JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **80**, 67–77 (1981)

Inducing Cooperation by Reciprocative Strategy in Non-Zero-Sum Games*

P. P. SHENOY AND P. L. YU

School of Business, The University of Kansas, Lawrence, Kansas 66045

Submitted by G. Leitmann

In this paper, we investigate the use of reciprocative strategy to induce cooperative behavior in non-zero-sum games. Reciprocative behavior is defined mathematically in the context of a two-person non-zero-sum game in which both the players have a common set of pure strategies. Conditions under which mutual cooperative behavior results when one of the players responds optimally to reciprocative behavior by the other player are described. Also, the desirability of playing the reciprocative strategy is investigated by stating conditions under which reciprocative strategy by one of the players or by both the players leading to mutual cooperative behavior is a Nash equilibrium outcome.

1. INTRODUCTION

It is frequently observed in real life that people tend to reciprocate their dealings with other people (Ref. [1], for instance). How you feel about a person or how you act towards a person depends to a large extent on how the other person feels about you or how he acts towards you. We shall call this the *principle of reciprocation*. Such behavior is considered by some as good wisdom, e.g., "Do unto others as you expect others to do unto you" is a familiar cliché that echoes this principle. Here we shall study the implications of reciprocative behavior in a situation characterized by partial conflict and partial cooperation.

Most conflict situations involving two parties can be characterized as having a cooperative and a competitive component. Familiar examples of such conflict situations are the arms race between USA and USSR, international trade, military conflicts between two nations. On a smaller scale, we have competition between business firms producing similar products, competition between different divisions of an organization, etc. Such situations involving partial conflict and partial cooperation can be modeled as a two-person non-zero-sum game that is played repeatedly over time.

* This research was partially supported by grants from the University General Research Fund of the University of Kansas to each of the authors.

0022-247X/81/030067–11$02.00/0

In this paper, we shall analyze the use of reciprocative behavior as a strategy to be used in a repeated non-zero-sum game. We shall indicate conditions under which reciprocative behavior by one of the players leads to a cooperative equilibrium point when the second player perceives such behavior by the first player as a fact that he cannot change.

In the next section, we shall formally describe a non-zero-sum game formulation of a partial competitive and partial cooperative game in which each player has two pure strategies—cooperate, and do not cooperate (with the other player). Reciprocative behavior is then defined rigorously as a particular kind of strategy in this two-person non-zero-sum game. In Section 3, we indicate the conditions under which reciprocative behavior by one player leads to an equilibrium point where both players choose the cooperative strategy on each play of the game. The desirability of playing the reciprocative strategy is examined in Section 4 and Section 5 summarizes the findings in this paper.

2. Reciprocative Behavior in a Non-Zero-Gum Game

Consider a two-person non-zero-sum game in which each player has two pure strategies—cooperate and do not cooperate. Such a game will be denoted by the matrix as shown below.

		Player II 0	Player II 1
Player I	0	(a_{00}, b_{00})	(a_{01}, b_{01})
	1	(a_{10}, b_{10})	(a_{11}, b_{11})

0 = do not cooperate
1 = cooperate

The a_{ij}'s are player I's payoffs and the b_{ij}'s are player II's payoffs. The relation between the a_{ij}'s and b_{ij}'s will determine the nature of the game. For example, in the prisoner's dilemma game, we have

$$a_{01} > a_{11} > a_{00} > a_{10} \tag{1}$$

and

$$b_{10} > b_{11} > b_{00} > b_{01} \tag{2}$$

in which case (a_{00}, b_{00}) becomes the unique Nash equilibrium point if the game is played exactly once [2]. Let $\tilde{x}_t$ denote the mixed strategy used by player I in time period t. A *mixed strategy* is a probability distribution on the set of pure strategies. The probability distribution $\tilde{x}_t$ can be specified by x_t, where $0 \leqslant x_t \leqslant 1$ and

$$P(\tilde{x}_t = 1) = x_t, \qquad P(\tilde{x}_t = 0) = 1 - x_t. \tag{3}$$

Without confusion, x_t will simply be called a mixed strategy. Similarly let y_t denote the mixed strategy used by player II in time period t. Let $X = \{x_t\}$ and $Y = \{y_t\}$ denote the infinite sequences of mixed strategies used by player I and II, respectively, over an infinite time horizon. Let $\tilde{h}_t$ denote the (prior) history of the strategies used by the players up to and including time period t, i.e.,

$$\tilde{h}_t = \{(\tilde{x}_1, \tilde{y}_1), (\tilde{x}_2, \tilde{y}_2), \ldots, (\tilde{x}_t, \tilde{y}_t)\}. \tag{4}$$

