Repetition and Reputation in the Prisoners' Dilemma

Suppose two people know they're going to be playing the Prisoners' Dilemma game repeatedly, many times in succession. We say the two people are engaged in a *continuing strategic relationship*. In this situation, it seems as if cooperation might be likely to occur. We might expect that at each play of the game we would see each player choose the "cooperate" action, the action that's strongly dominated when the game is instead played just once. Indeed, when subjects play the repeated PD game in laboratory experiments, this "cooperative" pattern of play is usually what we observe. And we also see cooperation in real-life continuing strategic relationships.

Why would we expect to see cooperative play? When the PD game is being played repeatedly, my cooperation can be construed as a signal that I will continue to cooperate in future plays. It establishes a *reputation* that I will cooperate, and this may induce the other player to cooperate as well, in order not to lose my cooperation. Moreover, the same argument applies to him: he may cooperate on the current play for the same reason I'm considering cooperation — in order to establish his own reputation, hoping it will induce me to cooperate in the future.

Furthermore, a *defection* by the other player can be *punished* by my choosing to defect at future plays, which will unambiguously harm him, compared to my cooperating at future plays. Similarly, he can punish me if I defect.

This pattern of behavior — establishing a reputation, and punishing defection — can only exist in the repeated setting. It's the *continuing* nature of the strategic relationship that makes reputation and punishment possible.

Example: The ***Tit-for-Tat strategy*** is defined as follows: Cooperate at the first play; then, on subsequent plays, mimic the action the other player chose on the immediately preceding play — cooperate if he cooperated on the preceding play, defect if he defected. An important fact to notice about this "strategy" is that it specifies exactly what action to take at every play of the game, in every possible eventuality.

Let's see how play will unfold if each player uses the Tit-for-Tat strategy to play the repeated Prisoners' Dilemma. At each play of the game, the players will choose actions a_1 and a_2 as follows:

	1	2	3	4	5	6	7	8	...
a_1	C	C	C	C	C	C	C	C	...
a_2	C	C	C	C	C	C	C	C	...

This looks promising! It appears that the Tit-for-Tat strategy (and perhaps other strategies as well) will allow the players to overcome the PD game's bad equilibrium, in spite of the fact that it's a strong (*i.e.*, dominant-strategy) equilibrium when the game is played just once. But is it really *an equilibrium* of the repeated game for each player to play Tit-for-Tat? If a player expects the other player to play Tit-for-Tat, will Tit-for-Tat be the best strategy for him to play in response? To answer this question we have to define the game that consists of playing the PD game repeatedly, *i.e.*, we must define the Repeated PD Game (the ***RPD game***, for short).

The Tit-for-Tat strategy, and the path of play we identified above when both players follow it, is a good starting point for defining this new game. Recall that a game is defined by a set of strategies for each player and a payoff function for each player. A ***strategy*** in the RPD game is a prescription that tells a player what action to take in every circumstance he could possibly face. For example, as pointed out above, the Tit-for-Tat strategy does exactly this. Thus, a strategy is a function that maps all possible circumstances into the set $\{C, D\}$ of actions in the PD game, called the ***stage game***.

Now suppose each player has selected a strategy s_i for playing the RPD game. The players could tell us their respective strategies and we could use this pair of strategies to determine exactly how play of the game would proceed. Just as we did above, where each player was using the Tit-for-Tat strategy, we would obtain a sequence $\{(a_1(t), a_2(t))\}_{t=1}^{T}$ telling us what actions in the PD game (*i.e.*, C or D) each player will choose at every stage of play in the RPD game. (We will consider only *finitely repeated* PD games, games that are played T times in succession.) From this sequence of actions, we can determine the sequence of outcomes, *i.e.*, the sequence of payoffs the players will receive in the sequence of PD games: $\{(x_1(t), x_2(t))\}_{t=1}^{T}$, where $x_i(t) := \pi_i(a_1(t), a_2(t))$ for each i and each t. Finally, a player's ***payoff*** in the RPD game, say $\overline{\pi}_i(s_1, s_2)$, is some aggregation of the sequence of stage game payoffs he receives from the sequence of action pairs that result from (s_1, s_2), the RPD strategies the players have chosen. For example, $\overline{\pi}_i(s_1, s_2)$ might be the sum of his stage-game payoffs, $\sum_1^T x_i(t)$, or it might be a discounted sum of his stage-game payoffs, $\sum_1^T \left(\frac{1}{1+\delta}\right)^t x_i(t)$.

