

# The Price of Justice: Altruistic Punishment and Human Cooperation

Based on Fehr & Gächter (Nature, 2002)

Timothy Nguyen

EC323: Behavioral Economics  
Professor Jawwad Noor

August 6, 2025

# The Puzzle of Cooperation

- Humans often help strangers, even at a personal cost.

# The Puzzle of Cooperation

- Humans often help strangers, even at a personal cost.
- This is a puzzle because standard economic models assume people are purely self-interested.

# The Puzzle of Cooperation

- Humans often help strangers, even at a personal cost.
- This is a puzzle because standard economic models assume people are purely self-interested.
- So, why do we punish "free-riders" when we get nothing out of it?

# The Puzzle of Cooperation

- Humans often help strangers, even at a personal cost.
- This is a puzzle because standard economic models assume people are purely self-interested.
- So, why do we punish "free-riders" when we get nothing out of it?
- This is called **altruistic punishment**.

# Connecting to Our Class: Social Preferences

This experiment challenges the basic assumption of pure self-interest.

- In our lectures, we've discussed how people are not always selfish. They have **social preferences**.

# Connecting to Our Class: Social Preferences

This experiment challenges the basic assumption of pure self-interest.

- In our lectures, we've discussed how people are not always selfish. They have **social preferences**.
- People care about fairness and what others get. This is sometimes called **inequity aversion**.

# Connecting to Our Class: Social Preferences

This experiment challenges the basic assumption of pure self-interest.

- In our lectures, we've discussed how people are not always selfish. They have **social preferences**.
- People care about fairness and what others get. This is sometimes called **inequity aversion**.
- Altruistic punishment is a powerful example of this: people are willing to *pay money* just to reduce the payoff of someone they think is unfair.



## Connecting to Our Class: Social Preferences

This experiment challenges the basic assumption of pure self-interest.

- In our lectures, we've discussed how people are not always selfish. They have **social preferences**.
- People care about fairness and what others get. This is sometimes called **inequity aversion**.
- Altruistic punishment is a powerful example of this: people are willing to *pay money* just to reduce the payoff of someone they think is unfair.
- This behavior shows that a sense of justice can be a stronger motivator than money.

# The Experiment: A "Public Goods" Game

## The Setup:

- Anonymous groups of 4.
- Each person gets 20 Money Units (MUs).
- You can invest in a group project.

# The Experiment: A "Public Goods" Game

## The Setup:

- Anonymous groups of 4.
- Each person gets 20 Money Units (MUs).
- You can invest in a group project.

## The Dilemma:

- For every 1 MU you invest, everyone in the group gets 0.4 MUs back.
- Your best strategy is to be a **free-rider**: invest nothing and just collect the rewards from others' investments.
- But if everyone free-rides, everyone is worse off!

# The Two Conditions

To see if punishment works, subjects were split into two scenarios.

- **Condition 1: No Punishment**

- You play the game, see what others invested, and that's it. You can't do anything about it.

# The Two Conditions

To see if punishment works, subjects were split into two scenarios.

- **Condition 1: No Punishment**

- You play the game, see what others invested, and that's it. You can't do anything about it.

- **Condition 2: Punishment Allowed**

- After investing, you have the option to punish others.
- **The Cost of Justice:** You can spend 1 MU to make someone else lose 3 MUs.
- This is "altruistic" because it costs you money and gives you no direct financial reward.

# Result 1: Punishment Boosts Cooperation

The effect of punishment was huge.

- **Without Punishment:**

Cooperation collapsed. People quickly learned to free-ride.

# Result 1: Punishment Boosts Cooperation

The effect of punishment was huge.

- **Without Punishment:**

Cooperation collapsed. People quickly learned to free-ride.

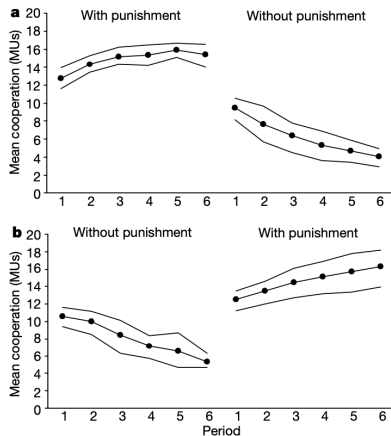
- **With Punishment:**

Cooperation was high and even *increased* over time.

# Result 1: Punishment Boosts Cooperation

The effect of punishment was huge.

- **Without Punishment:**  
Cooperation collapsed. People quickly learned to free-ride.
- **With Punishment:**  
Cooperation was high and even *increased* over time.
- The graph shows this clearly.  
The top line in (a) and the bottom-right line in (b) show cooperation thriving when punishment is possible.





## Result 2: Punishment Is Targeted

Punishment wasn't random.

- People consistently punished free-riders (those who contributed less than average).

## Result 2: Punishment Is Targeted

Punishment wasn't random.

- People consistently punished free-riders (those who contributed less than average).
- The more someone deviated from the average, the more heavily they were punished.

## Result 2: Punishment Is Targeted

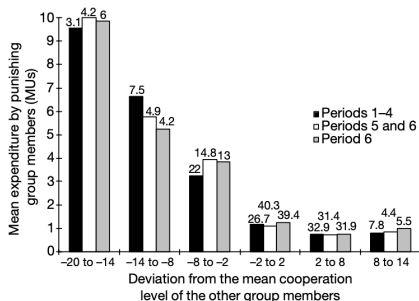
Punishment wasn't random.

- People consistently punished free-riders (those who contributed less than average).
- The more someone deviated from the average, the more heavily they were punished.
- This graph shows that the biggest free-riders (far left, "-20 to -14") received the most punishment.

## Result 2: Punishment Is Targeted

Punishment wasn't random.

- People consistently punished free-riders (those who contributed less than average).
- The more someone deviated from the average, the more heavily they were punished.
- This graph shows that the biggest free-riders (far left, "-20 to -14") received the most punishment.
- This suggests punishment is driven by a reaction to unfairness.



# Conclusions & Take-Aways

- Altruistic punishment is a powerful tool that helps enforce cooperation.

# Conclusions & Take-Aways

- Altruistic punishment is a powerful tool that helps enforce cooperation.
- This behavior is likely driven by our emotional hatred of unfairness, not a cold calculation of benefits.

# Conclusions & Take-Aways

- Altruistic punishment is a powerful tool that helps enforce cooperation.
- This behavior is likely driven by our emotional hatred of unfairness, not a cold calculation of benefits.
- It shows that "social preferences" can sometimes be more important than pure self-interest in making economic decisions.

# Conclusions & Take-Aways

- Altruistic punishment is a powerful tool that helps enforce cooperation.
- This behavior is likely driven by our emotional hatred of unfairness, not a cold calculation of benefits.
- It shows that "social preferences" can sometimes be more important than pure self-interest in making economic decisions.
- To understand cooperation, we need to understand our willingness to enforce fairness, even at a personal cost.



## My Reactions:

- It's fascinating that a "negative" emotion like anger can actually lead to better outcomes for the group.
- The simple design of the experiment makes the results very powerful and easy to understand.
- This research has huge real-world implications for everything from managing climate change to building online communities.

## Questions for Discussion:

- Can you think of real-world examples of altruistic punishment?
- Is this behavior truly "altruistic" if punishing others makes us feel good?
- When can this desire to punish unfairness go wrong?

**Thank You** Questions?