Prisoner's Dilemma: A Case Study

Ioanis Gaviotis, igaviotis@gmail.com

December 6, 2015

Abstract

When one starts to study Game Theory, a game called Prisoner's Dilemma is one of the first examples to be introduced.

This report was written just after I completed Stanford's on-line course on Game Theory. Its purpose is to investigate whatever aspects I found to be related to the Prisoner's Dilemma (PsD).

1 What is it?

The Prisoner's Dilemma involves two guys having been arrested by the police. Since there is not enough evidence to convict them, the police interrogator offers the following deal to each of them:

> If you testify that the other guy is guilty (defect), you will be set free and the other will be imprisoned for 3 years; unless he also defects, in which case you will be both imprisoned for 2 years. If you both keep your mouths shut (cooperate), each of you will serve 1 year in prison.

Every prisoner knows that the same offer that was made to him, has been made to his fellow convict. Moreover, each one knows that the other knowns that he knows, and so forth. This assumption about the game is called **common knowledge** and is a necessary prerequisite to have a **game of complete information**. Without common knowledge, the behavior of the prisoners could be different.

The situation is formalized as a **strategic game** $G = (P, A, u)$, where the two players in $P = \{p_1, p_2\}$ can play actions from $A = \{C, D\}$, where C stands for Cooperate with the other player (don't speak to the police) and D stands for Defect (make a deal with the police to testify against the other prisoner). After playing the players will get payoffs denoted by the utility function $u = (u_1, u_2)$, where $u_i : A \mapsto \Re$ is depicted in Table 1. Since players are awarded years in prison (they get punishment rather than reward), the payoffs take a minus sign, to denote that more years are worse (non-preferable) to less years.

		p_2 plays	
	u_1, u_2	C	D
p_1	C	-1,-1	-3, 0
plays	D	0,-3	**-2,-2**

Table 1: Years in prison for PsD

In case one is intimidated by the negative values for the utility function appearing in Table 1, one can add a constant number to all values and the essence of the game would still remain unaltered. For example, in Table 2 all payoffs are shifted by +3.

u_1, u_2	C	D
C	2,2	0,3
D	3,0	**1,1**

Table 2: Shifted values for the utility function

Actually, any **positive affine transformation** of the payoffs would not change the character of the game. So, changing all payoffs from x to $ax + b$, for any fixed real numbers $a > 0$ and b, would not make any difference for the game itself.

2 Domination

A player is always better off defecting, rather than collaborating, no matter what the other player does. For example, in Table 2, by playing D p_1 would gain:

- 3 rather than 2, had p_2 played C;
- 1 rather than 0, had p_2 played D.

Action profile (D,D) is called a **strictly dominant pure strategy**, as it is the only choice surviving **iterated removal of strictly dominated strategies**. In this respect, PsD is an easy to analyze game, because not only a preferred choice for players exists, but also it is unique. Contrast this to the Battle of Sexes game in Table 3, where although a couple detests going out on one's own, he would rather go to cinema and she to the ballet. In this game players don't have a single best response (pure) strategy, yet there is a mixed strategy, namely going together half of the times to the cinema and half to the ballet. Nash equilibria are marked as bold entries in the normal form representations.

↓He She→	Cinema	Ballet
Cinema	**4,2**	1,1
Ballet	1,1	**2,4**

Table 3: Battle of Sexes

Returning to PsD, its unique dominant strategy is necessarily its unique **pure strategy Nash equilibrium**. This fact saves us from the trouble of searching for mixed strategy Nash equilibria, which requires exponential time :-). Well, if PsD is so straightforward to analyze, why is it the most popular game to study?

2.1 Why dilemma?

What's really interesting about the Prisoner's Dilemma[1] is that, although it is common sense that collaboration would bring the best overall outcome for the two prisoners, i.e. a total of 2 years in prison by playing (C,C) in Table 1, our analysis contends that the best choice is to defect, which would cause the gang to spend at least 3 and most probably 4 years in jail.

Intuitively, this repels us from the strictly dominant and Nash equilibrium solution of all defecting and it is formalized with **Pareto dominance**. For all players, action profile (D,D) is worse than (C,C), because $\forall i, u_i(D,D) < u_i(C,C)$; in other words, all defecting is Pareto dominated by all collaborating. Unfortunately, there are three Pareto optimal solutions, namely (C,C), (C,D), (D,C), and the only solution that is *not* Pareto optimal is the overall best response (Nash equilibrium). Hence the singularity that incites the dilemma to the prisoners.

