INTERNATIONAL DAY OF MATHEMATICS
MARCH 14

The Evolution of Trust

This activity is an introduction to Game Theory. More specifically, it is an activity around the famous “prisoner’s dilemma” and the concepts of formal game, strategy, cooperation/defection, and collaboration. This activity analyzes a (simplified) competition situation between two opponent players mathematically, and gives conditions under which the players can develop a cooperation strategy to get mutual benefit, instead of applying a hostile competition strategy.

Game theory and this type of analysis has applications to economics, politics, conflict resolution, but also to evolutionary biology in the context of competing species.

This activity is based on and uses the online game *The Evolution of Trust,* by Nicky Case (https://ncase.me/trust/). The game, in turn, is based on the book *The Evolution of Cooperation* by Robert Axelrod (original English edition: Basic Books, 1984).

Participants:

Recommended ages: 14 and up. No prior mathematical knowledge is needed, but some logic arguments, such as identifying paradoxes and making deductions, may require some mature mathematical thinking. Although the activity is not about politics or ethics, implications can be drawn and spark ideological debate.

Preparations:

The first part of the activity requires a pile of coins and some pieces of paper. The number of coins is (ideally) 30 times the number of students. The “coins” can be actual money (with a face value low enough to be inexpensive), or they can be substituted with dry beans, chickpeas, pasta bits, nuts, or any other type of identical small objects that can be sourced inexpensively. The number of coins per student can be reduced by reducing the maximum number of games played per pair.

The second part of the workshop requires the use of a computer by the teacher, connected to a projector or screen so that all the students can watch the simulations and discuss together.

Introduction

Today, we will explore a branch of mathematics that may surprise you. It is called Game Theory, and it is applied to politics, economics, conflict resolution, and also to evolutionary biology. In Game Theory, we design formal *games,* which are situations in which two or more *players* must make rational decisions, trying to reach a *goal* that makes them win the game. Each player can develop a *strategy*, that is, a set of rules for making these decisions, depending on the environment of the game and also on the behavior of the other players.

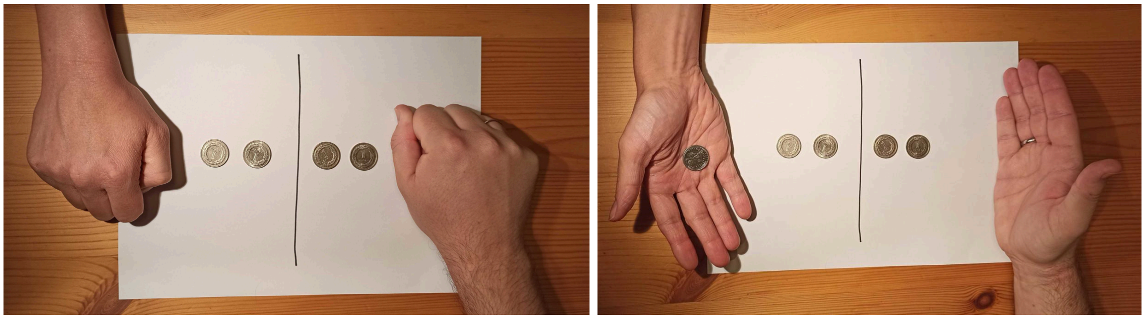
Chess or poker are games, but also retailers setting prices of goods in a competitive market, governments negotiating geopolitical conflicts, or biological species evolving to fit into an ecosystem are real-life examples that can be modeled as games.

We will play a very simple game that can, however, inspire some deep reflection.

The game (one round)

This is a game for two players. They must be sitting at a table, each facing the other. In the middle of the table, place a sheet of paper divided into two halves by a line drawn. The two halves are placed on the right and left sides of the players. The sheet of paper represents the Money Machine. The mechanism of the Machine is the following:

- Four coins are placed on the Machine, two on each side.
- Each player receives a coin.
- Each player can place their coin into the Machine (on their right side of the paper) or not. If player A puts a coin on its side of the machine, the resulting three coins (two from the Machine plus the one put by player A) are given to player B. If player A does not put a coin, player B receives nothing. Conversely, if player B puts a coin, player A receives three coins; otherwise, player A receives nothing.
- Both players must decide their actions at the same time. In order to do that, they, secretly under the table, put a coin or nothing in their right hand. Then both players place their right hands on the table and open them at the same time.
- Your goal in the game is to earn as many coins as possible.