The best way to gain an understanding of the Repeated Prisoners' Dilemma is to analyze a specific numerical example, an example in which there are only two stages of play.

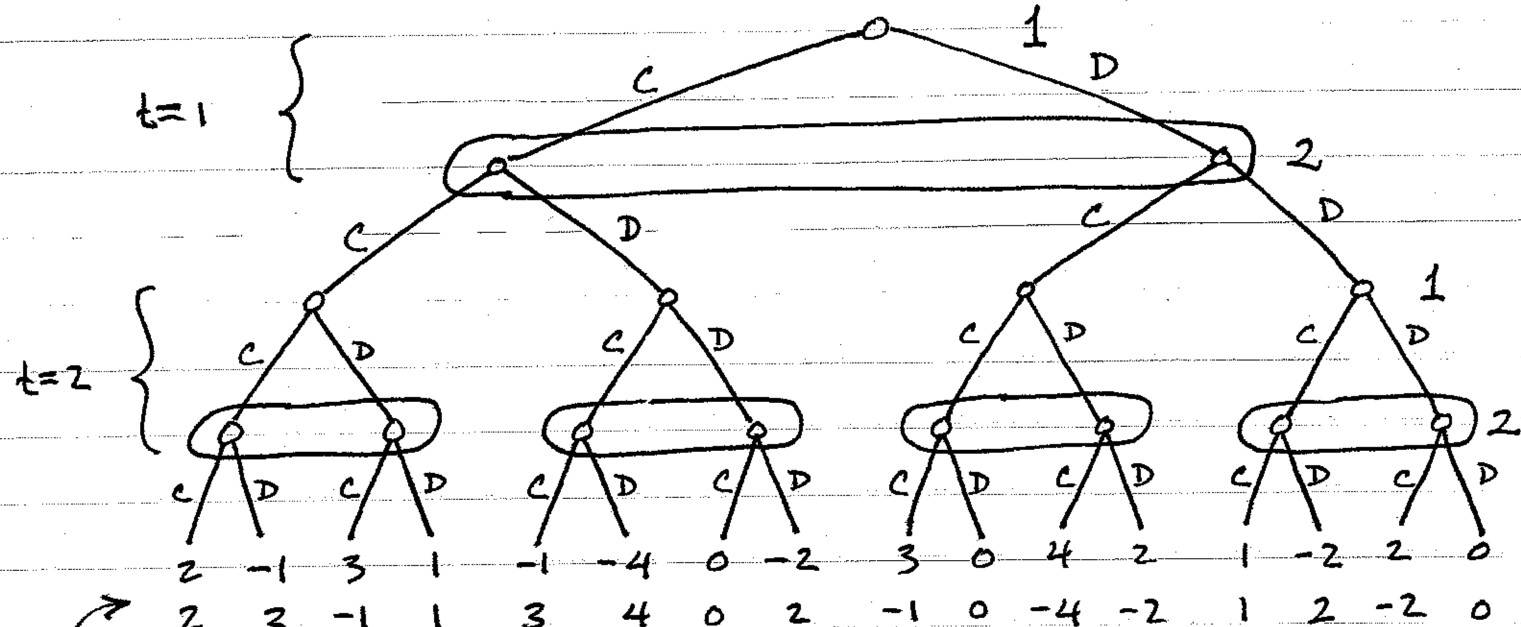
THE PRISONERS' DILEMMA PLAYED TWICE

(A TWO-STAGE REPEATED GAME)

THE BASIC, OR STAGE, GAME:

1 \ 2	C	D
C	1,1	-2,2
D	2,-2	0,0

THE GAME IS PLAYED TWICE, AT $t=1$ AND AGAIN AT $t=2$. CHOICES ARE SIMULTANEOUS AT $t=1$; THEN EACH OBSERVES THE $t=1$ CHOICES; THEN CHOICES ARE SIMULTANEOUS AT $t=2$.



(UNDISCOUNTED) SUM OF THE TWO STAGE-GAME PAYOFFS.

BEHAVIORAL STRATEGY FOR PLAYER 1:

(a) CHOOSE C OR D AT $t=1$. (TWO POSSIBLE CHOICES.)

(b) CHOOSE A FUNCTION $f_1 : \{C,D\}^2 \to \{C,D\}$ [i.e., $f_1 : (a_1,a_2) \mapsto a_1'$] FROM "HISTORIES" (a_1,a_2) TO ACTIONS a_1' AT $t=2$. (THERE ARE $2^4 = 16$ SUCH FUNCTIONS.)