2.2 Omertà

Defecting may collect a host of credentials, still from our experience we know that collaboration arises in many occasions due to ethical considerations, such as detesting to cause harm to your partner, or simply possible risks of reprisal. These factors can be incorporated in the values of the payoff function, so that it accurately reflects the overall consequences for the players.

For example, Mafia members abide to omertà, the code of silence. Breaking it could incur death, rendering collaboration as the only viable strategy. One concludes that omertà, not only eases the dilemma out of the prisoners, but also achieves the optimal solution for them as a group!

2.3 Security level

To change perspective, instead of opting for the best solution for maximizing payoff, prisoners wish to play it safe. Prisoner p_1 worries about the worst case if he chose to collaborate, call it **danger** $d_1(C) = \{D\}$ and its value is $u_1(C,D) = 0$. If p_1 defected, the danger is $d_1(D) = \{D\}$ and its value is $u_1(D,D) = 1$. Of those two dangers, clearly preferable is the latter, which would push the security-conscious player to defect. The other prisoner would choose analogously. Solution (D,D) is called the **maxmin strategy** and it is the preferable solution profile if security is top priority.

Defecting is the single maxmin strategy for PsD. Sticking to it ensures your safety, especially if you are unsure about the sanity of your opponents. It also screens you from their viciousness, if the job turned sour and they have hard feelings for you.

Alternatively, the **minmax strategy** prioritizes on keeping opponents' utility as low as possible, reflecting jealousy as dominant driver of behavior. For two-player games, such as PsD, the minmax and maxmin strategies coincide.

[1] More precisely, it should be named Prisoners' Dilemma.

3 Generalization of utility values

What can the payoff values of a two-player game be, so that it can be characterized as Prisoner's Dilemma? We already talked about applying any positive affine transformation, but we can go even further in generality.

First off, the police does not have any bias against any of the two prisoners—it is indifferent as to whom it will prosecute, therefore for all players i, it holds $u_i(x,y) = u_{-i}(y,x), \forall x,y \in A$. PsD is a **symmetrical game** where the identity of the player does not change the resulting game facing that player.

It is logical that the police puts its harshest punishment on the guy that has been nailed as criminal and he failed to incriminate the other. Therefore, he will get the lowest payoff, let's assume 0 without loss of generality, for the rest of the payoffs to be positive values. So, the 'victim' gets $u_1(C,D) = u_2(D,C) = 0$. At the other extreme, the police must offer high incentive to the snitch, let's call that payoff value the *bait* b, so $b = u_1(D,C) = u_2(C,D) > 0$. When both players defect, the common payoff for the *talkative* case is t. When players collaborate, the common payoff for the *reticent* case is r.

In order to enforce the strict dominance of (D,D), the inequality $b > r$ must hold. Furthermore, to effect the dilemma for the prisoners, being talkative must be worse than being reticent, therefore $r > t$.

u_1, u_2	C	D
C	r, r	$0, b$
D	$b, 0$	**t,t**

Table 4: Generalized values for any PsD game

Summarizing, a game with the payoff values of Table 4 iss PsD for any real values such that $0 < t < r < b$.

3.1 Price of Anarchy

The Price of Anarchy (PoA) is a measure of how much better would an optimal gameplay be, compared to the worse Nash equilibrium. More to the point, it shows whether players are collectively worse off by pursuing their personal interests, rather than having opted for their common welfare. Taking the utalitarian view, the **welfare** of a game play is defined as the sum of the utilities the players would gain, i.e. $W(s) = \sum_{i \in P} u_i(s)$[2]. PoA is then defined ([Mal11]) as

$$PoA(G) = \frac{\max_{s \in S} W(s)}{\min_{s \in N} W(s)}$$

where S is the set of all strategies and N is the set of Nash equilibria. Clearly, $PoA \geq 1$. When $PoA = 1$, as in the Battle of Sexes (Table 3), the game is 'benign' ([Rou15]), in the sense that personal and common interests do not collide. Let's focus on PsD now, in its generic form.

[2] Focusing on the least privileged player, the egalitarian view defines $W(s) = \min_{i \in P} u_i(s)$.

