FILE COPY

Chapter 20

CAN BOUNDED RATIONALITY RESOLVE THE PRISONERS' DILEMMA?*

ROY RADNER

AT&T Bell Laboratories, Murray Hill, NJ, USA

1. Introduction

If a Prisoners' Dilemma game is repeated finitely many times, all Cournot–Nash (non-cooperative) equilibria of the resulting sequential game have the property that the non-cooperative (and Pareto inferior) outcome occurs in each period, no matter how large the number of periods. On the other hand, if there are infinitely many periods, and the utility to each player is the long-run average of his one-period payoffs, then there are non-cooperative equilibria of the sequential game that produce the "cooperative" outcome (which is Pareto optimal) in each period.[1] This "discontinuity" of the set of equilibria as the number of periods increases without limit is uncomfortable on an intuitive level, since the larger the number of periods, the larger is the potential gain from cooperation, and the larger is the incentive for the players to signal their willingness to cooperate.

In these notes I explore three departures from the strict "rationality" of the Cournot–Nash equilibrium calculation. In the first departure, each player is uncertain about the degree of cooperativeness of the other player. In this case, one can give conditions under which the larger the number of total periods in the game, the longer the players remain cooperative (Section 4). In the second departure, if each player is satisfied to get within a small distance (in average utility) of the best response to the other player's strategy (epsilon-equilibrium),

*The views expressed here are the author's, and not necessarily those of AT & T Bell Laboratories or the National Science Foundation. The ideas presented here have been substantially inspired by discussions with R.W. Rosenthal. This research was partially supported by the National Science Foundation. A previous version of this paper was circulated in March 1979 as a Bell Laboratories Economic Research Center Discussion Paper.

[1] See Luce and Raiffa (1957, pp. 94–102) on the Prisoners' Dilemma; see also Aumann (1981) on infinite supergames.

then as the number of periods increases, the corresponding sets of equilibria include those with longer and longer cooperation (Section 5). In the third departure, each player is restricted to strategies that can be implemented by finite automata whose size has a given upper bound. In this case, if the number of periods is sufficiently large compared to the upper bound on the size of the automata, then there are equilibria in which the players cooperate throughout the game (Section 6). The last section interprets these results in terms of concepts of bounded rationality.

2. *The one-period "Prisoners' Dilemma" game*

Each player has two actions, labelled 0 and 1. Figure 1 shows the respective utility outcomes for the two players.

Utility to Player 1

Player 1's *action* \ Player 2's *action*	0	1
0	0	b
1	$-a$	c

Utility to Player 2

Player 1's *action* \ Player 2's *action*	0	1
0	0	$-a$
1	b	c

Figure 1

The numbers a, b and c satisfy the inequalities

$$a > 0, \qquad b > c > 0, \qquad c > \frac{b-a}{2}. \tag{1}$$

With these inequalities, for each player action 0 is better than action 1, no matter which action the other player takes. In other words, in the one-period game, taking action 0 is the dominant strategy. A pair of actions, one for each player, is called a *Cournot–Nash equilibrium* – or just *equilibrium* for short – if neither player can increase his utility by unilaterally changing his action. It is clear that (0, 0) is the only equilibrium and yields zero utility to each player. This will be called the *non-cooperative outcome*.

On the other hand, if each player takes action 1, then each player gets utility c, an outcome that is "Pareto superior" to the non-cooperative one. This will be called the *cooperative outcome*.

Since "side payments" are not allowed, the three outcomes (c, c), $(b, -a)$ and $(-a, b)$ are all Pareto optimal. The last two outcomes are not individually rational, however, since each player can guarantee himself at least zero utility, by taking action 0. Hence the outcome (c, c) is the only one in the core and is a reasonable target for cooperative behavior.

I emphasize that I assume that all the data of the game are common knowledge of the two players.

