Game Theory

6. Repeated Games

6.1. Example: Repeated Prisoners' Dilemma

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel and Robert Mattmüller

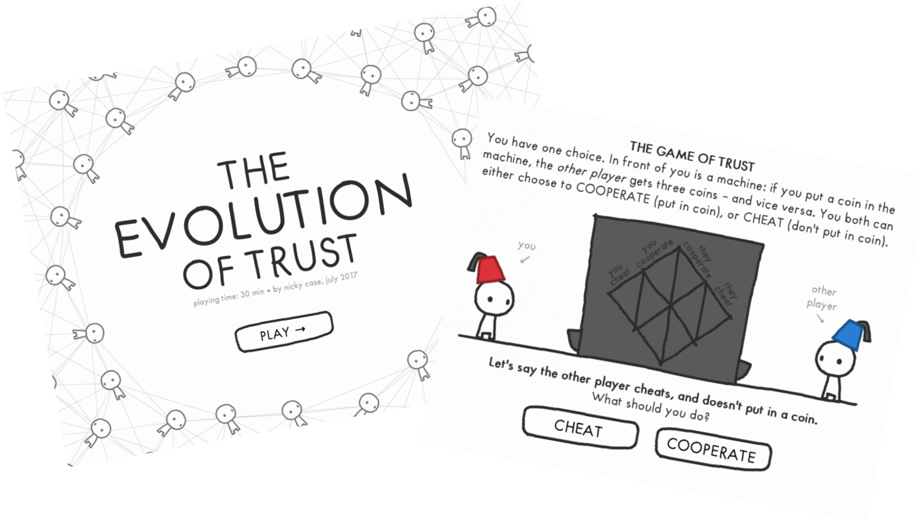
Summer semester 2020

Motivation and Intuition

- Remember: the Prisoner's Dilemma leads to the unsatisfying result (that both defect) because there is neither experience nor future encounters.
- What if the game is played repeatedly?
- Model this as an extensive game where in each turn, we repeat a given base game.
- Will social norms evolve?
- Will punishments, which can lead to short-term costs, nevertheless be played (are these potential punishments credible threats?)

See (and play!): The Evolution of Trust

(http://ncase.me/trust/)



Reminder: Prisoners' Dilemma

	C	D
C	$3,3$	$0,4$
D	$4,0$	$1,1$

(D,D) (i. e., both players defect) is the unique Nash equilibrium, the pair of maximinimizers and the pair of strictly dominant strategies.

So, in a single encounter, there is no argument for rationally playing C!

Repeated Prisoners' Dilemma

Terminology: The game that is played repeatedly is called the stage game.

Question: How often to repeat the stage game?

Possible answers:

- Finitely: repeat pre-specified number of times k.
- Infinitely: repeat infinitely often.
- Indefinitely: after each step, terminate with probability $0 < p < 1$.

Finitely Repeated Prisoners' Dilemma

Assume we play the prisoners' dilemma a pre-specified number of times k.

$\rightsquigarrow$ extensive game with perfect information and simultaneous moves

Finitely Repeated Prisoners' Dilemma

What will be a subgame perfect equilibirium?

Use backward induction:

- (D,D) is the NE in the last subgame, since this would be the only NE in the one-shot game.
- So, (D,D) will also be played in period $k-1$.
- $\ldots$
- So, the (only) subgame-perfect equilibrium and the only NE of this repeated game is $(D,D),(D,D),\ldots,(D,D)$.

$\rightsquigarrow$ players still defect all the time

$\rightsquigarrow$ allowing finitely many repetitions not really helpful!

Infinitely Repeated Prisoners' Dilemma

If we play the prisoners' dilemma infinitely often, we need to solve two problems:

1. How to define a strategy?
2. How to define the payoffs or preferences?

How to specify a strategy using only finite resources?

- In general: one could use an algorithm.
- Usually done in game theory: use Moore automata, i.e., finite state automata, where the inputs are actions of the other players, and in each state, a response action to the previous actions is generated.
- A Nash equilibrium is a profile of automata (strategies) such that no deviation is profitable.

