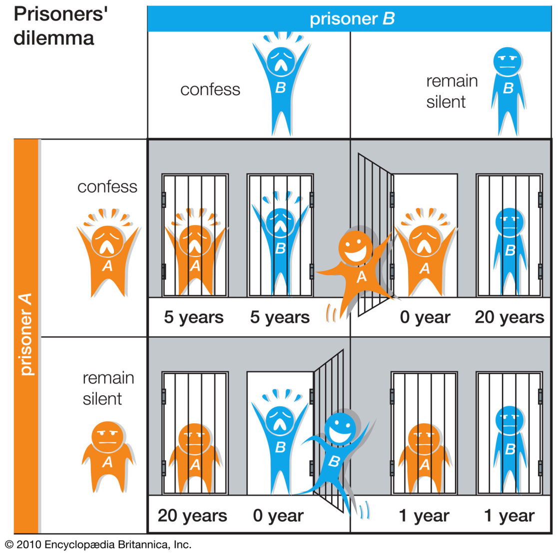
Multi-agent Planning

Decision Making in Large Populations





Rock-Paper-Scissors Game



Prisoners' Dilemma

Decision Making in Large Populations

Strategy Revision Mechanism

p : payoff

x: population state

$$\dot{q}(t) = \mathcal{G}\left(q(t), x(t)\right)$$
$$p(t) = \mathcal{H}\left(q(t), x(t)\right)$$

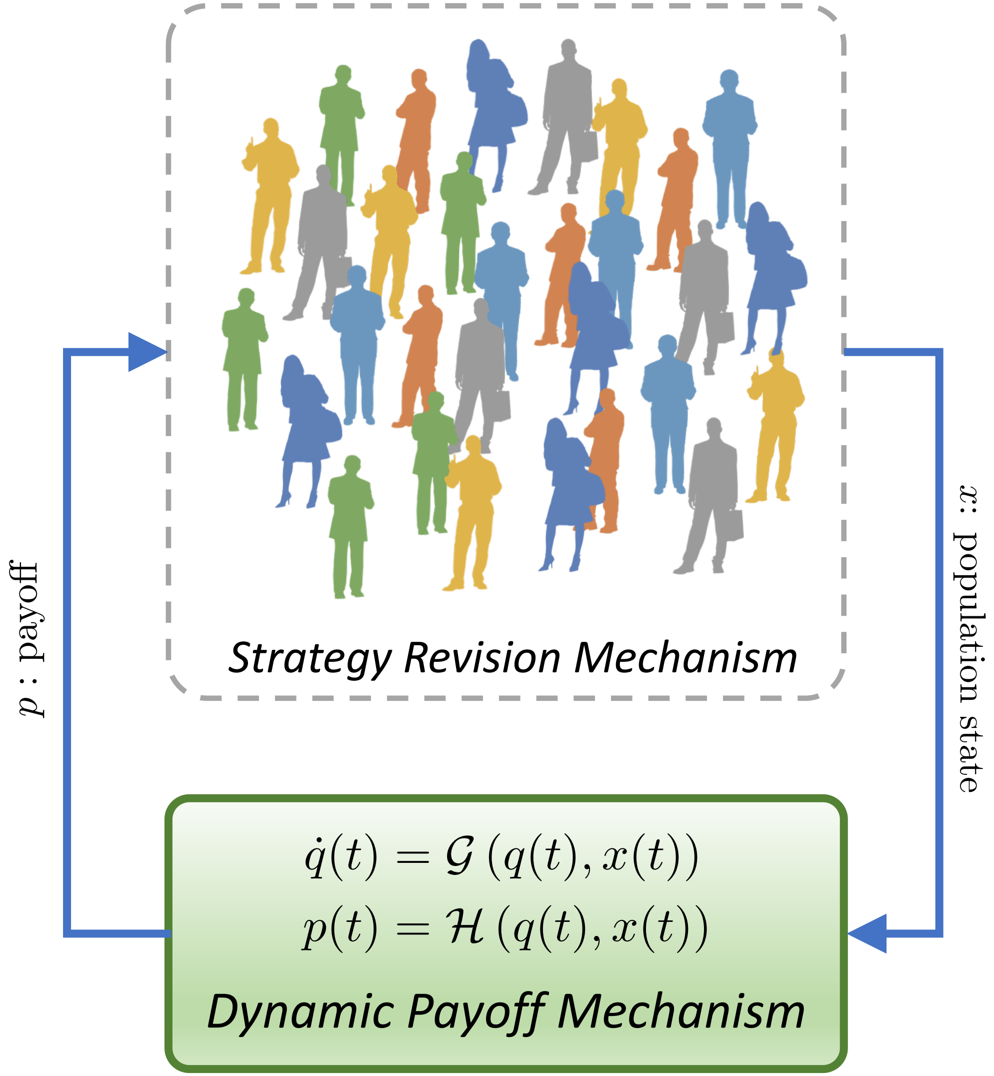
Dynamic Payoff Mechanism

- Dynamic payoff mechanism is constructed as

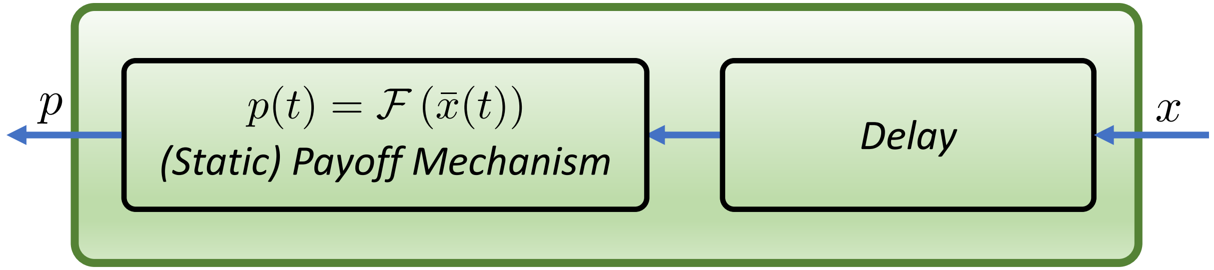
p ← *Learning Dynamics* ← $\bar{p}(t) = \mathcal{F}\left(x(t)\right)$ *(Static) Payoff Mechanism* ← x

  – Multi-agent learning/decision-making problem

Decision Making in Large Populations

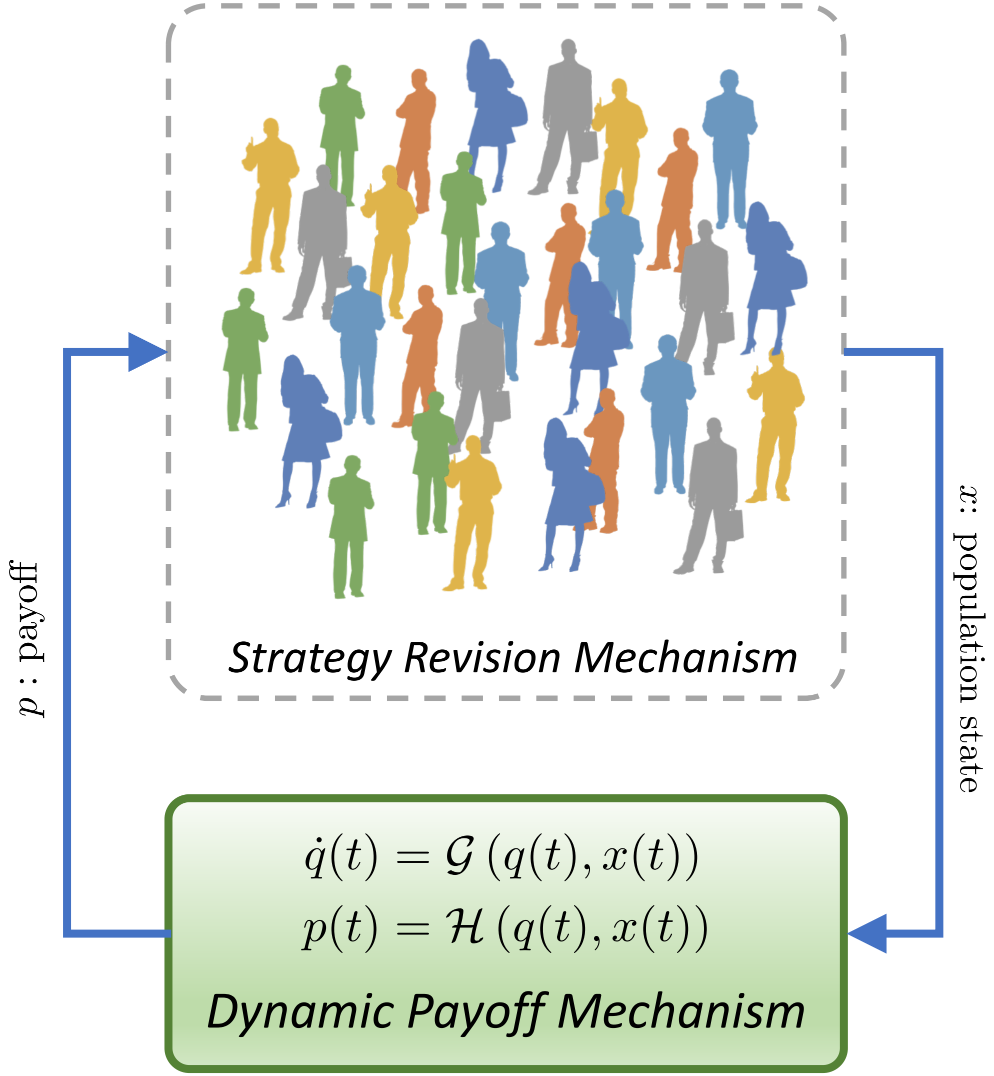


- Dynamic payoff mechanism is constructed as

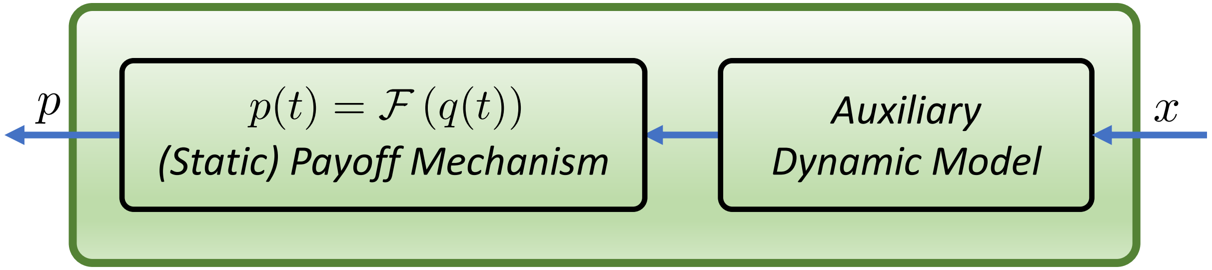


  - Traffic network routing problem

Decision Making in Large Populations



- Dynamic payoff mechanism is constructed as



  – Multi-robot manipulation problem

Decision Making in Large Populations

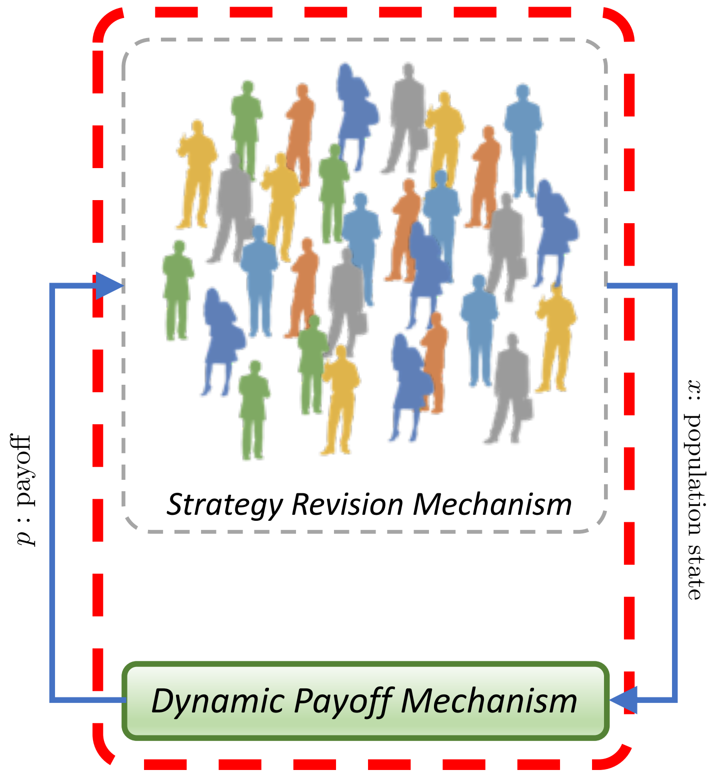
- **Notation:**
  - $\{1, \cdots, n\}$: Strategy set
  - $x = (x_1, \cdots, x_n) \in \mathbb{X}$: Mean population state, where

$$x_i = \lim_{N \to \infty} \frac{\#\text{ of agents adopting } i}{N}$$

  - $p = (p_1, \cdots, p_n) \in \mathbb{R}^n$: Deterministic payoff vector assigned to x

Decision Problem in Large Populations: Under what mean closed loop model, the population state and payoff vector converge to x^* and p^* for which

$$x_i^* > 0 \implies p_i^* = \max_{1 \le j \le \bar{s}} p_j^*$$



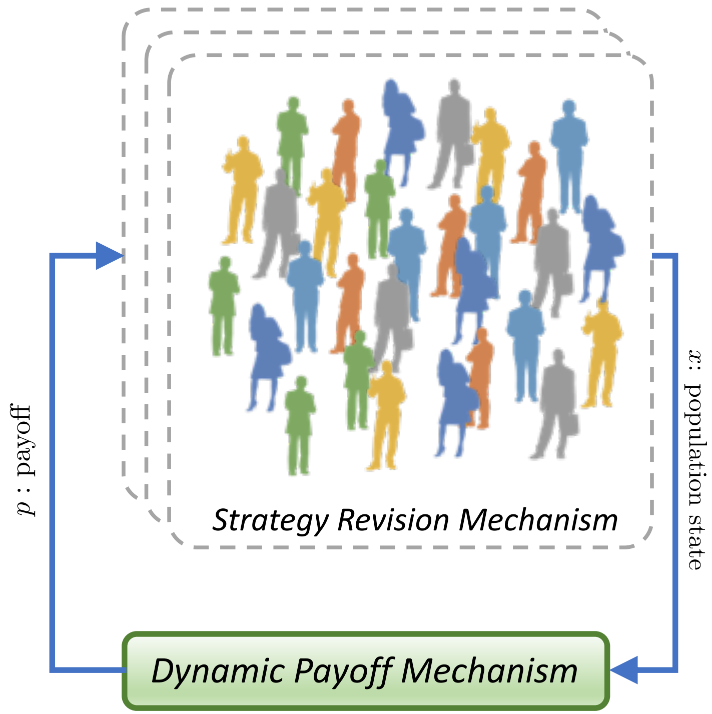
Mean Closed Loop Model

Multi-Population Case

- **Notation (for each population k):**
  - $\{1, \cdots, n^k\}$: Strategy set for population k
  - $x^k = (x_1^k, \cdots, x_{n^k}^k)$: Mean population state for population k, where

$$x_i^k = \lim_{N^k \to \infty} \frac{\#\text{ of agents in population } k \text{ adopting } i}{N^k}$$