Let h_t denote the observed value of $\tilde{h}_t$, i.e.,

$$h_t = \{(p_1, q_1), (p_2, q_2), \ldots, (p_t, q_t)\}, \tag{5}$$

where $p_t \in \{0, 1\}$ is an observed value of $\tilde{x}_t$ and $q_t \in \{0, 1\}$ is an observed value of $\tilde{y}_t$.

To incorporate the concept of the human learning process, we will assume that the mixed strategy used by a player at time t will be a function of the observed value of $\tilde{h}_{t-1}$. One such strategy is called reciprocative behavior. We say player II exhibits *perfect reciprocative behavior* iff

$$y_t(h_{t-1}) = \left(\sum_{n=1}^{t-1} p_n\right)\Big/(t-1) \qquad \text{for all} \quad t = 2, \ldots, \infty. \tag{6}$$

Note that reciprocative behavior implies that a player uses the entire history at any time t and that he gives equal weight to all the past actions of his opponent in forming his strategy at time t.

We can relax both these assumptions for greater flexibility as follows. First, we will assume that a player will consider only the past k observed actions of his opponent. Second, we will assume that a player will give more weight to the more recent actions of his opponent compared to the distant past. There are several ways of doing this. In particular, we will assume that a player will "discount the history" at discount rate γ, $0 \leqslant \gamma \leqslant 1$, i.e.,

$$\begin{aligned} y_t(h_{t-1}) &= \frac{p_{t-1} + \gamma^1 p_{t-2} + \cdots + \gamma^{k-1} p_{t-k}}{1 + \gamma_1 + \gamma^2 + \cdots + \gamma^{k-1}} \qquad \text{if} \quad t > k, \\ &= \frac{p_{t-1} + \gamma^1 p_{t-2} + \cdots + \gamma^{t-1} p_1}{1 + \gamma^1 + \cdots + \gamma^{t-1}} \qquad \text{if} \quad t \leqslant k, \end{aligned} \tag{7}$$

i.e.,

$$y_t(h_{t-1}) = \frac{1}{\zeta} \sum_{n=0}^{t_0 - 1} \gamma^n p_{t-n-1}, \tag{8}$$

where $t_0 = \min\{t, k\}$ and $\zeta = 1 + \gamma^1 + \cdots + \gamma^{t_0 - 1}$. We will refer to this type of behavior as *reciprocative behavior with k-step memory and memory discount rate* γ, i.e., (k, γ)-*reciprocative behavior* in short. Note that perfect reciprocative behavior is a special case of (k, γ)-reciprocative behavior with $k = \infty$ and $\gamma = 1$. In Section 3, assuming that player II will play the (k, γ)-reciprocative strategy as defined in (8), we determine the optimal response for player I. Since y_t is defined in terms of $p_1, ..., p_{t-1}$ (the observed ("posterior") values of $\tilde{x}_1, ..., \tilde{x}_{t-1}$), and our analysis in Section 3 is before the actual play of the game (a "pre-posterior" analysis), we shall substitute y_t defined in (8) with

$$\hat{y}_t = E[y_t] = \frac{1}{\zeta} \sum_{n=0}^{t_0 - 1} \gamma^n x_{t-n-1} \tag{9}$$

in devising the optimal response strategy.

