香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

人工智能學院
School of Artificial Intelligence

Large Language Models for Social Simulations

(AIE1902)

Unit 3: Giant Catalytic Effect of Altruists in Schelling's Segregation Model

Assistant professor, Zhao Zhanzhan 赵展展 (Instructor)

PhD student, Jia Xiao 贾晓 (Teaching assistant)

School of Artificial Intelligence

Overview

Two Extensions:

1. Catalyst Effect of Altruists

- Even a very small number of altruists can trigger the system into the optimal state.
- *“Small perturbations → Big changes”* in social dynamics.

2. Urban Investment Intervention

- City planners can reshape utility functions to guide individual choices.
- Shows how **institutional design** can improve social welfare.

Part 1

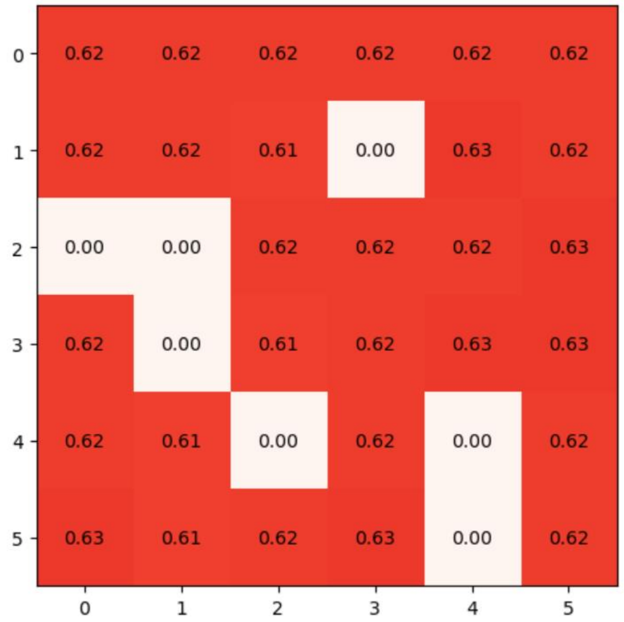
If only a small number of people are altruists🤗,

can they still improve the overall outcome?

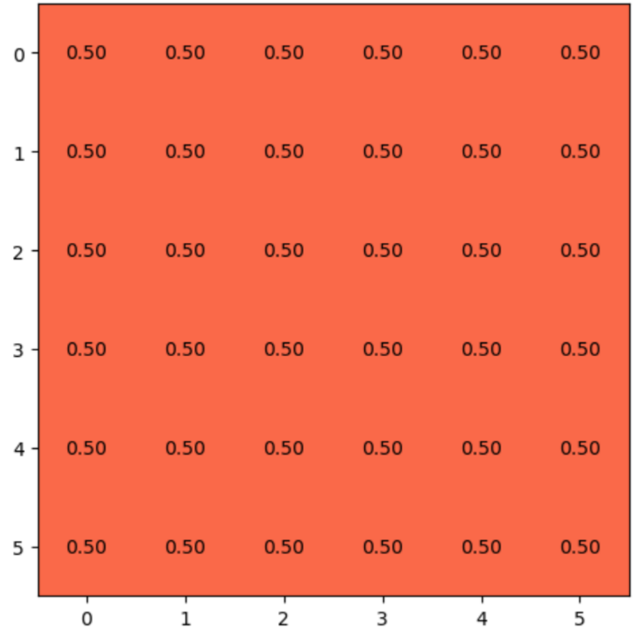
The "Catalyst Effect" of Altruists

From last class, we learned:

- If everyone is purely self-interested, the city tends to get crowded (low social welfare).



- If everyone cooperates to some extent (α > const), the city remains mid-level block density (high social welfare).



But here's the problem:

In reality, not everyone can be cooperative.

Research Question

We want to ask:

- When most people are still self-interested,
- What happens if we mix in a small group of altruists?

Intuition:

- It might help a little, but the effect should be limited.

Actual Finding:

Surprisingly, even a ***tiny*** *fraction* of altruists can ***catalyze*** the entire system into the optimal state! 🎉🎉🎉

Last Class: A Simplified City Model

The "city" is divided into Q blocks, each containing H housing units (cells).

Each housing unit can hold at most **one resident**.