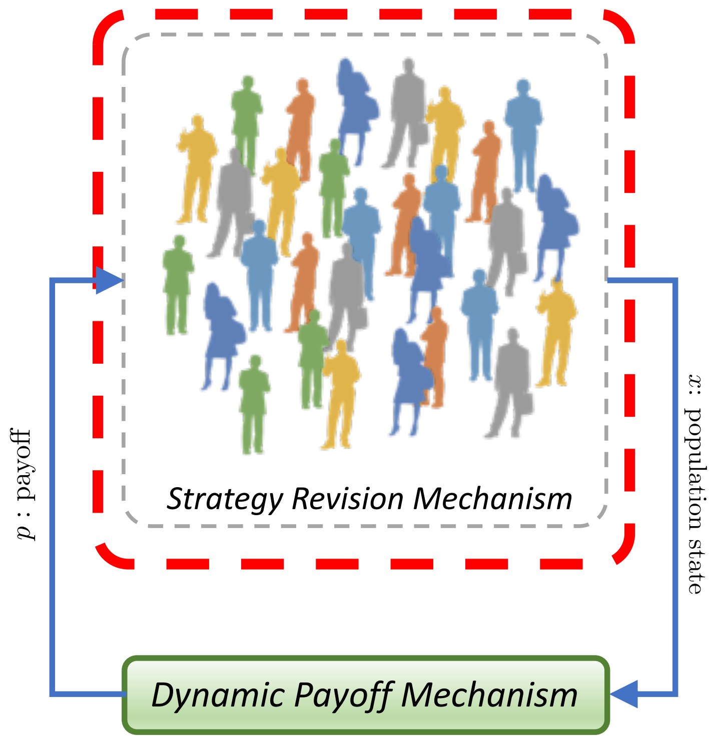
  - $p^k = (p_1^k, \cdots, p_{n^k}^k)$: Deterministic payoff vector assigned to x



Revision Protocol and Evolutionary Dynamics Model

- **Revision protocol $\mathcal{T}_{ji}(x,p)$:** The protocol describes the agents' switching rate from strategy j to strategy i, given the population state x and payoff p
- **Evolutionary dynamics model (EDM):**

$$\dot{x}_i(t) = \mathcal{V}_i\left(x(t), p(t)\right) = \underbrace{\sum_{j=1}^{n} x_j(t)\mathcal{T}_{ji}\left(x(t), p(t)\right)}_{\text{inflow to population of } i\text{-strategists}} - \underbrace{x_i(t)\sum_{j=1}^{n} \mathcal{T}_{ij}\left(x(t), p(t)\right)}_{\text{outflow from population of } i\text{-strategists}}$$



Revision Protocol and Evolutionary Dynamics Model

- **Revision protocol $\mathcal{T}_{ji}(x,p)$:** The protocol describes the agents' switching rate from strategy j to strategy i, given the population state x and payoff p
- **Evolutionary dynamics model (EDM):**

$$\begin{aligned}\dot{x}_i(t) &= \mathcal{V}_i\left(x(t),p(t)\right)\\ &= \sum_{j=1}^{n} x_j(t)\mathcal{T}_{ji}\left(x(t),p(t)\right) - x_i(t)\sum_{j=1}^{n}\mathcal{T}_{ij}\left(x(t),p(t)\right)\end{aligned}$$

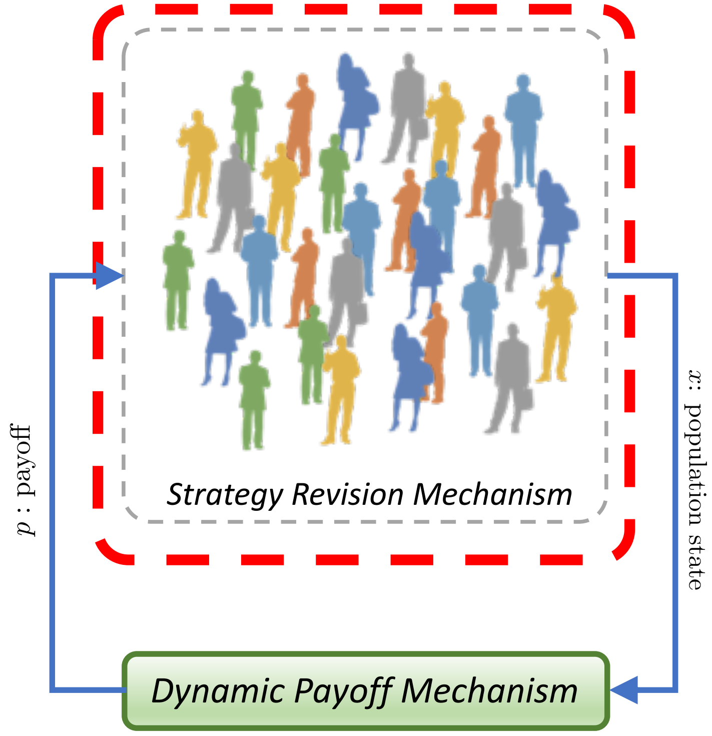
  - Examples:
    - BNN EDM:

$$\dot{x}_i(t) = \left[\hat{p}_i(t)\right]_+ - x_i(t)\sum_{j=1}^{n}\left[\hat{p}_j(t)\right]_+, \ 1 \le i \le n$$

$$(\hat{p}_i = p_i - \textstyle\sum_{j=1}^{n} p_j x_j\text{: excess payoff vector})$$

    - Logit EDM:

$$\dot{x}_i(t) = \frac{\exp\left(\eta^{-1}p_i(t)\right)}{\sum_{j=1}^{\bar{s}}\exp\left(\eta^{-1}p_j(t)\right)} - x_i(t), \ 1 \le i \le n$$



Payoff Dynamics Model

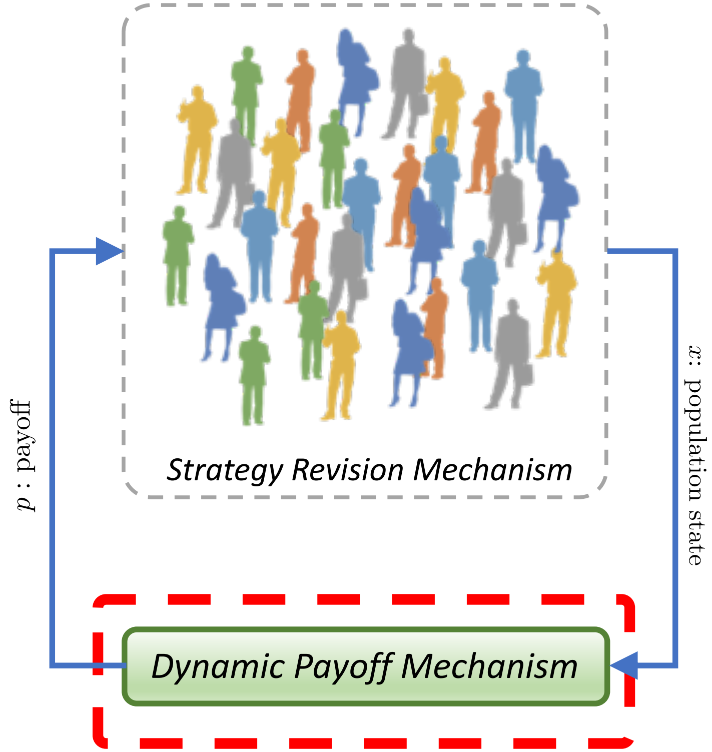
- **Payoff dynamics model (PDM):**

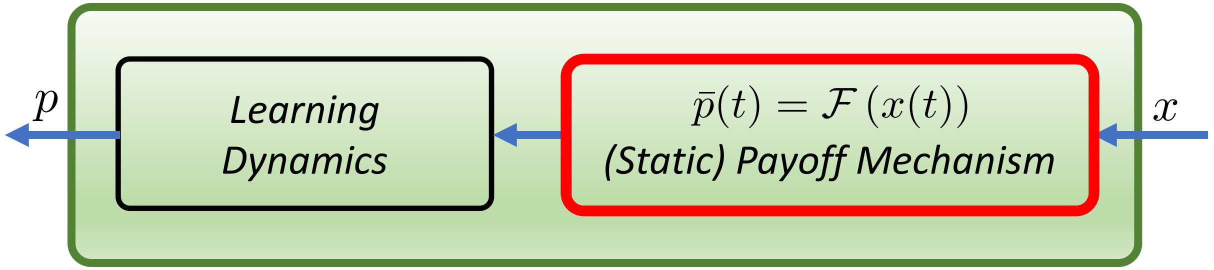
$$\dot{q}(t) = \mathcal{G}\left(q(t), x(t)\right)$$
$$p(t) = \mathcal{H}\left(q(t), x(t)\right)$$

where $\mathcal{G} : \mathbb{R}^n \times \mathbb{X} \to \mathbb{R}^n$ is Lipschitz continuous and $\mathcal{H} : \mathbb{R}^n \times \mathbb{X} \to \mathbb{R}^n$ is continuously differentiable and Lipschitz continuous.

- Stationary population game $\bar{\mathcal{F}}$: There is a continuous map $\bar{\mathcal{F}} : \mathbb{X} \to \mathbb{R}^n$ for which

$$\lim_{t\to\infty} \|\dot{x}(t)\| = 0 \implies \lim_{t\to\infty} \left\|p(t) - \bar{\mathcal{F}}\left(x(t)\right)\right\| = 0$$





Payoff Dynamics Model

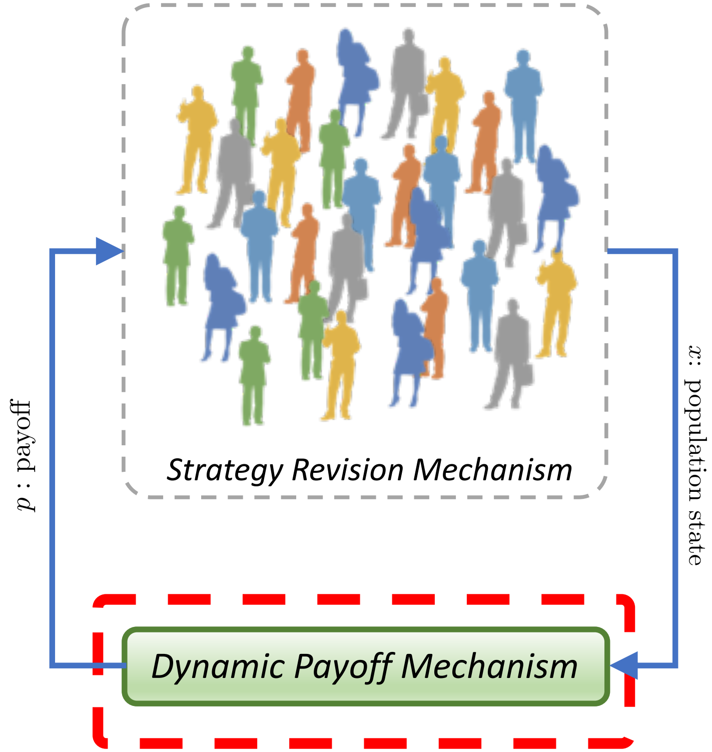
- **Payoff dynamics model (PDM):**

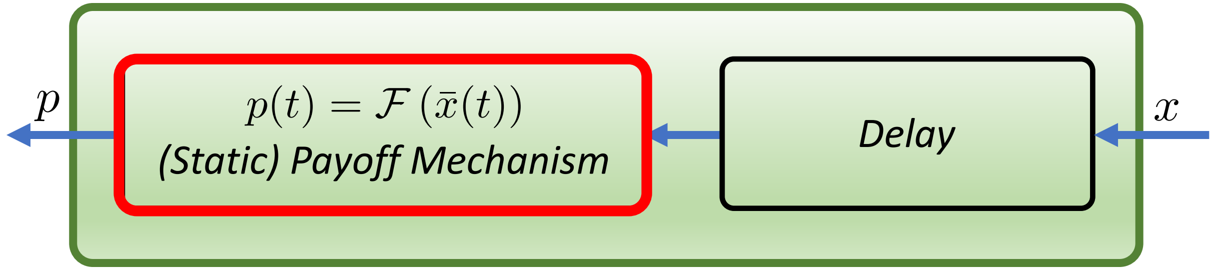
$$\dot{q}(t) = \mathcal{G}\left(q(t), x(t)\right)$$
$$p(t) = \mathcal{H}\left(q(t), x(t)\right)$$

where $\mathcal{G} : \mathbb{R}^n \times \mathbb{X} \to \mathbb{R}^n$ is Lipschitz continuous and $\mathcal{H} : \mathbb{R}^n \times \mathbb{X} \to \mathbb{R}^n$ is continuously differentiable and Lipschitz continuous.

  – Stationary population game $\bar{\mathcal{F}}$: There is a continuous map $\bar{\mathcal{F}} : \mathbb{X} \to \mathbb{R}^n$ for which

$$\lim_{t\to\infty} \|\dot{x}(t)\| = 0 \implies \lim_{t\to\infty} \left\|p(t) - \bar{\mathcal{F}}\left(x(t)\right)\right\| = 0$$





Payoff Dynamics Model

- **Payoff dynamics model (PDM):**

$$\dot{q}(t) = \mathcal{G}\left(q(t), x(t)\right)$$
$$p(t) = \mathcal{H}\left(q(t), x(t)\right)$$

where $\mathcal{G} : \mathbb{R}^n \times \mathbb{X} \to \mathbb{R}^n$ is Lipschitz continuous and $\mathcal{H} : \mathbb{R}^n \times \mathbb{X} \to \mathbb{R}^n$ is continuously differentiable and Lipschitz continuous.

  - Stationary population game $\bar{\mathcal{F}}$: There is a continuous map $\bar{\mathcal{F}} : \mathbb{X} \to \mathbb{R}^n$ for which

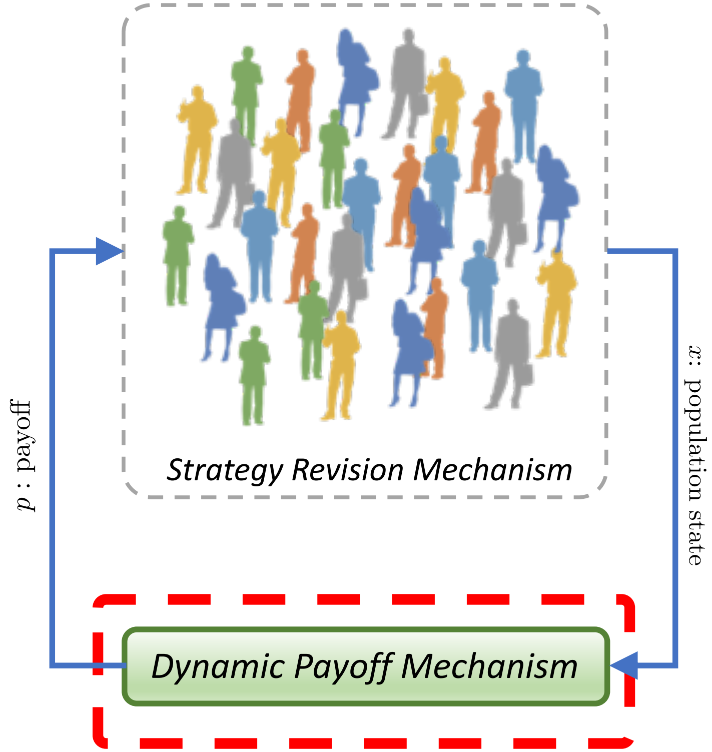
$$\lim_{t\to\infty} \|\dot{x}(t)\| = 0 \implies \lim_{t\to\infty} \left\|p(t) - \bar{\mathcal{F}}\left(x(t)\right)\right\| = 0$$

  - Examples:
    * Static PDM: Given $\mathcal{F} : \mathbb{X} \to \mathbb{R}^n$,

$$p(t) = \mathcal{F}(x(t))$$

    * Smoothing PDM: Given $\mathcal{F} : \mathbb{X} \to \mathbb{R}^n$ and $\alpha > 0$,

$$\dot{q}(t) = \alpha\Big(\mathcal{F}\left(x(t)\right) - q(t)\Big)$$
$$p(t) = q(t)$$