Infinitely Repeated Prisoners' Dilemma

Payoffs/Preferences

How to specify players' preferences using only finite resources?

- Use preferences from the stage game.
- Derive preferences over infinite repetitions using:
  - discounting future payoffs, or
  - limit of means criterion, or
  - overtaking criterion

Indefinitely Repeated Prisoners' Dilemma

Terminating with probability $0 < p < 1$ is equivalent to discounting future payoffs in infinitely repeated game with discount factor $1 - p$.

$\rightsquigarrow$ no need to study indefinitely repeated games separately

$\rightsquigarrow$ focus on infinitely repeated games

Summary

- Repeated games are extensive games with perfect information and simultaneous moves, in which a base strategic game (the stage game) is played in each round.
- Finitely, infinitely, or indefinitely many repetitions possible.
- Often, finitely many repetitions not helpful.

Game Theory

6. Repeated Games

6.2. Strategies and Preferences in Infinite Games

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel and Robert Mattmüller

Summer semester 2020

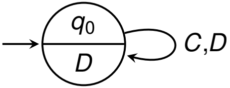
Some Possible Strategies

Using Moore automata, we can specify what to do in response to the new input (action played by others) and the state we are in. Since the automata are finite, this requires only finite memory!

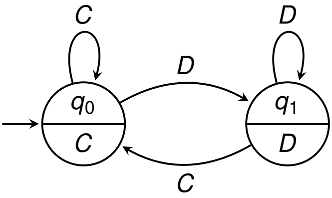
- Unconditionally cooperative: Always play C.
- Unconditonally uncooperative: Always play D.
- Tit-for-Tat: Start with C and then reply with C to each C and with D to each D.
- Grim: Start with C. After any play of D, play D in the future forever.
- Bipolar: Start with D and then always alternate between C and D.

Strategies as Automata

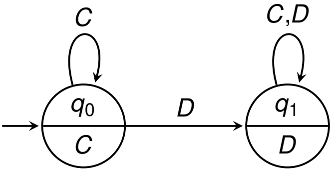
Uncooperative:



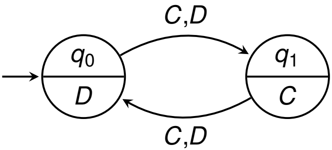
Tit-for-tat:



Grim:

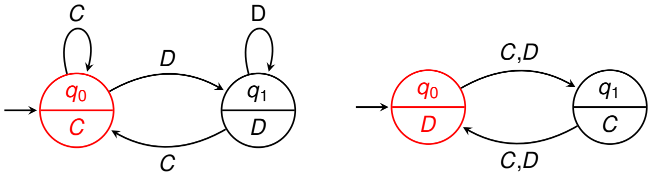


Bipolar:



Game Traces

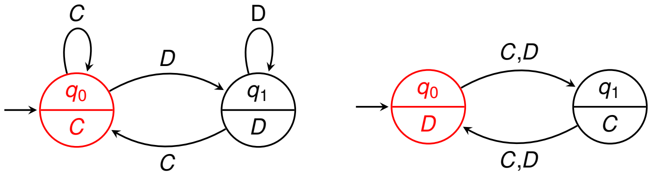
Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.



Round	Action	Utility	Accumulated payoff

Game Traces

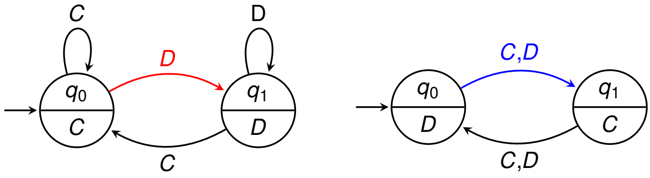
Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.



Round	Action	Utility	Accumulated payoff
1	(C,D)	(0,4)	(0,4)

Game Traces

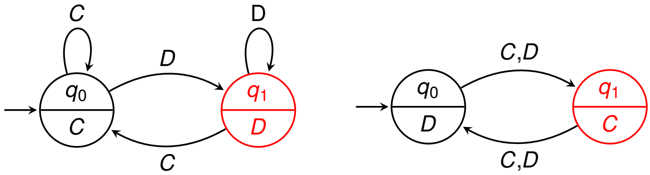
Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.