$\therefore$ THERE ARE $2 \times 16 = 32$ STRATEGIES AVAILABLE IN THE REPEATED GAME.

THE NORMAL-FORM REPRESENTATION:

(USING BEHAVIORAL STRATEGIES, AND IGNORING <u>REDUNDANT</u> DEPENDENCE, AT $t=2$, ON ONE'S OWN ACTION AT $t=1$)

C D	C-CC	C-CD	C-DC	C-DD	D-CC	D-CD	D-DC	D-DD
C-CC	CC,CC 2,2	CC,CC 2,2	CC,CD −1,3	CC,CD −1,3	CD,CC −1,3	CD,CC −1,3	CD,CD −4,4	CD,CD −4,4
C-CD	CC,CC 2,2	CC,CC 2,2	CC,CD −1,3	CC,CD −1,3	CD,DC 0,0	CD,DC 0,0	CD,DD −2,2	CD,DD −2,2
C-DC	CC,DC 3,−1	CC,DC 3,−1	CC,DD 1,1	CC,DD 1,1	CD,CC −1,3	CD,CC −1,3	CD,CD −4,4	CD,CD −4,4
C-DD	CC,DC 3,−1	CC,DC 3,−1	CC,DD 1,1	CC,DD 1,1	CD,DC 0,0	CD,DC 0,0	CD,DD −2,2	CD,DD −2,2
D-CC	DC,CC 3,−1	DC,CD 0,0	DC,CC 3,−1	DC,CD 0,0	DD,CC 1,1	DD,CD −2,2	DD,CC 1,1	DD,CD −2,2
D-CD	DC,CC 3,−1	DC,CD 0,0	DC,CC 3,−1	DC,CD 0,0	DD,DC 2,−2	DD,DD 0,0	DD,DC 2,−2	DD,DD 0,0
D-DC	DC,DC 4,−4	DC,DD 2,−2	DC,DC 4,−4	DC,DD 2,−2	DD,CC 1,1	DD,CD −2,2	DD,CC 1,1	DD,CD −2,2
D-DD	DC,DC 4,−4	DC,DC 2,−2	DC,DC 4,−4	DC,DD 2,−2	DD,DC 2,−2	DD,DD 0,0	DD,DC 2,−2	DD,DD 0,0

<u>NOTATION FOR STRATEGIES:</u> X-yz MEANS "X AT $t=1$; THEN, AT $t=2$, y IF OPPONENT CHOSE C AT $t=1$, BUT z IF HE CHOSE D." THUS, C-DC IS "COOPERATE AT $t=1$, THEN (AT $t=2$) DEFECT IF OPPONENT COOPERATED AT $t=1$, COOPERATE IF HE DEFECTED." NOTE THAT "TIT-FOR-TAT" IS C-CD.

<u>OUTCOMES</u> ARE DENOTED $(a_1(1), a_2(1); a_1(2), a_2(2))$, WHERE $a_i(t)$ IS i'S ACTION AT t. THUS, "CC, DD" MEANS THAT BOTH COOPERATED AT $t=1$ AND DEFECTED AT $t=2$.

<u>UTILITY</u> (OR PAYOFF) AT AN OUTCOME WE ARE TAKING TO BE THE (UNDISCOUNTED) SUM OF THE TWO PERIODS' PAYOFFS:

$$u_i(\underline{a}) := \pi_1(a_1(1), a_2(1)) + \pi_2(a_1(2), a_2(2)).$$

[NOTE THAT THIS FITS THE <u>GAME FORM</u> APPROACH.]

THERE ARE FOUR NASH EQUILIBRIA: THE FOUR STRATEGY PAIRS THAT RESULT IN EACH PLAYER DEFECTING AT EACH PERIOD (i.e., EACH PLAYER USES EITHER D-DD OR D-CD). EVERY NE YIELDS THE SAME SEQUENCE OF PLAY AND THUS THE SAME PAYOFFS.

BUT NOTE THAT NEITHER PLAYER HAS A DOMINANT STRATEGY. IN PARTICULAR, IF A PLAYER'S OPPONENT SHOULD BE USING "TIT-FOR-TAT" (i.e., C-CD), THEN IT IS BEST TO COOPERATE AT $t=1$ (TO "ESTABLISH A COOPERATIVE REPUTATION," THEREBY INDUCING THE OPPONENT TO COOPERATE AT $t=2$), AND THEN, OF COURSE, TO DEFECT AT $t=2$.

