بسم الله الرحمن الرحيم

Foundations of Game Theory for Electrical and Computer Engineering

Mohammad Hossein Manshaei

manshaei@gmail.com

1394

Let’s solve prisoners’ dilemma!

REPEATED GAMES

Introduction

- We focus on a class of games with several interactions, i.e., **Repeated Game**
- Our ultimate goal is to model and solve the problem of prisoners' dilemma with a repeated interaction, examples include
  - Several *real life situations* such as, friendship, marriage, and wars.
  - *Engineering Applications* such as, multiple access protocols in wireless communications, packet forwarding, and jamming.

A Brief Reminder

- A Nash Equilibrium (s_1*,s_2*,…,s_N*) is a **Sub-Game Perfect Nash Equilibrium (SPNE)** if it induces a Nash Equilibrium in every sub-game of the game
- We looked for the Nash equilibria in each of the sub-games, roll the payoffs back up, and then see what the optimal moves are higher up the tree (e.g., Mixed NE in the BoS Game)

The War of Attrition

- 2-Player game
- In each period of the game each chooses **Fight (F)** or **Quit (Q)**

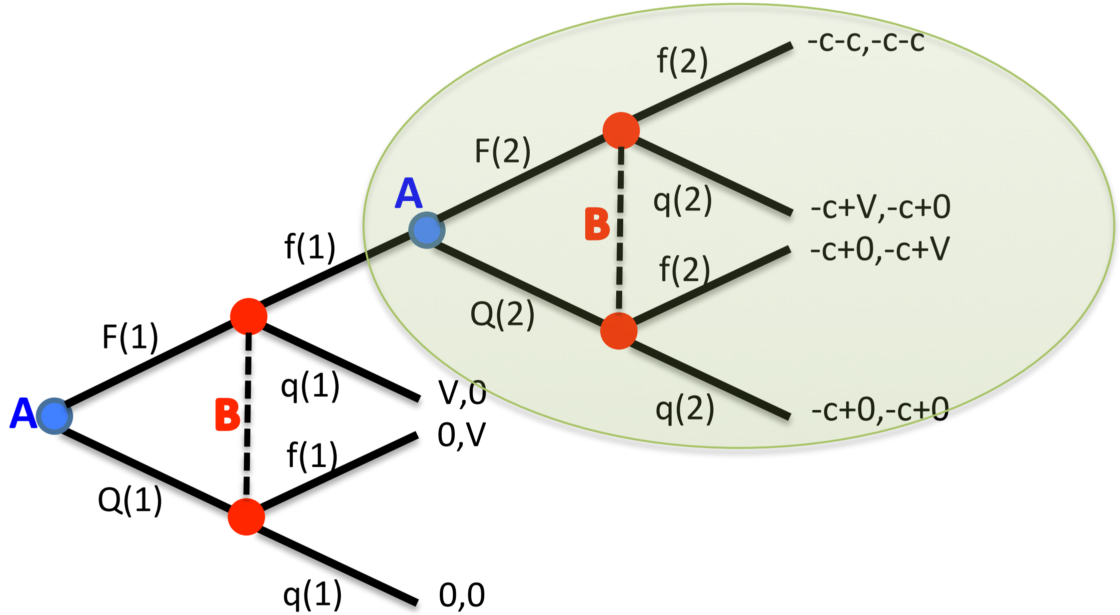
1. If the other player quits first, you win a prize **V**
2. Each period in which both **F**, each player pays cost **-c**
3. If both quit at once they get **0**

(Repeated TDMA Transmission)

- Example: First World War, The Battle for Broadband [IEEE Spectrum, 2005]

Let's Play This Game Repeatedly!

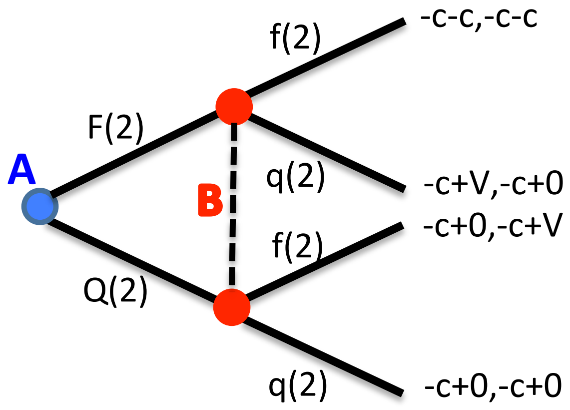
Two Period Game



Assumptions:

- We now focus on a two stage game
- Later we will play it for infinite stages
- We assume that v > c

