

Time Dependent Bounded Recall Strategies Are Enough to Play the Discounted Repeated Prisoners' Dilemma*

Mehmet Barlo
Sabancı University

Guilherme Carmona
Universidade Nova de Lisboa

September 30, 2006

Abstract

We show that for any discount factor, there is a natural number M such that all subgame perfect outcomes of the discounted repeated prisoners' dilemma (pure strategies) can be obtained by subgame perfect strategies with the following property: current play depends only on the number of the time-index and on the history of the last M periods. This strong result, however, does not suffice to establish the claim of the title of this study, because we also establish that for any natural number K , there exist a discount factor and an equilibrium outcome which can only be obtained by equilibrium strategies with recall strictly higher than K . On the other hand, for cases with high discount factors, we prove that payoffs that are bounded away from the common minmax return can be approximated in equilibrium with a bounded recall strategy whose order of recall is independent of the discount factor and the fineness of the approximation. Therefore, noticing that payoffs not bounded away from the minmax return can each be approximated by the minmax payoff (sustainable in equilibrium with 0-recall) enables us to conclude that bounded recall strategies are enough to play the prisoners' dilemma.

Journal of Economic Literature Classification Numbers: C72; C73; C79

Keywords: Repeated Prisoners' Dilemma; Memory; Bounded Rationality

Correspondent: Mehmet Barlo, FASS, Sabancı University Orhanlı, 34956 Tuzla Istanbul Turkey;

Email: barlo@sabanciuniv.edu; telephone: +(90) 216 483 9284; fax: +(90) 216 483 9250.

*We are grateful to Narayana Kocherlakota, and Andy McLennan for their valuable advice and support. Special thanks are due to Efe Ok and Hamid Sabourian for very detailed comments. We benefited from discussions in, and thank the participants of the Applied Microeconomics Workshop at the University of Minnesota, the Mathematical Economics Workshop at the Bilkent University, the Murat Sertel Memorial Conference on Economic Theory 2004 and the ASSET 2004 Conference. We thank John Hufstot for editorial assistance. Any errors that may remain are ours.

1 Introduction

Repeated games provide a framework in which long-term relationships can be analyzed. In particular, the repeated prisoners' dilemma is used to study whether or not it is possible for two individuals to cooperate even when they have a short-term incentive for not doing so. The well-known answer displays that if players are sufficiently patient, then cooperation in every period will be a (subgame perfect) equilibrium outcome. But, because players do not possess a short-term incentive to cooperate, in any equilibrium strategy that results in cooperation, current play must depend on past play.

In this paper, we ask how much players have to remember in order to play the discounted repeated prisoners' dilemma. In our view, this question is important inasmuch as we regard strategies that depend less on memory as being more attractive to players, which will be the case whenever remembering the past has costs.

We show that time dependent bounded recall strategies are enough to play the discounted repeated prisoners' dilemma with pure strategies. More precisely, we prove that for any discount factor strictly less than 1 there is a natural number M such that all equilibrium outcomes can be obtained by equilibrium strategies in which current play depends only on the number of the time-index and on what has transpired in the last M rounds of play. Thus, no equilibrium outcome requires an unbounded time dependent recall.¹

If conceivable discount factors are bounded away from 1, a stronger conclusion holds: There exists a limit on recall, M , with which M -recall strategies would be enough to play the prisoners' dilemma regardless of the particular value of the discount factor.² This result is significant because in the literature on repeated games (with pure strategies) not much is known about the equilibria when the discount factor is bounded away from 1. On the other hand, when the discount factor is not bounded away from 1, our first result cannot justify this stronger conclusion. This is because we also establish that for any natural number K , there exists a sufficiently high discount factor and an equilibrium outcome that can only be obtained by an equilibrium strategy with recall strictly greater than K .

¹Throughout this paper we will make a distinction between the terms "finite memory" and "bounded recall". Indeed, the former will denote the standard memory notion where the behavior is independent of the time-index.

²Alternatively, given M , our techniques can also be utilized to characterize the set of discount factors with which all equilibrium outcomes can be obtained with M -recall.

In order to analyze cases with high discount factors, we prove that bounded recall suffices to obtain the set of strictly individually rational payoffs that are bounded away from the common minmax return.³ In particular, for any $\varepsilon > 0$ there exists a natural number M_ε such that for any strictly individually rational payoff bounded away by ε from the minmax return, there is a sufficiently high discount factor with which this payoff can be arbitrarily approximated employing an M_ε -recall equilibrium strategy. Moreover, clearly any strictly individually rational payoff not bounded away by ε from the minmax return can each be ε -approximated by the minmax payoff itself, hence, requiring no recall. Consequently, our Theorems imply that time dependent bounded recall is enough to play the discounted repeated prisoners' dilemma.⁴

Although we have been able to obtain a strong result for the repeated prisoners' dilemma, it appears that our tools are not easily extended to other games. Indeed, in section 5 we identify a discounted repeated game in which bounded recall is not enough to support all equilibrium outcomes. Part of the difficulty of generalizing our main result to other games in the framework of this paper comes from restricting players to play pure strategies. In fact, using one-memory strategies, Barlo, Carmona, and Sabourian (2006) and Sabourian (1989) provide general characterizations for equilibrium payoffs when players have connected action spaces.