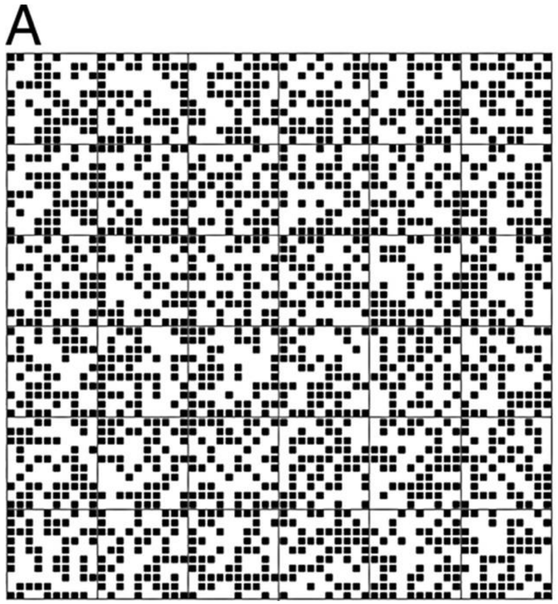
The **density** of block q is: $\rho_q = \dfrac{n_q}{H}$

where n_q, is the number of residents in block q.

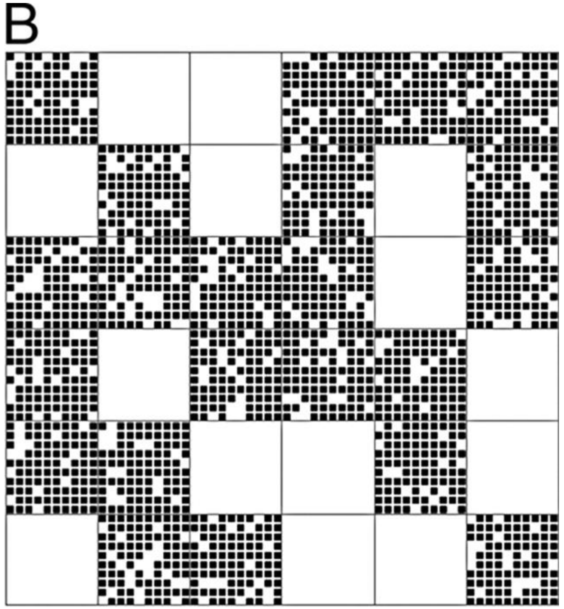
All residents share the **same utility function** $u(\rho_q)$,

which measures how satisfied they are with the density of their block.

Global utility / social welfare is the total satisfaction of all residents: $U = H \sum_q \rho_q \, u(\rho_q)$



Mixed state – residents evenly distributed across blocks.

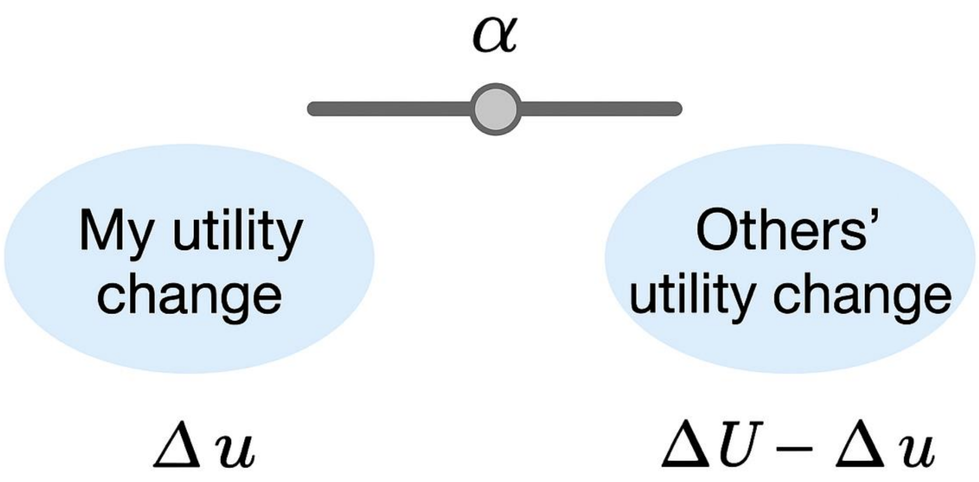


Segregated state – some blocks overcrowded, others empty.

