

Game Theory and Society

Blithering Genius

July 29, 2014

Contents

1	Game Theory	1
2	The Prisoner's Dilemma	1
3	The Iterated Prisoner's Dilemma	2
4	The Tragedy of the Commons	3
5	Using Coercion to Solve Problems of Cooperation	4
6	Conclusion	5

1 Game Theory

Game theory is a branch of mathematics that deals with decision making when there are multiple decision-makers, called “players”, and outcomes depend on the decisions of other players. Game theory is based on the metaphor of a game, in which the rules of play and the outcomes are well defined. The game metaphor is a very useful abstraction. Game theory is used to understand behavioral strategies in economics, evolutionary theory, politics, warfare and other domains. It is especially important for understanding how society works.

2 The Prisoner's Dilemma

The prisoner's dilemma is an important concept in game theory. It illustrates the benefits and challenges of creating cooperation between selfish individuals. The prisoner's dilemma is named after a thought experiment in which the police try to get two prisoners to confess and implicate each other in the crime. Unfortunately, that thought experiment is somewhat complicated and can be confusing. Rather than telling the story of the two prisoners, I will explain the prisoner's dilemma in abstract terms.

Here is a typical prisoner's dilemma. There are two players, A and B. Each has two choices: cooperate or defect. The scores are numbers that could represent dollars, megajoules, offspring or any other measure of value.

A's Move	B's Move	A's Score	B's Score
Cooperate	Cooperate	1	1
Cooperate	Defect	2	2
Defect	Cooperate	2	2
Defect	Defect	1	1

Let's consider the decision problem from A's perspective. If B cooperates, then A gets 1 for cooperating and 2 for defecting. If B defects, then A gets 2 for cooperating and 1 for defecting. In both cases, A improves his score by 1 if he defects. So, A should defect.

The same is true for B. Each player is better off defecting, regardless of what the other player does. So, both players will defect, and each will get 1.

If both players cooperated, they would each get 1. Both players would prefer that outcome to the one in which both defect and get 1. But each has an individual incentive to defect. So, unless there is some way to coordinate their actions, they will both defect and get 1.

The prisoner's dilemma illustrates the problem of creating cooperation between selfish individuals.

Here is an example of a prisoner's dilemma situation. Suppose that you are on a backpacking trip in a foreign country. Your backpack is heavy, and you would like to leave it while you walk around the city. You see a man sitting on the sidewalk, begging for spare change. You consider paying him to watch your backpack for an hour. The exchange would benefit both of you, but there is a problem. If you pay him up front, he might just take the money and run. If you offer to pay him afterward, he can't be sure that you will pay him when you get back. The risk of defection prevents cooperation.

Here is another example. This time it is between two societies, rather than two individuals. Consider two neighboring countries, A and B. Each has the option of building an army that could be used either to invade its neighbor or to defend itself from invasion. There is a cost to building and maintaining an army, but there is a greater benefit to invading an undefended country. If B does not build an army, then A can benefit by building an army and invading B. Likewise, if A does not build an army, then B can benefit by building an army and invading A. If either country builds an army, then the other country will need an army to defend itself from invasion. Thus, both countries must invest in building and maintaining armies just to maintain the status quo. Peaceful coexistence without armies would be better for both countries, but each is forced to build an army, because the other also has that option.

Cooperation between selfish players requires some coordination mechanism to prevent defection.

3 The Iterated Prisoner's Dilemma

Cooperation can emerge between selfish players if they interact regularly. The incentive to cooperate is the expectation of future reciprocity. Players will cooperate with each other to create an ongoing cooperative relationship. Once bootstrapped, a virtuous cycle can emerge. This is called "tit for tat".

The iterated prisoner's dilemma is simply a prisoner's dilemma that is repeated over time, between the same players. In addition to the immediate payoff, players also receive the expectation of future cooperation. $F(A)$ and $F(B)$ represent the long-term benefits of friendship with A and B, respectively.

A's Move	B's Move	A's Score	B's Score
Cooperate	Cooperate	$1 + F(B)$	$1 + F(A)$
Cooperate	Defect	2	2
Defect	Cooperate	2	2
Defect	Defect	1	1

Tit-for-tat only works when there isn't a huge benefit to defecting on the first move. A country can't disarm and then wait to see if its neighbor disarms too. A tit-for-tat relationship usually begins with exchanges of small favors. As trust develops, the exchanges increase in size. Although the decisions are simultaneous in the standard prisoner's dilemma game, tit-for-tat often involves exchanging favors at different points in time. Your friend helps you out today, and you help him out tomorrow. The basic principle — reciprocity — is the same.

Tit-for-tat doesn't work between strangers, or when interactions are anonymous. It depends on the players knowing each other and keeping track of past interactions. You can only have tit-for-tat relationships with people that you know personally and interact with on a regular basis.

People who associate tend to cooperate, and people who cooperate tend to associate. Thus, tit-for-tat causes people to cluster into small groups that are inwardly cooperative and outwardly competitive. But it doesn't scale up to large groups. It is limited to the number of people that you can frequently interact with and keep track of mentally, which is about 100. Large groups require some other mechanism of social cohesion.