Analysis of the Second Sub-game

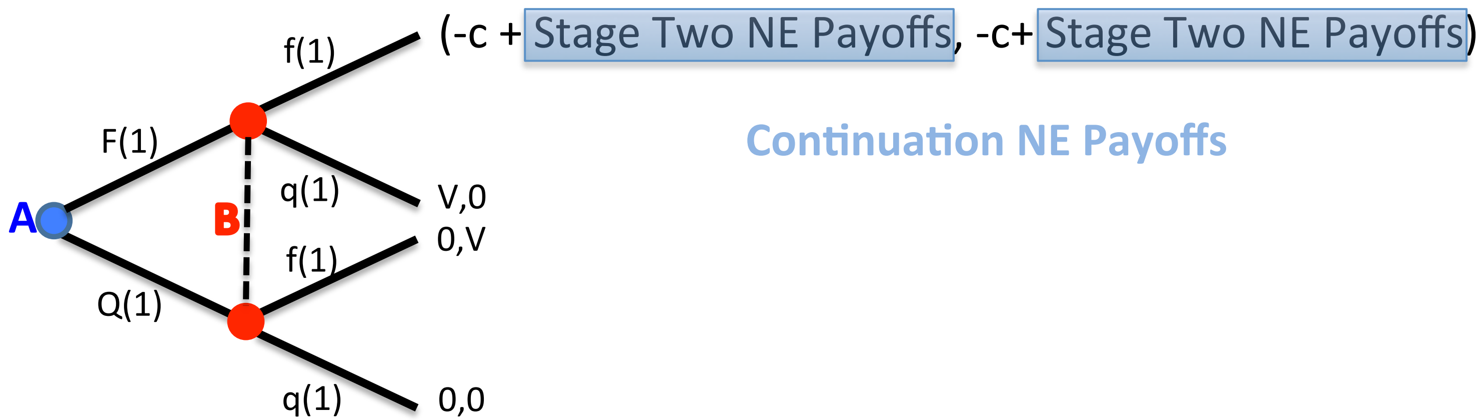


$\boxed{-c\ +}$ **A**

	B: f(2)	B: q(2)
F(2)	-c,-c	V,0
Q(2)	0,V	0,0

- Two Pure-Strategy NE in this sub-game
  - (F(2),q(2)) and (Q(2),f(2))
  - Payoffs are (V,0) and (0,V)
- Note that c (sunk costs) does not matter!

Roll the Payoffs Back Up to the First Sub-Game



B

A	f(1)	q(1)
F(1)	-c+V,-c+0	V,0
Q(1)	0,V	0,0

For **(F(2),q(2))** in stage 2 (tomorrow)
The NE is: (F(1),q(1))

B

A	f(1)	q(1)
F(1)	-c+0,-c+V	V,0
Q(1)	0,V	0,0

For **(Q(2),f(2))** in stage 2 (tomorrow)
The NE is (Q(1),f(1))

Pure Strategy SPNE (V>c)

- Two SPNE (Fighter vs Quitter)
  - [(F(1),F(2)),(q(1),q(2))]
  - [(Q(1),Q(2)),(f(1),f(2))]
- **Main Lesson:** If we know I am going to win tomorrow, then I win today
- **Our first intuition:** Rational players should involve fights in NE, but here one only involves to fight and other quits!
- What did we miss?

Second Sub-game Analysis

Mixed Strategy NE

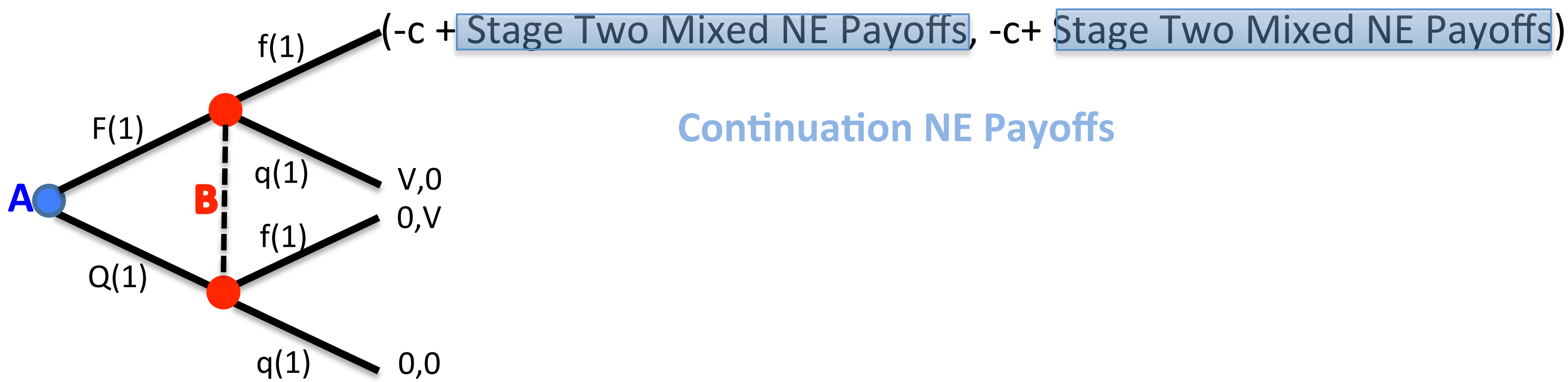
		f(2)	q(2)
		B	
$\boxed{-C+}$ **A**	F(2)	-c,-c	V,0
	Q(2)	0,V	0,0
		p	1-p

- If A Fights ➔ -cp + V(1-p)
- If A Quits➔ 0p + 0(1-p)