Round	Action	Utility	Accumulated payoff
1	(*C*,*D*)	(0,4)	(0,4)

Game Traces

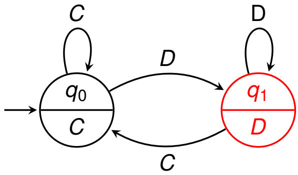
Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.

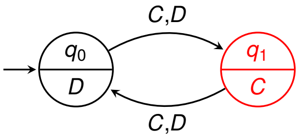


Round	Action	Utility	Accumulated payoff
1	(C,D)	(0,4)	(0,4)

Game Traces

Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.

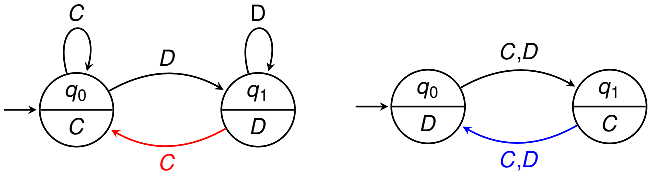




Round	Action	Utility	Accumulated payoff
1	(C,D)	(0,4)	(0,4)
2	(D,C)	(4,0)	(4,4)

Game Traces

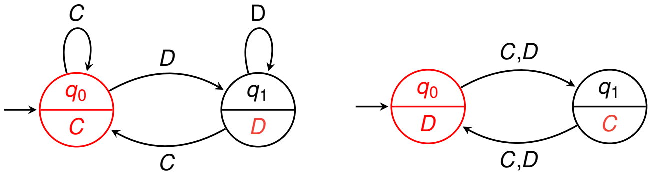
Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.



Round	Action	Utility	Accumulated payoff
1	(C,D)	(0,4)	(0,4)
2	(D,C)	(4,0)	(4,4)

Game Traces

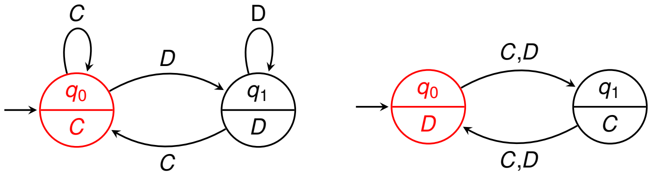
Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.



Round	Action	Utility	Accumulated payoff
1	(C,D)	(0,4)	(0,4)
2	(D,C)	(4,0)	(4,4)

Game Traces

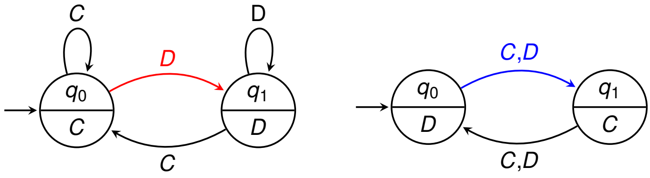
Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.



Round	Action	Utility	Accumulated payoff
1	(C,D)	(0,4)	(0,4)
2	(D,C)	(4,0)	(4,4)
3	(C,D)	(0,4)	(4,8)

Game Traces

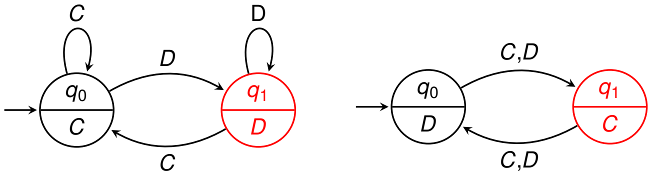
Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.



Round	Action	Utility	Accumulated payoff
1	(C,D)	(0,4)	(0,4)
2	(D,C)	(4,0)	(4,4)
3	(C,D)	(0,4)	(4,8)

Game Traces

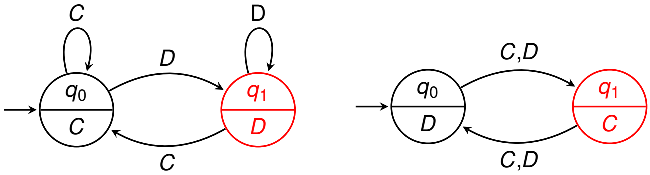
Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.