Prisoners' dilemma has been the focus of analysis in the complexity literature: Aumann (1981), following Simon's bounded rationality ideas, suggested that a strategy is more intuitive if in every period the behavior depends on a finite number of states, i.e., if it is implemented by a finite automata. Aumann's suggestion was then followed by Neyman (1985) and Rubinstein (1986), who pioneered the analysis of complexity in the prisoners' dilemma. Aumann (1981) also introduced bounded memory strategies as a special class of finite automata strategies, a concept first investigated by Kalai and Stanford (1988), Lehrer (1988) and Aumann and Sorin (1989). We refer the reader to Kalai (1990) for more details. In this study, we do not employ finite automata strategies to model limited memory. Instead, time dependent bounded recall separates knowledge of the time of

³A similar result has been obtained by Sabourian (1998) with the limits of the means payoff criterion and pure strategies: Any strictly individually rational payoff obtainable by a stationary outcome can be approximated with finite memory strategies. This important result, however, does not perform well in the analysis of the prisoners' dilemma, because the only strictly individually rational and stationary payoff is that of cooperative play which is sustainable in equilibrium with one-memory.

⁴It should be pointed out that all our conclusions hold whenever the stage game is an N -person normal form game, $N \in \mathbb{N}$, in which there is a strictly dominant common minmax action. The reason we have chosen to consider a specific prisoner's dilemma is to abstract from non-fruitful technicalities, and hence provide an easier reading. Indeed, such a generalization is conceptually straightforward, yet, cumbersome in notation.

play from that regarding the past moves, a notion used in Barlo and Carmona (2003) and Cole and Kocherlakota (2005). Consequently, even a time dependent zero-recall strategy might be of infinite complexity, which is due to the action plan varying across time. On the other hand, we argue that in situations in which agents can observe and use the time-index with no costs, time dependent bounded recall presents an appropriate notion of limited memory (see Barlo and Carmona (2003) for a more elaborate discussion).

2 Notation and Definitions

The prisoners' dilemma is described as a normal form game G with two players ($N = \{1, 2\}$), each of whom have two actions: $A_i = \{C, D\}$ for $i = 1, 2$. Players' payoff functions are described by the following table:

$1 \backslash 2$	C	D
C	3, 3	0, 4
D	4, 0	1, 1

We denote player i 's payoff function by $u_i : A \rightarrow \mathbb{R}$, for $i = 1, 2$, and $A = A_1 \times A_2$.

Let $\delta \in (0, 1)$ be the common discount factor. The *supergame* of G consists of an infinite sequence of repetitions of G taking place in periods $t = 1, 2, 3, \dots$. In period t the players make simultaneous moves denoted by $a_t^i \in \{C, D\}$ and then each player learns his opponent's move. We assume that players have complete information. For $k \geq 1$, a k -stage history is a k -length sequence $h_k = (a_1, \dots, a_k)$, where, for all $1 \leq t \leq k$, $a_t \in A$; the space of all k -stage histories is H_k , i.e., $H_k = A^k$ (the k -fold Cartesian product of A). We use e for the unique 0-stage history — it is a 0-length history that represents the beginning of the supergame. The set of all histories is defined by $H = \bigcup_{n=0}^{\infty} H_n$. For every $h \in H$, define $h^r \in A$ to be the projection of h onto its r^{th} coordinate. For every $h \in H$ we let $\ell(h)$ denote the *length of h* . For two positive length histories h and \bar{h} in H we define the *concatenation of h and \bar{h}* , in that order, to be the history $(h \cdot \bar{h})$ of length $\ell(h) + \ell(\bar{h})$: $(h \cdot \bar{h}) = (h^1, h^2, \dots, h^{\ell(h)}, \bar{h}^1, \bar{h}^2, \dots, \bar{h}^{\ell(\bar{h})})$. We follow the convention that $e \cdot h = h \cdot e = h$ for every $h \in H$. For a history $h \in H$ and an integer $0 \leq m \leq \ell(h) - 1$, the m -stage tail of h is denoted by $T^m(h) \in H : T^0(h) = e$ and $(T^m(h))^j = h^{\ell(h) - (m+1) + j}$ for $j = 1, 2, \dots, m$ and $1 \leq m \leq \ell(h) - 1$. We also follow the convention that $T^m(h) = h$, for all $m \geq \ell(h)$.

