# Prisoner's Dilemma (Matrix)

Teaching Note | MobLab x Harvard Business Impact

| | |
|---|---|
| **Primary Topic:** | Game Theory, Microeconomics, Organizational Behavior, Business Strategy |
| **Typical Duration:** | Prep & Instructions: 5 min \| Each game: 5-10 min \| Debrief: 15-20 min |
| **Group Size:** | Pairs (2 players); scales to any class size |
| **Difficulty Level:** | Introductory / Intermediate |
| **Prerequisites:** | No prior game theory required. Works well as a first exposure to strategic thinking and payoff matrices. |

## 1. Simulation Overview

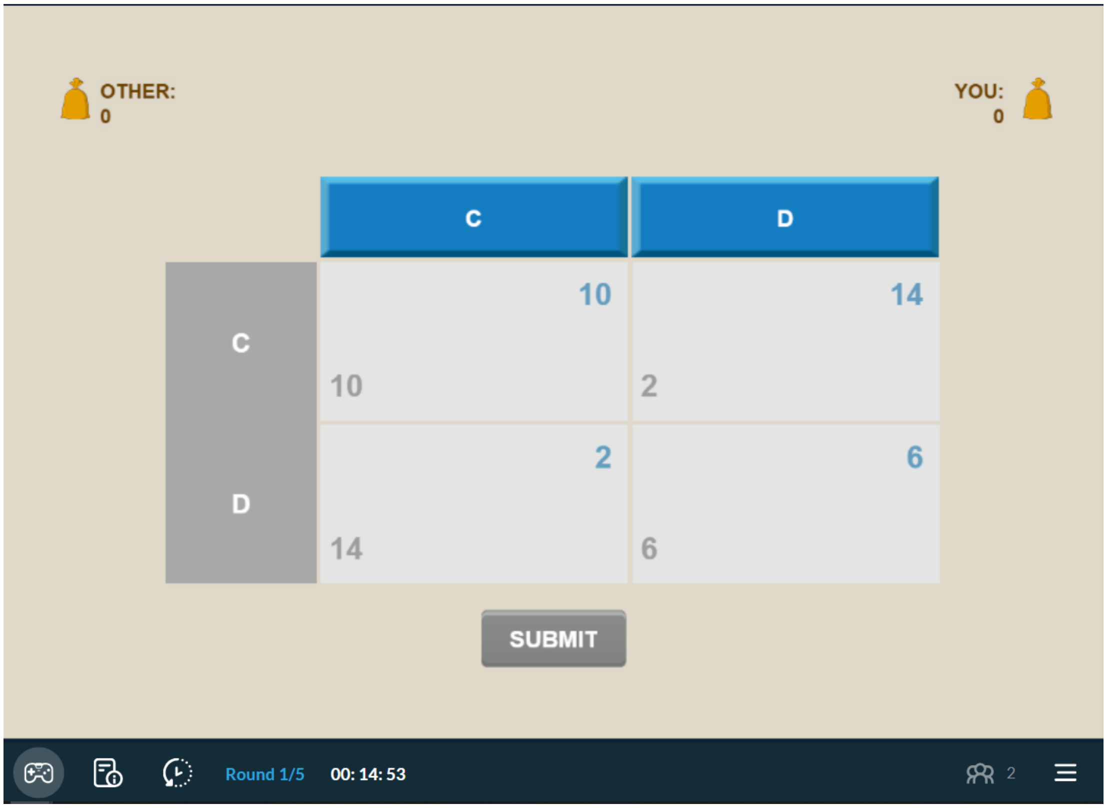
In this simulation, students are paired and play a classic Prisoner's Dilemma in matrix form. In each round, both players simultaneously choose either to Cooperate (C) or Defect (D). Neither player knows the other's choice before committing. Payoffs are revealed after each round, and the game repeats for a set number of rounds with the same partner.

The payoff structure creates the dilemma: defecting always yields a higher individual payoff regardless of what the partner does (defect is a dominant strategy), yet if both players defect, both earn less than if both had cooperated. This tension between individual rationality and collective welfare is the heart of the game, and students experience it firsthand — often going through cycles of attempted cooperation, betrayal, and retaliation before converging on the Nash equilibrium.

### The Payoff Matrix

The default payoff matrix is as follows. Your total payoff is the sum of earnings across all rounds.

| | Match Cooperates (C) | Match Defects (D) |
|---|---|---|
| **You Cooperate (C)** | You: $10 \| Match: $10 | You: $2 \| Match: $14 |
| **You Defect (D)** | You: $14 \| Match: $2 | You: $6 \| Match: $6 |




**Key Observation:** Switching from Cooperate to Defect gains you $4 (from $10 to $14, or from $2 to $6) — but it costs your partner $8 (from $10 to $2, or from $14 to $6). Defection is individually rational but collectively destructive.

