GAMES

444 Lecture 14

Brian Weatherson

2024-03-07

PRISONERS' DILEMMA

	C	**D**
C	3,3	0,5
D	5,0	1,1

Basic Challenge:

- Each player is better off defecting;
- The players are collectively better off if both cooperate.

TAKE TWO ON IN PERSON

- If it works - this is a 5-round Prisoners' Dilemma
- Go to https://veconlab.econ.virginia.edu
- Login as participant
- Session name: **pbw1**
- If this doesn't work, we'll go back to other things.

TRAGEDY OF THE COMMONS

- In a two-player setting, we normally call this Prisoners' Dilemma, or PD.
- In a multi-player setting it's sometimes called the Tragedy of the Commons.
- Though note this name traces back to Garret Hardin, who had some *problematic* associations.

TRAGEDY OF THE COMMONS

- The story (which is wildly ahistorical) is that everyone grazed their herds on the commons - which was a private good to get cheap food - but collectively this made the commons unusable.
- And in the standard story, private property was the solution to the tragedy.
- For the real story, see the work Eleanor Ostrom won the Nobel Prize for a few years back.

SOCIAL CHALLENGE

- In a PD, how do we get to cooperation?
- First question is whether in this case we should want to get to cooperation. (Compare price fixing.)
- Second question is whether we really are in a PD. (Compare walking on the right.)
- Let's assume that the answer in each case is *yes*, what do we do?

CHANGE THE PAYOUTS

One possible social response is to change the payouts.

- *Snitches get stiches* is kind of a version of this response.
- Hobbes's Leviathan who will kill you if you are anti-social is another.

CHANGE THE OPTIONS

Another is to make it just impossible for everyone to do the defecting move.

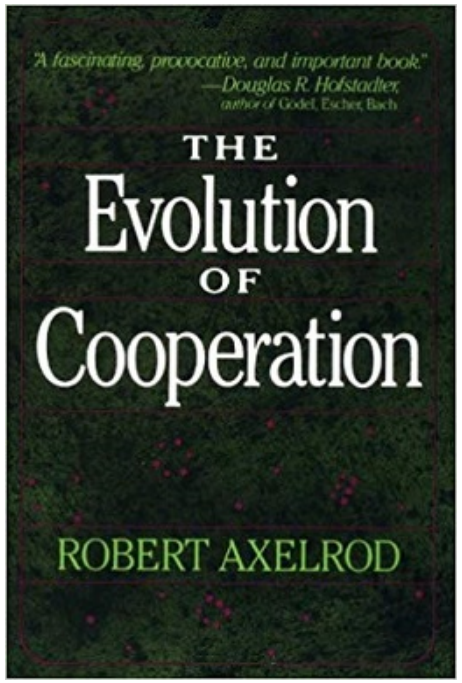
- Enclosures are sort of like this.
- The difference between making something expensive and making it impossible is a little vague, but it’s useful conceptually to think of them as separate options.

ITERATE THE GAME

- But the simplest way to handle this kind of problem is to iterate the game.
- Arguably it is in everyone's interests to be cooperative if they will have to interact with the other players repeatedly.

AXELROD

Robert Axelrod



Axelrod's Famous 1984 Book

FOUR PAPERS

- Effective Choice in the Prisoner's Dilemma, *Journal of Conflict Resolution* 24 (1980): 3-25.
- More Effective Choice in the Prisoner's Dilemma, *Journal of Conflict Resolution* 24 (1980): 379-403.
- The Emergence of Cooperation among Egoists, *The American Political Science Review* 75 (1981): 306-318.
- The Evolution of Cooperation with William Hamilton, *Science* 211 (1981): 1390-1396.

THE ONE SHOT GAME

Axelrod worked with this version of Prisoners' Dilemma (PD) (which you've seen before).

	C	**D**
C	3,3	0,5
D	5,0	1,1

The trick was that each pair of people would play an iterated version of this game.

INDEFINITE ITERATION

In the fancier version of the game, he didn't tell people how long the game would go.

- Instead he just said there was a probability of it ending after each round; if I recall 0.005.
- This was used to avoid backwards induction reasoning, but it was unnecessary.

BACKWARDS INDUCTION

Some of you might have come across this problem in the 5 round game.

- A big benefit of cooperating is triggering future cooperation.
- But in round 5 there's no future, so why cooperate then?
- If everyone knows that reasoning works, why cooperate in round 4? Etc back to never cooperating.
- This reasoning doesn't work in practice, and it's a fascinating question whether it works in theory.

THE TOURNAMENT

- There are n strategies submitted.
- The strategies are computer programs, which say what to do at each stage given the history of the game.
- Each will play k rounds of PD with each of the other n-1 strategies.
- Their payouts will add up over the $k(n-1)$ rounds and the one with the highest total will win.

COOPERATIVE AND COMPETITIVE

- This is not entirely a cooperative game; ultimately if I'm a strategy I want to win, and that means I want the other strategy I'm interaction with to lose.
- But in the short run there is much to be gained by improving our mutual position vs the other $n - 2$ strategies.
- So in the short run there is a benefit to cooperation, even if we're ultimately rivals.