It is assumed that at stage k each player knows h_k , that is, each player knows the actions that were played in all previous stages. As in Kalai and Stanford (1988), limited recall will be modeled by restricting the strategies that players are allowed to use, and not agents' knowledge of the history of the game.

For all $i \in N$, a *strategy* for player i is a function $f_i : H \rightarrow A_i$ mapping histories into actions. The set of player i 's strategies is denoted by F_i , and $F = \prod_{i \in N} F_i$ is the joint strategy space. Finally, a strategy vector is $f = (f_1, \dots, f_n)$. Given an individual strategy $f_i \in F_i$ and a history $h \in H$ we denote the *individual strategy induced at h* by $f_i|_h$. This strategy is defined pointwise on H : $(f_i|_h)(\bar{h}) = f_i(h \cdot \bar{h})$, for every $\bar{h} \in H$. We will use $(f|h)$ to denote $(f_1|_h, \dots, f_n|_h)$ for every $f \in S$ and $h \in H$. We let $F_i(f_i) = \{f_i|_h : h \in H\}$ and $F(f) = \{f|h : h \in H\}$.

Given a strategy of player i , $f_i \in F_i$, we say that f_i has *time dependent recall of order m* , $\text{rec}(f_i) = m$, if m is the smallest integer satisfying the property: $f_i(h) = f_i(\bar{h})$ for all $k \in \mathbb{N}$ and all $h, \bar{h} \in H_k$ with $T^m(h) = T^m(\bar{h})$. If such an m does not exist, we say that $\text{rec}(f_i) = \infty$. We let F_i^m be the set of all player i 's strategies with recall of order at most m , and $F^m = \prod_{i \in N} F_i^m$.

A strategy $f \in F$ induces an outcome $\pi(f)$ as follows: $\pi^1(f) = f(e)$, and $\pi^k(f) = f((\pi^t(f))_{t=1}^{k-1})$, for $k = 2, 3, \dots$. Thus, we have defined a function $\pi : F \rightarrow A^\infty$, where $A^\infty = A \times A \times \dots$. The payoff of player i in the supergame of G is, for $\delta \in (0, 1)$, the discounted sum of stage game payoffs:

$$U_i(f) = (1 - \delta) \sum_{k=1}^{\infty} \delta^{k-1} u_i(\pi^k(f)), \quad i = 1, 2.$$

Let $V_i : A^\infty \rightarrow \mathbb{R}$ be defined by $V_i(a^\infty) = (1 - \delta) \sum_{k=1}^{\infty} \delta^{k-1} u_i(a^k)$; then $U_i = V_i \circ \pi$.

A strategy vector $f \in F$ is a *Nash equilibrium* of the supergame of G if for all $i \in N$, $U_i(f) \geq U_i(\hat{f}_i, f_{-i})$ for all $\hat{f}_i \in F_i$. A strategy vector $f \in F$ is a *subgame perfect equilibrium* of the supergame of G if every $\bar{f} \in F(f)$ is a Nash equilibrium. We let $SPE(\delta)$ denote the set of subgame perfect equilibria and $E(\delta) = \pi(SPE(\delta))$ denote the *set of subgame perfect equilibrium outcomes*. Also, we let $E^m(\delta) = \pi(SPE(\delta) \cap F^m)$ be the set of subgame perfect equilibrium outcomes that can be obtained by subgame perfect equilibrium strategies with recall no greater than m .

3 Subgame Perfect Equilibria

Theorem 1 characterizes the set of subgame perfect outcomes in terms of outcomes obtainable by bounded recall strategies. Indeed, an outcome is subgame perfect if and only if it can be obtained by a subgame perfect equilibrium strategy with recall no greater than M , for some $M \in \mathbb{N}$.

Theorem 1 *For every $\delta \in [0, 1)$, there exists an $M \in \mathbb{N}$ such that $E(\delta) = E^M(\delta)$.*

The main idea of the proof of Theorem 1 is as follows. Consider any $\pi \in E(\delta)$. The M -recall strategy which we show to be an equilibrium inducing π is: If the last M periods of a given history agree with π , then player i chooses as specified by the outcome. Otherwise, player i chooses D .

This construction requires a player to distinguish, by remembering finite number of periods, if there were any deviations from the given equilibrium outcome π . Hence, first we have to analyze equilibrium outcomes which involve unbounded sequences of consecutive (D, D) s, because clearly such a path cannot be distinguished from a punishment path using finite recall.