3. *The several-period game*

Consider now a sequential, *T-period*, game in which the one-period game is repeated T times (T finite). The resulting utility to a player is the sum of the T one-period utilities. Let $A_i(t)$ denote the action of player i at date t ($1 \leq t \leq T$), i.e., during the tth one-period game. A pure strategy for player i is a sequence of functions, one for each date t; the function for date t determines i's action at t as a function of the previous actions. An *equilibrium* is a pair of strategies, one for each player, such that neither player can increase his total utility by unilaterally changing his strategy. It is easy to verify that in every equilibrium of the T-period game, each player takes action 0 at each date. This can be seen by "working backwards" since at the end of period t the players face a $(T - t)$-period game. The resulting total utility to each player is zero.

On the other hand, if each player were to take action 1 at each date, then the resulting total utility to each player would be Tc. The larger T is, the greater are the potential gains from cooperation. This suggests the following class of strategies for player i; for a given integer k, player i takes action 1 (i.e., "cooperates") as long as player j ($\neq i$) has been taking action 1, but not longer than k periods; thereafter player i takes action 0. Call this the (pure) strategy C_k. Formally define

$$\begin{aligned} D_j &\equiv \infty, && \text{if} \quad A_j(t) = 1 \quad \text{for all } t, \\ &\equiv \min\left\{ A_j(t) = 0 \right\}, && \text{otherwise}; \end{aligned} \tag{2}$$

then strategy C_k for player i is defined by

$$\begin{aligned} A_i(t) &= 1 \quad \text{if} \quad t \leq \min(D_j, k), \\ &= 0 \quad \text{if} \quad t > \min(D_j, k). \end{aligned} \tag{3}$$

Call C_k a *trigger strategy*, and let C denote the family of trigger strategies ($k \geq 0$). Note that if i uses C_0, then he always takes action 0.

Suppose that player 2 uses strategy C_k. If at some date t player 1 takes action 0, then player 2 will take action 0 at all subsequent dates, and it will be optimal for player 1 to do the same. Therefore, player 1's best response to C_k is some strategy with the property that

$$A_1(t) = 0 \quad \text{implies} \quad A_1(t') = 0 \quad \text{for all} \quad t' \geq t. \tag{4}$$

On the other hand, if at some date t player 2 (using C_k) takes action 0, then he will do so at all subsequent dates, and it will be optimal for player 1 to do the same. Therefore, player 1's best response to C_k also has the property that

$$A_2(t) = 0 \quad \text{implies} \quad A_1(t') = 0 \quad \text{for all} \quad t' > t. \tag{5}$$

Putting (4) and (5) together, it follows that 1's best response to C_k is a trigger strategy C_n, for some $n \geq 0$. It is straightforward to verify that if player 1 uses C_n and player 2 uses C_k, then the total utility to player 1 is

$$\begin{aligned} v_1(C_n, C_k) &= nc + b \quad \text{if} \quad n < k, \\ &= kc \qquad\quad \text{if} \quad n = k, \\ &= kc - a \quad \text{if} \quad n > k, \end{aligned} \tag{6}$$

(recall that $0 \leq n,\ k \leq T$). Hence player 1's best response to C_k is C_{k-1} if $k > 0$, and C_0 if $k = 0$. [This shows, incidentally, that the pair (C_0, C_0) is an equilibrium.]

4. Uncertainty

Now consider player 1's best response to a *mixture* of trigger strategies by player 2. This can also be interpreted as player 1's optimal strategy if he is certain that player 2 will use a trigger strategy, but is uncertain about which one. Let p_k denote the probability that player 2 uses trigger strategy C_k, and let p denote the probability vector with components p_k. Player 1's goal is to maximize his expected total utility, against p. Again, if at date t either $A_1(t)$ or $A_2(t)$ is 0, then $A_2(t')$ will be 0 for all $t' > t$. Hence player 1's best response to p must again have properties (4) and 5, and so it may be taken to be a (pure) trigger strategy.