V(1-p) = pc ➔ p=V/(V+c) and 1-p = c/(V+c)

- Mixed NE has both fight with probability = V/(V+c)
- Payoffs in this mixed NE = (0,0)
- Probability of fight **increases** in **V** and **decreases** in **c**

Roll the Payoffs Back Up to the First Sub-Game



Continuation NE Payoffs

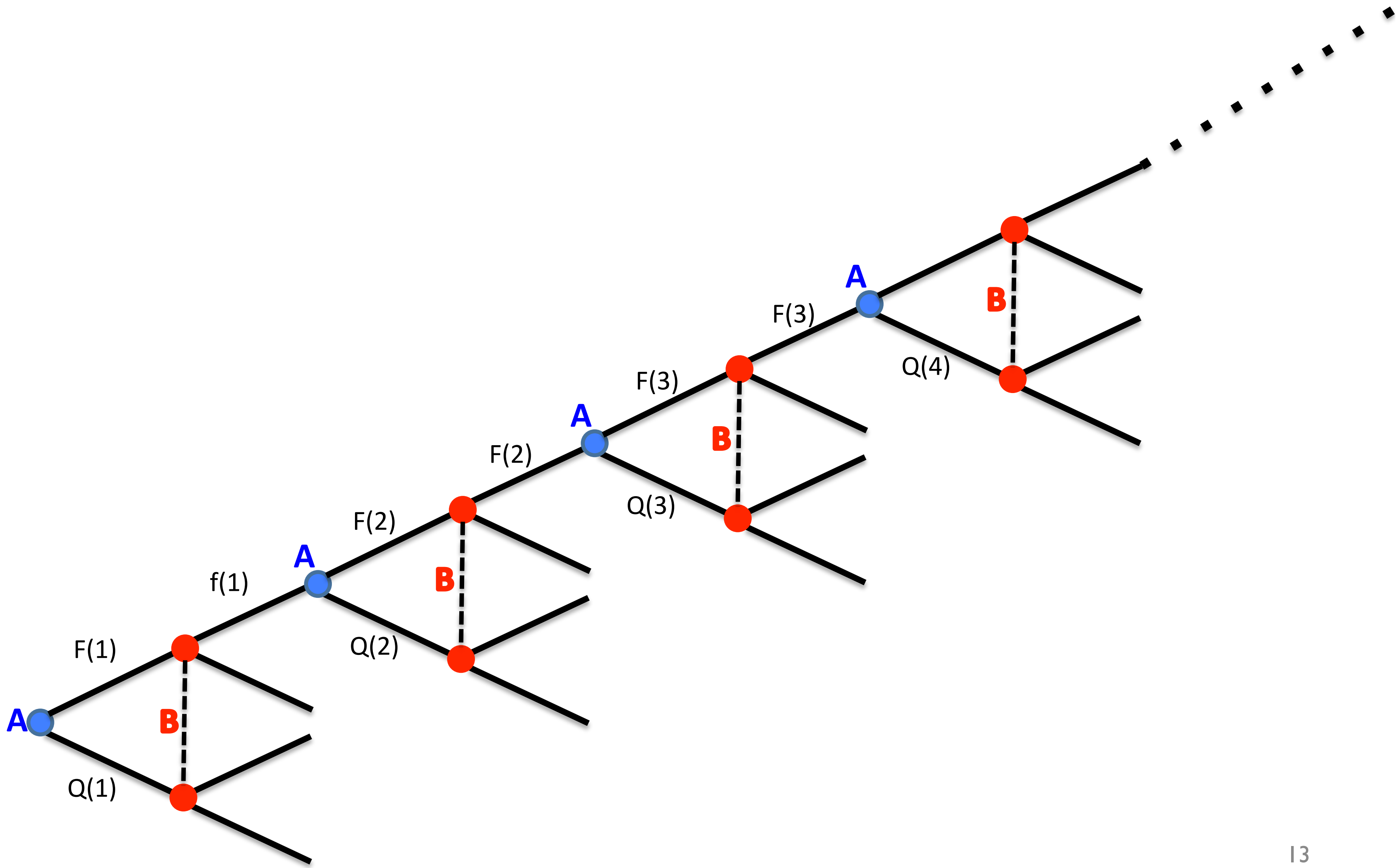
A \ B	f(1)	q(1)
F(1)	-c+0,-c+0	V,0
Q(1)	0,V	0,0