Lemma 1 below displays that: (1) If from some period on, (D, D) will be played for the remainder of the game, then that period must be the first period of the game; and (2) if an outcome π consists of $k \in \mathbb{N}$ consecutive (D, D) s followed by another outcome $\tilde{\pi}$ supported by an equilibrium strategy \tilde{f} , then there exists an equilibrium strategy f supporting π and $\text{rec}(f) = \text{rec}(\tilde{f})$. It should be pointed out that the outcome consisting of repeating the Nash equilibrium of the stage game does not require any recall. Then, in Lemma 2 we show that for outcomes different from those covered in Lemma 1, no equilibrium path will entail unbounded sequences of consecutive (D, D) s.

These properties are necessary for the strategy we use in the proof, since they guarantee that players can then conclude as to whether they are in the equilibrium path or in the punishment path. But they are not enough. The reason is that our construction critically depends on the following property: after a deviation from the given equilibrium outcome π , the behavior induced by the strategy we employ should result in a defection path for the remainder of the game. The example below is designed to demonstrate this point: Assume that the discount factor is sufficiently high, so that the following path is a subgame perfect outcome:

$$\pi = ((C, D), (D, D), (D, D), (C, D), (C, D), (C, C), (C, C), \dots, (C, C), \dots).$$

In this outcome, the longest consecutive sequence of (D, D) s is 2. Suppose that the recall figure to be used is 4, strictly more than enough to allow players to distinguish the equilibrium path from the punishment path. Assume that player 1 plays D in the first period. Hence, we wish the resulting outcome to be given by repeating (D, D) forever. But now consider player 1 deviating in the punishment phase and playing C in periods 4 and 5. Then the outcome of the first 5 periods is $((D, D), (D, D), (D, D), (C, D), (C, D))$. Since the history of the last 4 periods coincides with the

equilibrium path, players will play (C, C) forever. Hence, a unilateral deviation does not necessarily lead to (D, D) forever.

This problem is easily solved. The length of the longest consecutive sequence of D s played by any of the players in this outcome is 5. Then, one can easily show that players need to remember at most 6 periods. In general, if the length of the longest consecutive sequence of D s played by any of the players in an outcome is finite, then this outcome can be implemented with a time dependent bounded recall equilibrium. Our result then follows from Lemma 3 below, which proves that for outcomes not covered by Lemma 1, no equilibrium path will entail unbounded sequences of consecutive D s for any of the players. Thus, confusing instances such as the one above are eliminated by setting the recall size high enough to ensure that a deviating player cannot confuse the other.

Proof of Theorem 1. Our proof will employ three Lemmas:

Lemma 1 *Let $\pi \in E(\delta)$.*

1. *If for some $k \in \mathbb{N}$,*

$$\pi^t = (D, D) \quad \text{for all } t \geq k, \quad (1)$$

then, $\pi^t = (D, D)$ for all $t \in \mathbb{N}$.

2. *If for some $k \in \mathbb{N}$, and $\tilde{\pi} \in E(\delta)$,*

$$\pi^t = \begin{cases} (D, D) & \text{if } t \leq k, \\ \tilde{\pi}^{t-k} & \text{otherwise,} \end{cases} \quad (2)$$

and \tilde{f} is an equilibrium strategy such that $\pi(\tilde{f}) = \tilde{\pi}$, then there exists an equilibrium strategy f such that $\pi(f) = \pi$, and $\text{rec}(f) = \text{rec}(\tilde{f})$.

Proof of Lemma 1. Let $\pi \in E(\delta)$ satisfying condition (1) of Lemma 1. We may assume that k is the minimum $k' \in \mathbb{N}$ such that $t \geq k'$ implies $\pi^t = (D, D)$. If $k > 1$, there exists a player that plays C in period $k - 1 \in \mathbb{N}$. Clearly, this player has an incentive to deviate since the deviation induces the same outcome $((D, D) \text{ forever})$ that would occur had he not deviated.

Finally, for the second result define f_i as follows: for all $h \in H_n$, $f_i(h)$ equals D if $t < k$; and $\tilde{f}(T^{t-k}(h))$ otherwise. The result then follows immediately because in the first k periods players are playing a Nash equilibrium and so they do not need to remember the past. ■