## Learning Objectives

By the end of this simulation, students will be able to:

- **Dominant Strategy and Nash Equilibrium**: Identify Defect as the dominant strategy in a one-shot Prisoner's Dilemma, and explain why the (Defect, Defect) outcome is a Nash equilibrium even though (Cooperate, Cooperate) yields higher payoffs for both players.
- **The Efficiency-Incentive Tension**: Articulate the gap between the individually rational outcome and the socially efficient outcome, and explain why this gap is the defining feature of the Prisoner's Dilemma.
- **Repetition and Cooperation:** Explain how indefinite repetition, the threat of future punishment, and the possibility of communication can sustain cooperative outcomes that would unravel in a one-shot game.
- **End-Game Effects:** Predict and explain the end-game unraveling of cooperation in finitely repeated games, connecting the logic to backward induction.
- **Real-World Applications**: Identify Prisoner's Dilemma structures in business, policy, and everyday life — including price wars, arms races, environmental agreements, and team incentive problems.

## 2. Classroom Setup

| Configuration | Recommended Setting |
|---|---|
| **Role Assignment** | Random pairing each game (use Repeat > Random to reshuffle between sessions). |
| **Number of Rounds** | Game 1: 5 rounds, fixed partners, no chat (baseline). Game 2: 5 rounds, fixed partners, chat enabled. Game 3 (optional): Indefinitely repeated — set Minimum Rounds and Ending Probability below 100%. |
| **Chat** | Disabled for Game 1; Enabled for Game 2 onward. Chat significantly increases cooperation rates and generates rich debrief material. |
| **Indefinite Repetition** | Set Ending Probability < 100% to implement an indefinitely repeated game. With default payoffs, cooperation can be a Nash equilibrium (via grim trigger) when ending probability is 50% or less. |
| **Payoff Structure** | Per-round payoffs summed across all rounds. Default matrix: C/C = $10 each; C/D = $2 (cooperator) / $14 (defector); D/D = $6 each. |
| **Robot Players** | Available if group sizes are uneven. Robot strategies: Always C, Always D, Random, or Tit-for-Tat (mirrors partner's previous action). |
| **Technology Required** | Laptop or tablet with an up-to-date web browser |

## 3. Pedagogical Instructions

### Recommended Game Sequence

The Prisoner's Dilemma is most effective when run as a progression of two or three games that build on each other. The contrast between conditions is where the deepest learning happens.

- Game 1 — Finite, No Chat: 5 rounds, fixed partners, communication disabled. This is the baseline. Most pairs will end in mutual defection by the final round, especially if the number of rounds is announced in advance.
- Game 2 — Finite, With Chat: 5 rounds, fixed partners, chat enabled. Cooperation rates start higher. Watch for end-game defection even after explicit agreements — this generates powerful discussion about credible commitment.
- Game 3 (optional) — Indefinitely Repeated: Use Minimum Rounds and an Ending Probability of 50% or less. With chat enabled, sustained cooperation becomes achievable. Use the Compare button in Results to show students how their behavior changed across sessions.

> **Playlist Tip:** Build all games into a single MobLab playlist before class. Use the Copy Game tool to duplicate configurations quickly and name each activity clearly (e.g., 'PD Week 1 - No Chat', 'PD

Week 1 - Chat'). This keeps everything organized and lets you use the Compare button to show side-by-side results in the debrief.

## Pre-Game Briefing

Before Game 1, walk students through the payoff matrix carefully:

- Show the matrix on screen. Explain that both players choose simultaneously — neither knows the other's choice before committing.
- Walk through the four possible outcomes using concrete dollar amounts from the default matrix.
- Ask students: 'If your partner cooperates, which earns you more?' ($14 from D > $10 from C.) 'If your partner defects, which earns you more?' ($6 from D > $2 from C.) 'So what should you do?' Let them arrive at the dominant strategy themselves.
- Tell students the number of rounds in advance. This is intentional: knowing the endpoint changes strategy.

**Instructor Tip:** Do not tell students what to do or what the Nash equilibrium is before they play. Let them feel the pull of the dilemma — the temptation to defect, the hope that their partner cooperates, the sting of being betrayed. The lived experience makes the theory land much harder afterward.

## The Gameplay Experience

During Game 1 (no chat), watch for:

- Early rounds showing a mix of C/C and C/D outcomes as students test each other. Some pairs maintain cooperation for several rounds before one partner defects — note these for the debrief.
- A clear shift toward D/D in the final 1-2 rounds as the end-game unraveling sets in. This backward induction effect is one of the most striking results to highlight.
- The overall cooperation rate declining monotonically across rounds — a clean illustration of the theory.

