

# Repeated Prisoner's Dilemma with LLM Agents

Last-Round Signaling, Cooperation Dynamics & Chain-of-Thought

# Game Setting

Two LLM agents (Alice & Bob) play a 20-round Prisoner's Dilemma  
Implement via OpenRouter API

Payoff matrix:

	Cooperate	Defect
Cooperate	(3, 3)	(0, 5)
Defect	(5, 0)	(1, 1)

**Twist: On round 20 each agent is told "this is the last round."**

# Experimental Design

Seven runs total:

1. GPT-4o vs Claude-3
2. GPT-4o vs GPT-4o (self-play)

Observation:

Pure long-run cooperation → final-round betrayal

CoT logging

Prompts:

- Rounds 1–19: “Think long-term...”
- Round 20: “This is the last round...”
- **Recorded:** per-round moves, chain-of-thought, cumulative scores

# Run 1: GPT-4o vs GPT-4o

Round	Alice	Bob
1-19	Cooperate	Cooperate
20 (final round)	Cooperate	Defect

Final scores: 57 vs 62

One agent (Bob) exploits endgame signal.

# Run 2: GPT-4o vs GPT-4o

Round	Alice	Bob
1-19	Cooperate	Cooperate
20 (final round)	Defect	Defect

Final scores: 58 vs 58

Symmetric last-round defection — randomness in GPT-4o's decision (different from Run 1)

# Run 1: GPT-4o vs Claude-3

Round	Alice (GPT)	Bob (Claude)
1-19	Cooperate	Cooperate
20 (final round)	Defect	Cooperate

Looks like Claude-3 is more selfless than GPT-4o?

GPT-4o is smarter?

# Run 2: GPT-4o vs Claude-3

Round	Alice (GPT)	Bob (Claude)
1-10	Cooperate	Cooperate
11	Defect	Cooperate
12	Defect	Cooperate
13	Defect	Defect
14-15	Cooperate	Cooperate
16	Defect	Cooperate
17	Cooperate	Defect
18	Cooperate	Defect
19	Defect	Defect
20 (final round)	Defect	Cooperate

GPT-4o is more aggressive?

Claude-3 reacts slower or Claude-3 is more benign?

# Run 3: GPT-4o vs Claude-3

Round	Alice (GPT)	Bob (Claude)
1-19	Cooperate	Cooperate
20 (final round)	Defect	Cooperate

Run 3 same as Run 1, Run 4 same as Run 1

— Is Claude-3 more benign? Or is GPT-4o smarter?



# Chain of Thought Logging (GPT-4o)

In the Prisoner's Dilemma, I need to decide between cooperating or defecting. Let's break down the decision process:

1. **\*\*Understanding the Payoffs\*\***: In a single round of the Prisoner's Dilemma:
  - If both players cooperate, both receive a moderate reward.
  - If one cooperates and the other defects, the defector gets a high reward while the cooperator gets a low reward.
  - If both defect, they both get a punishment, but it's generally better than being the sole cooperator.
  
2. **\*\*Last Round Dynamics\*\***: Since this is the last round, future retaliations or benefits for cooperation in future rounds do not apply. The decision is entirely based on maximizing the payoff for this single round.

# Chain of Thought Logging (GPT-4o)

## 3. **Options and Outcomes**:

- If I **Cooperate**, and the other player also cooperates, we both get a decent payoff.
- If I **Cooperate**, but the other player defects, I get the worst outcome.
- If I **Defect**, and the other player cooperates, I get the best outcome while they get the worst.
- If I **Defect**, and the other player also defects, we get a payoff that is slightly better than if I had cooperated and they defected.

## 4. **Strategy Consideration**:

- Given that this is the last round, players are generally more inclined to defect since there's no future consequence.
  - Defecting ensures that I avoid the worst-case scenario of being the sole cooperator.
- In this single, final decision point where immediate payoff is the priority, **it's rational to choose the option that maximizes my individual guaranteed return regardless of the other player's action.**

Therefore, my decision is to **Defect**.

# Chain of Thought Logging (Claude-3)

Okay, let's think through this step-by-step:

1. This is the last round of a repeated Prisoner's Dilemma game. That means my choice now will not impact any future rounds or the other player's future behavior towards me.
  2. In a one-shot Prisoner's Dilemma, defecting always gives a higher payoff than cooperating, regardless of what the other player chooses.
  3. If I cooperate and the other player defects, I get the "sucker's payoff" which is the worst outcome for me. If I defect and they cooperate, I get the "temptation payoff" which is the best outcome for me.
  4. Even if the other player also chooses to defect, I still get a higher payoff by defecting than I would by cooperating.
  5. Since this is the last round, there is no benefit to establishing a cooperative relationship or punishing previous defection. The rational choice is to defect.
  6. Although **cooperating might feel better morally**, the goal as stated is to maximize my own payoff. Defecting is the strategy that best achieves that.
- Therefore, thinking purely in terms of maximizing my own payoff in this final round, I choose to Defect.

# Takeaways & Next Steps

- **Both Claude-3 and GPT-4o demonstrate rational strategic behavior**  
They tend to cooperate in earlier rounds to maintain mutual benefit.  
They defect in the final round, aligning with game-theoretic reasoning
- **Claude-3 exhibits a more benign and cooperative disposition across rounds**  
More likely to initiate or maintain cooperation, even when provoked.  
Shows greater hesitation toward defection unless strictly necessary.
- **Claude-3's chain-of-thought reasoning incorporates moral language**  
Explicitly acknowledges emotional conflict: “Although cooperating might feel better morally...”
- **Chain-of-thought explanations are not always aligned with final decisions**