3. Optimal Response to Reciprocative Strategy

Consider the two-person non-zero-sum game discussed in the previous section. Let A_t and B_t denote the expected payoff to players I and II, respectively, in the tth play if I plays mixed strategy x_t and II plays mixed strategy y_t. Then

$$\begin{aligned} A_t(x_t, y_t) = {} & (1 - x_t)(1 - y_t) a_{00} + (1 - x_t) y_t a_{01} \\ & + x_t(1 - y_t) a_{10} + x_t y_t a_{11}, \end{aligned} \tag{10}$$

$$\begin{aligned} B_t(x_t, y_t) = {} & (1 - x_t)(1 - y_t) b_{00} + (1 - x_t) y_t b_{01} \\ & + x_t(1 - y_t) b_{10} + x_t y_t b_{11} \end{aligned} \tag{11}$$

Now, there are several different ways in which a player may evaluate a stream of payoffs. Two possible evaluations of the stream $(A_1, A_2, ...)$ are

$$A(X, Y) = \sum_{t=1}^{\infty} \beta^t A_t(x_t, y_t), \tag{12}$$

where β, $0 < \beta < 1$, is called the (payoff) discount factor; and

$$\bar{A}(X, Y) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} A_t(x_t, y_t). \tag{13}$$

Expression (12) is called the discounted sum and expression (13) is called the long run average expected payoff per play. The discounted sum is more sensitive to the payoffs in the immediate future than the distant future whereas the long run average payoff per play is more sensitive to the steady-state payoffs.

Consider the case in which player II plays (k, γ)-reciprocative strategy as defined in (9). Player I perceives such behavior by player II as a fact that he is unable to change and proceeds to respond to such a strategy so as to maximize his payoff as given by expressions (12) or (13). To determine player I's optimal responding strategy, we shall first consider the case where player I restricts himself to playing a stationary strategy, i.e., playing the same mixed strategy at all plays of the game, i.e.,

$$x_t = x \qquad \text{for all } t.$$

Then player II's (k, γ) strategy y_t as given by (9) is

$$y_t = x \qquad \text{for all } t.$$

Hence, from (10), we have

$$A_t(x_t, y_t) = A_t(x) = x^2 a_{11} + x(1-x)a_{10} + (1-x)xa_{01} + (1-x)^2 a_{00}.$$

For national convenience, let

$$\begin{aligned} a_{01} - a_{11} &= \varepsilon_1, \\ a_{11} - a_{00} &= \varepsilon_2, \\ a_{00} - a_{10} &= \varepsilon_3. \end{aligned} \tag{14}$$

Then $A_t(x) = (\varepsilon_3 - \varepsilon_1)x^2 + (\varepsilon_1 + \varepsilon_2 - \varepsilon_3)x + (\varepsilon_3 + a_{10})$. Note that $A(X, Y)$ and $\bar{A}(X, Y)$ will be maximized at the same strategy which maximizes $A_t(x)$. To find the max $A_t(x)$, note that for $0 \leqslant x \leqslant 1$

$$\frac{d}{dx} A_t(x) = 2(\varepsilon_3 - \varepsilon_1)x + (\varepsilon_1 + \varepsilon_2 - \varepsilon_3), \tag{15}$$

$$\frac{d^2}{dx^2} A_t(x) = 2(\varepsilon_3 - \varepsilon_1), \tag{16}$$

and $A_t(1) = a_{11}$, $A_t(0) = a_{00}$.

Throughout the paper, we shall assume that mutual cooperative behavior results in a better outcome for both the players as compared to mutual non-cooperative behavior, i.e.,

$$a_{11} > a_{00}, \tag{17}$$

$$b_{11} > b_{00}. \tag{18}$$

We have

LEMMA 1.

(i) *If* $\varepsilon_1 \leqslant \varepsilon_3$, $A_t(x)$ *achieves its maximum at* $x = 1$.

(ii) *If* $\varepsilon_1 > \varepsilon_3$ *and* $\varepsilon_2 \geqslant \varepsilon_1 - \varepsilon_3$, $A_t(x)$ *achieves its maximum at* $x = 1$.

(iii) *If* $\varepsilon_1 > \varepsilon_3$ *and* $\varepsilon_2 < \varepsilon_1 - \varepsilon_3$, $A_t(x)$ *achieves its maximum at*

$$x = \frac{\varepsilon_2 + (\varepsilon_1 - \varepsilon_3)}{2(\varepsilon_1 - \varepsilon_3)} \, (<1).$$

Proof. See the Appendix.