Since the outcome consisting of perpetual (D, D) is obtained by the 0 – recall equilibrium strategy where players play defect in every period independent from the history, we need only

consider equilibrium outcomes that do not satisfy conditions (1) and (2). Let $\tilde{E}(\delta) \subset E(\delta)$ be the set of equilibrium outcomes that do not satisfy conditions (1) and (2) of Lemma 1. For $\pi \in \tilde{E}(\delta)$, let k_1 be the first period in which players play (D, D) ; if it does not exist, we let $k_1 = 0$. Let K_1 be the maximal $k \in \mathbb{N}$ satisfying $\pi^t = (D, D)$ for all $k_1 \leq t \leq k$, unless $k_1 = 0$, in which case we let $K_1 = 0$. We refer to the (finite) sequence $\{\pi^{k_1}, \dots, \pi^{K_1}\}$ as the *first sequence of consecutive (D, D) s under π* . Assuming that k_1, k_2, \dots, k_{n-1} and K_1, K_2, \dots, K_{n-1} have been defined, we let k_n be the first period after K_{n-1} in which players play (D, D) ; again, if it does not exist, we let $k_n = 0$. In this latter case let $K_n = 0$, while if $k_n > 0$ we let K_n be the maximal $k \in \mathbb{N}$ satisfying $\pi^t = (D, D)$ for all $k_n \leq t \leq k$. The sequence $\{\pi^{k_n}, \dots, \pi^{K_n}\}$ is called the *n^{th} sequence of consecutive (D, D) s under π* . Finally, let $M_n = M_n(\pi)$ be the *length of n^{th} sequence of consecutive (D, D) s under π* , i.e., $M_n = K_n - k_n + 1$. Note that either $k_1 = 0$, in which case $M_n(\pi) = 1$ for all $n \in \mathbb{N}$, or $k_1 > 1$.

Lemma 2 *There exists $B \in \mathbb{N}$ such that $\sup_{n \in \mathbb{N}} M_n(\pi) \leq B$ for all $\pi \in \tilde{E}(\delta)$.*

Proof. Suppose not. Then there exists $\pi \in \tilde{E}(\delta)$, and $n \in \mathbb{N}$ such that $1 - \delta > 4\delta^{M_n(\pi)+1}$ and $k_n > 1$. By the definitions of K_{n-1} and k_n , it follows that in period $k_n - 1$ the outcome is different from (D, D) . Let $i \in N$ be such that $\pi_i^{k_n-1} = C$. Suppose that player i deviates and plays D from period $k_n - 1$ onwards, giving rise to an outcome $\tilde{\pi}$. Note that we have $u_i(\tilde{\pi}^{k_n-1}) - u_i(\pi^{k_n-1}) = 1$, $u_i(\tilde{\pi}^k) - u_i(\pi^k) \geq 0$ for $k_n \leq k \leq K_n$ and $u_i(\tilde{\pi}^k) - u_i(\pi^k) \geq -4$ for $k > K_n$. Hence, $V_i(\tilde{\pi}) - V_i(\pi) \geq \delta^{k_n-2}((1 - \delta) - 4\delta^{M_n+1}) > 0$. Thus, player i has an incentive to deviate in period $k_n - 1$ from π , which contradicts the assumption that $\pi \in E(\delta)$. ■

Let $i \in \{1, 2\}$ and $\pi \in \tilde{E}(\delta)$; we will define the length of the n^{th} sequence of consecutive D s played by player i under π . Let k_1^i be the first period in which player i plays D ; if it does not exist, we let $k_1^i = 0$. Let K_1^i be the maximal $k \in \mathbb{N}$ satisfying $\pi_i^t = D$ for all $k_1^i \leq t \leq k$, unless $k_1^i = 0$, in which case we let $K_1^i = 0$. Assuming that $k_1^i, k_2^i, \dots, k_{n-1}^i$ and $K_1^i, K_2^i, \dots, K_{n-1}^i$ have been defined, we let k_n^i be the first period after K_{n-1}^i in which player i plays D ; again, if it does not exist, we let $k_n^i = 0$. In this latter case let $K_n^i = 0$, while if $k_n^i > 0$ we let K_n^i be the maximal $k \in \mathbb{N}$ satisfying $\pi_i^t = D$ for all $k_n^i \leq t \leq k$. Finally, let $M_n^i = M_n^i(\pi)$ be the *length of n^{th} sequence of consecutive D s played by player i under π* , i.e., $M_n^i = K_n^i - k_n^i + 1$.