← For the mixed NE in Period 2

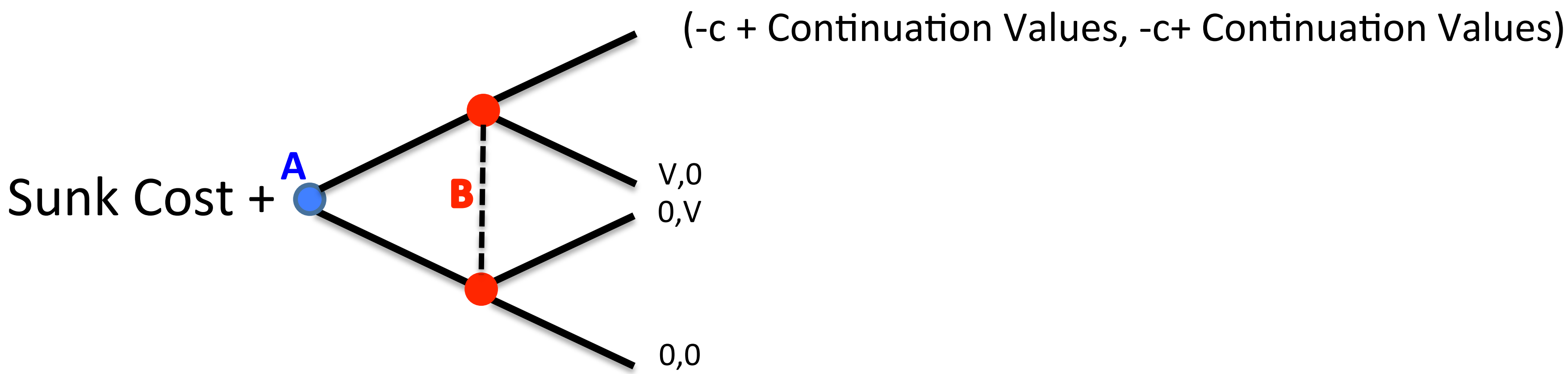
Conclusions

- When we rolled back, the matrix is the same as the second game (last sub-game)
- Same payoff matrix, so …
- Both Fights with $p=V/(V+c)$ ➔ Mixed SPE $[(p^*, p^*),(p^*, p^*)]$
  ➔ Expected payoff is 0
- We end up fighting in each game with probability of p^* (in infinite horizon)

Let’s Play Forever

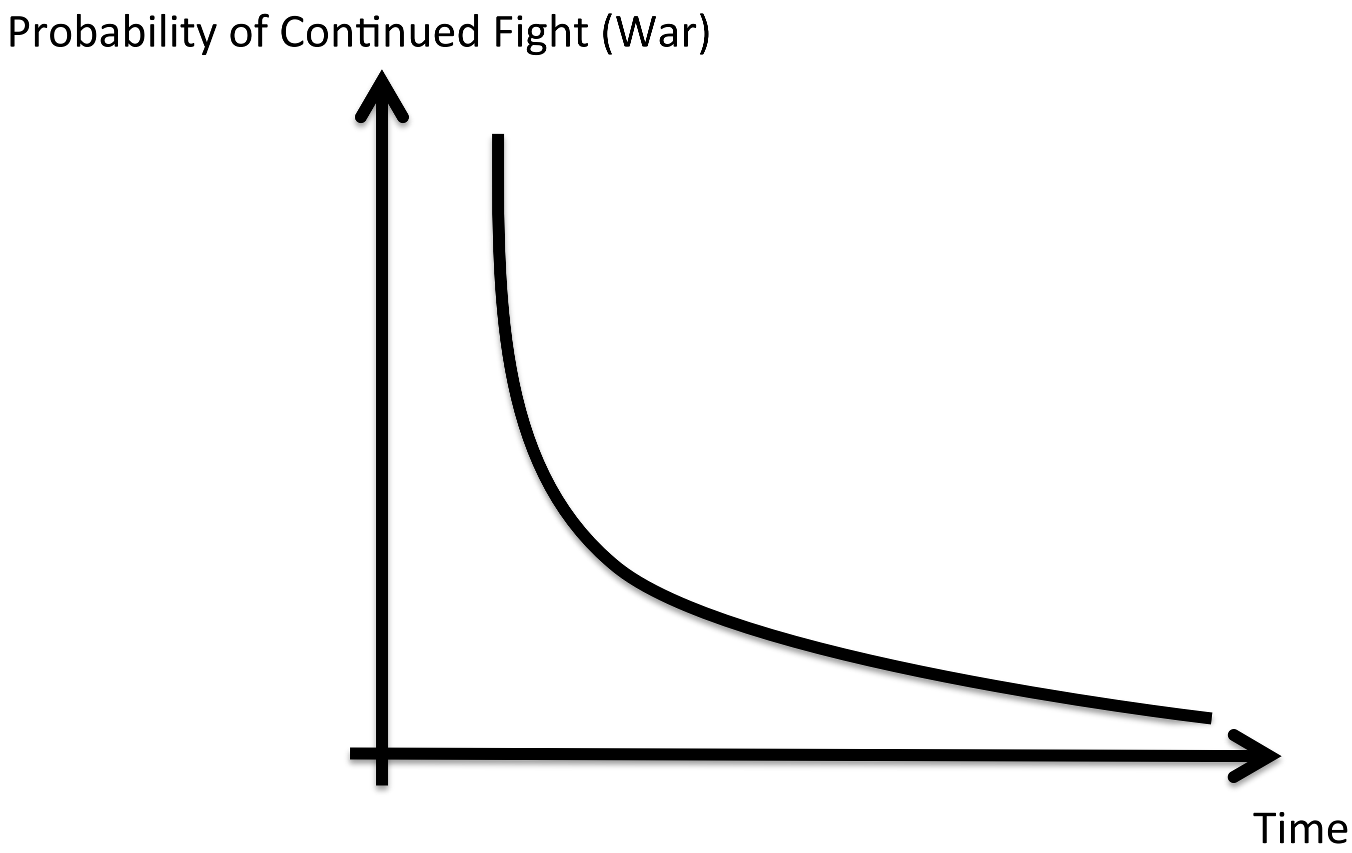


And Solution!