Last Class: Considering Myself and Others

At each step:

1. Randomly pick **one resident** and **one vacant unit**.
2. Calculate the "gain" G from moving.



Formula for gain: $G = \Delta u + \alpha(\Delta U - \Delta u)$

- Δu: Change in *my* own utility if I move.
- $\Delta U - \Delta u$: Change in *everyone else's* utility.
- A: Global *cooperation level* shared by all residents.

Special cases:

- $\alpha = 0$: Purely selfish – only care about my own utility.
- $\alpha = 1$: Fully cooperative – only care about total utility.

This setup allows us to smoothly adjust between **economic behavior** (self-interest) and **physical-system-like behavior** (global optimization).

香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen
人工智能學院
School of Artificial Intelligence

Model Recap

City setup:

•Q blocks, each with capacity H.

Utility function:

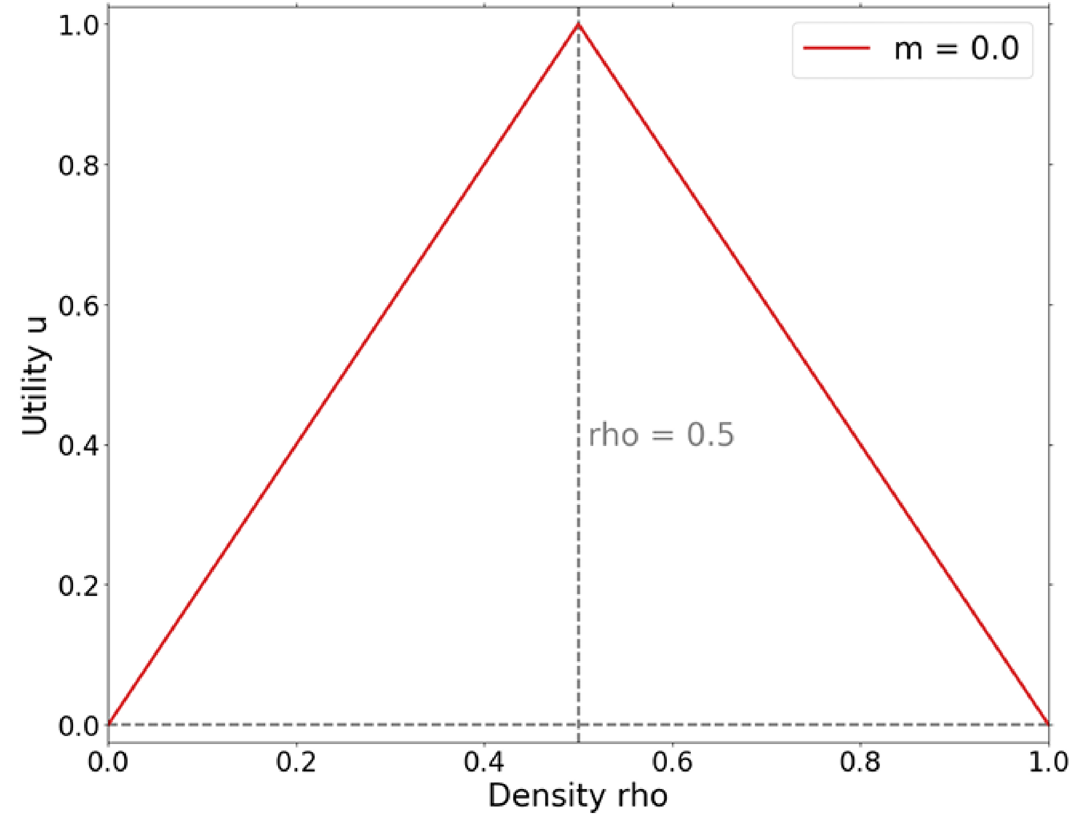
•Utility is highest (=1) when density ρ=0.5.

•Too sparse (ρ=0) or too crowded (ρ=1) → utility = 0

Two types of agents:

•**Self-interested:** Care only about their own Δu.

•**Altruists:** Care about the change in *total utility* ΔU.