The expected utility to player 1 if he uses C_n against p is

$$\begin{aligned} v_1(C_n, p) &= \sum_{k=0}^{n-1} p_k(kc - a) + p_n nc + \sum_{k=n+1}^{T} p_k(nc + b) \\ &= b \operatorname{prob}(k > n) - a \operatorname{prob}(k < n) + c \operatorname{E} \min(k, n). \end{aligned} \tag{7}$$

Taking the first differences of (7) one gets

$$\begin{aligned} v_1(C_{n+1}, p) - v_1(C_n, p) &= -ap_n - bp_{n+1} + c \operatorname{prob}(k > n) \\ &= -ap_n - (b - c)p_{n+1} + c \sum_{k > n+1} p_k. \end{aligned} \tag{8}$$

Let $\hat{n}$ denote player 1's optimal value of n, and let k^* be the largest k that has positive probability, i.e.,

$$k^* \equiv \max\{k\colon p_k > 0\}.$$

From (8), if $k^* > 0$, then

$$v_1(C_{k^*}, p) - v_1(C_{k^*-1}, p) = -ap_{k^*-1} - (b - c)p_{k^*} < 0.$$

Hence, if $k^* > 0$, then *player 1's optimal response* $C_{\hat{n}}$ to p must have $\hat{n} < k^*$.

From (7) one also gets

$$\begin{aligned} v_1(C_n, p) - v_1(C_0, p) = {} & -a \operatorname{prob}(k < n) - b \operatorname{prob}(0 < k \le n) \\ & + c \operatorname{E} \min(k, n). \end{aligned} \tag{9}$$

The sum of the first two terms on the right side of (9) cannot be less than $-(a + b)$. Hence

$$v_1(C_n, p) - v_1(C_0, p) \ge -(a + b) + c \operatorname{E} \min(k, n), \tag{10}$$

so a sufficient condition for $\hat{n}$ to be positive is that, for some $n > 0$,

$$\mathrm{E}\min(k, n) > \frac{a+b}{c}. \tag{11}$$

The left side of (11) is largest when $n = T$, so we can replace (11) by the condition

$$\mathrm{E}k > \frac{a+b}{c}. \tag{12}$$

Hence, *if* Ek *is sufficiently large, then player 1's best response* C_n *to* p *has* $n > 0$, *i.e., player 1 will be cooperative for at least one period.*

A similar argument leads to a sufficient condition that $\hat{n}$ be as large as any preassigned number. Suppose that $n \geq m$. If $A_2(1) = \cdots = A_2(m) = 1$, then player 1 faces a $(T - m)$-period game, with the original distribution p of k replaced by the conditional distribution of $(k - m)$ given that $k \geq m$. By condition (12) a sufficient condition for $(\hat{n} - m)$ to be positive (provided $\hat{n} \geq m$) is that

$$\mathrm{E}(k - m | k \geq m) > \frac{a+b}{c},$$

which is equivalent to

$$\mathrm{E}(k | k \geq m) > \frac{a+b}{c} + m.$$

It follows that, *for any* $m \geq 0$, *a sufficient condition for* $\hat{n} > m$ *is that*

$$\mathrm{E}(k | k \geq m') > \frac{a+b}{c} + m' \quad \text{for all } m' \text{ such that} \quad 0 \leq m' \leq m.$$

An example

Suppose that the probability distribution p is given by

$$p_k = \frac{(1-r)r^k}{1-r^{T+1}}, \qquad k = 0, \ldots, T,$$

where r is a real number with $0 < r < 1$. Recall that p_k is the probability that player 2 uses the trigger strategy C_k. One can verify that

$$\mathrm{E}\min(n,k) = \left(\frac{r}{1-r^{T+1}}\right)\left(\frac{1-r^n}{1-r} - nr^T\right),$$

$$\mathrm{E}k = \left(\frac{r}{1-r^{T+1}}\right)\left(\frac{1-r^T}{1-r} - Tr^T\right),$$

$$\operatorname{prob}(k \le n) = \frac{1-r^{n+1}}{1-r^{T+1}}.$$

Note that, for fixed r,

$$\lim_{T\to\infty} \mathrm{E}\,k = \frac{r}{1-r}.$$

Note also that

$$\operatorname{prob}(k \ne t+1 | k > t) = r,$$

so that r is the conditional probability that player 2 will take action 1 at least one more time, given that both players have used action 1 up through date t.