- ✧ Assume that we play in Stage "n" = 659870
- ✧ If mix in future then the continuation value is (0,0)
- ✧ This game is already analyzed (solved for mixed)
- ✧ Both mix with F with $p^{*}= V/(V+c)$
- ✧ And the fight will continue ....

Probability of Continued Fight



Long fight between rational players

Take away Messages

- We could sustain fighting by players that were **rational in a war of attrition.**
  - Break the analysis up into what we might call "stage games"
  - Break the payoffs up into: the payoffs that are associated with that stage and
    - Payoffs that are associated with the past (they're sunk, so they don't really matter)
    - Payoffs that are going to come in the future from future equilibrium play

REPEATED PRISONERS' DILEMMA

Relationships are Repeated

Not Contractual

- **Friendship**
  - If you are nice to me I will nice to you
- **Nations Relationships**
  - Visa issues
- **Exchange goods and Services**
  - Change fruit with petrol

Why Repeated Interactions?

In ongoing relationships the promise of **future rewards** and the **threat of future punishments** may sometimes provide **incentives** for **good behavior today**

Good News: Repeated interaction might get us out of prisoners' dilemma

Prisoners' Dilemma

		B: Coop	B: Defect
A	Coop	2,2	-1,3
A	Defect	3,-1	0,0

Let's Play This Game Repeatedly!

Repeated Solution

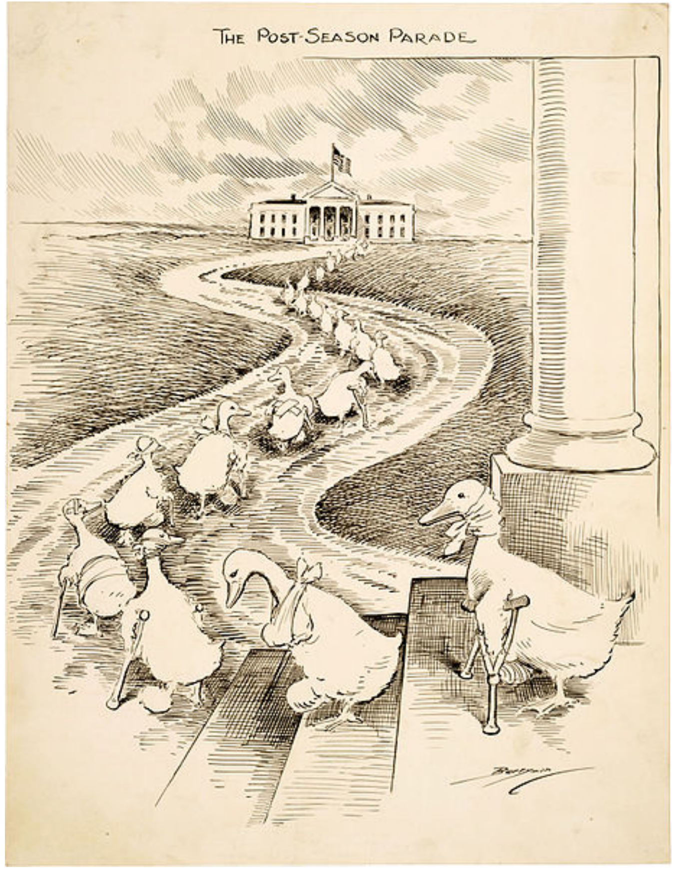
- The game at the last stage is just a simple Prisoners' Dilemma
- Then player should defect
- We put payoffs of tomorrow in the game of today and we have

		B Coop	Defect
A	Coop	2+0,2+0	-1+0,3+0
	Defect	3+0,-1+0	0+0,0+0

Repeated Solution

- It is again a PD game and we have again unraveling from back
- The problem is not solved!!!!

Example:



- Lame Duck Effect

- However, even a Finite game has
  Let's discuss an example

Another Example

	A	B	C
A	4,4	0,5	0,0
B	5,0	1,1	0,0
C	0,0	0,0	3,3

- We play twice
- We would like to sustain (A,A) Cooperation
- But (A,A) is not a NE in one-shot game
- The NE are (B,B) and (C,C)
  - There are some Mixed NE, but let's focus on pure NE

  ➔ We cannot sustain (A,A) in period 2

New Strategy

- **Play A, then:**
  - **Play C if (A,A) was played**
  - **Play B otherwise**
- Pay attention to *information sets*
  - It says what to do in all 1+9 information sets (we are fine!)
- Question: Is this a SPNE?
- Or can we sustain Nash behavior in all subgames?

Repeated Analysis

- For each activity in the first period we make a new sub-game
- In period 2:
  - After (A,A) this strategy induces (C,C)
    - Which is a NE☺
  - After the other choices in period 1, this strategy induces (B,B)
    - Which is a NE☺

We are fine with all subgames after the first move (i.e., 9 subgames)

What about the Whole Game?