Round	Action	Utility	Accumulated payoff
1	(C,D)	(0,4)	(0,4)
2	(D,C)	(4,0)	(4,4)
3	(C,D)	(0,4)	(4,8)

Game Traces

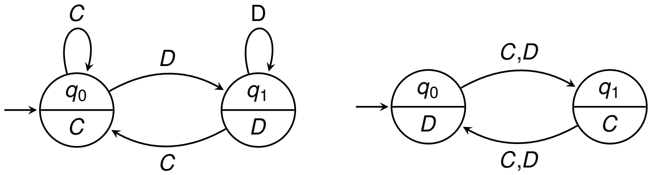
Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.



Round	Action	Utility	Accumulated payoff
1	(C,D)	(0,4)	(0,4)
2	(D,C)	(4,0)	(4,4)
3	(C,D)	(0,4)	(4,8)
4	(D,C)	(4,0)	(8,8)

Game Traces

Player 1 plays tit-for-tat, player 2 plays bipolar; 4 rounds.



Round	Action	Utility	Accumulated payoff
1	(C,D)	(0,4)	(0,4)
2	(D,C)	(4,0)	(4,4)
3	(C,D)	(0,4)	(4,8)
4	(D,C)	(4,0)	(8,8)

Preferences over Payoff Traces

How to define the payoff of an infinite game or whether to prefer one outcome over another one?

Given two infinite sequences $(v_i^t)_{t=1}^{\infty}$ and $(w_i^t)_{t=1}^{\infty}$ of payoffs, we will define when the first is preferred over the second by player i: $(v_i^t)_{t=1}^{\infty} \succsim_i (w_i^t)_{t=1}^{\infty}$.

(Strict preference $\succ_i$ and indifference $\approx_i$ defined similarly.)

Preferences over Payoff Traces

Option 1: Discounting

Discount future payoffs by a discount factor $\delta \in (0,1)$:

$$(v_i^t)_{t=1}^{\infty} \succsim_i^{\delta} (w_i^t)_{t=1}^{\infty} \qquad \text{iff} \qquad \sum_{t=1}^{\infty} \delta^{t-1}(v_i^t - w_i^t) \geq 0$$

$\sum_{t=1}^{\infty} \delta^{t-1} v_i^t$ is considered the payoff in the repeated game.

Example

- $(1,-1,0,0,\dots) \succ_i^{\delta} (0,0,0,0,\dots)$ for any $\delta \in (0,1)$
- $(-1,2,0,0,\dots) \succ_i^{\delta} (0,0,0,0,\dots)$ iff $\delta > 1/2$
- $(1,1,1,1,\dots) \approx_i^{\delta} (1/(1-\delta),0,0,0,\dots)$

Note: $\sum_{t=1}^{\infty} \delta^{t-1} = 1/(1-\delta)$

Preferences over Payoff Traces

Option 2: Limit of Means

Compare average payoffs in the limit (limit of means criterion):

$$(v_i^t)_{t=1}^{\infty} \succsim_i^{\mathrm{lom}} (w_i^t)_{t=1}^{\infty} \qquad \text{iff} \qquad \liminf_{T\to\infty} \sum_{t=1}^{T} (v_i^t - w_i^t)/T \geq 0$$

If $\lim_{T\to\infty} \sum_{t=1}^{T} v_i^t / T$ exists, this is considered the payoff in the repeated game.

Example

- $(1, -1, 0, 0, \dots) \approx_i^{\mathrm{lom}} (0, 0, 0, 0, \dots)$
- $(-1, 2, 0, 0, \dots) \approx_i^{\mathrm{lom}} (0, 0, 0, 0, \dots)$
- $(\underbrace{0, \dots, 0}_{m}, 1, 1, 1, \dots) \succ_i^{\mathrm{lom}} (1, 0, 0, 0, \dots)$ for all $m \in \mathbb{N}$

  (Note: For every δ there exists an m^* such that for all $m > m^*$ the preference under discounting is reversed.)