Mean Closed Loop Model

- **Mean closed loop model:**

$$\left.\begin{aligned}\dot{q}(t) &= \mathcal{G}\left(q(t), x(t)\right) \\ p(t) &= \mathcal{H}\left(q(t), x(t)\right)\end{aligned}\right\} \text{(PDM)}$$

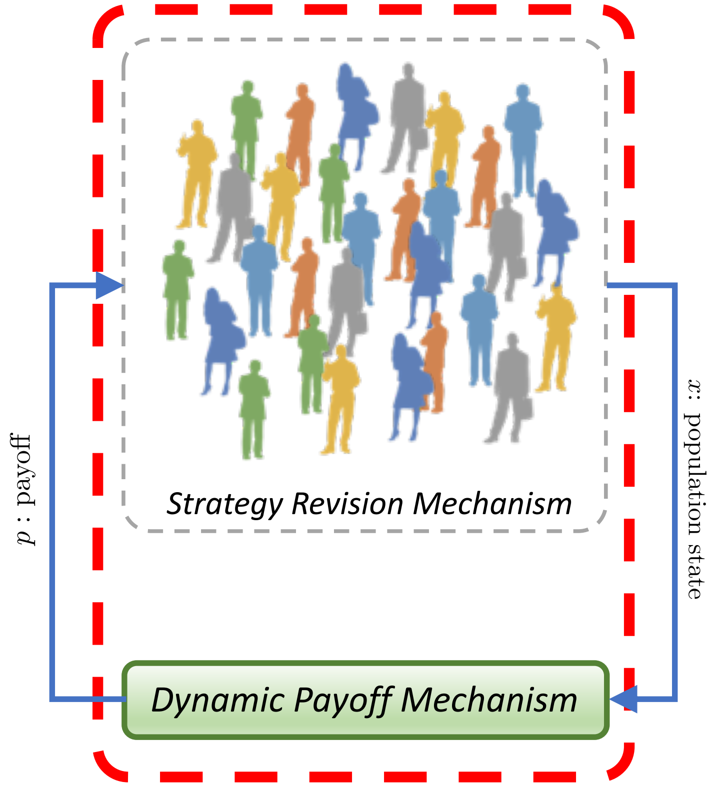
$$\dot{x}(t) = \mathcal{V}\left(x(t), p(t)\right) \quad \text{— (EDM)}$$

  - Equilibrium set:

$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \left\{z \in \mathbb{X} \,\middle|\, \mathcal{V}\left(z, \bar{\mathcal{F}}(z)\right) = 0\right\}$$

    $\mathbb{E}\left(\bar{\mathcal{F}}\right)$ is a projection on $\mathbb{X}$ of the set $\mathbb{A}$ of stationary points of the mean closed loop model.

    * $\mathbb{A} = \left\{(s, z) \in \mathbb{R}^n \times \mathbb{X} \,\middle|\, \mathcal{V}\left(z, \mathcal{H}\left(s, z\right)\right) = 0 \text{ and } \mathcal{H}(s, z) = \bar{\mathcal{F}}(z)\right\}$



– $\bar{\mathcal{F}}$: stationary population game

Mean Closed Loop Model

- **Mean closed loop model:**

$$\dot{q}(t) = \mathcal{G}\left(q(t), x(t)\right)$$
$$p(t) = \mathcal{H}\left(q(t), x(t)\right)$$
$$\dot{x}(t) = \mathcal{V}\left(x(t), p(t)\right)$$

  - Equilibrium set:

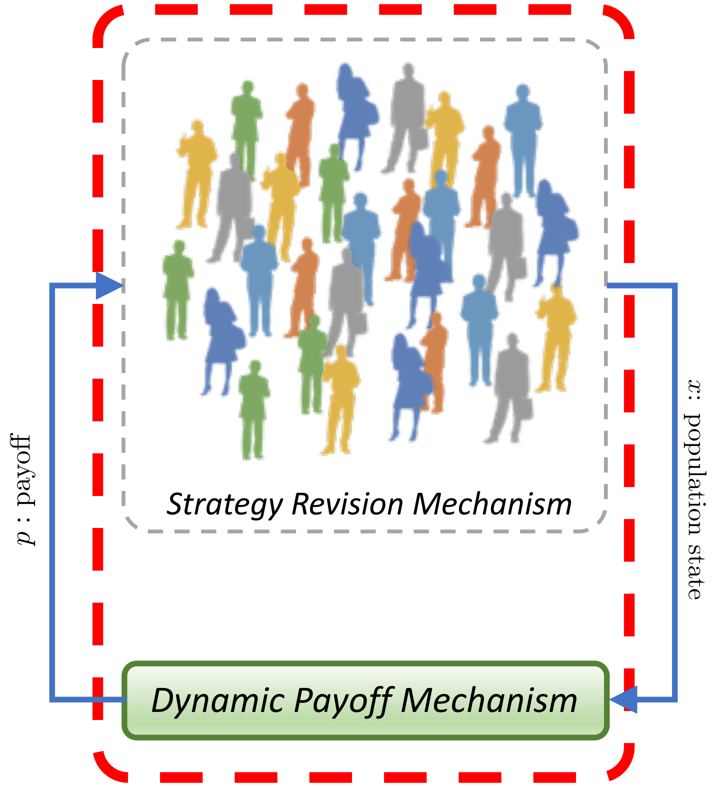
$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \left\{z \in \mathbb{X} \,\middle|\, \mathcal{V}\left(z, \bar{\mathcal{F}}(z)\right) = 0\right\}$$

  In this study,

$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \begin{cases} \mathbb{NE}\left(\bar{\mathcal{F}}\right) & \text{for IPC and EPT EDMs (e.g. BNN EDM)} \\ \mathbb{PE}\left(\bar{\mathcal{F}}, \mathcal{Q}\right) & \text{for PBR EDMs}^{\dagger} \text{ (e.g. Logit EDM)} \end{cases}$$

  * $\mathbb{NE}\left(\bar{\mathcal{F}}\right) = \left\{z \in \mathbb{X} \,|\, z_i > 0 \implies \bar{\mathcal{F}}_i(z) = \max_{1 \le j \le n} \bar{\mathcal{F}}_j(z)\right\}$
  * $\mathbb{PE}\left(\bar{\mathcal{F}}, \mathcal{Q}\right) = \left\{z \in \mathbb{X} \,|\, z_i > 0 \implies \tilde{\mathcal{F}}_i(z) = \max_{1 \le j \le n} \tilde{\mathcal{F}}_j(z)\right\}$,
    where $\tilde{\mathcal{F}} = \bar{\mathcal{F}} - \nabla \mathcal{Q}$.

  † PBR EDMs: $\dot{x}(t) = \operatorname{argmax}_{y \in \mathbb{X}} \left(\sum_{j=1}^{n} y_j p_j(t) - \mathcal{Q}(y)\right) - x(t)$



– $\bar{\mathcal{F}}$: stationary population game

Stability Concept

- **Mean closed loop model:**

$$\begin{aligned}\dot{q}(t) &= \mathcal{G}\left(q(t), x(t)\right)\\ p(t) &= \mathcal{H}\left(q(t), x(t)\right)\\ \dot{x}(t) &= \mathcal{V}\left(x(t), p(t)\right)\end{aligned}$$

  - Equilibrium set:

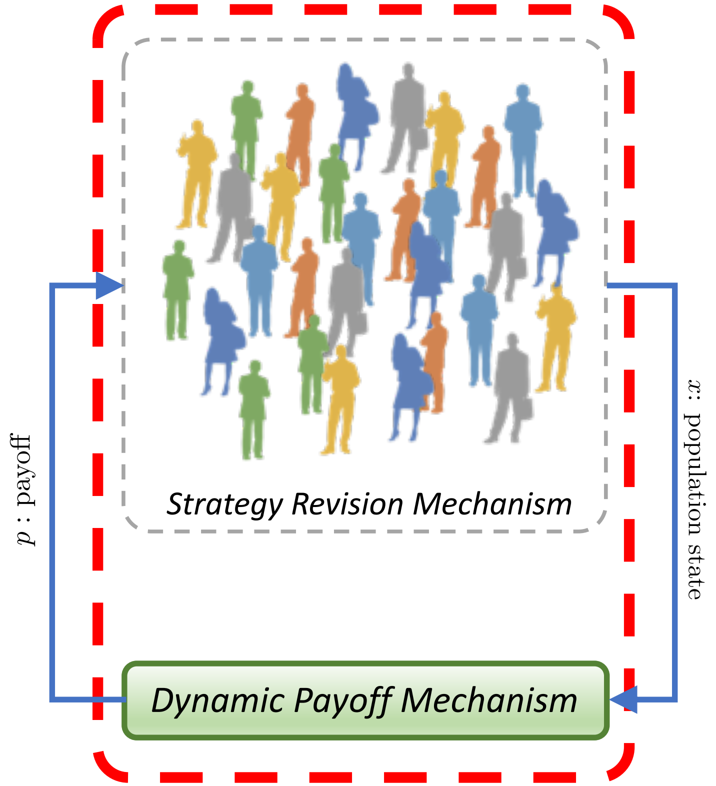
$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \left\{z \in \mathbb{X} \,\middle|\, \mathcal{V}\left(z, \bar{\mathcal{F}}(z)\right) = 0\right\}$$

  In this study,

$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \begin{cases}\mathbb{NE}\left(\bar{\mathcal{F}}\right) & \text{for IPC and EPT EDMs}\\ \mathbb{PE}\left(\bar{\mathcal{F}}, \mathcal{Q}\right) & \text{for PBR EDMs}^{\dagger}\end{cases}$$

(Global Asymptotic Stability) *The set $\mathbb{E}(\bar{\mathcal{F}})$ is globally asymptotically stable if the set $\mathbb{A}$ of stationary points is globally attractive and Lyapunov stable.*

$*$ $\mathbb{A} = \left\{(s, z) \in \mathbb{R}^n \times \mathbb{X} \,\middle|\, \mathcal{V}\left(z, \mathcal{H}\left(s, z\right)\right) = 0 \text{ and } \mathcal{H}(s, z) = \bar{\mathcal{F}}(z)\right\}$



– $\bar{\mathcal{F}}$: stationary population game

Stability Concept

- **Mean closed loop model:**

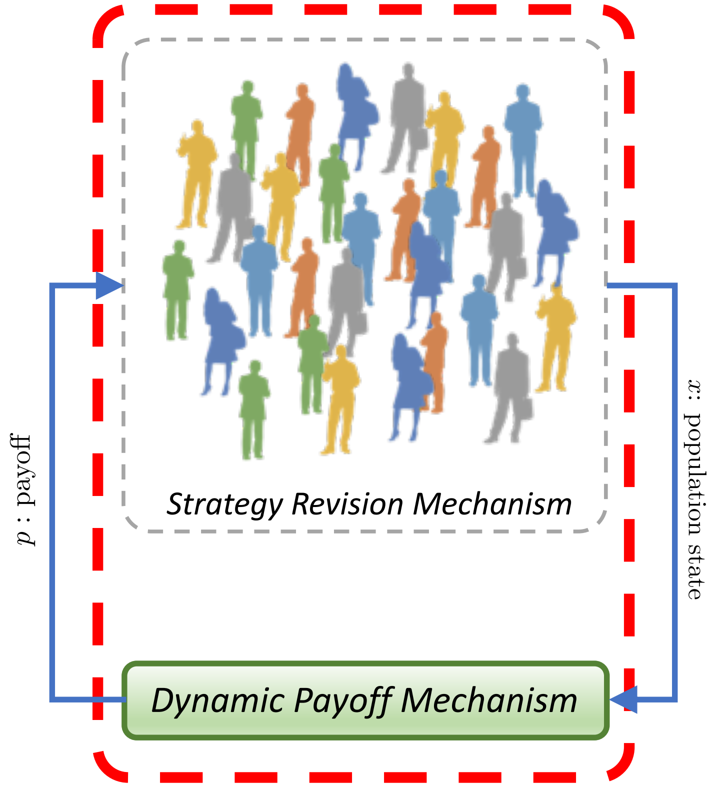
$$\begin{aligned}\dot{q}(t) &= \mathcal{G}\left(q(t), x(t)\right)\\ p(t) &= \mathcal{H}\left(q(t), x(t)\right)\\ \dot{x}(t) &= \mathcal{V}\left(x(t), p(t)\right)\end{aligned}$$

  - Equilibrium set:

$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \left\{z \in \mathbb{X} \,\middle|\, \mathcal{V}\left(z, \bar{\mathcal{F}}(z)\right) = 0\right\}$$

    In this study,

$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \begin{cases}\mathbb{NE}\left(\bar{\mathcal{F}}\right) & \text{for IPC and EPT EDMs}\\ \mathbb{PE}\left(\bar{\mathcal{F}}, \mathcal{Q}\right) & \text{for PBR EDMs}^{\dagger}\end{cases}$$



– $\bar{\mathcal{F}}$: stationary population game

***(Global Asymptotic Stability)** The set $\mathbb{E}(\bar{\mathcal{F}})$ is globally asymptotically stable if the set $\mathbb{A}$ of stationary points is globally attractive and Lyapunov stable.*

When does the equilibrium set attain global asymptotic stability ?

Applications to Finite Population Decision Problems

- Multi-agent learning/decision-making problem

Cost to minimize: $\mathbb{E}\left[\int_0^{\infty} f\left(X^N(\tau)\right)\, \mathrm{d}\tau\right]$

Instantaneous reward: $r(t) = \nabla f\left(X^N(t)\right)$

- Multi-robot manipulation

Model on waste volume:

$$\dot{Q}_i^N(t) = \begin{cases} -RX_i^N(t) + w_i & \text{if } Q_i^N(t) > 0 \\ \max\left\{0, -RX_i^N(t) + w_i\right\} & \text{otherwise} \end{cases}$$

  - $Q_i^N(t)$: *Waste volume in cell i*
  - $X_i^N(t)$: *Portion of the robots working in cell i*
  - w_i: *Increase rate of the waste in cell i*

Finite Population Approximation

- **Notation for the finite population case:**
  - $\{1, \cdots, n\}$: Strategy set
  - $X^N = (X_1^N, \cdots, X_n^N)$: Population state defined as

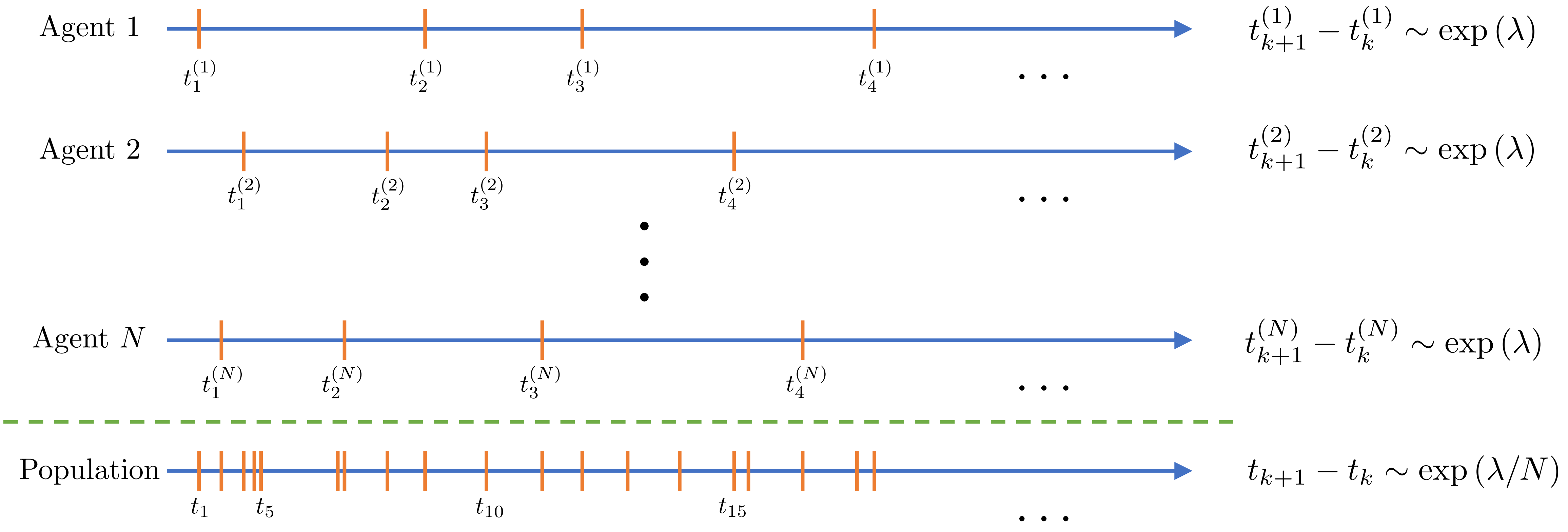
$$X_i^N = \frac{\#\text{ of agents adopting } i}{N}$$