During Game 2 (chat enabled), watch for:

- Cooperation starting much higher as students make explicit agreements. Collect examples from the chat log — these are often funny, revealing, and pedagogically gold.
- End-game defection even after explicit promises. Students who cooperated for four rounds and then defect on round 5 produce some of the best debrief conversations about credible commitment, cheap talk, and trust.
- Variation between groups: some will maintain C/C all five rounds; others will unravel completely. This heterogeneity is useful — ask different groups to explain their experience.

**Instructor Tip:** The chat log CSV file is a goldmine. Download it after class and pull the most interesting exchanges for your next lecture. Student quotes like 'If you defect I will defect every round' or 'I gotta do what I gotta do to get the victory' open up rich discussions about the meaning of threats, promises, and reputation.

### Facilitator Checklist

1. Build the game sequence into a playlist before class and test it using Preview Activity.
2. Announce the number of rounds to students before each game begins.
3. For Game 1: disable chat, run 5 rounds, note the final-round defection rate.
4. Between games: do not debrief yet. Let each game's results sit until all sessions are complete.
5. For Game 2: enable chat, run 5 rounds, download the chat log CSV afterward.
6. For Game 3 (if running): set Ending Probability to 40-50%, enable chat.
7. Use Repeat > Random between games to reshuffle partners.
8. Navigate to Results and open the Compare view before the debrief to show sessions side by side.

## 4. Debriefing & Discussion

Plan for 15-20 minutes of discussion. Begin with the results graphs to anchor the conversation in what actually happened, then pull students toward the theory. The contrast between game sessions is the most powerful teaching tool — use the Compare button to show cooperation rates side by side.

### Suggested Discussion Questions

- In Game 1, what happened to cooperation rates over the five rounds? Why did more pairs end up at Defect/Defect by the final round? Did you know this was going to happen?
- Walk me through your reasoning in round 1. What did you choose, and why? Now walk me through your reasoning in round 5. What changed?
- In Game 2 with chat, did you make any agreements with your partner? Did everyone keep their agreements? What happened when they did not — and why did that happen even when both players would have been better off cooperating?
- Defect is always the individually rational choice in a single round. Yet Cooperate/Cooperate earns both players more than Defect/Defect. How can something be individually rational and collectively irrational at the same time?
- In the indefinitely repeated game, some pairs maintained cooperation throughout. What allowed them to do that? What would happen to their strategy if they knew exactly when the game would end?
- Where do you see Prisoner's Dilemma structures in the real world? What features of those situations make cooperation more or less likely to emerge?

### Theoretical Connections

This experiment provides a direct, lived demonstration of several foundational concepts in game theory and economics:

- Dominant Strategy: Defect strictly dominates Cooperate — it yields a higher payoff regardless of what the partner does. Students experience this pull firsthand rather than just reading about it.
- Nash Equilibrium: The (Defect, Defect) outcome is a Nash equilibrium — neither player can improve their payoff by unilaterally switching. Yet it is Pareto-inferior to (Cooperate, Cooperate), illustrating that Nash equilibria need not be efficient.
- Backward Induction and End-Game Effects: In a finitely repeated game with a known endpoint, cooperation unravels from the final round backward. The data from round 5 typically shows a clear spike in defection that illustrates this logic without the need for a formal proof.
- The Folk Theorem: Indefinite repetition can support cooperative outcomes as Nash equilibria if players are sufficiently patient. With default payoffs, this holds when the ending probability is 50% or less — a condition students can calculate and verify against their own behavior.
- Credible Commitment and Cheap Talk: Chat-enabled games reveal the limits of verbal promises. Because defection is always tempting, promises to cooperate are cheap talk unless backed by a credible punishment strategy (such as grim trigger or Tit-for-Tat).
- Tit-for-Tat: The robot strategy that mirrors the partner's previous action is one of the most robust strategies in iterated Prisoner's Dilemma research. Discussing why Tit-for-Tat performs well connects the experiment to Axelrod's famous tournaments.