Key Result 1: Threshold Effect

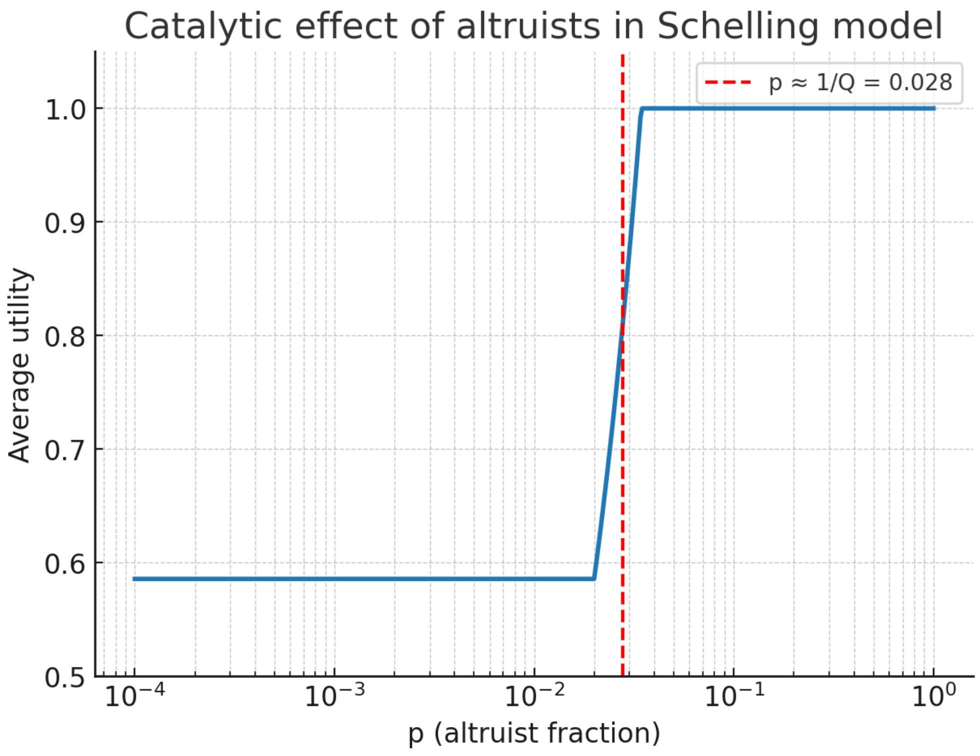
If the fraction of altruists in the population is about $p \approx \frac{1}{Q}$,

that is already enough to ***trigger*** the whole system into the optimal mixed state.

Example:
In a city with 36 blocks, only about **3% altruists** are needed.

This is **not** a smooth, linear improvement—

it's a sudden *jump* to high social welfare!



香港中文大學(深圳) The Chinese University of Hong Kong, Shenzhen
人工智能學院 School of Artificial Intelligence

Key Result 2: The Mechanism

Without altruists:

•Self-interested agents cluster together → blocks become overcrowded → low efficiency / social welfare.

With altruists ("sacrifice" behavior):

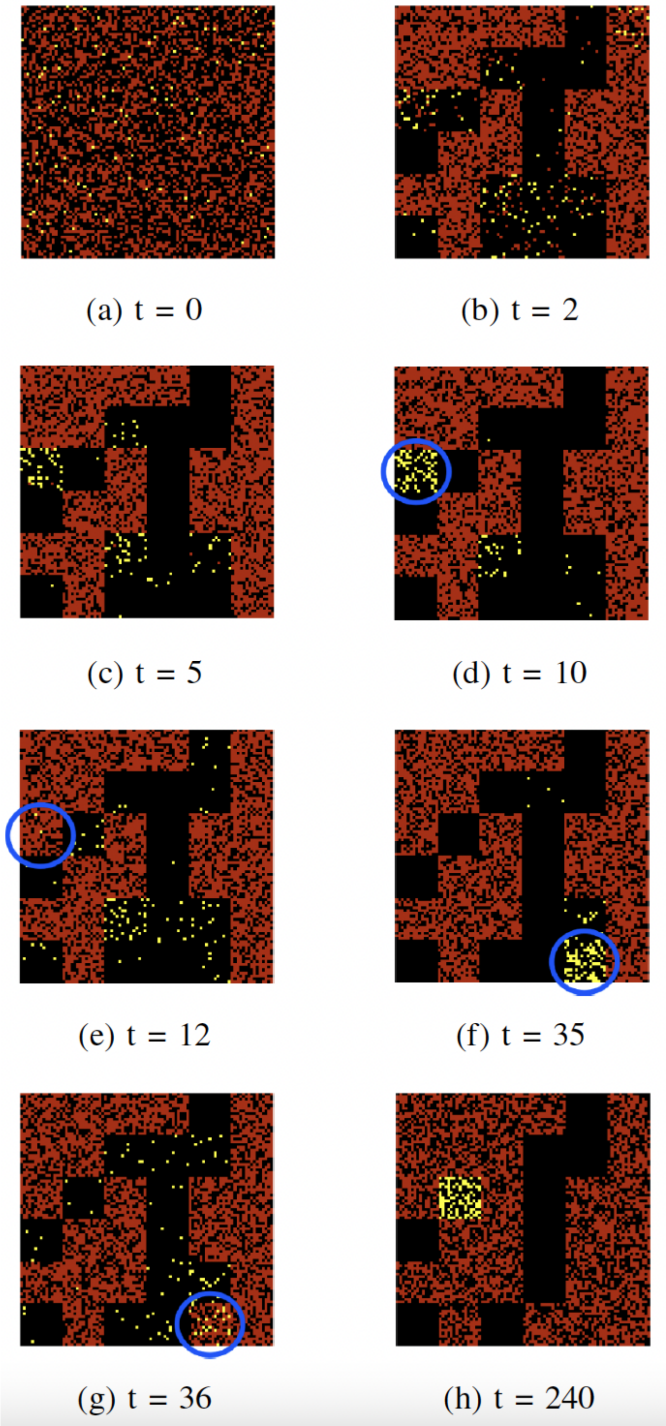
•They *choose* to leave crowded blocks and move to sparser ones.