  - $P^N = (P_1^N, \cdots, P_n^N)$: Payoff vector assigned to X^N

- **Poisson clock and jump times for individual agents:**
  - Γ_α: Poisson process (Poisson Clock) associated with each agent α
  - $\mathbb{T}_\alpha$: Jump times of Γ_α defined as $\mathbb{T}_\alpha = \{t_1, t_2, \cdots\}$ for which
    * At each $t_k \in \mathbb{T}_\alpha$, the agent can revise its strategy
    * $t_{k+1} - t_k \sim \exp(\lambda)$ $\left(\mathbf{E}\left[t_{k+1} - t_k\right] = \lambda\right)$

Finite Population Approximation

- **Poisson clock and jump times for population:**
  - Γ: Poisson process (Poisson Clock) associated with the entire population defined by $\Gamma = \sum_{\alpha=1}^{N} \Gamma_\alpha$
  - $\mathbb{T}$: Jump times of Γ defined as $\mathbb{T} = \{t_1, t_2, \cdots\}$ for which
    - At each $t_k \in \mathbb{T}$, one (randomly selected) agent revises its strategy
    - $t_{k+1} - t_k \sim \exp(\lambda/N)$ $\left(\mathbf{E}\left[t_{k+1} - t_k\right] = \lambda/N\right)$



Finite Population Approximation

Suppose that at every $t \in \mathbb{T}$, (Q^N, P^N, X^N) changes according the the following models.

– Payoff dynamics model (for finite population):

$$\lim_{\delta \to 0} \mathbf{E}\left[\frac{1}{\delta} \left(Q_i^N(t+\delta) - Q_i^N(t) \right) \Big| \, (X^N(t), Q^N(t)) = (z, s) \right] = \mathcal{G}_i(s, z)$$

where $P^N(t) = \mathcal{H}\left(Q^N(t), X^N(t)\right)$.

– Evolutionary dynamics model (for finite population):

$$\lim_{\delta \to 0} \mathbf{E}\left[\frac{1}{\delta} \left(X_i^N(t+\delta) - X_i^N(t) \right) \Big| \, (X^N(t), P^N(t)) = (z, r) \right] = \mathcal{V}_i(z, r)$$

Mean closed loop model:

$$\begin{aligned} \dot{q}(t) &= \mathcal{G}\left(q(t), x(t)\right) \\ p(t) &= \mathcal{H}\left(q(t), x(t)\right) \\ \dot{x}(t) &= \mathcal{V}\left(x(t), p(t)\right) \end{aligned}$$

– $\bar{\mathcal{F}}$: Stationary population game

– $\mathbb{E}\left(\bar{\mathcal{F}}\right)$: Equilibrium set

Finite Population Approximation

Suppose that at every $t \in \mathbb{T}$, $\left(Q^N, P^N, X^N\right)$ changes according the the following models.

- Payoff dynamics model (for finite population):

$$\lim_{\delta \to 0} \mathbf{E}\left[\frac{1}{\delta}\left(Q_i^N(t+\delta) - Q_i^N(t)\right) \,\middle|\, \left(X^N(t), Q^N(t)\right) = (z, s)\right] = \mathcal{G}_i(s, z)$$

  where $P^N(t) = \mathcal{H}\left(Q^N(t), X^N(t)\right)$.

- Evolutionary dynamics model (for finite population):

$$\lim_{\delta \to 0} \mathbf{E}\left[\frac{1}{\delta}\left(X_i^N(t+\delta) - X_i^N(t)\right) \,\middle|\, \left(X^N(t), P^N(t)\right) = (z, r)\right] = \mathcal{V}_i(z, r)$$

Mean closed loop model:

$$\begin{aligned} \dot{q}(t) &= \mathcal{G}\left(q(t), x(t)\right) \\ p(t) &= \mathcal{H}\left(q(t), x(t)\right) \\ \dot{x}(t) &= \mathcal{V}\left(x(t), p(t)\right) \end{aligned}$$

- $\bar{\mathcal{F}}$: Stationary population game
- $\mathbb{E}\left(\bar{\mathcal{F}}\right)$: Equilibrium set

If $\left(X^N(0), Q^N(0)\right) = (x(0), q(0))$, *for any* $\bar{t}, \delta > 0$, *it holds that*

$$\mathbf{P}\left(\sup_{0 \le t \le \bar{t}} \left\| \left(X^N(t), Q^N(t)\right) - (x(t), q(t)) \right\| > \delta\right) \xrightarrow{N \to \infty} 0$$

Ref. on Finite Population Approximation:
* R. T. Boylan, Journal of Economic Theory, 1992
* M. Benaim and J. W. Weibull, Econometrica, 2003

Finite Population Approximation

Suppose that at every $t \in \mathbb{T}$, $\left(Q^N, P^N, X^N\right)$ changes according the the following models.

– Payoff dynamics model (for finite population):

$$\lim_{\delta \to 0} \mathbf{E}\left[\frac{1}{\delta}\left(Q_i^N(t+\delta) - Q_i^N(t)\right) \,\middle|\, \left(X^N(t), Q^N(t)\right) = (z, s)\right] = \mathcal{G}_i(s, z)$$

where $P^N(t) = \mathcal{H}\left(Q^N(t), X^N(t)\right)$.

– Evolutionary dynamics model (for finite population):

$$\lim_{\delta \to 0} \mathbf{E}\left[\frac{1}{\delta}\left(X_i^N(t+\delta) - X_i^N(t)\right) \,\middle|\, \left(X^N(t), P^N(t)\right) = (z, r)\right] = \mathcal{V}_i(z, r)$$

– Exit time from a set $\mathbb{U} \subset \mathbb{R}^n \times \mathbb{R}^n$:

$$\tau^N(\mathbb{U}) = \inf\{\tau \geq 0 \,|\, \left(X^N(\tau), Q^N(\tau)\right) \notin \mathbb{U}\}$$

If $\mathbb{E}\left(\bar{\mathcal{F}}\right)$ is asymptotically stable, for any neighborhood set $\mathbb{U}$ of $\mathbb{A}$, it holds that

$$\mathbf{P}\left(\liminf_{N \to \infty} \tau^N(\mathbb{U}) = +\infty\right) = 1 \text{ if } \left(X^N(0), Q^N(t)\right) \in \mathbb{U}$$

$* \ \mathbb{A} = \left\{(s, z) \in \mathbb{R}^n \times \mathbb{X} \,\middle|\, \mathcal{V}\left(z, \mathcal{H}(s, z)\right) = 0 \text{ and } \mathcal{H}(s, z) = \bar{\mathcal{F}}(z)\right\}$

Mean closed loop model:

$$\dot{q}(t) = \mathcal{G}\left(q(t), x(t)\right)$$
$$p(t) = \mathcal{H}\left(q(t), x(t)\right)$$
$$\dot{x}(t) = \mathcal{V}\left(x(t), p(t)\right)$$

– $\bar{\mathcal{F}}$: Stationary population game

– $\mathbb{E}\left(\bar{\mathcal{F}}\right)$: Equilibrium set

Ref. on Finite Population Approximation:
* R. T. Boylan, Journal of Economic Theory, 1992
* M. Benaim and J. W. Weibull, Econometrica, 2003

Applications – multi-agent learning/decision-making

Consider that, at each time instance $t \in \mathbb{T}$, one agent is uniformly randomly selected.

1. For every i, the agent examines $r_i(t)$ and updates

$$Q_i^N(t) \leftarrow \frac{\alpha}{N} Q_i^N(t) + \left(1 - \frac{\alpha}{N}\right) r_i(t)$$

   In expectation, it holds that

$$\lim_{\delta \to 0} \mathbf{E}\left[\frac{1}{\delta}\left(Q_i^N(t+\delta) - Q_i^N(t)\right) \,\middle|\, \left(X^N(t), Q^N(t)\right) = (z, s)\right] = \alpha\left(\nabla_i f(z) - s_i\right)$$

2. The agent revises to strategy j with probability $\frac{\exp P_j^N(t)}{\sum_{l=1}^n \exp P_l^N(t)}$, where $P^N(t) = Q^N(t)$.
   In expectation, it holds that

$$\lim_{\delta \to 0} \mathbf{E}\left[\frac{1}{\delta}\left(X_j^N(t+\delta) - X_j^N(t)\right) \,\middle|\, \left(X^N(t), P^N(t)\right) = (z, r)\right] = \frac{\exp P_j^N(t)}{\sum_{l=1}^n \exp P_l^N(t)} - z_j$$

Cost to minimize: $\mathbb{E}\left[\int_0^\infty f\left(X^N(\tau)\right)\, \mathrm{d}\tau\right]$

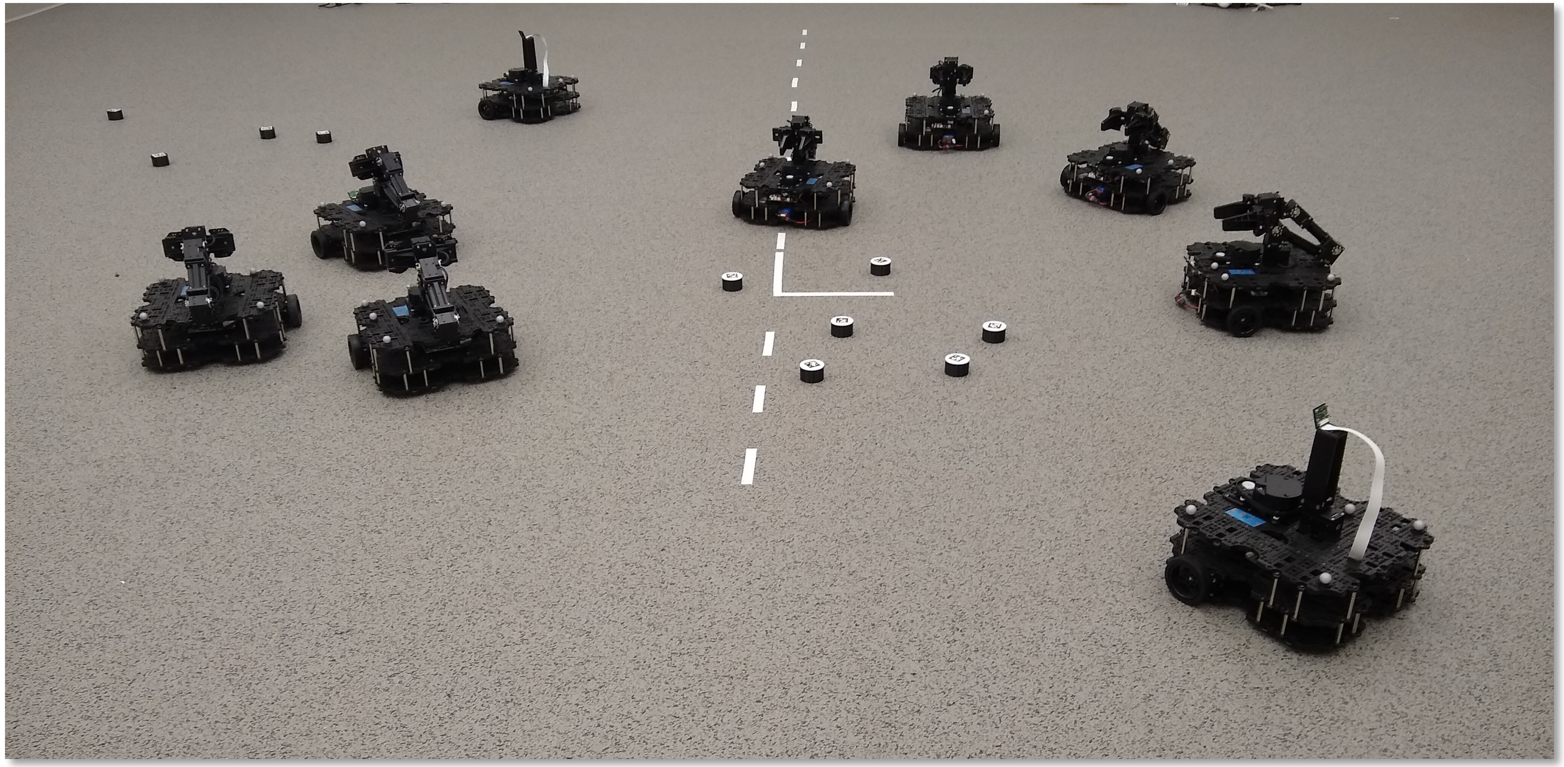
Instantaneous reward: $r(t) = \nabla f\left(X^N(t)\right)$

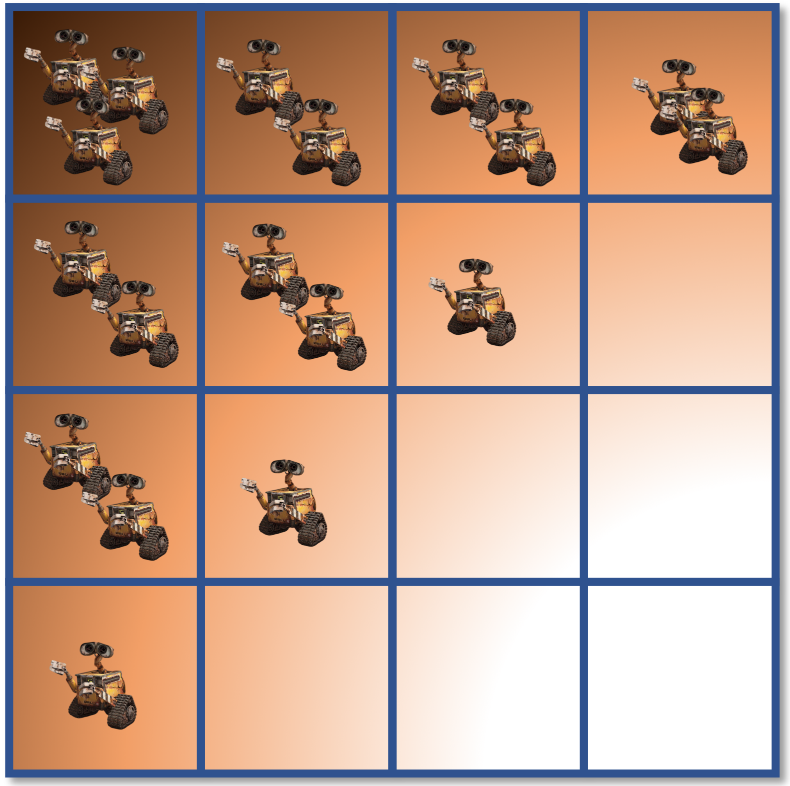
Mean closed loop model:

$$\dot{q}_i(t) = \alpha\left(\nabla f_i(x(t)) - q_i(t)\right)$$
$$p_i(t) = q_i(t) \qquad \textit{(Smoothing PDM)}$$

$$\dot{x}_i(t) = \frac{\exp p_i(t)}{\sum_{j=1}^n \exp p_j(t)} - x_i(t) \qquad \textit{(Logit EDM)}$$