THEOREM 1. *If*

(i) *player II is committed to playing the* (k, γ)*-reciprocative strategy regardless of what player I does,*

(ii) *player I is aware and is convinced of this commitment,*

(iii) *player I is restricted to playing a stationary strategy, and*

(iv) *player I wishes to maximize either the discounted sum of payoffs or the long run average payoff per play, then player I will play the cooperative strategy at all plays of the game if and only if either*

(a) $\varepsilon_1 \leqslant \varepsilon_3$

or

(b) $\varepsilon_1 > \varepsilon_3$ *and* $\varepsilon_2 \geqslant \varepsilon_1 - \varepsilon_3$.

Proof. The proof is a direct consequence of Lemma 1.

Theorem 1 gives conditions on the payoffs of player I for reciprocative behavior by player II to lead to mutual cooperative behavior by both the players. Note that the conditions indicated by Theorem 1 are independent of k and γ (a result of the assumption that player I is restricted to playing a stationary strategy).

Next, consider the case in which player I is not restricted to playing a stationary strategy. In this case, we have the following result:

THEOREM 2. *If*

(i) *player II is committed to playing the* (k, γ)*-reciprocative strategy regardless of what player I does,*

(ii) *player I is aware and is convinced of this commitment, and*

(iii) *player I wishes to maximize the discounted sum of payoffs, then a sufficient condition for player I to play the cooperative strategy all the time is either*

(a) $\varepsilon_1 \leqslant \varepsilon_3$ *and* $(\beta\gamma^0 + \cdots + \beta^k\gamma^{k-1})/(\gamma^0 + \cdots + \gamma^{k-1}) > \varepsilon_3/(\varepsilon_1 + \varepsilon_2)$

or

(b) $\varepsilon_1 > \varepsilon_3$ *and* $(\beta\gamma^0 + \cdots + \beta^k\gamma^{k-1})/(\gamma^0 + \cdots + \gamma^{k-1}) > \varepsilon_1/(\varepsilon_3 + \varepsilon_2)$.

Proof. See the Appendix.

Note that Theorems 1 and 2 are "consistent" since

$$(\beta\gamma^0 + \cdots + \beta^k\gamma^{k-1})/(\gamma^0 + \cdots + \gamma^{k-1}) < 1.$$

Theorems 1 and 2 give sufficient conditions for player I to play the cooperative strategy on all plays of the game (leading to mutual cooperative behavior on all plays of the game) assuming that

(i) player II is committed to playing (k, γ)-reciprocative strategy regardless of what player I does and

(ii) player I is aware and is convinced of this commitment.

In other words, player II (by announcing his intention to play the reciprocative strategy regardless of what player I does) can induce player I to play the cooperative strategy all the time (assuming player I is convinced about player II's intentions) if the conditions in Theorem 1 or 2 are satisfied.

In the above analysis, the roles of players I and II were chosen arbitrarily. Similar results can be derived if the roles of players I and II are interchanged.

4. Reciprocative Behavior as a Nash Equilibrium Strategy

In the last section, we determined some sufficient conditions that enabled, say, player II to induce cooperative behavior by player I in all plays of the game leading to mutual cooperative behavior on all plays of the game. However, if mutual cooperative behavior does not yield the best payoff to, say, player II, then player II may not have the motivation to play the (k, γ)-reciprocative strategy. However, if mutual cooperative behavior results in the best outcome for, say, player II, then there is a strong motivation for player II to announce that he is committed to playing the (k, γ)-reciprocative strategy regardless of what player I does. If player I is indeed convinced of player II's intentions then under the conditions indicated in Theorem 1 or 2, player I will respond by playing the cooperative strategy on all plays even though mutual cooperative behavior may not yield the best outcome for him (player I). Thus we have:

COROLLARY 1. *Given the conditions of Theorem* 1 *or* 2 *and if mutual reciprocative behavior results in the best outcome for player II, i.e.,*

$$b_{11} = \max_{i,j} b_{ij},$$

then the optimal response by player I to player II's (k, γ)*-reciprocative strategy leading to mutual cooperative behavior on all plays will be a Nash equilibrium point of the game.*

A similar result will hold if we interchange the roles of players I and II. Next, we may ask the question: when is (k, γ)-reciprocative behavior by both players a Nash equilibrium point? The answer clearly follows from Theorem 2 and is stated explicitly in Corollary 2 below.