Using (7) and (8), one can verify that

$$v_1(C_n, p) = \frac{b(r^{n+1} - r^{T+1})}{1-r^{T+1}} - \frac{a(1-r^n)}{1-r^{T+1}} + \left(\frac{cr}{1-r^{T+1}}\right)\left(\frac{1-r^n}{1-r} - nr^T\right),$$

$$\Delta v_1(C_n, p) \equiv v_1(C_{n+1}, p) - v_1(C_n, p) = \frac{r^n[cr - (br+a)(1-r)] - cr^{T+1}}{1-r^{T+1}}.$$

Define

$$K \equiv cr - (br+a)(1-r),$$

and suppose that

$$K > 0; \tag{13}$$

then $\Delta v_1(C_n, p)$ is decreasing in n. For sufficiently large T, $\Delta v_1(C_n, p) > 0$, so that the optimal n, say $\hat{n}$, is approximately the solution of $\Delta v_1(C_n, p) = 0$, or

$$Kr^n = cr^{T+1}.$$

Hence, approximately,

$$\hat{n} = T + 1 - \frac{\log(K/c)}{\log r}, \tag{14}$$

provided of course this gives something less than or equal to T, otherwise $\hat{n} = T$. Note that in (14) $(\hat{n} - T)$ does not depend on T, in other words, player 1 continues to "cooperate" as long as player 2 does, until the horizon T becomes sufficiently close.

In this example, the parameter r measures the tendency of player 2 to cooperate, or player 1's subjective estimates of that tendency. Player 1's best response is to cooperate initially only if condition (13) is satisfied, i.e., only if r is sufficiently different from zero (given a, b and c). In particular, condition (13) does not depend on T. If condition (13) is satisfied, then player 1 will cooperate until reaching a fixed time from the end of the game.

5. *Epsilon-equilibria*

Consider again the sequential game of Section 3, except that each player's utility for the game is defined to be his total utility divided by the number of periods. An *ε-equilibrium* is a pair of strategies, one for each player, such that each player's strategy is within ε of being the best response to the other player's strategy.

From (6), it is easy to check that, for any positive ε, there is a T_0 such that, for all $T \geq T_0$ and all $k \leq T$, the pair (C_k, C_k) of trigger strategies is an ε-equilibrium. However, using (6) and (7) one can construct larger sets of ε-equilibria. Recall that player 1's best response to the pure trigger strategy C_k is C_{k-1}. From (6) one has, for $k > 0$,

$$\begin{aligned} v_1(C_{k-1}, C_k) - v_1(C_n, C_k) &= (k - 1 - n)c, \quad && n < k, \\ &= b - c, && n = k, \\ &= a + b - c, && n > k. \end{aligned} \tag{15}$$

Hence player 1's *average utility* for C_n is within ε of being optimal against C_k

if and only if

$$\begin{aligned} n &\geq k - 1 - \frac{T\varepsilon}{c}, & n &\leq k - 1, \\ T &\geq \frac{b - c}{\varepsilon}, & n &= k, \\ T &\geq \frac{a + b - c}{\varepsilon}, & n &> k. \end{aligned} \tag{16}$$

Suppose that

$$T \geq \frac{a + b - c}{\varepsilon}. \tag{17}$$

Then, given (17), (16) is equivalent to

$$k - n \leq 1 + \frac{T\varepsilon}{c}. \tag{18}$$

Hence, given (17), (C_n, C_k) is an ε-equilibrium if and only if

$$\frac{k - n}{T} \leq \frac{\varepsilon}{c} + \frac{1}{T}. \tag{19}$$

Using (7), one can also show that, if both players use mixed trigger strategies, with random parameters n and k, respectively, then a *sufficient* condition for an ε-equilibrium is

$$\frac{\max(\mathrm{E}n, \mathrm{E}k) - \mathrm{E}\min(n, k)}{T} \leq \frac{\varepsilon}{c} - \frac{a + b}{Tc}. \tag{20}$$

As an alternative to the above definition of epsilon-equilibrium, one may suppose that at each date each player is satisfied with a strategy that gets within epsilon of the maximum *average utility for the rest of the game*, given the other player's strategy. The preceding analysis of this section, e.g., (19), can be directly reinterpreted for this definition, with the result that there are epsilon-equilibria in which the players cooperate when there are sufficiently many periods remaining, and then stop cooperating when the end of the game is sufficiently close.