Lemma 3 *For every $i = 1, 2$, there exists $B^i \in \mathbb{N}$ such that $\sup_{n \in \mathbb{N}} M_n^i(\pi) \leq B^i$ for all $\pi \in \tilde{E}(\delta)$.*

Proof. Suppose not. Let B be given by Lemma 2. Then there exists $i \in \{1, 2\}$, $\pi \in \tilde{E}(\delta)$, and $n \in \mathbb{N}$ such that $1 - \delta > 4\delta^{M_n^i(\pi)-B}$ and $M_n^i(\pi) > B$. This last condition implies that there exists

$k_n^i \leq k \leq K_n^i$ such that $\pi^k \neq (D, D)$ and so, $\pi_{-i}^k = C$. In fact, $k \leq k_n^i + B$. Suppose that player $-i$ deviates and plays D from period k onwards, giving rise to an outcome $\tilde{\pi}$. Note that we have $u_i(\tilde{\pi}^k) - u_i(\pi^k) = 1$, $u_i(\tilde{\pi}^t) - u_i(\pi^t) \geq 0$ for $k \leq t \leq K_n^i$ and $u_i(\tilde{\pi}^t) - u_i(\pi^t) \geq -4$ for $t > K_n^i$. Hence, $V_i(\tilde{\pi}) - V_i(\pi) \geq \delta^{k-1}((1 - \delta) - 4\delta^{K_n^i - k + 1}) \geq \delta^{k-1}((1 - \delta) - 4\delta^{M_n^i - B}) > 0$. Thus, player $-i$ has an incentive to deviate in period k from π , which contradicts the assumption that $\pi \in E(\delta)$. ■

We finally show that all equilibrium outcomes in $\tilde{E}(\delta)$ can be obtained by a bounded recall equilibrium strategy. Let $\pi \in \tilde{E}(\delta)$. Let B, B^1, B^2 be as in Lemmas 2 and 3, and let $B^* = \max\{B, B^1, B^2\}$. Thus, we have that M_n, M_n^1 and M_n^2 are all less than or equal to B^* for all $n \in \mathbb{N}$. Define f_i as follows: start by playing according to the path π ; i.e., $f_i(e) = \pi^1$. For an arbitrary $h \in H_k, k \in \mathbb{N}$, let:

$$f_i(h) = \begin{cases} \pi_i^{k+1} & \text{if } T^{B^*+1}(h) = (\pi_{k-B}, \dots, \pi_k) \\ D & \text{otherwise.} \end{cases}$$

This definition implies that if player i deviates unilaterally from π in period k , then (D, D) will be played in all periods $t \geq k + 1$. This follows because player $-i$ will be able to observe the deviation on the first $B^* + 1$ periods after it has occurred, which will lead to a sequence in which he plays D for $B^* + 1$ periods; after that, since along the equilibrium path player $-i$ plays at most B^* consecutive D s, it follows that the $(B^* + 1)$ -stage tail of the history following the deviation can never be consistent with equilibrium play; thus, he keeps playing D forever.

Obviously, $\text{rec}(f) = B^* + 1$ and $\pi(f) = \pi$. Furthermore, since $\pi \in E(\delta)$, then f is a subgame perfect equilibrium. Finally, setting $M = B^* + 1$ finishes the proof of Theorem 1. ■

4 High Discount Factor

The analysis of players' behavior when constrained to use at most M -recall strategies is intriguing. Theorem 1 provides a partial answer to this question: If there is an upper bound on the set of discount factors expected, then there exists a level of recall M , so that any equilibrium outcome can be obtained with equilibrium strategies with recall less than or equal to M .

However, when the set of plausible discount factors cannot be restricted, and thus, a discount factor not bounded away from 1 can be expected, Theorem 1 cannot be employed to identify an upper bound for the recall requirement. Indeed, the next Theorem shows that for any natural number K , there exists a discount factor and equilibrium outcome which can be supported only by equilibrium strategies requiring strictly more than K recall.

For all $\delta \in [0, 1)$, let $M(\delta) = \min\{M \in \mathbb{N} : E(\delta) = E^M(\delta)\}$, and note that it is well defined in the light of Theorem 1.

Theorem 2 $\lim_{\delta \rightarrow 1} M(\delta) = \infty$.

Proof. Let $\delta \in (0, 1)$ be such that $\delta \geq 1/3$. We claim that $M(\delta) > \ln[(1 - \delta)/2]/\ln \delta - 1$, from which the conclusion follows.

Let M be the greatest integer smaller or equal to $\ln[(1 - \delta)/2]/\ln \delta$. In particular,

$$\frac{\ln[(1 - \delta)/2]}{\ln \delta} - 1 < M \leq \frac{\ln[(1 - \delta)/2]}{\ln \delta}. \quad (3)$$

Define an outcome π as follows:

$$\pi^t = \begin{cases} (C, C) & \text{if } t = 1, \\ (D, D) & \text{if } 2 \leq t \leq M, \\ (C, C) & \text{if } t \geq M + 1. \end{cases}$$

Due to $\delta \geq 1/3$, no player wants to deviate in period $n \geq M + 1$. Furthermore, no player wants to deviate in period $2 \leq n \leq M$ because in those periods he is playing a best-reply. Hence, in order to show that π is an equilibrium outcome, it is enough to show that no player wants to deviate in the first period.