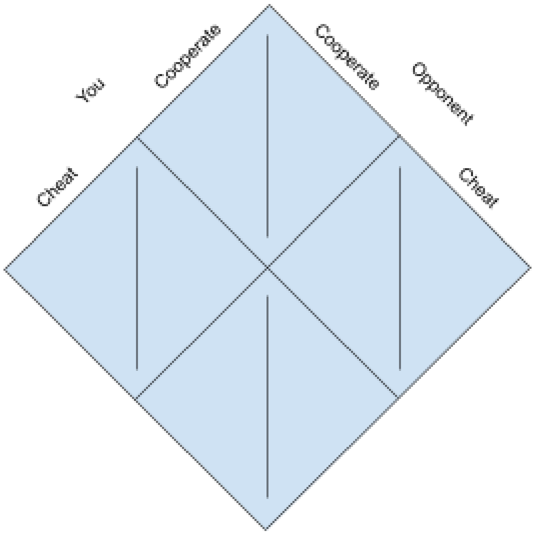


You must pay one coin, so your opponent wins three coins.

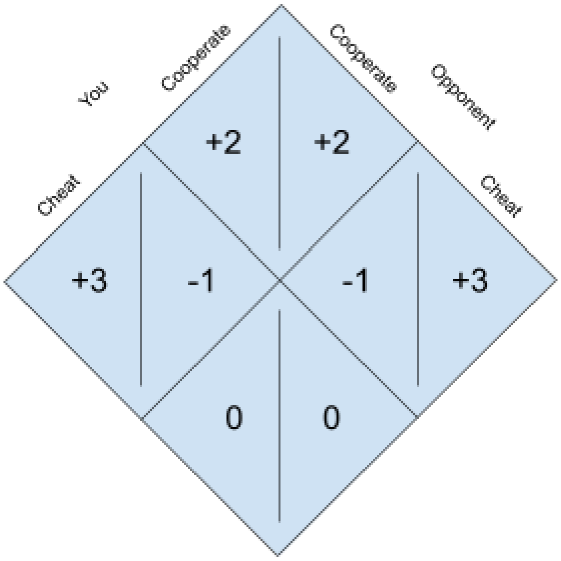
One round analysis. Prisoner’s dilemma

Collect the impressions and strategies from the students.

Let’s call the two actions that you can take “Collaborate” (place a coin and reward your opponent) and “Cheat” (not to place a coin and give no reward to your opponent). Make a double-entry table that shows the four possible outcomes of the game.



On each of the four possible outcomes, place the rewards that each player gets. A negative value means the player loses money.



What is the best strategy?

- If your opponent cooperates, you can either win +2 (if you cooperate) or you can win +3 (if you cheat), so it’s better to cheat.
- If your opponent cheats, you can either lose -1 (if you cooperate), or you can win 0 (if you cheat), so it’s better to cheat.

Therefore, in any case, it is better to cheat. However, the same logic applies to your opponent, so your opponent also decides that it is better to cheat. Therefore, you both cheat, and you don’t win any money. If you both cooperated, the outcome would be better for both of you (+2 to each). Hence the dilemma.

This game is also called the Prisoner’s dilemma, because of a different but analogous formulation: Two thieves are caught in prison. Each one can either accuse the other or remain silent. If both are silent (cooperate), they both get a 1-year prison sentence. If one accuses and the other remains silent, the accuser is released from prison, and the other receives a 3-year sentence. If both accuse, they both receive a 2-year sentence. Each prisoner can see that, for any action of their opponent, it’s better to cheat. However, both reach to the same conclusion, both cheat, and thus each of them receives a 2-year sentence while they could have received just a 1-year sentence if they had cooperated.

Iterated game

It seems that in one game, it is best to cheat even if you could do better, because of the risk of being fooled. You could try to convince your opponent with words, but you can never be sure that they will keep their word. However, the situation changes if we play repeatedly. Then you can base your decisions on the previous actions of your opponent, in addition to the rules of the game.