W	C	D
C	$2r$	b
D	b	$2t$

Table 5: Welfare for PsD

Substituting the values for PsD, where the denominator for the single Nash equilibrium is $W(D, D)$ and taking into account the necessary inequality between parameters

$$PoA(PsD) = \frac{\max\{2r, b, 2t\}}{2t} = \max\{\frac{r}{t}, \frac{b}{2t}\} > 1$$

The value of PoA always being greater than 1 numerically depicts the prisoners' dilemma. Moreover, since the bait b may become as big as the police wants, PoA is unbounded. Other than the game being intristically devious, a large value of PoA may be an indication that Nash equilibria, albeit rational, may after all *not* be the desired strategy of game playing ([Rou07]).

4 Extensive game

Until now we assume that both players make their choice, without any of them knowing the choice of the other. When their choices are done sequentially, we are talking about an **extensive form game**, which is modeled with the decision tree of Figure 1. In this variant of the game without loss of generality, p_1 chooses first his action, p_2 is informed of p_1's action and then plays accordingly. Payoffs appear as leaves of the tree.

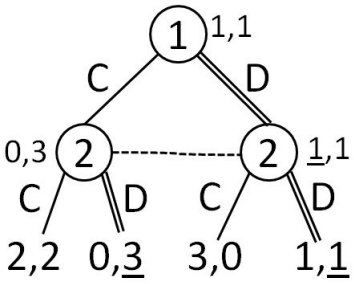

Figure 1: Extensive form of PsD

Applying the **backward induction** algorithm on the extensive form tree, the **subgame perfect equilibrium** is (D,D) denoted by the double line from the root to the leaves of the tree. Indeed, as p_1 is the first player to play, he would refrain from playing C, because it is obvious that p_2 would respond with D to get 3 utility points. Therefore, p_1 will choose D to ensure that –no matter what p_2 plays– he will get at least 1 point. When p_2's turn to play comes, he will prefer D which awards him 1 point.

Concluding the discussion about the sequential PsD game, nothing changes from the concurrent game: (D,D) is the single strategy to follow.

5 Finitely repeated game

Let's study what happens when the PsD game is repeated for a finite number of times, say t. In each repetition of the game the prisoners' moves happen concurrently, so neither of them knows what his fellow will choose. Still, they both know all players' choices made during all previous games. When they need to make a choice, they may reminisce all game history and play accordingly. To discover the strategy for the best outcome, we build the full binary decision tree of Figure 2, which includes all 2^{2t} possible strings of length $2t$ containing {C, D}. Using backward induction we locate the subgame perfect equilibrium consisting consistently of D's and offering utility t.

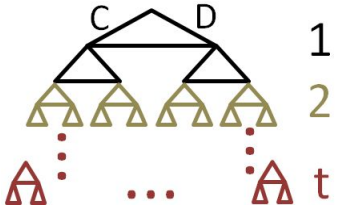

Figure 2: Repeated PsD

In reality, however, repeated offenders do not know a priori when their career will conclude. Being unregenerate criminals, they will keep breaking the law, to be arrested (hopefully) some of those times. In situations where the base game is repeated indefinitely, players can not discern when the game's end is coming. This is effectively modeled as a game with infinite repetitions. The corresponding decision tree is infinitely deep and no leaf is reachable to initiate backward induction.

6 Infinitely repeated game

For infinitely repeated PsD future payoffs get lower each time by a **discount factor** $d \in [0, 1)$, reflecting the fact that immediate gains are relatively more important than distant ones—after all the game may end prematurely, if the sheriff shoots the criminal. Furthermore, prospective prisoners tend to value youth years out of jail more than elderly years. If for no other reason, a youngster will surely serve all his imprisonment years, whereas an senior convict may get a parole or perish...

We explained previously the inequality that characterizes the utility values of PsD. For the repeated version of the game, however, we need to take some extra caution. Some canny prisoners may decide to alternate collaboration and defection splitting the bait in the long run ([Ken07]). To avoid this and keep the character of the game, therefore, we additionally insist that $r > b/2$h[3].

In order to study infinitely repeated PsD, we need to study specific strategies. A strategy prescribes the next action for each player taking into consideration the game's history.

6.1 Vendetta

In the absence of any history, the player initially cooperates for his first move as a gesture of goodwill. From then on, as long as the opponent cooperates, the player continues to cooperate. Should the opponent choose to defect,

[3] An implication of this extra condition is that the price of anarchy cannot get indefinitely large; it can only reach r/t, which is still > 1.

would infuriate the player and trigger reprisal. In the next and all subsequent moves the player plays defection forever. This is called **grim-trigger** strategy.