•This makes sparse blocks gradually more attractive → self-interested agents move in.

•Overcrowded blocks get diluted → overall utility increases.

Bottom line:

Altruists keep repeating this process, acting like a *catalyst* that drives the whole system toward efficiency / high social welfare.



(a) t = 0 (b) t = 2 (c) t = 5 (d) t = 10 (e) t = 12 (f) t = 35 (g) t = 36 (h) t = 240

Why Call It "Catalysis"?

In chemistry:

- A catalyst is *not consumed* but can *accelerate* a reaction.

In the model:

- Altruists don't disappear or "get tired,"
- Yet they continuously guide self-interested agents toward better configurations.

Impact of a single altruist:

- As the system size Q increases,
  the *marginal effect* of each altruist becomes even larger.

Summary

Core Finding: A very small number of altruists can *catalyze* the whole system toward the optimal state.

Key Insight:

- Individual self-interest may cause collective failure.
- But just a few "pioneers" — or clever institutional design — can shift the entire system.

Real-World Connections:

- Public policy
- Green behaviors
- Volunteer actions

👉 **The altruism of a few may be the *spark* for large-scale social transformation.**

Part 2

What if city planners could invest in certain blocks?

Extension: Urban Investment 💴

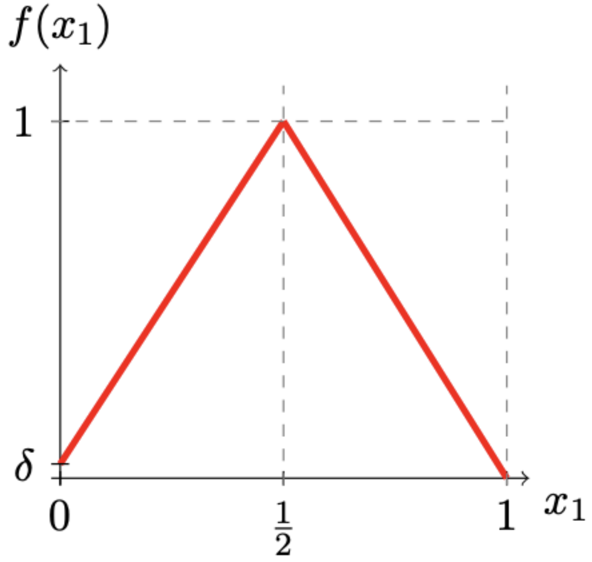
Planner intervention (city-level):