Each player starts with 10 coins, and there is also a pile of 40 coins (the Bank) that refills the machine[1]. After a round, the Bank resets the machine, so it contains four coins again. The game is played several times, and the goal is to accumulate as many coins as possible.

Note that, again, you can make your opponent receive coins, but it costs you money. Similarly, your opponent can also give you a reward at a cost. Your opponent does not know what you will play in the next round, but he/she remembers your previous actions.

Play several rounds and try to devise a strategy. After playing with a fellow student, switch to playing with another. Finally, share your findings and the strategies you were playing with the rest of the class.

Note the strategies on the blackboard. Some of the strategies that may appear:

- CHEATER: Always cheats.
- COOPERATOR: Always cooperates.
- COPYCAT: Starts cooperating, then does the same as the opponent did in the previous round.

[1] This is for 10 rounds. For n rounds, each of the two players needs to start with n coins and the machine with 4n coins to account for the extreme cases that the players or the machine exhaust all their money.

- GRUDGER: Starts cooperating and keeps cooperating, but if the opponent cheats, they will always cheat from then on.
- DETECTIVE: Plays some rounds to test the opponent (for instance, cooperates, cheats, cooperates, cooperates). Then, if the opponent never cheats, they will exploit and always cheat. If the opponent cheats at any point, they will be cautious and play like COPYCAT.

If any other strategy appears, write it down on the blackboard and give it a nice name.

Discuss which strategy is best.

Key insight: There is no best strategy in absolute terms.

Exercise: Prove that no strategy is the best independently of the opponent's strategy. Hint: Imagine your opponent is CHEATER. What is the best strategy? Imagine your opponent is GRUDGER. What is the best strategy?

In conclusion, there is no way to guarantee the best outcome independently of the strategy of your opponent. This is different from, for instance, chess, where you should always try to play the move that gives you the most advantage.

Key insight: Iterating the game is necessary to develop some trust. Past history helps you to evaluate your opponent. Also, the possibility of a future encounter is needed, so the game should be played for an undetermined number of rounds.

Exercise: Prove that if the players know the number of rounds to be played, the original dilemma resurfaces; that is, you can show that the best strategy is to always cheat. Hint: In the last round, there is no future to influence, so the best strategy is to cheat (as with only one round remaining). Once you and your opponent have decided to cheat in the last round, the last round that matters is the penultimate, so it pays off to cheat there. By induction, both players will always cheat. However, discuss the validity of this inductive argument.

Computer simulations

Now it's time to use the computer to simulate hundreds of games and analyze the results. This will shed some light on which strategy is best.

The teacher opens the computer and the projector for the whole class, and goes to the web game *The Evolution of Trust* by Nicky Case (https://ncase.me/trust/). You can directly jump to section **3. One tournament**.

Here, we compare five strategies by comparing every possible match between two opponents with different strategies.

Note: Place your bet, or make every student place a bet, as said in the game. You can analyze every match between two players. You can ignore the comments about the WWI story.

Key insight: The winner is COPYCAT. This strategy, also called TIT FOR TAT, can be seen as a moral principle: do to others as the others do to you. It combines a willingness to cooperate, but at the same time, it defends itself against abuses from others. Crucially, note that the fact that this strategy is effective is a mathematical fact, not anything imposed by any human ethic.

Evolution

Now, let us simulate what would happen in a scenario with many players, each using a single strategy. Some of the players will perform better (winners) than others (losers). Let's say that the losers decide to change strategy and copy the strategy of the winners. What would happen in the long run? Will the population evolve to a state in which everyone applies the same strategy?

In the online game, go to **4. Repeated Tournament**. Follow the explanation slowly to understand why COPYCAT wins the tournament again.