4 The Tragedy of the Commons

The tragedy of the commons is similar to the prisoner's dilemma, but it involves the relationship between an individual and a collective, rather than two individuals.

The table below shows the costs and benefits of an action to both the individual and the collective. In this example, the individual receives a benefit of 1 for doing some action, and the collective benefit is also 1. That means the individual receives all the benefit of his action, because the collective benefit equals the individual benefit, and he is part of the collective. The collective cost of his action is 2, but he only pays a fraction of that cost, $2/N$, where N is the number of people in the collective.

	Individual	Collective
Benefit	1	1
Cost	$2/N$	2

The total cost exceeds the total benefit, but if $N > 2$, each individual has an incentive to do the action, because his individual benefit is greater than his individual cost. The result is that each individual will do the action, and the net payoff to each individual will be 1. Everyone is harmed by everyone doing the action.

Imagine a crowded room in which everyone speaks loudly to be heard over the noise of everyone else speaking loudly. Each individual would prefer everyone to speak quietly, but each individual has an incentive to speak loudly. So, they all speak loudly. That is a simple example of a tragedy of the commons.

Here is another example. Suppose that a group of people lives next to a lake that could provide them with nice, clean drinking water. The lake could also be used to clean their dishes and wash

their clothes. For an individual, washing his own clothes and dishes in the lake has little impact on him. He could wash his stuff in one place, and then walk a few feet away and dip his pot in the lake to get fresh water. But if everyone uses the lake to wash their stuff, then the water will be undrinkable. It would be better for everyone to only use the lake for drinking water. However, each individual has an incentive to pollute the lake.

Scale matters in this example, as it does with tit-for-tat. Suppose that the lake is small, and the number of people is also small. When N is small, each individual pays a higher fraction of the cost himself, and he also has an incentive to police the actions of others, because their individual actions have a greater impact on the cost he pays. When the group is small, the individual's actions will be remembered by others, and he will want a reputation as a cooperator. These things can make it better to cooperate than defect. But as the number of people increases, the incremental effect of an individual's actions on any other individual, including himself, becomes negligible. There will be no incentive to curb his actions or police the actions of others.

The free rider problem is similar to the tragedy of the commons, but it involves the absence of an incentive to be productive, rather than the absence of an incentive to not be destructive. For example, the welfare state allows individuals to benefit at the expense of others, without making a positive contribution to society. Another example is when a society allows people to opt out of military service, but receive the protection of the military. Free riders get a "free ride" at the expense of others.

We are used to thinking of tyranny as something imposed on individuals by society, but tyranny can also be imposed on society by individuals.

There are two types of collectivism: top-down and bottom-up. In top-down collectivism, the benefits of individual action are collectivized, while the costs are paid by individuals. Communism is an example of top-down collectivism. In bottom-up collectivism, the costs of individual action are collectivized, while the benefits are received by individuals. The tragedy of the commons is an example of bottom-up collectivism. In top-down collectivism, there is no incentive to be productive. In bottom-up collectivism, there is no incentive to not be destructive. For society to work, there must be individual incentives to be productive and not destructive.

5 Using Coercion to Solve Problems of Cooperation

The prisoner's dilemma and the tragedy of the commons are problems of cooperation. To create cooperation, we must solve problems of cooperation.

In small groups, tit-for-tat solves problems of cooperation. To create cooperation on a large scale, it must be imposed on individuals by coercion. The state is a mechanism for representing the collective interests of the members of society. It makes cooperation possible between strangers by punishing defection.

The following table shows how the expectation of punishment solves the prisoner's dilemma:

A's Move	B's Move	A's Score	B's Score
Cooperate	Cooperate	1	1
Cooperate	Defect	2	2 E(P)
Defect	Cooperate	2 E(P)	2
Defect	Defect	1 E(P)	1 E(P)

$E(P)$ represents the expected penalty that will be imposed on a player for defecting. It is an expectation, because the player might not get caught and punished. Suppose that $E(P) = 2$. In that case, there is an incentive to cooperate, even in a single game with no potential for tit-for-tat.

To be effective, the punishment must be significantly worse than the potential benefit of the crime, because there is a chance of not getting caught, and so the expected penalty is lower than the actual penalty if caught. For example, the punishment for theft is usually more than simply replacing the stolen item.

The tragedy of the commons is solved by the state in a similar way:

	Individual	Collective
Benefit	1	1
Cost	$2/N + E(P)$	$2 + E(P)$

The state uses punishment to create incentives that align the interests of the individual with the interests of the collective. If $E(P) > 1$, the individual has no incentive to commit the socially harmful action.

6 Conclusion

The necessity of the state raises two important questions:

- How is state power created?
- What prevents it from being abused?

I'll leave those questions for another essay. They are too big to address here.

Given that human nature is selfish, modern society is almost a miracle. I can go to the store and buy groceries from strangers. When I walk home on crowded streets, no one hits me over the head and steals my groceries. We take these things for granted, partly because we view them as how things should be. But it is no small task to explain how they are possible.

Society is based on cooperation between selfish individuals, not altruism. Game theory formalizes problems of cooperation, and explains how we can solve them. Understanding game theory is necessary for understanding society.