Conforming with π yields a payoff of $3(1 - \delta) + (1 - \delta) \sum_{k=2}^M \delta^{k-1} + 3\delta^M$, while deviation to D in period 1 followed by (D, D) forever yields $4(1 - \delta) + (1 - \delta) \sum_{k=2}^M \delta^{k-1} + \delta^M$. Hence, the gain from conforming is $2\delta^M - (1 - \delta)$, which must be greater or equal to 0 due to (3).

The next and final step is to show that if f is an equilibrium strategy such that $\pi(f) = \pi$, then player i must play D for at least M periods after a deviation in period 1 by player $-i$. In order to prove this claim, let m be the number of times that player i plays D after the deviation. In order to reach a contradiction, assume that $m \leq M - 1$. Then if player $-i$ deviates by playing D in every period, his payoff will be at least $4(1 - \delta) + \delta(1 - \delta^{m-1}) + (1 - \delta)4\delta^m + \delta^{m+1}$, which equals $4 - 3\delta + 3\delta^m(1 - \delta)$, because in period $m + 1$ player i plays C and player $-i$ can guarantee himself a payoff of 1 in the following periods. Since $3\delta^m(1 - \delta)$ decreases with m , it follows that if a deviation is profitable when $m = M - 1$, it will be profitable for all $m < M - 1$. Thus, consider $m = M - 1$. So, (D, D) will be played in periods $2 \leq t \leq M$. Hence, conforming with π yields a payoff of $3(1 - \delta) + (1 - \delta) \sum_{k=2}^M \delta^{k-1} + 3(1 - \delta)\delta^M + 3\delta^{M+1}$, while deviation to D in period 1 by player $-i$ followed by D forever yields to him at least $4(1 - \delta) + (1 - \delta) \sum_{k=2}^M \delta^{k-1} + 4(1 - \delta)\delta^M + \delta^{M+1}$. Thus,

the gain from deviation will be at least $(1 - \delta) + (1 - \delta)\delta^M - 2\delta^{M+1} > (1 - \delta) - 2\delta^{M+1}$, which is strictly greater than 0 due to (3).

Thus, in period $M + 1$, players must know the full history (h_1, \dots, h_M) in order to distinguish whether they are in the equilibrium path or in the punishment path. Hence, $\text{rec}(f) \geq M$. ■

The negative result of Theorem 2 also holds with the limit of the means payoff criteria, because the following outcome is both subgame perfect and of infinite recall: Consider a path $\pi = (s^1, s^2, s^3, \dots)$, where s^k , $k \in \mathbb{N}$, is a loop consisting of k consecutive (C, C) s followed by k consecutive (D, D) s. But, the payoff of this path can also be obtained with a 2-recall equilibrium strategy: Cooperate in the first two periods, and every other period for which the last two periods is given by $(C, C), (D, D)$; otherwise, defect. Thus, even though the payoff of the more complicated outcome can be obtained with a recall of size 2, the outcome itself cannot be induced by any bounded recall equilibrium strategy.

Theorem 3 generalizes this observation to the case of discounting and presents a positive result in terms of payoffs: the Folk Theorem for the discounted repeated prisoner's dilemma with bounded recall strategies. It needs to be pointed out that the only addition to the standard requirements is that the strictly individually rational payoffs under consideration must be bounded away from the common minmax payoff.

Let $\mathcal{U}^\varepsilon = \{u \in \text{co}(u(A)) \mid u_i > 1 + \varepsilon \text{ for all } i \in N\}$. We refer to \mathcal{U}^ε as the set of ε -strictly individually rational payoffs. Moreover, $\mathcal{U}_{\mathbb{Q}}^\varepsilon$ denotes $u \in \mathcal{U}^\varepsilon$ such that $u = \sum_{a \in A} \lambda_a u(a)$ where $\lambda_a \in \mathbb{Q}_+$ for all $a \in A$, $\sum_{a \in A} \lambda_a = 1$, and \mathbb{Q}_+ refers to the set of non-negative rational numbers.

Theorem 3 *For all $\varepsilon > 0$, there exists $M_\varepsilon \in \mathbb{N}$ such that for all $u \in \mathcal{U}^\varepsilon$, and all $\zeta > 0$, there is $\delta^* \in [0, 1)$ with the following property: For all $\delta > \delta^*$, there exists an equilibrium strategy f with $\text{rec}(f) \leq M_\varepsilon$, and $\|U(f) - u\| < \zeta$. Moreover, if $u \in \mathcal{U}_{\mathbb{Q}}^\varepsilon$, the above conclusion holds with $\zeta = 0$.*

Theorem 3 establishes that if the strictly individually rational payoffs under analysis are bounded away from the common minmax return, then they can arbitrarily be approximated in equilibrium with a bounded recall strategy whose order of recall is independent of the discount factor and fineness of the approximation.⁵ Therefore, Theorem 3 displays that identifying the amount with which strictly individually rational payoffs under analysis are bounded away from the minmax return, suffices to manufacture an upper bound on recall. Moreover, clearly any payoff not bounded

⁵Moreover, no approximation is needed when the payoff under consideration can be obtained as a convex combination with rational multipliers.

away from the minmax return (by $\varepsilon > 0$) can be (ε) -approximated in equilibrium with 0-recall (by repeating (D, D) forever).