Overtaking criterion:

$$(v_i^t)_{t=1}^{\infty} \succsim_i^{\mathrm{ot}} (w_i^t)_{t=1}^{\infty} \qquad \text{iff} \qquad \liminf_{T \to \infty} \sum_{t=1}^{T} (v_i^t - w_i^t) \geq 0$$

Example

- $(1, -1, 0, 0, \ldots) \approx_i^{\mathrm{ot}} (0, 0, 0, 0, \ldots)$
- $(-1, 2, 0, 0, \ldots) \succ_i^{\mathrm{ot}} (0, 0, 0, 0, \ldots)$

Infinitly Repeated Games Formally

Definition (Infinitely repeated game of G)

Let $G = \langle N, (A_i)_{i \in N}, (u_i)_{i \in N} \rangle$ be a strategic game with $A = \prod_{i \in N} A_i$. An infinitely repeated game of G is an extensive game with perfect information and simultaneous moves $\langle N, H, P, (\succsim_i)_{i \in N} \rangle$ in which

- $H = \{\langle\rangle\} \cup (\bigcup_{t=1}^{\infty} A^t) \cup A^{\infty}$,
- $P(h) = N$ for all nonterminal histories $h \in H$, and
- $\succsim_i$ is a preference relation on A^{∞} that is based on discounting, limit of means or overtaking.

Summary

- Strategies in infinitely repeated games are described using finite Moore automata.
- For preferences over the outcomes of infinitely repeated games, different preference criteria are possible:
  - discounting
  - limit of means
  - overtaking

Game Theory

6. Repeated Games

6.3. Analysis of the Infinitely Repeated Prisoners' Dilemma

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel and Robert Mattmüller

Summer semester 2020

Let us consider the Grim strategy. Is (Grim, Grim) a Nash equilibrium under all the preference criteria?

Recall: (Grim, Grim) leads to outcome $(C,C),(C,C),(C,C),\dots$.

Limit of means and overtaking criterion: (Grim, Grim) is a Nash equilibrium! Any deviation will result in getting ≤ 1 instead of 3 infinitely often.

Discounting: This is a bit more complicated.

- If a player gets v for every round, he will accumulate the following payoff:

$$v + \delta v + \delta^2 v + \ldots = \sum_{i=1}^{\infty} \delta^{i-1} v$$

- Since we know that $\sum_{i=0}^{\infty} \delta^i = \frac{1}{1-\delta}$ (for $0 < \delta < 1$), we have:

$$\sum_{i=1}^{\infty} \delta^{i-1} v = v \sum_{i=0}^{\infty} \delta^i = \frac{v}{1-\delta}$$

UNI FREIBURG

Assume that both players play Grim for the first $k-1$ rounds. Then player 1 deviates and plays D once. In the remainder he must play D in order to get at least 1 in each round.

- Starting in round k, he receives:

$$4+\delta+\delta^2+\ldots = 3+\sum_{i=0}^{\infty}\delta^i = 3+\frac{1}{1-\delta}$$

- If he had not deviated, the accumulated payoff starting at round k would have been:

$$3+3\delta+3\delta^2+\ldots = 3\sum_{i=0}^{\infty}\delta^i = \frac{3}{1-\delta}$$

- $\rightsquigarrow$ deviation is profitable iff $3+1/(1-\delta) > 3/(1-\delta)$ iff $\delta < 1/3$.
- $\rightsquigarrow$ Grim is a NE strategy for $\delta \geq 1/3$.

Under which preference criteria is (Tit-for-tat,Tit-for-tat) an equilibrium?

Recall: (Tit-for-tat, Tit-for-tat) leads to outcome $(C,C),(C,C),(C,C),\dots$.

Limit of means: Finitely many deviations do not change the payoff profile in the limit. Infinitely many deviations lead to lower payoff. So Tit-for-tat is an NE strategy under this preference criterion.