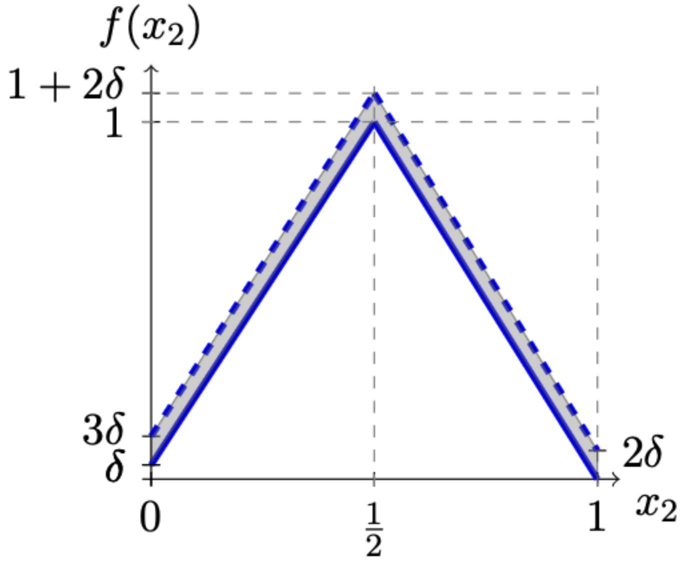
- The planner can change the utility function of certain blocks.
- This makes certain blocks *q more attractive*.

Mathematical expression:

$$g_q(\rho) \geq f_q(\rho)$$

Investment raises the attractiveness of block q.





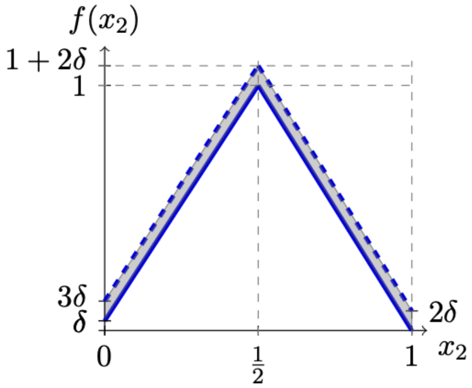
Effects of Investment

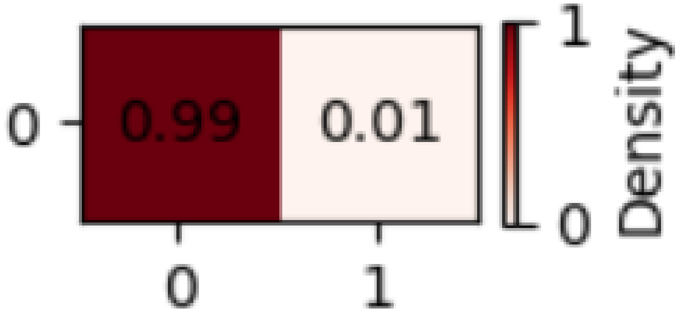
Individuals remain self-interested: They only respor to utility differences.

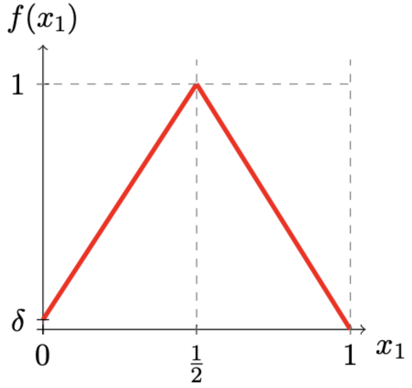
But with investment: Some blocks become more attractive.

Result:

- Migration paths of individuals are indirectly altered.
- At the macro level, social welfare gets changed