COROLLARY 2. *If player I plays the* (k_1, γ_1)*-reciprocative behavior strategy with* $x_1 = 1$ (*i.e., cooperative pure strategy on the first play*) *and player II plays the* (k_2, γ_2)*-reciprocative behavior strategy with* $y_1 = 1$, *then this pair of strategies will be a Nash equilibrium point if*

(i) *the payoffs* a_{ij}*'s of player I satisfy either condition* (a) *or* (b) *of Theorem* 2 *with* $k = k_2$, $\gamma = \gamma_2$, $\beta = \beta_1 =$ *payoff discount rate for player I*; *and*

(ii) *the payoffs* b_{ij}*'s of player II satisfy conditions similar to those given for player I above with* $k = k_1$, $\gamma = \gamma_1$ *and* $\beta = \beta_2 =$ *payoff discount rate for player II.*

CONCLUSIONS

Reciprocative behavior was defined rigorously in the context of a two-person non-cooperative non-zero-sum game situation, a situation characterized by partial conflict and partial cooperation. In some games, a player can employ the reciprocative strategy to induce the other player to play the cooperative strategy in all plays of the game. Also conditions for reciprocative behavior to lead to a Nash equilibrium outcome were given.

6. Appendix

Proof of Lemma 1.

(i) If $\varepsilon_1 < \varepsilon_3$, then $(d^2/dx^2)A_t(x) > 0$ for all $0 \leqslant x \leqslant 1$, i.e., $A_t(x)$ is concave downward ($\cup$) everywhere. Hence $A_t(x)$ achieves its maximum at one of the endpoints of the interval $[0, 1]$. Since $A_t(1) > A_t(0)$, the conclusion follows.

If $\varepsilon_1 = \varepsilon_3$ then $A_t(x) = \varepsilon_2 x + \varepsilon_3 + a_{10}$. Since $\varepsilon_2 > 0$, the result follows.

(ii) and (iii) If $\varepsilon_1 > \varepsilon_3$, then $(d^2/dx^2)A_t(x) < 0$ for all $0 \leqslant x \leqslant 1$, i.e., $A_t(x)$ is concave upwards ($\cap$) everywhere. The maximum of $A_t(x)$ occurs at the point where $(d/dx)A_t(x) = 0$, i.e., at

$$x^* = \frac{\varepsilon_2 + (\varepsilon_1 - \varepsilon_3)}{2(\varepsilon_1 - \varepsilon_3)}.$$

Hence if $\varepsilon_2 \geqslant \varepsilon_1 - \varepsilon_3$, then $x^* \geqslant 1$, and the result in (ii) follows. If $\varepsilon_2 < \varepsilon_1 - \varepsilon_3$, then $x^* < 1$ and the assertion in (iii) is proved. ∎

Proof of Theorem 2.

$$A_t(x_t, y_t) = x_t y_t(\varepsilon_3 - \varepsilon_1) - x_t \varepsilon_3 + y_t(\varepsilon_1 + \varepsilon_2) + \varepsilon_3 + a_{00}$$

and from (12), we have

$$A(X, Y) = \sum_{t=1}^{\infty} \beta^t [A_t(x_t, y_t)].$$

Substituting for y_t by $\hat{y}_t$, from (9), we get

$$A(\{x_t\}) = \sum_{t=1}^{\infty} \beta^t \left[x_t \frac{(\sum_{n=0}^{t_0-1} \gamma^n x_{t-n-1})}{\zeta} (\varepsilon_3 - \varepsilon_1) - x_t \varepsilon_3 + \frac{(\sum_{n=0}^{t_0-1} \gamma^n x_{t-n-1})}{\zeta} (\varepsilon_1 + \varepsilon_2) + \varepsilon_3 + a_{00} \right],$$

where $\zeta = 1 + \gamma + \gamma^2 + \cdots + \gamma^{t_0-1}$ and $t_0 = \min\{t, k\}$. Then

$$\frac{\partial}{\partial x_m} A(\{x_t\}) = \frac{\beta^m}{\sum_{n=0}^{t_1-1} \gamma^n} \left(\sum_{n=0}^{t_1-1} \gamma^n x_{m-n-1} \right) (\varepsilon_3 - \varepsilon_1) - \beta^m \varepsilon_3 + \sum_{t=m+1}^{m+k} \beta^t \left[\frac{x_t \gamma^{t-m-1}(\varepsilon_3 - \varepsilon_1)}{\zeta} + \frac{\gamma^{t-m-1}(\varepsilon_1 + \varepsilon_2)}{\zeta} \right],$$

where $t_1 = \min\{m, k\}$.