Overtaking: Even only one deviation leads to a payoff of 5 over two rounds instead of 6. So, in no case, a deviation can lead to a better payoff.

Infinitely Repeated Prisoners' Dilemma

Tit-for-tat vs. Tit-for-tat

Discounting: Deviating only in one move in round k and then returning to being cooperative leads to $4 + 0 + \ldots$ instead of $3 + \delta 3 + \ldots$ in round k.

- $\rightsquigarrow$ deviation is profitable iff $4 > 3 + \delta 3$ iff $\delta < 1/3$ (This is the best case for a deviation!)
- $\rightsquigarrow$ Tit-for-tat is a NE strategy for $\delta \geq 1/3$.

- With outcome $(C,C),(C,C),\ldots$:
  - (Grim, Grim)
  - (Tit-for-tat, Tit-for-tat)
  - (Grim, Tit-for-tat)
- With outcome $(D,D),(D,D),\ldots$:
  - (Always-defect, Always-defect)

Summary

- In the repeated Prisoners' Dilemma, it is possible to play Nash Equilibrium strategies that result in infinite (C, C) sequences, i. e., infinite cooperation.
- Outlook: can also be studied from an evolutionary perspective – which strategies survive if whole populations of players are considered?

Game Theory

6. Repeated Games

6.4. Punishments and Enforceable Outcomes

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel and Robert Mattmüller

Summer semester 2020

Punishments

Observe that the Nash equilibrium strategies are based on being able to punish a deviating player.

Definition (Minmax payoff)

Let $G = \langle N, (A_i)_{i\in N}, (u_i)_{i\in N}\rangle$ a strategic game. Player i's minimax payoff in G, also written as $v_i(G)$, is the lowest payoff that the other players can force upon player i:

$$v_i(G) = \min_{a_{-i}\in A_{-i}} \max_{a_i\in A_i} u_i(a_{-i}, a_i).$$

The idea is that the other players all punish a deviating player in the next round(s) and allow him only to get $v_i(G)$.

Enforceable Payoffs

Definition (Feasible payoff profile)

Given a strategic game $G = \langle N, (A_i)_{i \in N}, (u_i)_{i \in N} \rangle$, a vector $v \in \mathbb{R}^N$ is called payoff profile of G if there exists $a \in A$ such that $v = u(a)$. $v \in \mathbb{R}^N$ is called feasible payoff profile if there exists a vector $(\alpha_a)_{a \in A} \in \mathbb{Q}^A$ with $\sum \alpha_a = 1$ and $v = \sum \alpha_a u(a)$.

Note: Such payoffs can be generated in a repeated game by playing a for β_a rounds in a set of γ games with $\gamma = \sum_{a \in A} \beta_a$ and $\alpha_a = \beta_a / \gamma$.

Definition (Enforceable payoff)

Given a strategic game $G = \langle N, (A_i)_{i \in N}, (u_i)_{i \in N} \rangle$, a payoff profile w with $w_i \geq v_i(G)$ for all $i \in N$ is called enforceable. If $w_i > v_i(G)$ for all $i \in N$, it is said to be strictly enforceable.

Many Different Equilibria ...

UNI FREIBURG

- Using the concept of enforceable payoffs, one can construct many different Nash equilibria and payoff profiles for the repeated prisoners' dilemma!
- E. g., $(^{10}/_{3}, 2)$ is a feasible payoff profile, because $4 \times (C, C)$ and $2 \times (D, C)$ leads to $(^{(4\times3+2\times4)}/_{6}, ^{(4\times3+2\times0)}/_{6})$.
- Construct two automata that implement this repeated sequence and in case of deviation revert to playing D.
- These two automata implement Nash equilibrium strategies, since deviating leads to a payoff of 1 instead of $^{10}/_{3}$ or 2!
- Folk theorems stating that all enforceable outcomes are reachable have been proven for the general case.

Summary

- In the repeated prisoners' dilemma, it is possible to achieve any feasible payoff profile under the limit of means criterion.