THE PRISONER'S DILEMMA

Game theory can be applied in situations in which one player's gain is not necessarily the other player's loss. These games are called "partial conflict games." One problem of this type is called "The Prisoner's Dilemma" and the problem is stated below.

A district attorney suspects two persons, recently taken into custody, of together having committed a serious crime. Although the suspects are known "almost certainly" to be guilty of the crime, the district attorney does not have sufficient evidence to convict either one.

To each prisoner, confined alone in a cell and unable to communicate with the other, the district attorney makes the following offer:

"Tommorow I will visit you in your cell and give you the opportunity to admit your guilt. If you confess and the other prisoner does also, you will each receive a 7 year jail term. If you confess and the other prisoner fails to confess, you will go free in exchange for the evidence you have provided, and the other prisoner will face a 10 year jail term. If neither of you confesses, each of you faces a 3 year jail term as a result of a list of minor charges for which we already have evidence. To whichever prisoner I visit second, I will reveal nothing about the results of my first visit."

Since the game is not a zero-sum game, the payoff matrix must contain information about the payoffs to both of the player. To accomplish this, each entry in the payoff matrix is a pair, in which the first entry is the payoff to Player A (whose strategies are listed down the left side of the matrix) and the second entry is the payoff to Player B (whose strategies are listed across the top.) The payoff matrix for this problem is written below. The number of years in jail is given as a negative so that the higher the payoff value, the more a prisoner desires it.

		Prisoner B	
		Keep quiet (C)	Rat (D)
Prisoner A	Keep quiet (C)	(-3,-3)	(-10,0)
	Rat (D)	(0.-10)	(-7,-7)

The prisoner's dilemma is more than just an interesting hypothetical situation. It is also a model for a general type of decision situation. The characteristics of such a decision problem, which we will henceforth call a "PD problem," are as follows:

1. There are two decision makers, each with the same pair of choices, designated by C (for cooperation with the other player) and D (for defecting from an agreement with the other player.)

2. Each player, acting alone, gains more from choice D no matter what the other player chooses.

3. Each player gains more if the other player chooses C no matter what choice he himself makes.

4. Both players gain more if they both choose C than if they both choose D.

When the consequences of a PD problem are expressed in terms of numerical payoffs and displayed in a matrix, the following terminology is useful:

r is the *reward* payoff to each player when both choose to cooperate.

p is the *penalty* payoff to each player when both choose to defect.

t is the payoff that creates a *temptation* for one player to choose to defect if the other cooperates.

s is the payoff to the *sucker* who chooses to cooperate when the other defects.

In order for the 4 characteristics of a PD problem to be satisfied, the values above must satisfy the set of inequalities $s < p < r < t$.

The prisoner's dilemma problem seems to have no good solution. Each player is tempted to defect in order to obtain a better solution for himself, but when both players do this, they obtain an outcome that is worse instead of better. The solution to the problem must involve a third party with the ability to change the values in the payoff matrix, so that defection is no longer the dominant strategy. In particular, t and p must be decreased, or s and r must be increased, or both.

One way to do this is to introduce a value v, called the reward/penalty value which is to be added to any payoff resulting from cooperation and subtracted from any payoff resulting from defection. The minimum reward/penalty value which will ensure cooperation is calculated as follows:

1. Calculate $\frac{1}{2}(t-r)$ and $\frac{1}{2}(p-s)$.
2. Choose the larger of these two numbers.