

The Prisoner's Dilemma and the Logic of Collective Action

Class notes by Gregory Gleason (2002)

One way to abstractly imagine the interaction among countries is to think of it in terms of game theory. Game theory is a subfield of mathematics and logic that makes simple assumptions about the logical structure of a situation and then considers the distribution of incentives within the situation. Real world situations are usually more complex, involving players who have myriad motivations, high levels of uncertainty, poor judgment or information, and in general may not be able to calculate as well as game theoreticians in a surgical environment. Nevertheless, sometimes the metaphor of games can be useful by revealing the skeletal structure of a situation.

One simple class of games includes so-called *zero-sum games* such as checkers or chess. In these games there are only three outcomes, win, lose or draw. In each case the "utility" (or value) attached to the outcomes is zero. One side wins (+1), the other side loses (-1), so the outcome of utilities is zero. A more complex set of situations is called "mixed motive games." These are ones in which the outcome depends on what one player calculates that the other player will do. The prisoner's dilemma, a heuristic device invented by the mathematician Albert Tucker, is just such a mixed motive game.

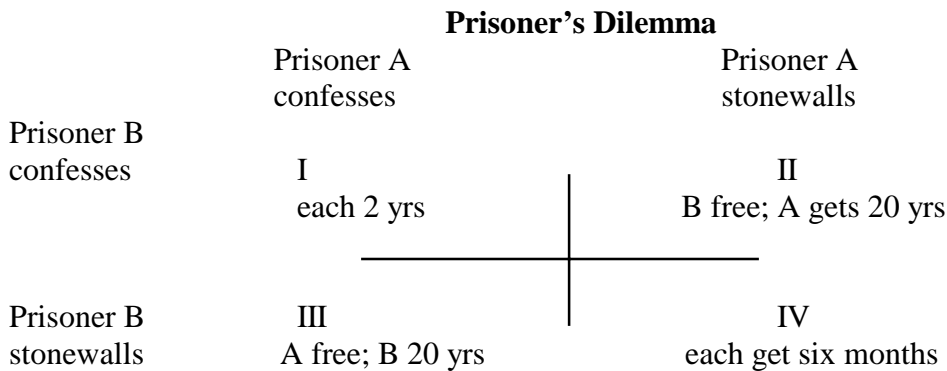
Imagine a situation in which there are two prisoners who have been put in jail for suspicion of armed robbery. The two suspects were apprehended with weapons, so the prosecutor has enough evidence to obtain a conviction on a charge of carrying weapons. But he does not have enough evidence to obtain a conviction on the more serious charge of armed robbery without an admission of guilt from at least one of the prisoners. So the prosecutor devises a strategy to induce at least one of the suspects to confess to the crime.

The prosecutor calls prisoner # 1 into a private consultation and offers him a reduced sentence if he confesses to the crime. He says, "I need a confession and I want you to confess. To make it to your advantage, I can offer you a reduced or even a suspended sentence if you confess." Prisoner #1 thinks for a moment and then asks what his partner in crime, prisoner #2, intends to do. The prosecutor responds, "I don't think prisoner # 2 will confess. But I need a confession. If you confess and he doesn't, I can offer you a suspended sentence. You won't serve any time at all." Prisoner # 1 thinks this over and asks, "but what if #2 confesses?" "Then we will have two admissions and proof that you are guilty so you will have to serve time, but since you are being cooperative, you will both get reduced sentences of only get two years a piece," answers the prosecutor. "And what if I say nothing?" prisoner 2 questions further. "Then you will both be charged with a weapons violation and you will probably get only six months," answers the prosecutor. "But, remember, if #2 confesses and you do not," the prosecutor then warns him, "you will get 20 years and he will get a suspended sentence." After this meeting, the prosecutor independently meets with prisoner # 2 and offers him the same options.

What is the "rational" thing for the prisoners to do?

I	#A confesses	#B confesses	both get 2 years (total= 4 years)
II	#A stonewalls	#B confesses	# 1 serves 20 years, # 2 goes free (total=20 years)
III	#A confesses	#B stonewalls	# 2 serves 20 years, # 1 goes free (total=20 years)
IV	#A stonewalls	#B is stonewalls	both get 6 months (total =1 year)

In terms of a diagram of the possible outcomes of the situation, the prisoner's dilemma looks like this:



What is the rational action for each prisoner? Let us set aside the morality of the situation for the moment (many of us would agree that the “right” thing to do would be to confess to a crime that one committed) and just consider the best course of action from the perspective of the prisoners who would like to get out of confinement as quickly as possible.

The *optimal* outcome in this situation is IV because the total amount of time served is only one year compared with the other two outcomes (four years and twenty years). Yet many people in this situation would be moved to confess, on the assumption that remaining silent could result in a twenty year sentence. But if rational actors select this option, it is not the optimal one but one less than optimal, that is *suboptimal*. Trust and cooperation fall by the wayside and the prisoners are trapped in a suboptimal outcome. This dilemma illustrates how, even when rational actors can appreciate the importance of cooperation, they find themselves trapped in situations where they cannot get to their desired cooperative goal.

There are many real world situations that seem to have the form of the prisoner's dilemma. Consider the nuclear arms rivalry between India and Pakistan. India and Pakistan have a historical animosity over borders. They find themselves in the grip of an arms race. Consider the logic of the situation from the point of view of the decision to “go nuclear” in the arms race. What are the preferences of the countries? In other words, how do they evaluate the possible outcomes? The best outcome would be that if one but not the other had nuclear weapons (the expense of building the weapons would

be offset by the leverage they provided). The second best outcome would be for neither to go nuclear. In this case there would be no leverage but also no expense. The third best would be for both to develop nuclear weapons (no leverage and expense). The fourth would be for one to forgo and the other to have nuclear weapons and be subject to intimidation.

In innumerable bilateral situations in international affairs, payoff matrices often resemble those of the prisoner's dilemma. This illustration may suggest why the "logic of the situation" so often leads the parties involved to suboptimal outcomes. Why do countries go to war when they realize there is so much to lose and so little to gain? This illustration would suggest they sometimes do so because it is a suboptimal outcome that they feel imposed upon them by the logic of the situation.