Distrust

Although it seems that COPYCAT is a winning strategy, this is a very delicate game. Things could go wrong for several reasons:

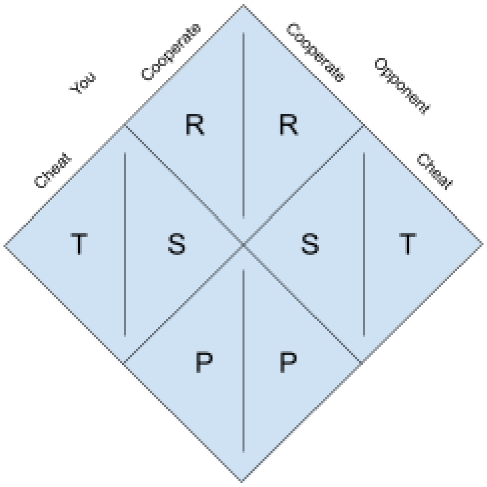
- Lack of enough interactions. The simulations show that with fewer rounds per match, COPYCAT is no longer a winning strategy.

Key insight: With few interactions, for example, just one round, there is no time to build enough trust to collaborate, and the CHEATER can get better rewards.

- Not enough incentive to cooperate. Change the payoffs and see that the winning strategy changes.

Key insight: This is a non-zero-sum game. This means both players can have a win-win situation, which fosters collaboration. In this game, it comes from the fact that the machine is producing new money. The opposite is a zero-sum game, in which everything that one player wins must be a loss for the other player. Then it is impossible to develop a collaboration because there is no common good to achieve.

Exercise. Take the trade-offs on the chart



where:
R = reward for mutual cooperation
S = sucker's payoff
T = temptation to cheat
P = punishment for mutual cheating

Part 1: Order the four quantities to encourage cooperation.
Answer:

$$T > R > P > S$$

Part 2: Find a condition so that cooperation is better than winning half of the time and losing half of the time (lack of cooperation by each player exploiting the other by turns)
Answer:

$$R > \frac{T+S}{2}$$

These two equations define a "prisoner's dilemma game".

Mistakes

Let's assume that a player has a small probability of making an error, causing them to do the opposite of what they intend according to their strategy.

In this case, for instance, COPYCAT would enter a chain of mutual retaliations forever.

Introduce some new strategies that deal with random mistakes

- COPYKITTEN: Similar to COPYCAT, it will copy the last move of the opponent if it is "cooperate", but will cheat only if the two last moves of the opponent were cheats. It is more forgiving than COPYCAT.

- SIMPLETON: Starts by cooperating. Then, if the opponent cooperates, they repeat the previous move; if the opponent cheats, they do the opposite.
- RANDOM: Chooses randomly at 50/50 chances.

Try tournaments with these new players.

Sandbox

Try to find a configuration of the game (number and type of players, payoffs, tournament rules), without all players initially identical, so that:

- RANDOM wins.
- COOPERATOR wins.
- Two different types of players survive.
- Three different types of players survive.
- Explore on your own

Conclusions

Recap all the key insights and open a discussion/debate.

Some key points to foster collaboration are:

- Repeated interactions
  - Can't collaborate if we don't have time to learn from each other.
  - Can't collaborate if there is no future to influence.
- Possibility of win-win situations.
  - You can't collaborate in a zero-sum game. But most games in real life are non-zero-sum games, allowing all players to win something.
  - Recall the mathematical constraints on T,R,S,P.
- Low miscommunication
  - It pays off to be forgiving and overcome some small miscommunication.
- There is no best strategy independent of the other players, you must adapt to the environment.
- Don't be envious
  - You can't collaborate if you expect the best outcome for you individually, or if your goal is to outperform others.
- Don't be the first to cheat
  - You must react if you are cheated on, but cheating first develops mistrust, and collaboration needs to be restored later.
- Reciprocate cooperation and cheating
  - Lack of response will lead to abuse or mistrust.
- Don't be too clever
  - Let other players understand your strategy. This develops trust since the outcome is more predictable.

Mathematical background and resources

This activity is based on the game The Evolution of Trust, by Nicky Case (https://ncase.me/trust/).

The game is based on the book *The Evolution of Cooperation* by Robert Axelrod (1984) and itssequel *The Complexity of Cooperation* (1997) by the same author.

More resources on https://ncase.me/trust/notes/

Create and Share!
Share the participants' findings using the hashtags **#idm314trust** and **#idm314**.