THE FIRST TOURNAMENT

- Axelrod advertised the first round of his tournament, and called for submissions.
- This was far from trivial in pre-internet days, and he only got 13 submissions.
- In the first tournament he said that $\mathbf{k}$ would be 100, but no one actually exploited that fact.

QUESTION

What kind of strategy would you endorse?

Would the programming language make a difference?

THE WINNER

Tit-for-Tat

TIT-FOR-TAT

Two rules.

1. Play C at round 1.
2. In all subsequent rounds, do whatever the other player just did.

THE SECOND TOURNAMENT

- So Axelrod wrote this up, including saying who won.
- He called for more submissions, and now got 66.
- Some of these were typed, some came to Ann Arbor on the huge magnetic disks that were used way back then.
- He ran the tournament again, this time with a random number of rounds.
- And Tit-for-Tat won again.

LOGIC AND VICTORY

- This doesn’t mean Tit-for-Tat is the best strategy.
- Indeed, in each tournament it was easy in retrospect to describe strategies that would have beaten everyone, including TFT, if they had been entered.
- But still, it’s pretty impressive.

FOUR FEATURES

Tit-for-Tat has five striking characteristics, each of which was positively correlated with success in the tournaments.

- Nice
- Provocable
- Forgiving
- Not envious
- Simple

NICE

The clearest distinction in the tournament was between strategies that were Nice and those that were Nasty.

- By definition, a strategy is Nice iff it is never the first to defect.
- You don’t have to be very nice in the intuitive sense to count as Nice.

GRIM TRIGGER

Here is one nice strategy, one Axelrod calls Grim Trigger.

1. Cooperate on move 1.
2. If the other player ever defects, defect on every subsequent move.

This strategy did really badly; it was the worst Nice strategy in round 2. But still many Nasty strategies did worse.

NICE STRATEGIES

- In the evolutionary versions of the game, there can be a tendency for strategies to tend towards being Nice.
- Then evolution stops, because when two Nice strategies meet, the payout is inevitably 3k to each.
- Although the best strategies are all Nice, it is how they interact with Nasty strategies that determines who wins.

PROVOCABLE

- It’s bad to get pushed around.
- Nasty strategies are always looking for how much they can get away with.
- So you want to send a clear message that defections will not be tolerated.
- Obviously TFT does that.

FORGIVING

- But you don’t want to be Grim Trigger.
- It’s bad to be pushed around, but it’s not much better to end up in all defect land.
- You need a way back to all cooperate land.
- TFT has that, though notably it isn’t perfect at this.
- TFT can get into CD-DC-CD-etc cycles with a bunch of strategies.

NOT ENVIOUS

- In any interaction, TFT never does better than who it is playing with.
- Yet it comes out first overall.
- This is kind of amazing.
- It just does not care at all about winning against who it is facing off with.

NOT ENVIOUS TO A FAULT

- Note that TFT doesn't always do that well in **evolutionary games**. (If we get time we'll come back to what I mean by this.)
- This is because it might take this a bit too far.
- It doesn't look to exploit weaknesses in opponents.

SIMPLE

- Other strategies try to figure out what their rivals are doing.
- They normally get this wrong.
- Or they try and send complex signals.
- These are usually misinterpreted.
- TFT keeps things simple, and doesn't lose points messing around looking for any edges.

VARIANT GAMES

- The most interesting variant to me is the one where a strategy only gets implemented with probability 0.99 on each move.
- Sometimes there are performance errors.
- TFT does terribly in this; it can’t get out of randomly generated defection cycles.

VARIANT GAMES

- In this kind of game you need to be a bit more forgiving.
- But also you can try to get away with a bit more; if the other person will treat a defection as random, you can plan a few.

WHAT DIFFERENCE DOES IT MAKE

- Same game as before - a 5-round Prisoners' Dilemma
- Go to https://veconlab.econ.virginia.edu
- Login as participant
- Session name: **pbw2**
- If this doesn't work, we'll go back to other things.

THREE KINDS OF GAME THEORY

1. Rational Choice
2. Evolutionary
3. Experimental

RATIONAL CHOICE

Assume common knowledge of rationality, see what you can figure out about how the game will go.

This is **primarily** what Bonanno's book is about.

There is no standard name for this, it's often just called 'game theory'. Like with Apple product naming, it's very unhelpful to have one variant have a null modifier, so I use expressions like 'rational choice game theory'.

EVOLUTIONARY

Assume strategies are hard-wired and interactions are random.

Also assume that populations of each kind (i.e., each strategy) in generation (n+1) are proportional to

1. Population of that kind in generation n;
2. Average score of that kind in generation n.

In this game, you might have 100 generations where each generation is 100 rounds of PD.

EXPERIMENTAL

Put people (and as always by people here we usually mean undergrads at fancy universities) in labs and see what happens.

PD is interesting to all three kinds.

FOR NEXT TIME

Next week we'll go over two central notions of rational choice game theory:

- Iterated deletion
- Backwards induction