6. Repeated games played using finite automata

Another way to model the boundedness of the players' rationality is to assume that the players are restricted to implementing their sequential strategies by means of finite automata of bounded size. This approach has recently been explored by Neyman (1985), Rubinstein (1985) and Ben-Porath (1985); I present here Neyman's simplest result.[2]

For the present purpose it suffices to define the finite automata to be used by a player (say player 1) as follows. Let $\boldsymbol{M}_1$ be a finite set, called the *set of alternative states* of the automaton. If at date t, the state of player 1's automaton is $M_1(t)$, and player 2's action is $A_2(t)$, then at date $t+1$ the state of player 1's automaton will be

$$M_1(t+1) = f_1[M_1(t), A_2(t)],$$

and player 1's action will be

$$A_1(t+1) = g_1[M_1(t+1)].$$

The quadruple $[\boldsymbol{M}_1, f_1, g_1, M_1(1)]$ constitutes player 1's automaton, where f_1 is a function from $\boldsymbol{M}_1 \times \{0,1\}$ into $\boldsymbol{M}_1$ (called the *next-state function*), g is a function from $\boldsymbol{M}_1$ into $\{0,1\}$ (called the *output function*), and $M_1(1)$ in $\boldsymbol{M}_1$ is the *initial state*. The number of elements in $\boldsymbol{M}_1$ is called the *size* of the automaton.

An automaton implements a strategy for a player, and any strategy for a T-period game can be implemented by a finite automaton. On the other hand, if T is too large compared to m, there will be strategies of the T-period game that cannot be implemented by an automaton of size m. For example, to implement a trigger strategy C_k, with $k < T$, an automaton must be able to count from 1 to k, and hence must have at least k states. On the other hand, the trigger strategy C_T is implementable by a two-state automaton as follows:

$$\begin{aligned} &\boldsymbol{M} = \{0,1\}, && M(1) = 1, \\ &f(1,1) = 1, && f(M,s) = 0 \quad \text{otherwise}, \\ &g(1) = 1, && g(0) = 0. \end{aligned}$$

[2] The suggestion that bounded rationality be modelled by finite automaton theory was made by Marschak and McGuire (1971) in unpublished lecture notes. An early exploration of this in the context of the Prisoners' Dilemma is reported in Aumann (1981).

Note that if a strategy is implementable by an automaton of size m', then for any $m > m'$ it is also implementable by an automaton of size m.

I shall say that a pair of strategies is an *automaton-equilibrium of order m* if (1) each of the two strategies is implementable by an automaton of size m, and (2) neither player can increase his total utility by unilaterally changing to another strategy that is also implementable by an automaton of size m. I shall now show:

Proposition (Neyman). If

$$2 \leq m \leq T - \frac{b-c}{c},$$

then the strategy pair (C_T, C_T) is an automaton-equilibrium of size m.

To prove this result, first note that player 1 can improve on C_T (against C_T) only by playing 0 at some date n. After playing 0 at date n, the best thing player 1 can do is to play 0 until the end of the game. The resulting gain to player 1 (relative to using C_T) is

$$(b-c) - (T-n)c,$$

which is strictly positive if and only if

$$n > T - \frac{b-c}{c}. \tag{21}$$

Hence a strategy for player 1 is better against C_T than C_T is only if player 1 plays 0 for the first time at a date n that satisfies (21).

Second, note that in order to implement a strategy in which player 1 plays 0 for the first time at date n (against C_T), player 1's automaton must have at least n elements. Suppose to the contrary that the automaton had m elements, with $m < n$. Then a state will be repeated by date $(m + 1)$ at the latest. Recall that as long as player 1 plays 1, player 2 will also play 1. Hence once a state of player 1's automaton were repeated, player 1 would play 1 for the rest of the game, a contradiction. The proposition now follows from (21) and the fact that C_T can be implemented by a two-state automaton.

Rubinstein (1984) also considers strategies that can be implemented by finite automata, but otherwise his assumptions and results are rather different. There are infinitely many repetitions, and each player cares lexicographically about his average payoff and the number of states of his automaton that are used infinitely often. A key assumption is that states that are not used in the actual play of the game are omitted from the automaton. He finds that, roughly speaking, all "equilibria" in this model are Pareto inefficient.