Broader Connections: Braess's Paradox

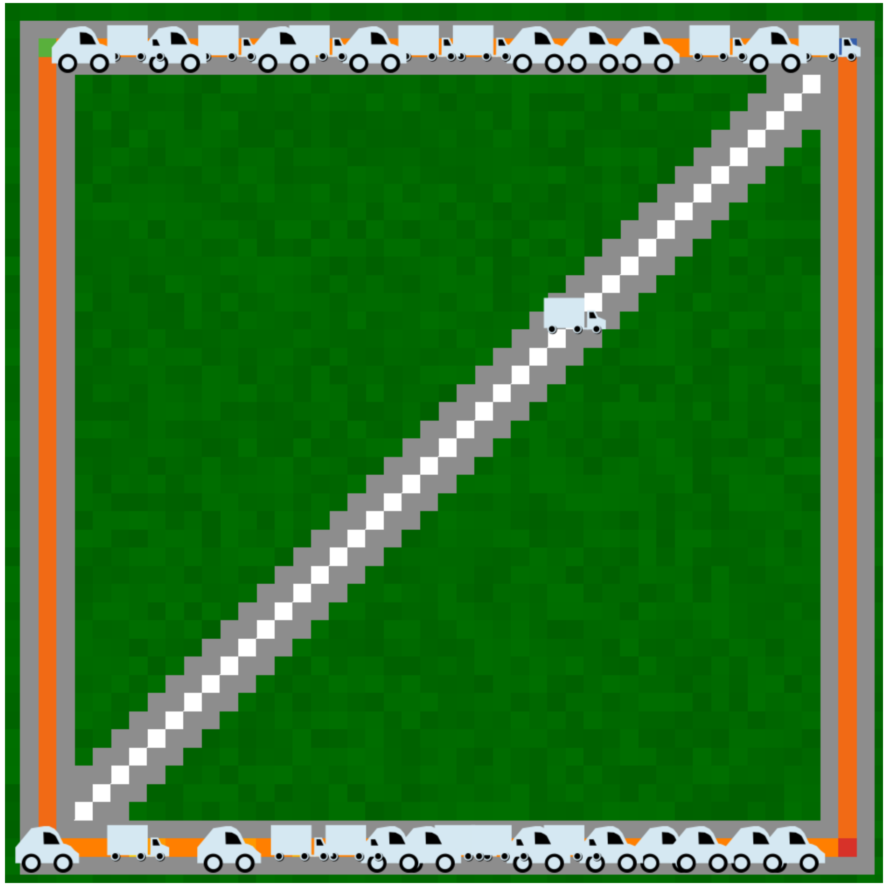
Sometimes, *"better conditions"* can actually make outcomes worse.

Example:

Adding an extra road can make traffic jams *worse*.

Why?

- Each driver chooses what seems best for themselves.
- But collectively, this creates a *less efficient* traffic pattern.



Broader Connections: **Congestion Games**

Each individual pursues their own *shortest path*.

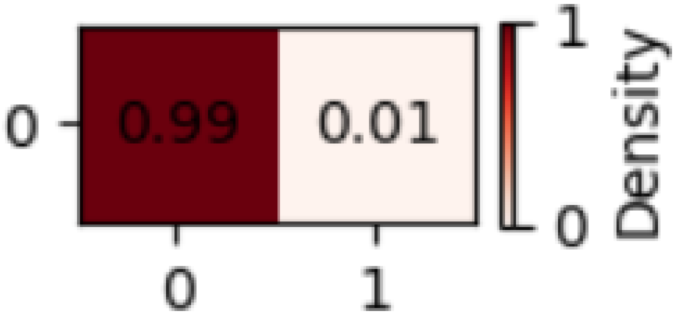
But when everyone acts this way, the overall outcome is *inefficient (low social welfare)*.

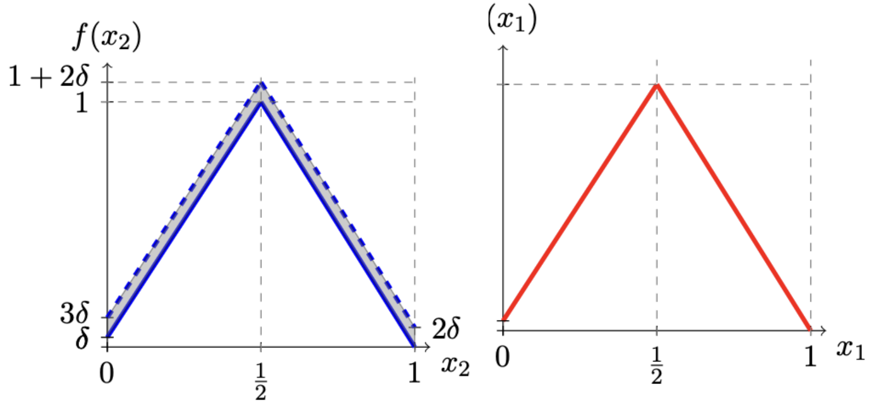
Lesson:

What is *best for individuals* is not always *best for the group*.

Question to think about:

Can we make smart investments to improve social welfare despite egoists (selfish individuals)?





香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen
人工智能學院
School of Artificial Intelligence

Thank You!

Thank you for your attention and participation.

📫 **Contact Information:**

zhanzhanzhao@cuhk.edu.cn (Dr. Zhao Zhanzhan)
225080011@link.cuhk.edu.cn (Mr. Jia Xiao)