- In the whole game:
  - A → u(A,A) + u(C,C) = 4 + 3 = 7
  - If Defect (an obvious defection, check the rest at home☺):
    - B → u (B,A) + u (B,B) = 5 + 1 = 6

Temptation to defect (cheat) today ≤
(Value of rewards <u>tomorrow</u> – Value of punishment <u>tomorrow</u>)

5 - 4 ≤ 3 – 1 → 1 ≤ 2

➔ Temptation is outweighed by the difference between the value of the reward and the value of the punishment

Important Lesson

If a "Stage Game" has more than one NE, then we **may be able** to use the prospect of playing different equilibria tomorrow to **provide incentives** (rewards and punishments) **for cooperating today**

A Brief Comment!

- There may be a problem of renegotiation
  - Between two stages they negotiate to switch to (C,C) ➔ Then there is no incentive to cooperate in the first stage!
  - E.g., 2008 crisis (bail out or bankruptcy)

Repeated Prisoners' Dilemma

		B: Coop	B: Defect
A	Coop	2,2	-1,3
	Defect	3,-1	0,0

- Each round we toss a coin twice and decide to follow the game or not
- Main difference: We do not know the end

Grim Trigger Strategy

		B Coop	Defect
A	Coop	2,2	-1,3
	Defect	3,-1	0,0

- The game continue with probability of δ *(toss a coin two times)*
- **Play C then**
  - **Play C if no one has played D**
  - **Play D otherwise**

Is it NE?

- *Temptation to defect (cheat) <u>today</u> ≤ (Value of rewards <u>tomorrow</u> – Value of punishment <u>tomorrow</u>)*

- 3-2 $\overset{?}{\leq}$ δ (u(C,C) forever – u (D,D) forever)

➢ Why δ ?
  - Because the game may end
  - You need money today!

Is it NE?

$3-2 \overset{?}{\leq} \delta$ (u(C,C) forever – u (D,D) forever)

- u(D,D) forever = 0
- u(C,C) forever = $2 + 2\delta + 2\delta^2 + 2\delta^3 + \ldots$

$= 2/(1-\delta)$

Is it NE?

- Is Grim Trigger a Nash Equilibrium?

$$1 \leq [2/(1-\delta)-0]\ \delta$$

$$\delta \geq 1/3$$

Any other NE?

- What about playing D now, then C, then D forever?
- (D,C), (C,D), (D,D), (D,D), …
- $u = 3 + (-1)\ \delta\ + 0 + 0 + 0 + \ldots = 3 - \delta$
- It is even worse comparing to all D after the first D (D,D,D,…)

→ Punishment (D,D) forever is a SPNE

Lesson

We can get cooperation in Prisoner's Dilemma using the *Grim Trigger Strategy* (as a SPNE) provided $\delta \geq 1/3$

General Lesson

For an ongoing relationship to provide incentives for good behavior today, it helps for there to be a **high probability that the relationship will continue**

(weight you put on the future)

One Period Punishment

- Play C to start, then
  - Play C if either (C,C) or (D,D) were played last
  - Play D if either (C,D) or (D,C) were played last

Is this a SPNE?

- *Temptation to defect (cheat) today ≤ (Value of promise <u>tomorrow</u> – Value of the threat <u>tomorrow</u>)*

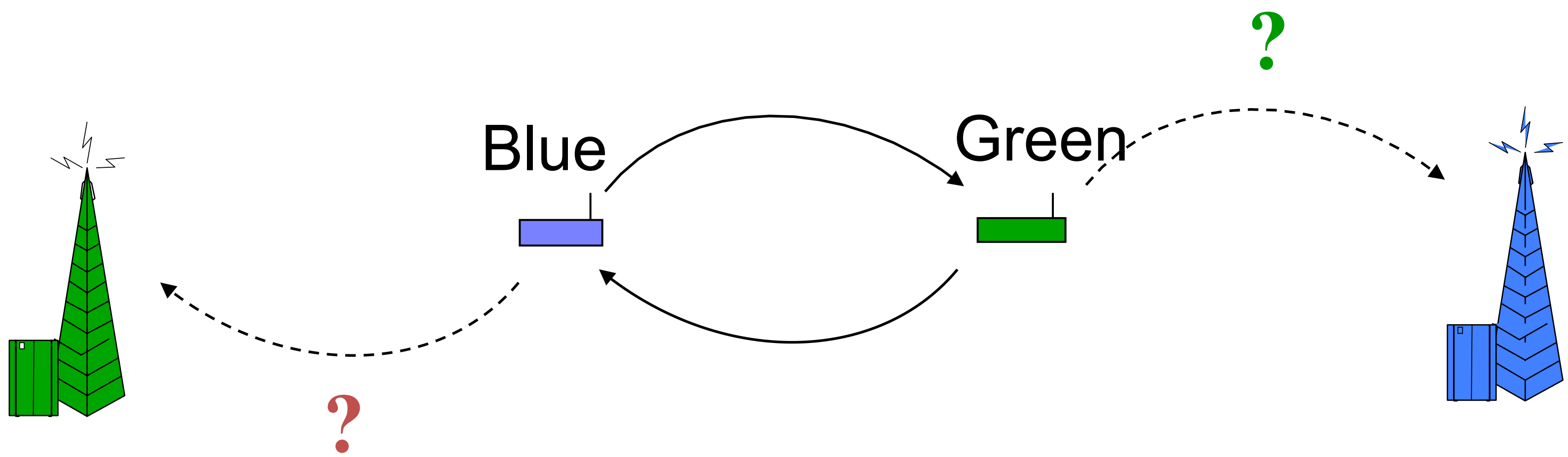
$$3-2 \overset{?}{\leq} [\,(2\text{ “forever”}) - (\text{value of “0” for tomorrow and then “2” forever starting the next day})]\ \delta$$