THE ROW PLAYER'S BEST-RESPONSE FUNCTION:

	C-CC	C-CD	C-DC	C-DD	D-CC	D-CD	D-DC	D-DD
C-CC								
C-CD						• 0		
C-DC		• 3						
C-DD		• 3				• 0		
D-CC								
D-CD					• 2	• 0	• 2	• 0
D-DC	• 4		• 4	• 2				
D-DD	• 4		• 4	• 2	• 2	• 0	• 2	• 0

NOTE:

① ROW HAS NO DOMINANT STRATEGY. D-DD IS "ALMOST" DOMINANT, BUT IF COLUMN CHOOSES C-CD, THEN ROW IS BETTER OFF TO CHOOSE C-DC OR C-DD, EACH OF WHICH "ESTABLISHES A REPUTATION" FOR CHOOSING C AT $t=1$.

② IF ROW BELIEVES THERE IS AT LEAST A $\frac{2}{3}$ CHANCE THAT COLUMN IS PLAYING C-CD, THEN ROW'S BEST CHOICE IS C-DD. SIMILARLY FOR COLUMN.
(THE CALCULATION FOR THIS IS ON THE FOLLOWING PAGE.)

BELIEFS BY ROW ABOUT COLUMN'S LIKELY PLAY THAT WILL SUPPORT 1ST-STAGE COOPERATION BY ROW:

t: THE STRATEGY C-CD (TIT-FOR-TAT)
d: THE STRATEGY D-DD
c: THE STRATEGY C-DD

p: ROW'S BELIEF ABOUT THE PROBABILITY THAT COLUMN IS PLAYING TIT-FOR-TAT.

$E\bar{\pi}_d$ AND $E\bar{\pi}_c$: ROW'S EXPECTED PAYOFFS FROM c AND d.

$$E\bar{\pi}_c = p\bar{\pi}_R(c,t) + (1-p)\bar{\pi}_R(c,d)$$

$$= p[\pi_R(C,C)+\pi_R(D,C)] + (1-p)[\pi_R(C,D)+\pi_R(D,D)]$$

$$= p[1+2] + (1-p)[-2+0] = 3p + (1-p)(-2)$$

$$= 5p-2.$$

$$E\bar{\pi}_d = p\bar{\pi}_R(d,t) + (1-p)\bar{\pi}_R(d,d)$$

$$= p[\pi_R(D,C)+\pi_R(D,D)] + (1-p)[\pi_R(D,D)+\pi_R(D,D)]$$

$$= p[2+0] + (1-p)[0+0] = 2p$$

$$\therefore E\bar{\pi}_c > E\bar{\pi}_d \Leftrightarrow 5p-2 > 2p$$

i.e., $3p > 2$; i.e., $\boxed{p > \frac{2}{3}}$

THE REPEATED PRISONERS' DILEMMA

SUPPOSE THE TWO PLAYERS KNEW THAT THEY WOULD BE PLAYING THE SAME PRISONERS' DILEMMA (PD) GAME AGAINST ONE ANOTHER MANY TIMES — SAY, T TIMES. IT SEEMS AS IF IT WOULD BE A GOOD IDEA TO COOPERATE (i.e., TO CHOOSE C) AT EACH PLAY t UNTIL t IS NEAR T, AS LONG AS ONE'S OPPONENT IS DOING THE SAME. COULD THIS OCCUR IN AN EQUILIBRIUM OF AN APPROPRIATELY FORMULATED "SUPERGAME" DEFINED AS "THE REPEATED PD GAME (REPEATED T TIMES)"?

LET A_i BE THE STRATEGY (i.e., ACTION) SET $\{C, D\}$ FOR EACH i, AND π_i THE PAYOFF FUNCTION, IN THE PD GAME PLAYED AT EACH "STAGE" t OF THE REPETITION; IN PARTICULAR, FOR $i = 1, 2$:

(1) D IS A STRONGLY DOMINANT STRATEGY FOR π_i, AND

(2) $\pi_i(C, C) > \pi_i(D, D)$.

THIS "STAGE GAME" IS TO BE PLAYED T TIMES ($t = 1, \ldots, T$), WITH THE ACTIONS AT t OBSERVABLE PRIOR TO PLAY $t+1$.