Let's consider the utility values for the generalized PsD of Table 4. Players start with cooperation and if this keeps on forever, the total utility is

$$U_{GT}(C) = r + r \cdot d + r \cdot d^2 + r \cdot d^3 + \ldots = \frac{r}{1-d}.$$

If some player attempts defection[4], he will grab the initial bait, but he will trigger eternal defection, reaping

$$U_{GT}(D) = b + t \cdot d + t \cdot d^2 + t \cdot d^3 + \ldots = b + t \cdot \frac{d}{1-d}.$$

Comparing behaviors within grim-trigger will reveal

$$U_{GT}(C) \geq U_{GT}(D) \Leftrightarrow d_{GT} \geq \frac{b-r}{b-t}.$$

Under grim-trigger not deviating from cooperation is preferable as long as the discount factor is sufficiently large. At last, cooperation gets the prize—remember it had only succeeded in Pareto domination. Interesting outcome, taking into account that defection rules the finite repetition case.

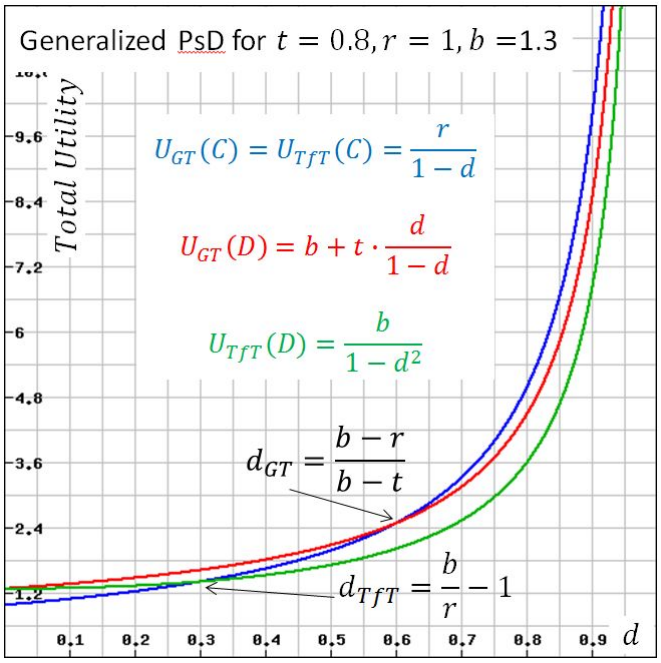

Figure 4: Comparison of behaviors within grim-trigger and Tit-for-Tat strategies

6.2 Forgiveness

A variation of grim-trigger is to punish your opponent just once after him defecting on you (Figure 3). This forgiving strategy is called **Tit-for-Tat**[5]. Let's study alternative behaviors within Tit-for-Tat.

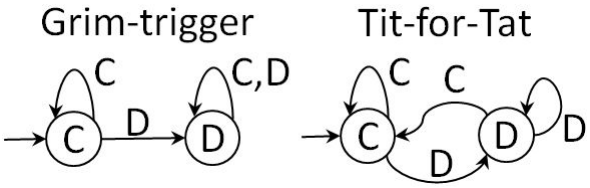

Figure 3: Automata for trigger strategies

If both players start off with (C, C), they will cooperate forever accumulating $U_{TfT}(C) = \frac{r}{1-d}$. If they both start off with (D, D), they will loop into defection forever, settling for $\frac{t}{1-d}$, which is always worse than $U_{TfT}(C)$. If they start off with (C, D), actions alternate ad infinitum. Player that initiated defection earns

$$U_{TfT}(D) = b + 0 \cdot d + b \cdot d^2 + 0 \cdot d^3 + \ldots = \frac{b}{1-d^2}.$$

For cooperation to be viable within Tit-for-Tat, there must be

$$U_{TfT}(C) \geq U_{TfT}(D) \Leftrightarrow d_{TfT} \geq \frac{b}{r} - 1.$$

Figure 4 shows the merit of cooperation as far as trigger strategies are concerned.

[4] At which point of the game defection first occurred is irrelevant: before it, utility is the same for both players; after it, the above analysis applies.

[5] Meaning: A blow or some other retaliation in return for an injury from another.

Simulations have shown that Tit-for-Tat is a particularly good strategy: it consistently achieves excellent scores among a multitude of competitors (robustness) and even tops among groups of the fittest opponents who have survived consecutive competitions ([Axe80]).