$$3-2 \leq [2/(1-\delta) - 2\delta/(1-\delta)]\ \delta$$

$$\tfrac{1}{2} \leq \delta$$

Shorter punishments need more weight (δ) on future

The Repeated Forwarder's Dilemma



Blue \ Green	Forward	Drop
Forward	(1-c, 1-c)	(-c, 1)
Drop	(1, -c)	(0, 0)

stage payoff

NE in Finite Repeated FD

In the finite-horizon Repeated Forwarder's Dilemma, the strategy profile (All-D, All-D) is a Nash equilibrium.

Payoffs in the Repeated Game FD

- Finite-horizon vs. infinite-horizon games
- Myopic vs. long-sighted repeated game

myopic: $\bar{u}_i = u_i(t+1)$

long-sighted finite: $\bar{u}_i = \sum_{t=0}^{T} u_i(t)$

long-sighted infinite: $\bar{u}_i = \sum_{t=0}^{\infty} u_i(t)$

payoff with discounting: $\bar{u}_i = \sum_{t=0}^{\infty} u_i(t) \cdot \delta^t$

$0 < \delta \leq 1$ is the discounting factor

Strategies in the Repeated Game FD

- usually, history-1 strategies, based on different inputs:
  - others' behavior: $m_i\left(t+1\right)=s_i\left[m_{-i}\left(t\right)\right]$
  - others' and own behavior: $m_i\left(t+1\right)=s_i\left[m_i\left(t\right),m_{-i}\left(t\right)\right]$
  - payoff: $m_i\left(t+1\right)=s_i\left[u_i\left(t\right)\right]$

Example strategies in the Forwarder's Dilemma:

Blue (t)	initial move	F	D	strategy name
Green (t+1)	F	F	F	AllC
	F	F	D	Tit-For-Tat (TFT)
	D	D	D	AllD
	F	D	F	Anti-TFT

Analysis of the Repeated Forwarder's Dilemma (1/3)

Infinite game with discounting: $\overline{u}_i = \sum_{t=0}^{\infty} u_i(t) \cdot \delta^t$

Blue strategy	Green strategy	Blue utility	Green utility
AllD	AllD	0	0
AllD	TFT	1	-c
AllD	AllC	1/(1-δ)	-c/(1-δ)
AllC	AllC	(1-c)/(1-δ)	(1-c)/(1-δ)
AllC	TFT	(1-c)/(1-δ)	(1-c)/(1-δ)
TFT	TFT	(1-c)/(1-δ)	(1-c)/(1-δ)

Discount Factor Interpretation:

1. The player cares more about the near term payoff than in the long term payoff
2. The player has no preferences, but the game ends with probability of 1- δ in each stage

Analysis of the Repeated Forwarder's Dilemma (2/3)

Blue strategy	Green strategy	Blue utility	Green utility
AllD	AllD	0	0
AllD	TFT	1	-c
AllD	AllC	$1/(1-\delta)$	$-c/(1-\delta)$
AllC	AllC	$(1-c)/(1-\delta)$	$(1-c)/(1-\delta)$
AllC	TFT	$(1-c)/(1-\delta)$	$(1-c)/(1-\delta)$
TFT	TFT	$(1-c)/(1-\delta)$	$(1-c)/(1-\delta)$

- AllC receives a high payoff with itself and TFT, but
- AllD exploits AllC
- AllD performs poor with itself
- TFT performs well with AllC and itself, and
- TFT retaliates the defection of AllD

TFT is the best strategy if δ is high enough!

NE in Infinite Repeated FD

In the Repeated Forwarder's Dilemma, if both players play AllD, it is a Nash equilibrium.

NE in Infinite Repeated FD

In the Repeated Forwarder's Dilemma, both players playing TFT is a Nash equilibrium (if $\delta > c$).

Sketch of Proof:

If one deviate in stage t, then its payoff is:

$(1-\delta)\,[(1+\delta+\delta^2 \ldots + \delta^{t-1})(1-c) + \delta^t] =$

$1-c+\delta^t(c-\delta) \rightarrow$

Hence if "$\delta > c$" there is no temptation to deviate

Or (i.e., other approach):

$1-(1-c) \leq \delta$ (u(C,C) forever – u (D,D) forever)

$c \leq \delta\,((1-c)/(1-\delta) - 0) \rightarrow \delta > c$

Pareto-optimal in Repeated FD

The Nash equilibrium s_{Blue} = TFT and s_{Green} = TFT is Pareto-optimal
(but s_{Blue} = AllD and s_{Green} = AllD is not) !