Case (*i*): $\varepsilon_1 > \varepsilon_1$. In this case $\partial A(\{x_t\})/\partial x_m$ will take a minimum value when

$$x_{m-t_1} = \cdots = x_{m-2} = x_{m-1} = x_{m+1} = \cdots = x_{m+k} = 0$$

At these values of x, we have

$$\frac{\partial A(\{x_t\})}{\partial x_m} = -\beta^m \varepsilon_3 + \sum_{t=m+1}^{m+k} \frac{\beta^t \gamma^{t-m-1}(\varepsilon_1 + \varepsilon_2)}{\zeta}.$$

Hence

$$\left.\frac{\partial A(\{x_t\})}{\partial x_m}\right|_{\min} > 0$$

if

$$\sum_{t=m+1}^{m+k} \frac{\beta^t \gamma^{t-m-1}(\varepsilon_1 + \varepsilon_2)}{\zeta} > \beta^m \varepsilon_3,$$

i.e.,

$$\frac{1}{\zeta} \sum_{t=1}^{k} \beta^t \gamma^{t-1} > \frac{\varepsilon_3}{\varepsilon_1 + \varepsilon_2},$$

i.e.,

$$\frac{\beta\gamma^0 + \beta^2\gamma^1 + \cdots + \beta^k\gamma^{k-1}}{\gamma^0 + \gamma^1 + \cdots + \gamma^{k-1}} > \frac{\varepsilon_3}{\varepsilon_1 + \varepsilon_2}. \tag{19}$$

Case (ii): $\varepsilon_3 < \varepsilon_1$. In this case $(\partial/\partial x_m) A(\{x_t\})$ is minimum when

$$x_{m-t_1} = \cdots = x_{m-2} = x_{m-1} = x_{m+1} = x_{m+2} = \cdots = x_{m+k} = 1.$$

At these values of x, we have

$$\begin{aligned}\frac{\partial A(\{x_t\})}{\partial x_m} &= \beta^m(\varepsilon_3 - \varepsilon_1) - \beta^m \varepsilon_3 \\ &\quad + \sum_{t=m+1}^{m+k} \beta^t \left[\frac{\gamma^{t-m-1}(\varepsilon_3 - \varepsilon_1 + \varepsilon_1 + \varepsilon_2)}{\zeta}\right] \\ &= -\beta^m \varepsilon_1 + \sum_{t=m+1}^{m+k} \beta^t \left[\frac{\gamma^{t-m-1}}{\zeta}\right](\varepsilon_3 + \varepsilon_2).\end{aligned}$$

Hence, (at these values of x)

$$\left.\frac{\partial A(\{x_t\})}{\partial x_m}\right|_{\min} > 0 \quad \text{if } \frac{\beta^1\gamma^0 + \beta^2\gamma^1 + \cdots + \beta^k\gamma^{k-1}}{\gamma^0 + \gamma^1 + \cdots + \gamma^{k-1}} > \frac{\varepsilon_1}{\varepsilon_2 + \varepsilon_3}. \tag{20}$$

In view of (19) and (20) we see that under the conditions of the theorem,

$$\frac{\partial A(\{x_t\})}{\partial x_m} > 0 \qquad \text{for all} \quad x_m \in [0, 1]$$

and hence player I can maximize his discounted sum of payoffs only if he plays the cooperative strategy (i.e., $x_m = 1$) on the mth play. Since m was picked arbitrarily, the result in Theorem 2 follows. ∎

References

1. J. L. Freedman, J. M. Carlsmith, and D. O. Sears, "Social Psychology," Prentice–Hall, Englewood Cliffs, N.J., 1974.
2. J. F. Nash, Non-cooperative games, *Ann. of Math.* **54** (1951), 286–291.