Applications – multi-robot manipulation





* Joint work with Naomi Leonard (Princeton University)

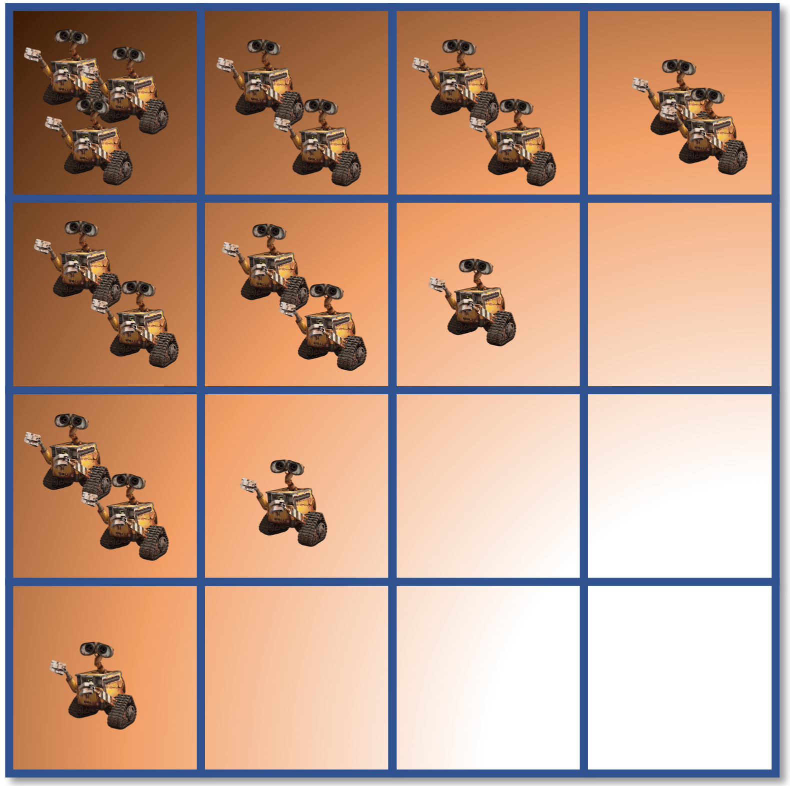
Applications – multi-robot manipulation

- Modeling:

$$\dot{Q}_i^N(t) = \begin{cases} -RX_i^N(t) + w_i & \text{if } Q_i^N(t) > 0 \\ \max\left\{0, -RX_i^N(t) + w_i\right\} & \text{otherwise} \end{cases}$$

  - $Q_i^N(t)$: Waste volume in cell i
  - $X_i^N(t)$: Portion of the robots working in cell i
  - w_i: Increase rate of the waste in cell i

- Planning policy: At each time instance $t \in \mathbb{T}$, a (uniformly) randomly selected robot moves to cell j with probability $\frac{\exp P_j(t)}{\sum_{l=1}^n \exp P_l(t)}$, where $P^N(t) = Q^N(t)$.



Mean closed loop model:

$$\dot{q}_i(t) = \begin{cases} -Rx_i(t) + w_i & \text{if } q_i(t) > 0 \\ \max\left\{0, -Rx_i(t) + w_i\right\} & \text{otherwise} \end{cases}$$

$$p_i(t) = q_i(t)$$

$$\dot{x}_i(t) = \frac{\exp p_i(t)}{\sum_{j=1}^n \exp p_j(t)} - x_i(t)$$

Recap

- **Mean closed loop model:**

$$\left.\begin{aligned}\dot{q}(t) &= \mathcal{G}\left(q(t), x(t)\right)\\ p(t) &= \mathcal{H}\left(q(t), x(t)\right)\end{aligned}\right\} \text{(PDM)}$$

$$\dot{x}(t) = \mathcal{V}\left(x(t), p(t)\right) \quad \text{— (EDM)}$$

  – Equilibrium set:

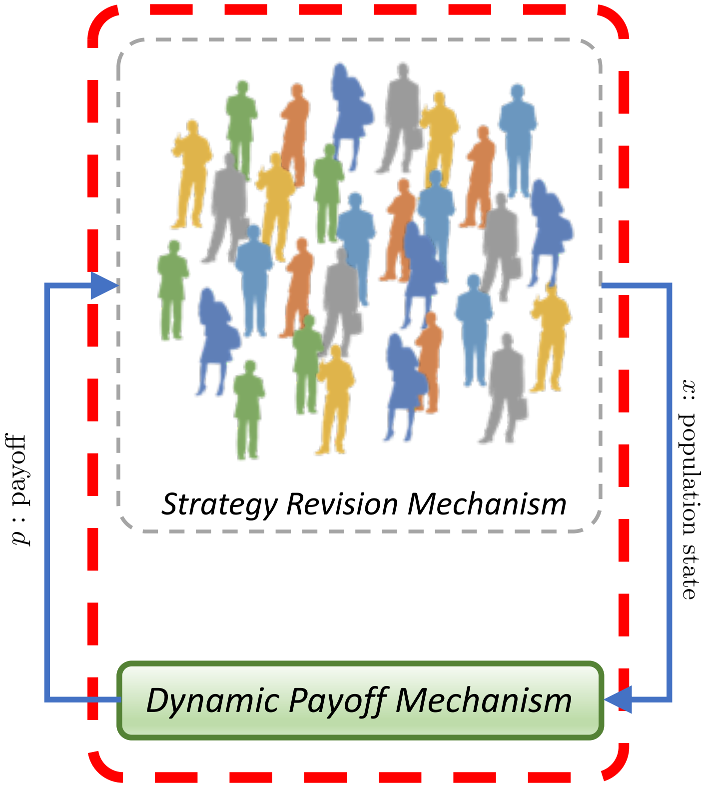
$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \left\{ z \in \mathbb{X} \,\middle|\, \mathcal{V}\left(z, \bar{\mathcal{F}}(z)\right) = 0 \right\}$$

  In this study,

$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \begin{cases} \mathbb{NE}\left(\bar{\mathcal{F}}\right) & \text{for EPT and IPC EDMs} \\ \mathbb{PE}\left(\bar{\mathcal{F}}, \mathcal{Q}\right) & \text{for PBR EDMs}^{\dagger} \end{cases}$$

(Global Asymptotic Stability) *The set $\mathbb{E}(\bar{\mathcal{F}})$ is globally asymptotically stable if the set $\mathbb{A}$ of stationary points is globally attractive and Lyapunov stable.*

When does the equilibrium set attain global asymptotic stability ?



– $\bar{\mathcal{F}}$: stationary population game

Passivity-based Stability Analysis

- **δ-Passivity for evolutionary dynamics models:**

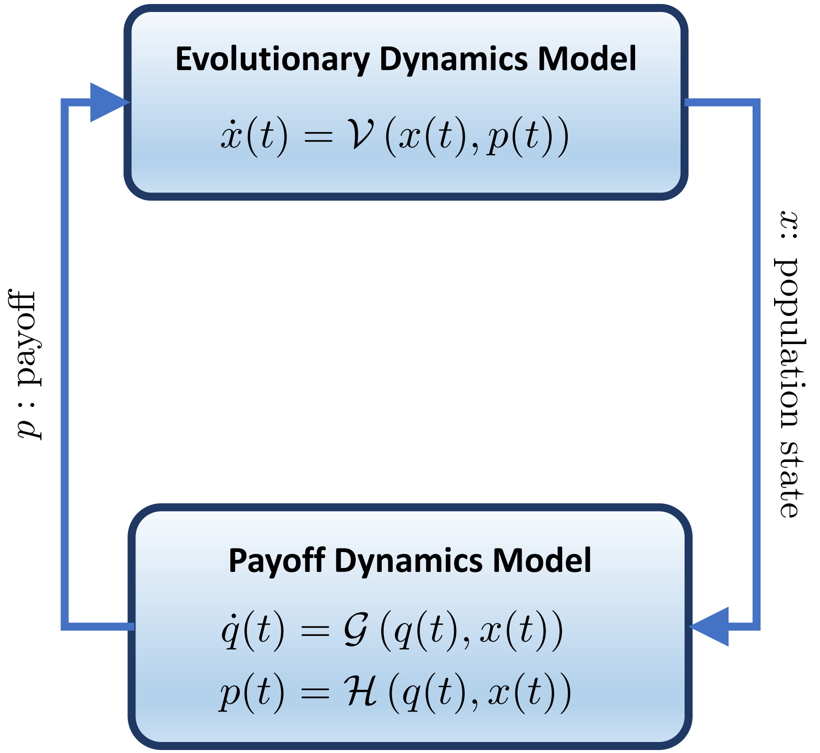
$$\mathcal{S}\big(x(t), p(t)\big) - \mathcal{S}\big(x(t_0), p(t_0)\big) \le \int_{t_0}^{t} \dot{x}^T(\tau)\dot{p}(\tau)\,\mathrm{d}\tau,\ \ t \ge t_0$$

  - $\mathcal{S} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_+$: δ-storage function (continuously differentiable map)

- **δ-Antipassivity for payoff dynamics models:**

$$\mathcal{L}\left(x(t), q(t)\right) - \mathcal{L}\left(x(t_0), q(t_0)\right) \le -\int_{t_0}^{t} \dot{x}^T(\tau)\dot{p}(\tau)\,\mathrm{d}\tau,\ \ t \ge t_0$$

  - $\mathcal{L} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_+$: δ-antistorage function (continuously differentiable map)



- $\bar{\mathcal{F}}$: Stationary population game
- $\mathbb{E}\left(\bar{\mathcal{F}}\right)$: Equilibrium set $\left(\mathbb{NE}\left(\bar{\mathcal{F}}\right)\ \text{or}\ \mathbb{PE}\left(\bar{\mathcal{F}}, \mathcal{Q}\right)\right)$

Passivity-based Stability Analysis

- δ-Passivity for evol

$$\mathcal{S}\left(x(t), p(t$$

  – $\mathcal{S} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}$

- δ-Antipassivity for

$$\mathcal{L}\left(x(t), q(t)\right.$$

  – $\mathcal{L} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_+$

ary Dynamics Model

$\mathcal{V}\left(x(t), p(t)\right)$

x: population state

Dynamics Model

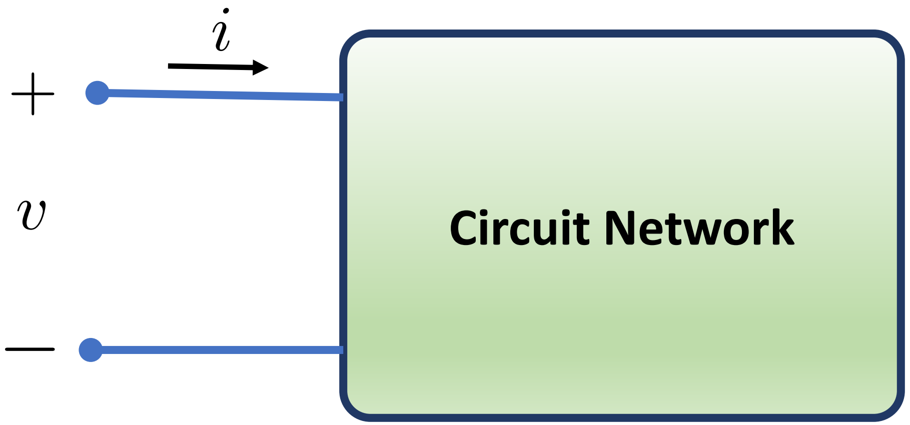
$\mathcal{G}\left(q(t), x(t)\right)$

$\mathcal{H}\left(q(t), x(t)\right)$

nary population game

quilibrium set

or $\mathbb{PE}\left(\bar{\mathcal{F}}, \mathcal{Q}\right)$)



Passivity in the circuit network means the change in stored energy is less than the supplied energy:

$$E(t) - E(0) \leq \int_0^t v(\tau) i(\tau) \, \mathrm{d}\tau$$

Passivity-based Stability Analysis

- **δ-Passivity for evolutionary dynamics models:**

$$\mathcal{S}\big(x(t), p(t)\big) - \mathcal{S}\big(x(t_0), p(t_0)\big) \leq \int_{t_0}^{t} \dot{x}^T(\tau)\dot{p}(\tau)\,\mathrm{d}\tau,\ \ t \geq t_0$$

  - $\mathcal{S} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_+$: δ-storage function (continuously differentiable map)

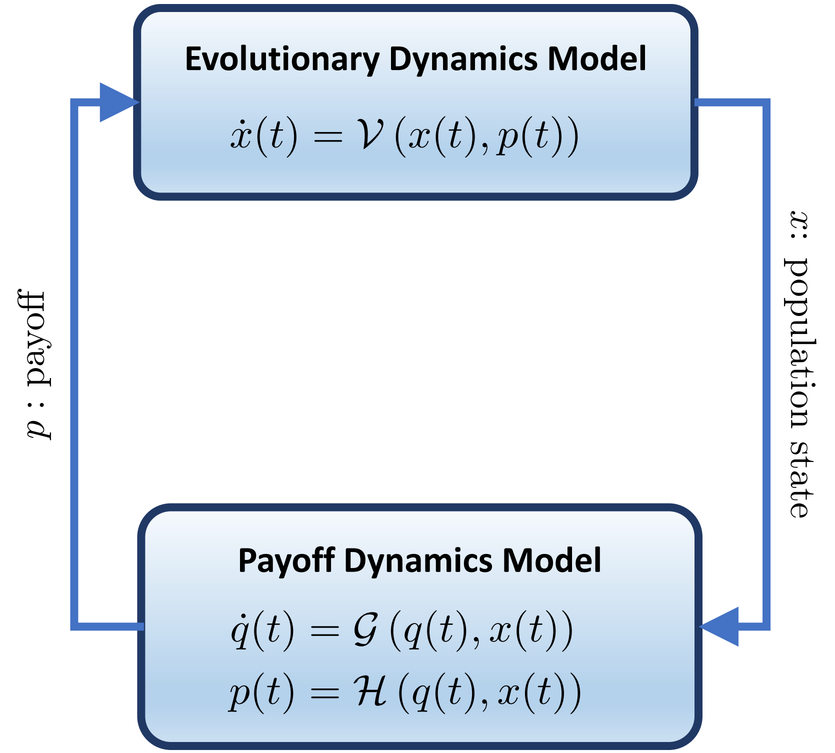
- **δ-Antipassivity for payoff dynamics models:**

$$\mathcal{L}\left(x(t), q(t)\right) - \mathcal{L}\left(x(t_0), q(t_0)\right) \leq -\int_{t_0}^{t} \dot{x}^T(\tau)\dot{p}(\tau)\,\mathrm{d}\tau,\ \ t \geq t_0$$

  - $\mathcal{L} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_+$: δ-antistorage function (continuously differentiable map)

Stability Result:

$$\boxed{\mathcal{S}\left(x(t), p(t)\right) + \mathcal{L}\left(x(t), q(t)\right) \overset{t\to\infty}{\longrightarrow} 0 \implies \mathbb{E}\left(\bar{\mathcal{F}}\right) \ \textit{is asymptotically stable}}$$



- $\bar{\mathcal{F}}$: Stationary population game
- $\mathbb{E}\left(\bar{\mathcal{F}}\right)$: Equilibrium set $\left(\mathbb{NE}\left(\bar{\mathcal{F}}\right) \text{ or } \mathbb{PE}\left(\bar{\mathcal{F}}, \mathcal{Q}\right)\right)$

Passivity-based Stability Analysis

- δ-Passivity for e

$\mathcal{S}(x(t),$

– $\mathcal{S} : \mathbb{X} \times \mathbb{R}^n \rightarrow$

- δ-Antipassivity f

$\mathcal{L}(x(t), q$

– $\mathcal{L} : \mathbb{X} \times \mathbb{R}^n \rightarrow$

Stability Res

$\mathcal{S}(x(t), p(t)) +$

arXiv:1903.02018

Payoff Dynamics Model and Evolutionary Dynamics Model: Feedback and Convergence to Equilibria