Axelrod ([Axe84]) has stated necessary conditions for a successful strategy:

Nice Not defect before opponent does

Retaliating React to opponent's bad behavior

Forgiving Fall back to cooperation after retaliation

Non-envious Not strive to score more than the opponent

6.3 Other strategies

Figure 5 shows **Win-Stay-Lose-Shift** or Pavlov strategy ([Nov93]), an altruistic trigger strategy which beats Tit-for-Tat, but does not score as much as it. This strategy decides for next move taking into account only the last play: if it had a good outcome, repeat your move; otherwise, alternate.

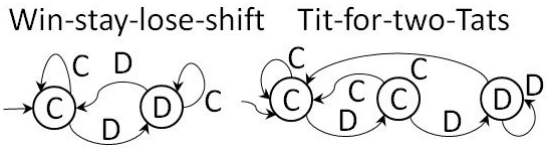

Figure 5: Automata for other strategies

Tit-for-two-Tats triggers defection only if opponent plays two consequtive defections; thus it is more forgiving than the original Tit-for-Tat condoning single, possibly accidental defections which could cause long runs of mutual backbiting in the original Tit-for-Tat.

In the 2004 tournament, a devious approach was taken by a team from the University of Southampton. They conquered the three first positions of the final ratings exploiting teamwork. Agents were able to recognize teammates and would sacrifice themselves to boost their team's score. This methodology had been suggested by evolutionary biologist Richard Dawkings ([Daw76]) as commonly occurring in nature to effect gene survival by exhibiting selfishness among members of the same gene against their competitors.

7 Brief history

Prisoner's Dilemma was introduced in 1950 as a model of cooperation and conflict by mathematicians Merrill Flood and Melvin Dresher working as strategic analysts at RAND Corporation. Their aim was to question the validity of Nash's equilibrium, whose theorem they had just heard[6]. Albert Tucker, one of Nash's professors, gave the prisoner's interpretation and named it.

Tit-for-Tat was Anatol Rapoport's winning solution to Robert Axelrod's computer tournaments ([Axe84]) in the 1980s for programs competing against each other to achieve the best PsD score.

References

[Axe80] Axelrod Robert, More Effective Choice in the Prisoner's Dilemma, *Journal of Conflict Resolution*, vol. 24, no. 3, pp. 379-403, 1980

[Axe84] Axelrod Robert, *The Evolution of Cooperation*, Basic Books, 1984

[Das09] Daskalakis Constantinos, et all, The Complexity Theory of Computing a Nash Equilibrium, *Comm. of the ACM*, vol. 52, no. 2, pp. 89-97, 2009

[Daw76] Dawkings Richard, *The Selfish Gene*, Oxford Univ. Press, 1976

[Goe00] Goeree Jacob, Holt Charles, Ten Little Treasures of Game Theory and Ten Intuitive Contradictions, *Univ. of Virginia*, Feb. 2000

[Ken07] Kendall Graham, Yao Xin, Chong Siang Yew, *The Iterated Prosoners' Dilemma 20 Years on*, World Sci. Publ., 2007

[Kou99] Koutsoupias Elias, Papadimitriou Christos, Worst-case Equilibria, *STACS*, vol. 1563, pp. 404-413, 1999

[Ley08] Leyton-Brown Kevin, Shoham Yoav, *Essentials of Game Theory*, Morgan&Claypool Publ., 2008

[Mal11] Malkevitch Joe, The Price of Anarchy, *AMS*, web essay, 2011

[Nov93] Novak Martin, Sigmund Karl, A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoners Dilemma game, *Nature*, vol. 364, pp. 5658, 1993

[Rou07] Roughgarden Tim, Tardos Eva, Introduction to the Inefficiency of Equilibria, ch. 17 in *Algorithmic Game Theory*, 2007

[Rou15] Roughgarden Tim, Intrinsic Robustness of the Price of Anarchy, *Journal of the ACM*, vol. 62, no. 5, 2015

[Sho08] Shoham Yoav, Computer Science and Game Theory, *Comm. of the ACM*, vol. 51, no. 8, pp. 75-79, 2008

[Sho10] Shoham Yoav, Leyton-Brown Kevin, *Multiagent Systems*, PDF, 2010

[Wat13] Watson Joel, *Strategy An introduction to Game Theory*, W.W. Norton & Co., 2013

[6] In fact, Nash ([Goe00]) responded to them: "The flaw in the experiment as a test of equilibrium point theory is that the experiment really amounts to having the players play one large multi-move game. One cannot just as well think of the thing as a sequence of independent games... There is too mush interaction...".