Infinitely repeated games with strategies implemented by finite automata have also been studied by Ben-Porath (1985).

7. Some interpretations

In the Prisoners' Dilemma game, repeated finitely many times, the "rationality" of the non-cooperative equilibrium seems compelling because the game does not explicitly provide any mechanism for enforcing "cooperative" agreements. On the other hand, there is another "rationality" in the cooperative outcome, because it is Pareto superior. Each player might well be uncertain about which kind of rationality the other player will adopt, or expect. (In another terminology, each player may be uncertain about how "trustworthy" the other player is.) In the family of trigger strategies, C_k, studied in Sections 3 and 4, the parameter k may be interpreted as a degree of cooperativeness, with $k = 0$ corresponding to completely non-cooperative behavior, and $k = T$ to completely cooperative behavior. Thus, we seem to have a genuine conflict between two modes of rationality, which (to my knowledge) is not resolved by any more general theory of rationality. In this case, it is natural to regard a player's degree of cooperativeness as a "behavioral" parameter, about which the other player might be uncertain.

Uncertainty about the kind of rationality the other player will adopt is different from uncertainty about the other data of the game, e.g., about the other player's payoffs. Kreps, Milgrom, Roberts and Wilson (1982) have shown that with the latter kind of uncertainty there can be equilibria of the finitely repeated game that result in cooperative outcomes. Note, however, that throughout this paper I have assumed that the data of the game in Section 2, and the number of repetitions, constitute common knowledge of the two players.

The concept of epsilon-equilibrium (Section 5) provides a formalization of a different kind of bounded rationality. In this case, each player is satisfied to get "close" to an optimal response to the other player's strategy, without actually achieving a complete optimum. In as simple a game as the Prisoners' Dilemma, it may be difficult to see why the players would behave this way if they were really imbued with the ideal of Cournot–Nash rationality. However, in a very complex game the costs and difficulties of computation, information processing, etc. might well lead the player to fall short of optimal responses. [For applications of the epsilon-equilibrium concept in the context of repeated games, see Radner (1980) on oligopoly theory, and Radner (1981) on principal–agent theory.]

The "complexity" of a strategy is modelled in an explicit way in Section 6, using the model of a finite automaton. It is assumed that the players can only use strategies that can be implemented by finite automata, and the boundedness of rationality is expressed by a bound on the size of the automaton, i.e., the number of internal states. Although this is an intriguing idea, one may question whether the size of an automaton is – in an economic or psychological sense – the most appropriate measure of the complexity of the corresponding strategy. For example, this measure does not take account of the "complexity" of the two functions (the next-state function and the output function) that also form part of the characterization of the automaton. Thus more refined measures of complexity may be useful in pursuing this new and important direction of research.

References

Aumann, R.J., 1981, Survey of repeated games, in: Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern (Bibliographisches Institut, Mannheim/Wien/Zurich) 11–42.

Ben-Porath, E., 1985, Repeated games with bounded complexity, Conference on Repeated Games, Israel Academy of Sciences, Jerusalem, Unpublished.

Kreps, D., P. Milgrom, J. Roberts and R. Wilson, 1982, Rational cooperation in the finitely repeated Prisoners' Dilemma, Journal of Economic Theory 27, 245–252.

Luce, R.D. and H. Raiffa, 1957, Games and Decisions (Wiley, New York).

Marschak, T.A. and C.B. McGuire, 1971, Economic models for organization design, April, Unpublished.

Neyman, A., 1985, Bounded complexity justifies cooperation in the finitely repeated Prisoners' Dilemma, Conference on Repeated Games, Israel Academy of Science, Jerusalem, Unpublished.

Radner, R., 1980, Collusive behavior in noncooperative epsilon-equilibria in oligopolies with long but finite lives, Journal of Economic Theory 22, 136–154.

Radner, R., 1981, Monitoring cooperative agreements in a repeated principal–agent relationship, Econometrica 49, 1127–1148.

Rubinstein, A., 1984, Finite automata play the repeated Prisoners' Dilemma, The Journal of Economic Theory, forthcoming.