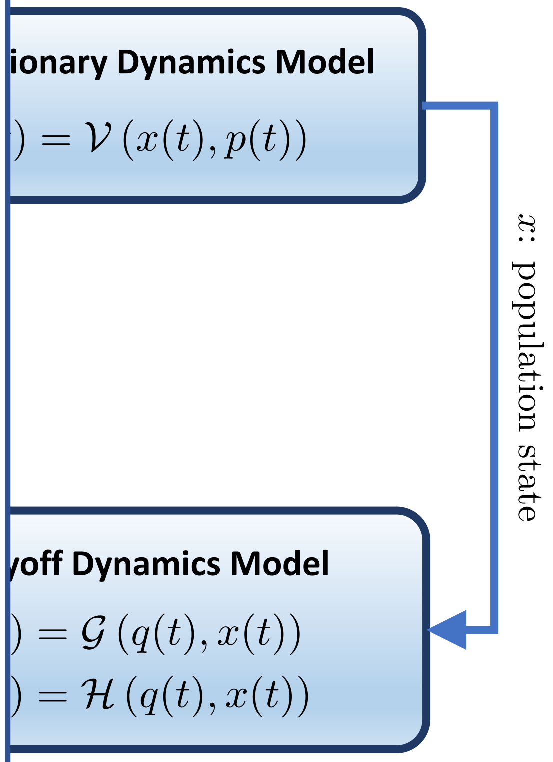
Shinkyu Park, Nuno C. Martins, and Jeff S. Shamma

***Abstract*—We consider that at every instant each member of a population, which we refer to as an agent, selects one strategy out of a finite set. The agents are nondescript, and their strategy choices are described by the so-called population state vector, whose entries are the portions of the population selecting each strategy. Likewise, each entry of the so-called payoff vector is the reward attributed to a strategy. We consider that a finite-dimensional nonlinear dynamical system, denoted as payoff dynamics model (PDM), specifies a mechanism that determines the payoff as a causal map of the population state. A bounded-rationality protocol, which is often inspired on evolutionary biology principles, governs how each agent continually revises its strategy based on complete or partial knowledge of the population state and payoff. The population is protocol-homogeneous but is otherwise strategy-heterogeneous considering that the agents are allowed to select distinct strategies concurrently. A stochastic mechanism determines the instants when agents revise their strategies, but we consider that the population is large enough that, with high probability, the population state can be approximated with arbitrary accuracy uniformly over any finite horizon by a so-called (deterministic) mean population state. We propose an approach that takes advantage of passivity principles to obtain sufficient conditions determining, for a given protocol and PDM, when the mean population state is guaranteed to converge to a meaningful set of equilibria, which could be either an appropriately defined extension of Nash's for the PDM or a perturbed version of it. By generalizing and unifying previous work, our framework also provides a foundation for future work. We specialize our results for well-known protocols and a class of PDM that includes as particular cases the payoff mechanisms proposed in previous work to model learning, inertia, and anticipation.**

methods put forth in this article can be readily adapted to the multi-population case.

We adopt the evolutionary dynamic paradigm well-documented in [1]–[3] according to which the members of the population, which we call agents, repeatedly revise their strategy choices according to a given mechanism modeled by a so-called revision protocol (protocol for short). The population is protocol-homogeneous because all of its agents follow the same revision mechanism, but is otherwise strategy-heterogeneous considering that the agents are allowed to concurrently select distinct strategies. Although mixed strategies are not allowed as each agent chooses exactly one strategy at every instant, the protocol may be probabilistic when the revision mechanism involves randomization. The identity of each agent is unimportant, and consequently all the information that is relevant, at any given instant, is encapsulated in the so-called *population state* vector whose entries quantify the portions of the population adopting each strategy. Henceforth, we refer to the set of all possible population states, or equivalently the state space, as the strategy profile set. Hence, every strategy profile is a vector with n nonnegative entries adding up to the so-called population mass m.

In most related prior work, a so-called *population game* [4] determines at any given instant the n-dimensional payoff vector, which quantifies the reward associated with each strategy, as a function of the population state. A population game may represent a pricing scheme that is implemented by a



ationary population game

Equilibrium set

$\bar{\mathcal{F}})$ or $\mathbb{PE}(\bar{\mathcal{F}}, \mathcal{Q}))$

Passivity-based Stability Analysis

- **δ-Passivity with surplus η^* for evolutionary dynamics models:**

$$\mathcal{S}\big(x(t), p(t)\big) - \mathcal{S}\big(x(t_0), p(t_0)\big) \leq \int_{t_0}^{t} \left[\dot{x}^T(\tau)\dot{p}(\tau) - \eta\, \dot{x}^T(t)\dot{x}(t)\right] \mathrm{d}\tau,\ \ t \geq t_0$$

  - $\mathcal{S} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_+$: δ-storage function (continuously differentiable map)
  - η: Positive constant $(\eta \leq \eta^*)$

- **δ-Antipassivity with deficit ν^* for payoff dynamics models:**

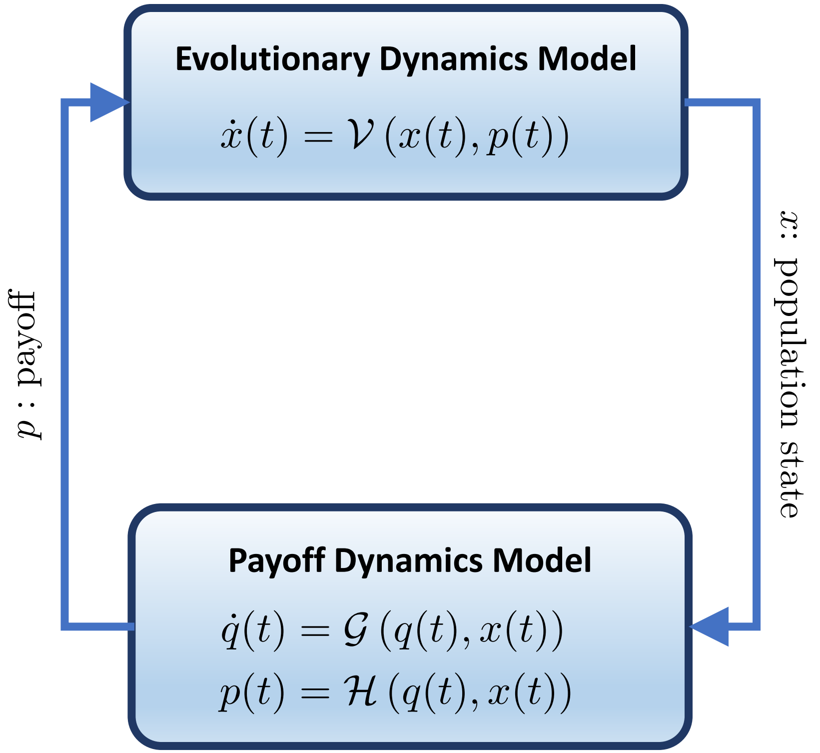
$$\mathcal{L}\left(x(t), q(t)\right) - \mathcal{L}\left(x(t_0), q(t_0)\right) \leq \int_{t_0}^{t} \left[-\dot{x}^T(\tau)\dot{p}(\tau) + \nu\, \dot{x}^T(\tau)x(\tau)\right] \mathrm{d}\tau,\ \ t \geq t_0$$

  - $\mathcal{L} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_+$: δ-antistorage function (continuously differentiable map)
  - ν: Positive constant $(\nu \leq \nu^*)$

Stability Result:

If surplus of passivity in EDM exceeds lack of passivity in PDM ($\nu^* < \eta^*$),

$$\mathcal{S}\left(x(t), p(t)\right) + \mathcal{L}\left(x(t), q(t)\right) \xrightarrow{t\to\infty} 0 \implies \mathbb{E}\left(\bar{\mathcal{F}}\right) \ \textit{is asymptotically stable}$$



- $\bar{\mathcal{F}}$: Stationary population game
- $\mathbb{E}\left(\bar{\mathcal{F}}\right)$: Equilibrium set $\left(\mathbb{NE}\left(\bar{\mathcal{F}}\right)\ \text{or}\ \mathbb{PE}\left(\bar{\mathcal{F}}, \mathcal{Q}\right)\right)$

Weak Passivity and Stability Results

- **δ-Passivity for evolutionary dynamics models:**

$$\mathcal{S}\big(x(t),p(t)\big) - \mathcal{S}\big(x(t_0),p(t_0)\big) \leq \int_{t_0}^{t} \dot{x}^T(\tau)\dot{p}(\tau)\,\mathrm{d}\tau,\ t \geq t_0$$

  – $\mathcal{S} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_+$: δ-storage function (continuously differentiable map)

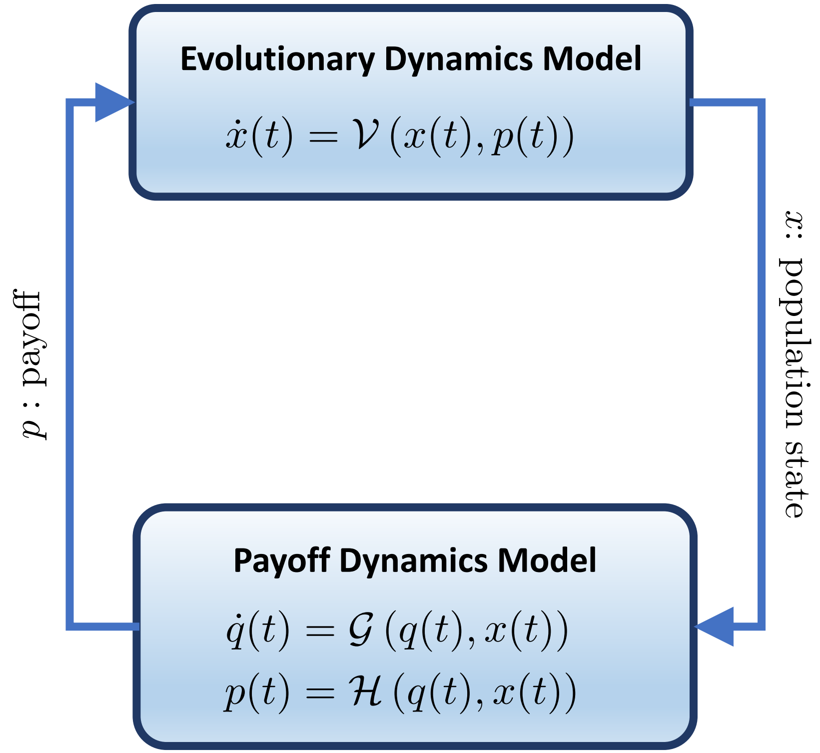
- **Weak δ-antipassivity for payoff dynamics models:**

$$\alpha \leq -\int_{t_0}^{t} \dot{x}^T(\tau)\dot{p}(\tau)\,\mathrm{d}\tau,\ t \geq t_0$$

  – α : Constant

Stability Result:

$$\boxed{\begin{array}{l}\mathcal{S}\left(x(t),p(t)\right) \overset{t\to\infty}{\longrightarrow} 0 \\ \quad \Longrightarrow \mathbb{E}\left(\bar{\mathcal{F}}\right)\ \textit{is attractive (but not necessarily Lyapunov stable)}\end{array}}$$



– $\bar{\mathcal{F}}$: Stationary population game

– $\mathbb{E}\left(\bar{\mathcal{F}}\right)$: Equilibrium set $\left(\mathbb{NE}\left(\bar{\mathcal{F}}\right) \text{ or } \mathbb{PE}\left(\bar{\mathcal{F}},\mathcal{Q}\right)\right)$

STABILITY ANALYSIS IN EVOLUTIONARY GAME THEORY

- **Evolutionary dynamics and stable (contractive) games:**

$$\dot{x}_i(t) = \sum_{j=1}^{n} x_j(t)\mathcal{T}_{ji}\left(x(t), \mathcal{F}(x(t))\right) - x_i(t)\sum_{j=1}^{n} \mathcal{T}_{ij}\left(x(t), \mathcal{F}(x(t))\right)$$

  – Population game $\mathcal{F} : \mathbb{X} \to \mathbb{R}^n$ is called <u>contractive</u> if

$$(y - z)^T \left(\mathcal{F}(y) - \mathcal{F}(z)\right) \leq 0, \ \forall y, z \in \mathbb{X}$$

    ∗ (Example) Concave potential game:

$$\mathcal{F}(z) = \nabla f(z), \ \forall z \in \mathbb{X}$$

    where $f : \mathbb{X} \to \mathbb{R}$ is a concave function.

  – Under EPT, IPC, PBR dynamics, the Nash equilibria are asymptotically stable.

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 144 (2009) 1665–1693.e4

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Stable games and their dynamics

Josef Hofbauer [a], William H. Sandholm [b,*]

[a] *Department of Mathematics, University of Vienna, Nordbergstrasse 15, A-1090 Vienna, Austria*

[b] *Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, USA*

Received 20 December 2006; final version received 14 January 2009; accepted 26 January 2009

Available online 23 February 2009

Abstract

We study a class of population games called *stable games*. These games are characterized by *self-defeating externalities*: when agents revise their strategies, the improvements in the payoffs of strategies to which revising agents are switching are always exceeded by the improvements in the payoffs of strategies which revising agents are abandoning. We prove that the set of Nash equilibria of a stable game is globally asymptotically stable under a wide range of evolutionary dynamics. Convergence results for stable games are not as general as those for potential games: in addition to monotonicity of the dynamics, integrability of the agents' revision protocols plays a key role.

JEL classification: C72; C73

Keywords: Population games; Evolutionarily stable strategies; Evolutionary dynamics; Global stability; Lyapunov functions

Stability Analysis in Evolutionary Game Theory

- **Evolutionary dynamics and stable (contractive) games:**

$$\dot{x}_i(t) = \sum_{j=1}^{n} x_j(t)\mathcal{T}_{ji}\left(x(t), p(t)\right) - x_i(t)\sum_{j=1}^{n}\mathcal{T}_{ij}\left(x(t), p(t)\right)$$

$$p_i(t) = \mathcal{F}_i(x(t))$$

  - Population game $\mathcal{F} : \mathbb{X} \to \mathbb{R}^n$ is called contractive if

$$(y-z)^T\left(\mathcal{F}(y) - \mathcal{F}(z)\right) \leq 0, \ \forall y, z \in \mathbb{X}$$

    - Contractive population games are δ-antipassive.
  - Under EPT, IPC, PBR dynamics, the Nash equilibria are asymptotically stable.
    - EPT, IPC, PBR dynamics are δ-passive.

Games **2013**, *4*, 561-583; doi:10.3390/g4040561

OPEN ACCESS
games
ISSN 2073-4336
www.mdpi.com/journal/games

Article

Population Games, Stable Games, and Passivity

Michael J. Fox and Jeff S. Shamma *

School of Electrical and Computer Engineering, Georgia Institute of Technology, 777 Atlantic Drive NW, Atlanta, GA 30332, USA

* Author to whom correspondence should be addressed; E-mail: shamma@gatech.edu.

Received: 4 April 2013; in revised form: 3 September 2013 / Accepted: 26 September 2013 / Published: 7 October 2013

Abstract: The class of "stable games", introduced by Hofbauer and Sandholm in 2009, has the attractive property of admitting global convergence to equilibria under many evolutionary dynamics. We show that stable games can be identified as a special case of the feedback-system-theoretic notion of a "passive" dynamical system. Motivated by this observation, we develop a notion of passivity for evolutionary dynamics that complements the definition of the class of stable games. Since interconnections of passive dynamical systems exhibit stable behavior, we can make conclusions about passive evolutionary dynamics coupled with stable games. We show how established evolutionary dynamics qualify as passive dynamical systems. Moreover, we exploit the flexibility of the definition of passive dynamical systems to analyze generalizations of stable games and evolutionary dynamics that include forecasting heuristics as well as certain games with memory.