Sketch of Proof:
There is no way for a player to go above his normalized payoff of 1-c without hurting his opponent's payoff

Formal Definition!

EQUILIBRIUM IN INFINITE REPEATED GAME

Minmax Value

The minmax value is the lowest stage payoff that the opponents of player i can force him to obtain with punishments, provided that i plays the best response against them.

$$\underline{u_i} = \min_{s_{-i}} \left[\max_{s_i} u_i(s_i, s_{-i}) \right]$$

Enforceable Payoff Profile

A payoff profile $u = (u_1,u_2,\ldots,u_n)$ is **enforceable** if $u_i \geq \underline{u}_i$

Feasible Payoff Profile

In n-player game $G=(N,S,u)$, a payoff profile u is **feasible** if there exist *rational, non-negative values* α_j such that for all j, we can express u_i as $\Sigma_{j \in |S|} \alpha_j u_i(j)$, with $\Sigma_{j \in S} \alpha_j = 1$.

Example of Feasible Profile

Blue \ Green		
	(2,0)	(0,0)
	(0,0)	(0,2)

- (1,1) is a feasible payoff given that we assign 0.5 to the strategy profile over diagonal.
- (2,2) is not feasible

Theorem

Player i's normalized payoff is at least equal to Minmax value in any Nash equilibrium of the infinitely repeated game, regardless of the level of the discount factor

Intuition: a player playing All-D will obtain a (normalized) payoff of at least 0

Folk Theorem

(Infinitely Repeated Game with Average Rewards)

Consider any n-player game G and any payoff vector $(u_1, u_2, \ldots, u_n)$.

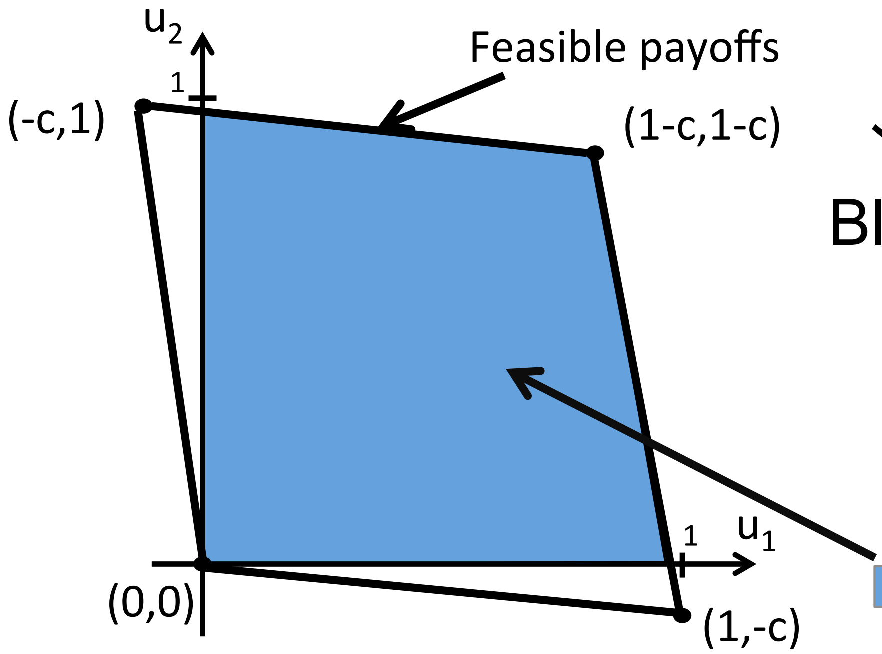
1. If u is the payoff in any Nash equilibrium of the infinitely repeated G with average rewards, then for each player i, u_i is enforceable.
2. If u is both feasible and enforceable, then u is the payoff in **some Nash equilibrium** of the infinitely repeated G with average rewards.

Folk Theorem

Infinitely Repeated Game with discounting Factor

For every feasible payoff vector
$u = \{u_i\}_i$ with $u_i > \underline{u_i}$ (i.e., it is enforceable as well),
there exists a discounting factor $\underline{\delta} < 1$,
such that for all $\underline{\delta} < \delta < 1$,
there is a Nash equilibrium with payoff u.

Feasible Payoffs in the Repeated Forwarder's Dilemma



Blue \ Green	C	D
C	(1-c,1-c)	(-c,1)
D	(1,-c)	(0,0)

Note that p_i can obtain at least his minmax value in any stage (*enforceable payoffs are always* higher than the minmax payoff)

Intuition and Example!

If the game is long enough, the gain obtained by a player by deviating once is outweighed by the loss in every subsequent period, when loss is due to the punishment (minmax) strategy of the other players.

Example: In infinite repeated FD, a player is deterred from deviating, because the short term gain obtained by the deviation (1 instead of 1-c) is outweighed by the risk of being minmaxed (for example using the Trigger strategy) by the other player (provided that $c < \delta$).