HOWEVER WE MODEL THE "SUPERGAME," A <u>PLAY</u> OF IT WILL RESULT IN A T-TUPLE OF PAIRS (a_1, a_2) OF ACTIONS IN $\{C, D\}^2$: $((a_1(1), a_2(1)), \ldots, (a_1(T), a_2(T))) \in (\{C, D\}^2)^T$.
[NOTE THAT <u>STRATEGIES</u> IN THE SUPERGAME ARE MUCH MORE COMPLEX THAN THIS.]

EACH SUCH T-TUPLE $\underline{a} \in (\{C, D\}^2)^T$ WILL RESULT IN A T-TUPLE OF OUTCOMES $\underline{x} = (x_1(t), x_2(t))_{t=1}^T$, WHERE $x_i(t) = \pi_i(s_1(t), s_2(t))$. SUPPOSE THAT EACH PLAYER'S UTILITY FUNCTION OVER SUCH OUTCOMES $\underline{x}$ SATISFIES

$$\frac{\partial u_i}{\partial x_i(t)} > 0, \qquad t = 1, \ldots, T.$$

WHICH T-TUPLES $\underline{a}$ COULD OCCUR IN AN EQUILIBRIUM OF THIS (SUPER)GAME? COULD WE HAVE $a_i(T) = C$ FOR EITHER $i=1$ OR $i=2$? NO, BECAUSE AT T EACH PLAYER FACES JUST THE ORIGINAL SINGLE-PLAY, OR "ONE-SHOT," PD GAME — $\underline{a}(T)$ AFFECTS ONLY $\underline{x}(T)$. EQUILIBRIUM THEREFORE REQUIRES THAT $\underline{a}(T) = (D, D)$.

COULD WE HAVE $a_i(T-1) = C$ FOR EITHER i? NO, BECAUSE IN AN EQUILIBRIUM EACH KNOWS THAT $\underline{x}(T)$ WILL NOT BE AFFECTED BY $a_i(T-1)$ — AND CERTAINLY $\underline{x}(1), \ldots, \underline{x}(T-2)$ WILL ALL BE UNAFFECTED — SO ONLY $\underline{x}(T-1)$ WILL BE AFFECTED, AGAIN LEAVING US WITH A ONESHOT PD GAME. THUS, $\underline{a}(T-1) = (D, D)$.

WE MAKE THE SAME ARGUMENT RECURSIVELY (OR "INDUCTIVELY") FOR EACH t, AND WE FIND THAT IN AN EQUILIBRIUM THE ONLY $\underline{a}$ THAT COULD OCCUR IS $\underline{a} = ((D,D), (D,D), \ldots, (D,D))$. THIS ARGUMENT, INVOLVING THE "UNRAVELLING" OF ANY OTHER "PLAY-PATH" $\underline{a}$ IN THE REPEATED PD GAME IS REFERRED TO AS "BACKWARD INDUCTION."

The Repeated Prisoners' Dilemma: Summary

1. The repeated game consists of playing the PD game T times – say 2 times, or 100 times. We've confined our attention to finitely repeated PD games, *i.e.*, ones in which the game is to be repeated a finite number of times, and the number of repetitions is known to each player.

2. The repeated game is itself a single game, which we call the RPD (Repeated Prisoners' Dilemma). Each successive PD game is a stage of the RPD. When a new stage is to be played the players have information about the prior plays: what each player played at all prior stages (and therefore the resulting payoffs obtained at all prior stages).

3. A strategy in the RPD game is a complete contingent plan specifying what moves one will make as a function of the information one has received. Even for a two-stage RPD the number of strategies is surprisingly large. The resulting normal form (strategic form) game is the game in which each player's strategy set is the set of all these RPD strategies. A Nash equilibrium (NE) is defined as always for a normal form game.

4. In the RPD game it is not a dominant strategy to always choose Defect.

5. There are multiple Nash equilibria of the RPD.

6. Each NE of the RPD results in the same path of play: each player defects at every stage. This can be established directly, via the extremely large payoff table, or alternatively by a proof that works backward from the final stage, called a backward induction argument.

7. By allowing that one or both players aren't completely certain of the other player's "rationality" or payoff function, it has been shown that cooperation for many stages, followed by defection near the final stage, is a NE. In other words, with some incomplete information on the players' part, cooperation instead of defection can be supported as a NE of the RPD.

8. Cooperation can also be supported as a NE if the PD game is repeated infinitely often, or (in the finitely repeated PD) if the players are uncertain about which stage will be the final stage.