## 5. Results & Data Interpretation

### Reading the Results Dashboard

Four views are available in the Results tab:

- Graphs Tab — Action Frequencies by Round: Shows the incidence of the three outcomes per round (C/C, D/D, and C/D or D/C mixed). This is the primary view for the debrief. The typical pattern — C/C declining and D/D rising across rounds — is immediately visible. Use the legend checkboxes to focus on one outcome at a time.
- Tables Tab — Likelihood of Each Action by Round: Displays the percentage of students choosing C or D in each round. Useful for precise numbers, but note that aggregate percentages can mask heterogeneity — half the class at C/C and half at D/D will show 50% C in the table, which looks the same as everyone playing mixed strategies.
- Raw Data Tab: Provides each player's choice, group, and round in a table. Download the CSV for detailed analysis or to extract chat transcripts from chat-enabled sessions.
- Compare Button: After running multiple sessions, use Compare to display results side by side. Showing Game 1 (no chat) next to Game 2 (with chat) is one of the most effective debrief visuals — the difference in cooperation rates is typically striking.


## Common Outcome Patterns

- Declining Cooperation (expected — finite game, no chat): C/C starts moderately high in round 1 and declines toward D/D by round 5. The final-round spike in defection is typically sharp and reliable. Use this pattern to motivate the backward induction argument.
- Higher Cooperation with Chat (expected — finite game, with chat): C/C starts noticeably higher in round 1 and the overall cooperation rate is elevated, but end-game defection still occurs. This gap between 'with' and 'without' chat is the key comparison.
- Sustained Cooperation (indefinitely repeated): Some pairs maintain C/C throughout, particularly when chat is enabled. Highlight these pairs and ask them to explain their strategy — they often describe something close to grim trigger without knowing the term.
- Heterogeneity Between Groups: Even under identical conditions, some pairs cooperate and others defect. This is pedagogically valuable — it opens discussion about what factors (trust, communication style, risk tolerance) affect equilibrium selection.

## 6. Extensions & Variations

### Parameter Adjustments

- Custom Payoff Matrix: The Instructor Specified Matrix game allows you to design your own payoff structure. Use this to explore variations where defection is more or less tempting, or to test whether students recognize a Prisoner's Dilemma when the payoffs are disguised.
- Announce vs. Conceal the Number of Rounds: Concealing the endpoint from students prevents backward induction and typically produces higher cooperation rates in finite games. The contrast between announced and concealed endpoints is a clean experimental manipulation.
- Vary the Ending Probability: In the indefinitely repeated version, adjust the ending probability to shift the theoretical threshold for cooperative equilibria. This lets students see their own behavior change in response to changing incentives.
- Replay with Role Rotation: Use Repeat > Rotate Role to let students experience both sides of the matrix across sessions, building a fuller understanding of the strategic situation.
- Blank Survey Pairing: Run a short MobLab survey after the game to check conceptual understanding — for example, ask students to identify the dominant strategy or calculate the minimum discount factor needed for cooperation. Reflection questions immediately after play significantly improve retention.

### Cross-Disciplinary Connections

- Business Strategy — Price Wars and Collusion: Oligopolists face a Prisoner's Dilemma when deciding whether to undercut competitors on price. The experiment makes visceral why collusion is tempting, why it is unstable, and why antitrust law focuses on facilitating practices that help firms maintain cooperative outcomes.
- Organizational Behavior — Team Incentives: When individual and team rewards are misaligned, employees face a Prisoner's Dilemma between shirking and contributing. The chat-enabled session illustrates how communication, monitoring, and norms can shift behavior.
- International Relations — Arms Races and Treaties: Nations choosing between arming and disarming face the same payoff structure. The indefinitely repeated game with chat maps naturally onto the role of international institutions, verification mechanisms, and credible commitment in treaty compliance.
- Environmental Economics — Common Pool Resources: Firms deciding whether to pollute or abate face a Prisoner's Dilemma when abatement is costly and pollution has diffuse negative externalities. The experiment motivates the case for regulation as a cooperation-enforcing mechanism.

## 7. References & Further Reading

The following resources pair well with this experiment:

- Axelrod, R. (1984). The Evolution of Cooperation. Basic Books. The foundational study of iterated Prisoner's Dilemma tournaments and the success of Tit-for-Tat — highly accessible for undergraduates.
- Dixit, A. & Nalebuff, B. (2008). The Art of Strategy. W.W. Norton. Chapters on repeated games and cooperation offer an applied, non-technical treatment that connects directly to the classroom experiment.
- Osborne, M.J. (2003). An Introduction to Game Theory. Oxford University Press. Formal treatment of dominant strategies, Nash equilibrium, and the Folk Theorem for more advanced courses.
- Mankiw, N.G. (current edition). Principles of Economics. Cengage. Chapter on Oligopoly covers the Prisoner's Dilemma as applied to firm competition — a natural pairing for introductory courses.
- Fudenberg, D. & Tirole, J. (1991). Game Theory. MIT Press. Advanced treatment of repeated games and Folk Theorem for graduate-level courses.