Keywords: population games; evolutionary games; passivity theory

Stability Analysis in Evolutionary Game Theory

- **Evolutionary dynamics model and payoff dynamics model:**

$$\dot{x}_i(t) = \sum_{j=1}^{n} x_j(t)\mathcal{T}_{ji}\left(x(t), p(t)\right) - x_i(t)\sum_{j=1}^{n}\mathcal{T}_{ij}\left(x(t), p(t)\right)$$

$$\boxed{\begin{aligned} \dot{q}_i(t) &= \mathcal{G}_i\left(q(t), x(t)\right) \\ p_i(t) &= \mathcal{H}_i(q(t), x(t)) \end{aligned}}$$

Payoff Dynamics Model

- **δ-Passivity for evolutionary dynamics models:**

$$\mathcal{S}\big(x(t), p(t)\big) - \mathcal{S}\big(x(t_0), p(t_0)\big) \leq \int_{t_0}^{t} \dot{x}^T(\tau)\dot{p}(\tau)\,\mathrm{d}\tau,\ t \geq t_0$$

- **δ-Antipassivity for payoff dynamics models:**

$$\mathcal{L}\left(x(t), q(t)\right) - \mathcal{L}\left(x(t_0), q(t_0)\right) \leq -\int_{t_0}^{t} \dot{x}^T(\tau)\dot{p}(\tau)\,\mathrm{d}\tau,\ t \geq t_0$$

$$\boxed{\begin{aligned} &\mathcal{S}\left(x(t), p(t)\right) + \mathcal{L}\left(x(t), q(t)\right) \xrightarrow{t\to\infty} 0 \\ &\implies \mathbb{E}\left(\bar{\mathcal{F}}\right) \text{ is asymptotically stable} \end{aligned}}$$

arXiv:1903.02018

Payoff Dynamics Model and Evolutionary Dynamics Model: Feedback and Convergence to Equilibria

Shinkyu Park, Nuno C. Martins, and Jeff S. Shamma

***Abstract*—We consider that at every instant each member of a population, which we refer to as an agent, selects one strategy out of a finite set. The agents are nondescript, and their strategy choices are described by the so-called population state vector, whose entries are the portions of the population selecting each strategy. Likewise, each entry of the so-called payoff vector is the reward attributed to a strategy. We consider that a finite-dimensional nonlinear dynamical system, denoted as payoff dynamics model (PDM), specifies a mechanism that determines the payoff as a causal map of the population state. A bounded-rationality protocol, which is often inspired on evolutionary biology principles, governs how each agent continually revises its strategy based on complete or partial knowledge of the population state and payoff. The population is protocol-homogeneous but is otherwise strategy-heterogeneous considering that the agents are allowed to select distinct strategies concurrently. A stochastic mechanism determines the instants when agents revise their strategies, but we consider that the population is large enough that, with high probability, the population state can be approximated with arbitrary accuracy uniformly over any finite horizon by a so-called (deterministic) mean population state. We propose an approach that takes advantage of passivity principles to obtain sufficient conditions determining, for a given protocol and PDM, when the mean population state is guaranteed to converge to a meaningful set of equilibria, which could be either an appropriately defined extension of Nash's for the PDM or a perturbed version of it. By generalizing and unifying previous work, our framework also provides a foundation for future work. We specialize our results for well-known protocols and a class of PDM that includes as particular cases the payoff mechanisms proposed in previous work to model learning, inertia, and anticipation.**

methods put forth in this article can be readily adapted to the multi-population case.

We adopt the evolutionary dynamic paradigm well-documented in [1]–[3] according to which the members of the population, which we call agents, repeatedly revise their strategy choices according to a given mechanism modeled by a so-called revision protocol (protocol for short). The population is protocol-homogeneous because all of its agents follow the same revision mechanism, but is otherwise strategy-heterogeneous considering that the agents are allowed to concurrently select distinct strategies. Although mixed strategies are not allowed as each agent chooses exactly one strategy at every instant, the protocol may be probabilistic when the revision mechanism involves randomization. The identity of each agent is unimportant, and consequently all the information that is relevant, at any given instant, is encapsulated in the so-called *population state* vector whose entries quantify the portions of the population adopting each strategy. Henceforth, we refer to the set of all possible population states, or equivalently the state space, as the strategy profile set. Hence, every strategy profile is a vector with n nonnegative entries adding up to the so-called population mass m.

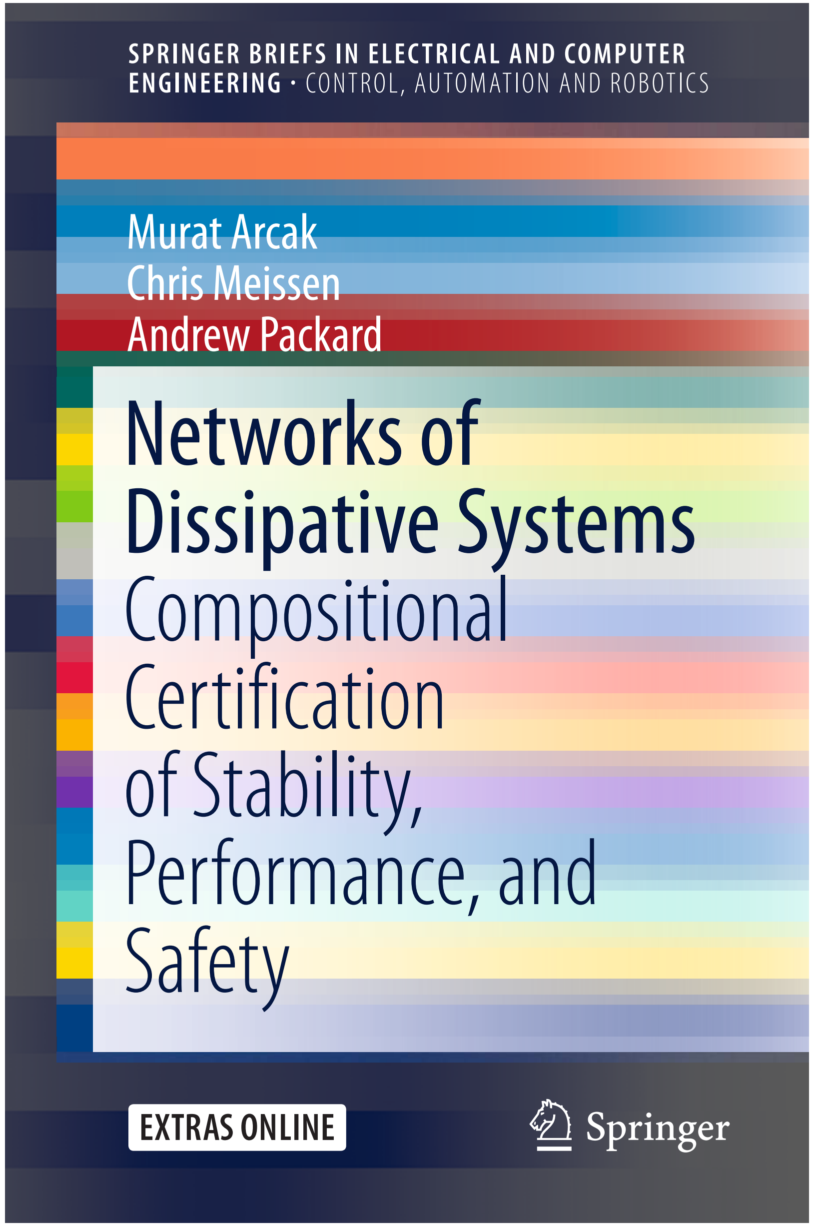
In most related prior work, a so-called *population game* [4] determines at any given instant the n-dimensional payoff vector, which quantifies the reward associated with each strategy, as a function of the population state. A population game may represent a pricing scheme that is implemented by a

Passivity Approaches in Evolutionary Game Theory

- **Equilibrium-independent passivity:**

$$\mathcal{S}_{p^*}(x(t), p(t)) - \mathcal{S}_{p^*}(x(t_0), p(t_0)) \leq \int_{t_0}^{t} (x(\tau) - x^*)^T (p(\tau) - p^*) \, \mathrm{d}\tau$$

holds for every stationary point (x^*, p^*) of EDM.



Passivity Approaches in Evolutionary Game Theory

- **Equilibrium-independent passivity:**

$$\mathcal{S}_{p^*}(x(t), p(t)) - \mathcal{S}_{p^*}(x(t_0), p(t_0)) \leq \int_{t_0}^{t} \left(x(\tau) - x^*\right)^T \left(p(\tau) - p^*\right) \, \mathrm{d}\tau$$

holds for every stationary point (x^*, p^*) of EDM.

- **Exponentially-discounted RL dynamics:**

$$\begin{aligned} \dot{p}_i(t) &= \alpha \left(\mathcal{F}_i(x(t)) - p_i(t)\right) \\ x_i(t) &= \frac{\exp \eta^{-1} p_i(t)}{\sum_{j=1}^{n} \exp \eta^{-1} p_j(t)} \end{aligned}, \; 1 \leq i \leq n$$

  – Convergence to the equilibrium state $p_i^* = \mathcal{F}_i\left(x^*\right)$, $x_i^* = \frac{\exp \eta^{-1} p_i^*}{\sum_{j=1}^{n} \exp \eta^{-1} p_j^*}$, $1 \leq i \leq n$

On Passivity, Reinforcement Learning and Higher-Order Learning in Multi-Agent Finite Games

Bolin Gao and Lacra Pavel

***Abstract*—In this paper, we propose a passivity-based methodology for analysis and design of reinforcement learning in multi-agent finite games. Starting from a known exponentially-discounted reinforcement learning scheme, we show that convergence to a Nash distribution can be shown in the class of games characterized by the monotonicity property of their (negative) payoff. We further exploit passivity to propose a class of higher-order schemes that preserve convergence properties, can improve the speed of convergence and can even converge in cases whereby their first-order counterpart fail to converge. We demonstrate these properties through numerical simulations for several representative games.**

I. Introduction

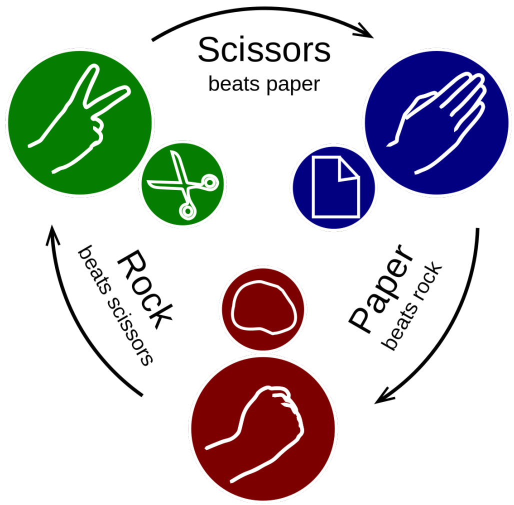
In multi-agent reinforcement learning in games, an agent repeatedly adjusts his strategy in response to a stream of payoffs which are in turn dependent on the actions of the other agents. The objective is for agents to arrive at a strategy profile that yields the best obtainable outcome for each of them, under information-constraints or in the presence of noise. Whether such an action profile can be reached depends on the learning process used by the collection of these agents. Learning in finite games typically involves discrete-time stochastic processes, where stochasticity arises from the agents' randomized choices, [1]–[7]. A key approach to analyzing such processes is based on the ordinary differential equation (ODE) method of stochastic approximation, a method which relates their behaviour to that of a "mean field" ODE, [8]. Motivated by this, we follow [5], [9], [10], [12] and consider a reinforcement learning scheme directly in continuous-time. By so doing, we are able to focus on the relationship between reinforcement learning, convex analysis, and passivity. [1]

Our starting point is a variant of the continuous-time exponentially-discounted learning (EXP-D-RL) in [5], version Nash distribution (logit equilibria) in 2-player zero-sum and partnership (potential) games. It is also related to the scheme proposed in [6] for traffic games in a repeated-game setup and shown to converge to a Nash distribution in N-player potential games. The strategy dynamics induced by this score dynamics is analyzed in [5] and proved to converge towards logit equilibria in potential games. With the exception of [3], the majority of these works have focused exclusively on convergence in potential games, [4]–[6], [13]. In contrast, less attention has been paid to stable games [14], [15], which encompass zero-sum games, potential games with concave payoffs and include well-known examples from evolutionary game theory such as the Rock-Paper-Scissors (RPS) game.

Motivated by the above, in this paper, we exploit passivity techniques and the natural monotonicity property associated with this class of games to show convergence to a Nash distribution. The use of passivity to investigate game dynamics was first proposed in [16] for population games, based on the notion of δ-passivity. The authors showed that certain game dynamics and the class of stable games, naturally satisfy this type of passivity. The coupling between a δ-passive system with a stable game implies a stable behaviour in the closed-loop solution. [17] showed that if an evolutionary dynamic does not satisfy a passivity property, then it is possible to construct a higher-order stable game that results in instability. Here we use an alternative concept, that of equilibrium-independent passivity, [18]. This allows us to directly address convergence towards equilibria. We note that equilibrium-independent passivity was used recently in [19] for continuous-kernel games, to relax the assumption on perfect-information on other players' actions.

Contributions. Our contributions are twofold: we show (1) that a passivity framework can be used to prove convergence

Simulations



Rock-Paper-Scissors Game

- Strategy set: $\{1, 2, 3\}$
- Population state: $x(t) = (x_1(t),\ x_2(t),\ x_3(t))$
- Payoff vector: $p(t) = (p_1(t),\ p_2(t),\ p_3(t))$

- Strategy Revision Mechanism:
  - BNN EDM:

$$\dot{x}_i(t) = [\hat{p}_i(t)]_+ - x_i(t) \sum_{j=1}^{3} [\hat{p}_j(t)]_+ \,, \ \ 1 \leq i \leq 3$$

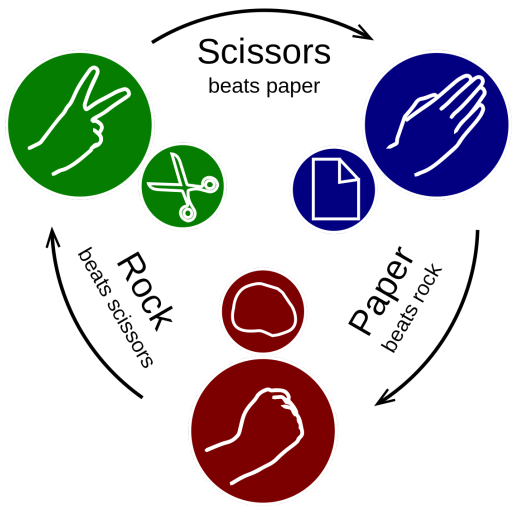
BNN EDM is δ-passive

  - Logit EDM:

$$\dot{x}_i(t) = \frac{\exp 0.5\, p_i(t)}{\sum_{j=1}^{3} \exp 0.5\, p_j(t)} - x_i(t), \ \ 1 \leq i \leq 3$$

Logit EDM is δ-passive with surplus 2

Simulations



Rock-Paper-Scissors Game

- Strategy set: $\{1, 2, 3\}$
- Population state: $x(t) = (x_1(t),\ x_2(t),\ x_3(t))$
- Payoff vector: $p(t) = (p_1(t),\ p_2(t),\ p_3(t))$

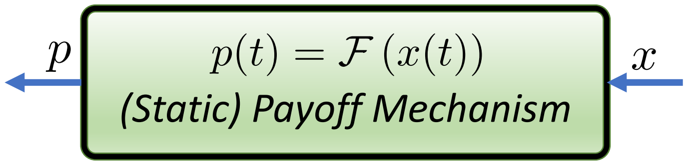
- Payoff Mechanism:
  - Static PDM:

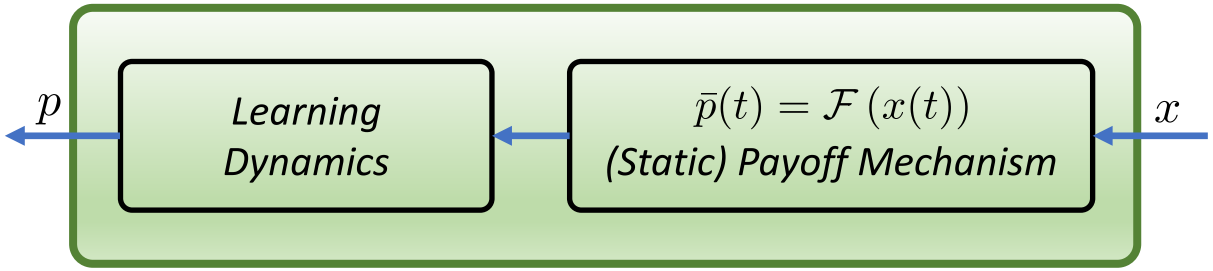
$$p(t) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} x(t)$$

  - Smoothing PDM:

$$\dot{q}(t) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} x(t) - q(t)$$
$$p(t) = q(t)$$

Simulations





- Payoff Mechanism:
  - Static PDM:

$$p(t) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} x(t)$$

Static PDM is δ-antipassive

  - Smoothing PDM:

$$\dot{q}(t) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} x(t) - q(t)$$

$$p(t) = q(t)$$

Smoothing PDM is (weak) δ-antipassive with deficit $\nu(< 2)$

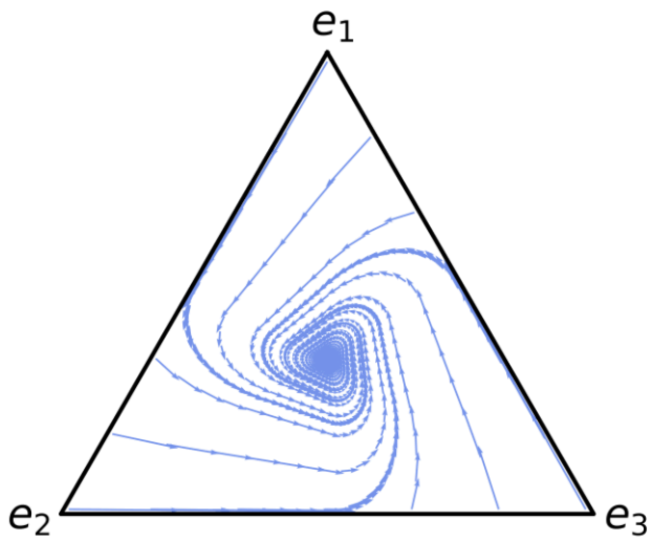
Simulations

- **Mean closed loop model**
  - Static PDM:

$$p(t) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} x(t) \qquad (\delta\text{-}antipassive)$$

  - BNN EDM:

$$\dot{x}_i(t) = [\hat{p}_i(t)]_+ - x_i(t) \sum_{j=1}^{3} [\hat{p}_j(t)]_+ \,, \ 1 \leq i \leq 3 \qquad (\delta\text{-}passive)$$



Social State Trajectories

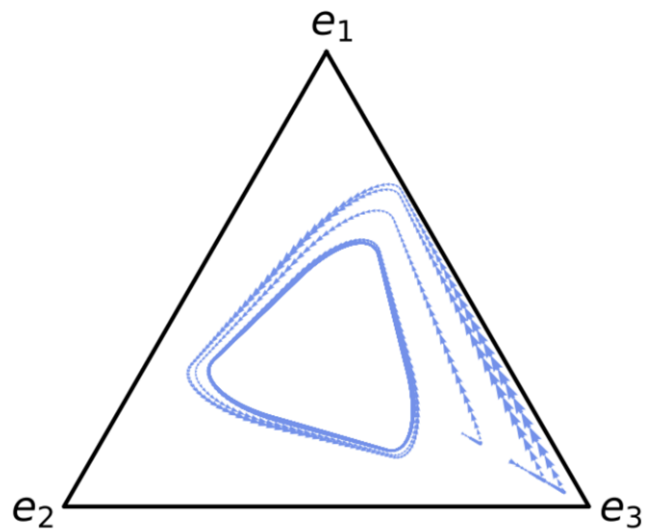
- **Mean closed loop model**
  - Smoothing PDM:

$$\dot{q}(t) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} x(t) - q(t)$$
$$p(t) = q(t)$$

(δ-antipassive with deficit (< 2))

  - BNN EDM:

$$\dot{x}_i(t) = [\hat{p}_i(t)]_+ - x_i(t) \sum_{j=1}^{3} [\hat{p}_j(t)]_+ \,, \ 1 \leq i \leq 3 \qquad (\delta\text{-}passive)$$



Social State Trajectories

Simulations

- **Mean closed loop model**
  - Static PDM:

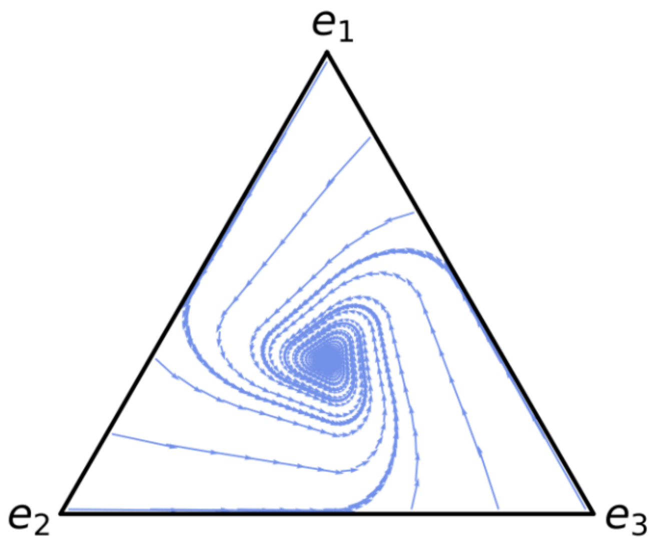
$$p(t) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} x(t)$$

(δ-antipassive)

  - BNN EDM:

$$\dot{x}_i(t) = [\hat{p}_i(t)]_+ - x_i(t) \sum_{j=1}^{3} [\hat{p}_j(t)]_+ \,, \; 1 \le i \le 3$$

(δ-passive)



Social State Trajectories

- **Mean closed loop model**
  - Smoothing PDM:

$$\dot{q}(t) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} x(t) - q(t)$$

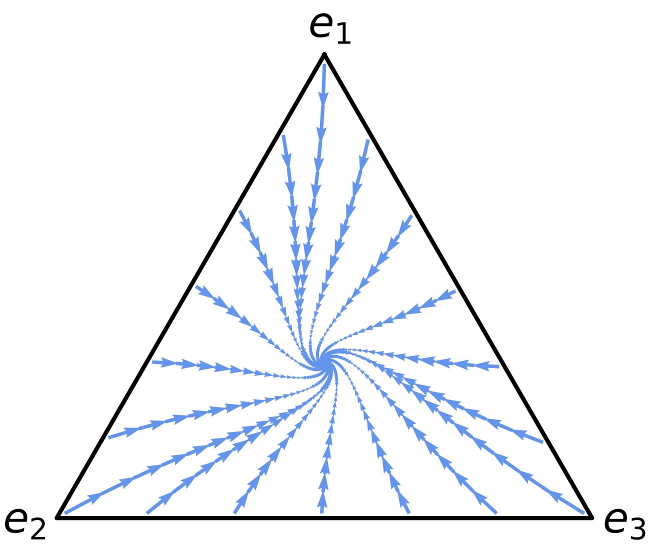
$$p(t) = q(t)$$

(δ-antipassive with deficit (< 2))

  - Logit EDM:

$$\dot{x}_i(t) = \frac{\exp 0.5\, p_i(t)}{\sum_{j=1}^{3} \exp 0.5\, p_j(t)} - x_i(t), \; 1 \le i \le 3$$

(δ-passive with surplus 2)



Social State Trajectories

Summary

- **Mean closed loop model:**

$$\left.\begin{aligned}\dot{q}(t) &= \mathcal{G}\left(q(t), x(t)\right)\\ p(t) &= \mathcal{H}\left(q(t), x(t)\right)\end{aligned}\right\}\ \text{(PDM)}$$

$$\dot{x}(t) = \mathcal{V}\left(x(t), p(t)\right)\ \text{— (EDM)}$$

  - Equilibrium set:

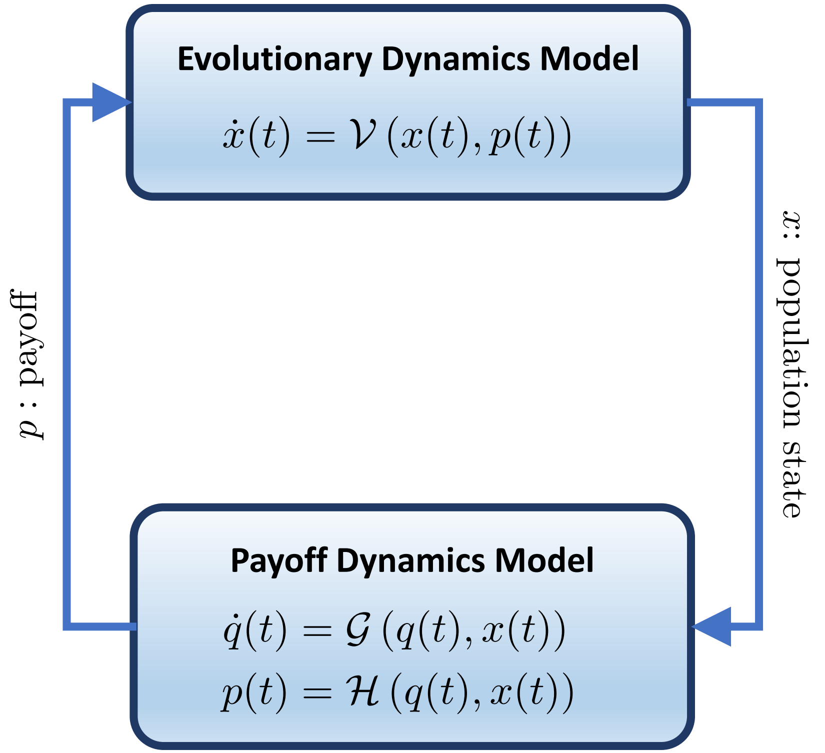
$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \left\{z \in \mathbb{X} \,\middle|\, \mathcal{V}\left(z, \bar{\mathcal{F}}(z)\right) = 0\right\}$$

    In this study,

$$\mathbb{E}\left(\bar{\mathcal{F}}\right) = \begin{cases}\mathbb{NE}\left(\bar{\mathcal{F}}\right) & \text{for EPT and IPC EDMs}\\ \mathbb{PE}\left(\bar{\mathcal{F}}, \mathcal{Q}\right) & \text{for PBR EDMs}^{\dagger}\end{cases}$$

(Global Asymptotic Stability) *The set $\mathbb{E}(\bar{\mathcal{F}})$ is globally asymptotically stable if the set $\mathbb{A}$ of stationary points is globally attractive and Lyapunov stable.*

When does the equilibrium set attain global asymptotic stability ?



- $\bar{\mathcal{F}}$: Stationary population game
- $\mathbb{E}\left(\bar{\mathcal{F}}\right)$: Equilibrium set

Summary

- **δ-Passivity for evolutionary dynamics models:**

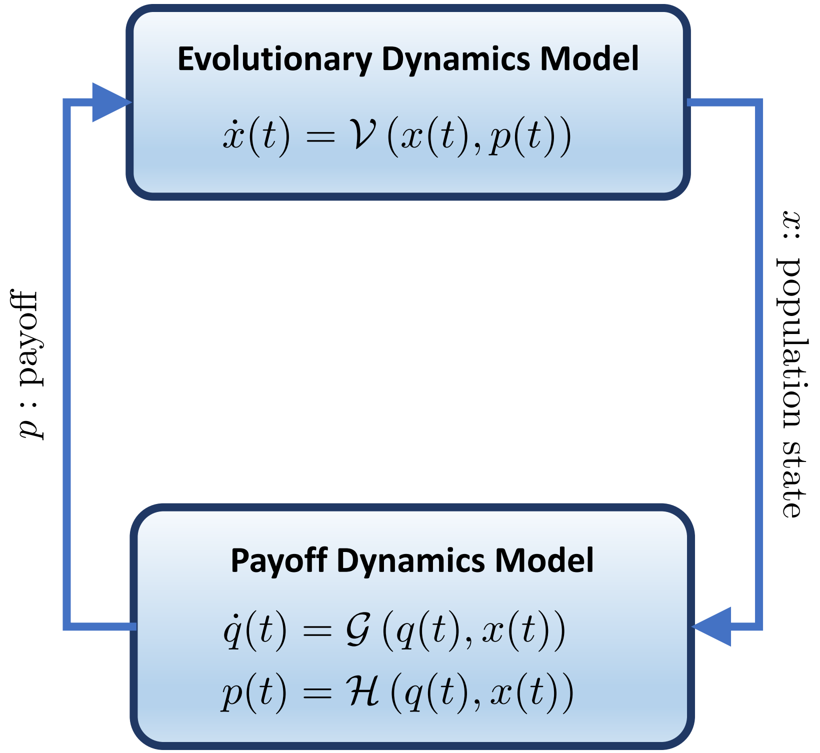
$$\mathcal{S}\big(x(t),p(t)\big) - \mathcal{S}\big(x(t_0),p(t_0)\big) \leq \int_{t_0}^{t} \dot{x}^T(\tau)\dot{p}(\tau)\,\mathrm{d}\tau,\ \ t \geq t_0$$

  - $\mathcal{S} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_+$: δ-storage function (continuously differentiable map)

- **δ-Antipassivity for payoff dynamics models:**

$$\mathcal{L}\left(x(t),q(t)\right) - \mathcal{L}\left(x(t_0),q(t_0)\right) \leq -\int_{t_0}^{t} \dot{x}^T(\tau)\dot{p}(\tau)\,\mathrm{d}\tau,\ \ t \geq t_0$$

  - $\mathcal{L} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_+$: δ-antistorage function (continuously differentiable map)



- $\bar{\mathcal{F}}$: Stationary population game
- $\mathbb{E}\left(\bar{\mathcal{F}}\right)$: Equilibrium set

$$\boxed{\mathcal{S}\left(x(t),p(t)\right) + \mathcal{L}\left(x(t),q(t)\right) \overset{t\to\infty}{\longrightarrow} 0 \implies \mathbb{E}\left(\bar{\mathcal{F}}\right)\ \ \textit{is asymptotically stable}}$$

Summary

- Multi-agent learning/decision-making problem

Cost to minimize: $\mathbb{E}\left[\int_0^\infty f\left(X^N(\tau)\right)\, \mathrm{d}\tau\right]$

Instantaneous reward: $r(t) = \nabla f\left(X^N(t)\right)$

- Multi-robot manipulation

Model on waste volume:

$$\dot{Q}_i^N(t) = \begin{cases} -RX_i^N(t) + w_i & \text{if } Q_i^N(t) > 0 \\ \max\left\{0, -RX_i^N(t) + w_i\right\} & \text{otherwise} \end{cases}$$

  - $Q_i^N(t)$: *Waste volume in cell i*
  - $X_i^N(t)$: *Portion of the robots working in cell i*
  - w_i: *Increase rate of the waste in cell i*

Thank You!
Questions?

Acknowledgement:

AFOSR
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

School of Engineering and Applied Science
PRINCETON

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of Science and Technology