Consequently, Theorems 1 and 3 imply that time dependent bounded recall is enough to play the discounted repeated prisoners' dilemma.

Proof of Theorem 3. It suffices to prove that for all $\varepsilon > 0$, there exists $M_\varepsilon \in \mathbb{N}$ such that for all $u \in \mathcal{U}^\varepsilon$ and for all $\zeta > 0$, there is an outcome π consisting of repetition of a loop with no more than M_ε consecutive D s and a payoff ζ close to u . Because then, the rest of the argument would follow standard techniques introduced and employed by Fudenberg and Maskin (1985).

Let $\varepsilon > 0$, $\gamma = \varepsilon/2$, $\zeta > 0$ and $u \in \mathcal{U}^\varepsilon$. Then, for all $w \in \mathcal{U}^\gamma$ there exist $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4) \in \Delta$ such that $w = \lambda_1(1, 1) + \lambda_2(4, 0) + \lambda_3(0, 4) + \lambda_4(3, 3)$ and $(\lambda_1 + \lambda_2 + \lambda_3)/\lambda_4 < (3 - \gamma)/\gamma$. In fact, if $w_1 = w_2$, let $\lambda_2 = \lambda_3 = 0$, and so, $w_1 > 1 + \gamma$ implies that $\lambda_4 > \gamma/2$. Hence, $(1 - \lambda_4)/\lambda_4 < (2 - \gamma)/\gamma < (3 - \gamma)/\gamma$. If, instead, $w_2 > w_1$, then let $\lambda_2 = 0$ and it follows from $w_1 > 1 + \gamma$ that $\lambda_4 > \gamma/3$. This implies that $(1 - \lambda_4)/\lambda_4 < (3 - \gamma)/\gamma$. Since the case $w_1 > w_2$ is just symmetric, the proof of the above claim is completed. Define $(3 - \gamma)/\gamma = (6 - \varepsilon)/\varepsilon$ and define $M_\varepsilon = \max\{1, (6 - \varepsilon)/\varepsilon\}$. We may assume, without loss of generality, that $\zeta < \gamma$. Let v be such that $\|v - u\| < \zeta$. Then, it follows that $v \in \mathcal{U}^\gamma$. Moreover, v can be chosen so that $v = \lambda_1(1, 1) + \lambda_2(4, 0) + \lambda_3(0, 4) + \lambda_4(3, 3)$, where λ_j is a rational number for all $j = 1, 2, 3, 4$ and $(1 - \lambda_4)/\lambda_4 < M_\varepsilon$. At this stage it is important to point out that, if $u \in \mathcal{U}_Q^\varepsilon$, then we may choose λ so that $\zeta = 0$. Hence, the following also establishes the second part of the Theorem.

We proceed to the definition of a path π that consists of a repetition of a loop and yields an average payoff of v . Let q and (p_1, \dots, p_4) be such that $p_j/q = \lambda_j$ for all $j = 1, 2, 3, 4$. Then, the loop has length of q and involves p_1 occurrences of (D, D) , p_2 of (C, D) , p_3 of (D, C) , and p_4 of (C, C) . Let $\theta = (p_1 + p_2 + p_3)/p_4 = (\lambda_1 + \lambda_2 + \lambda_3)/\lambda_4$. The path π is defined as follows. If $\theta < 1$, let $\alpha = 1/\theta$, let m be the greatest integer smaller or equal to α . Then, define the loop in the following way: let $t_1 = 1$ and $t_j = t_{j-1} + m + 1$ for all $j = 1, \dots, p_1 + p_2 + p_3$ and let $\pi^t = (C, C)$ for all $t \in \{1, \dots, q\} \setminus \{t_1, \dots, t_{p_1+p_2+p_3}\}$. Finally, insert the p_1 (D, D) s, the p_2 (D, C) s and the p_3 (C, D) s arbitrarily in periods $t \in \{t_1, \dots, t_{p_1+p_2+p_3}\}$. Clearly, the number of consecutive D s for any one of the players is bounded above by 1. If $\theta \geq 1$, let $m \leq M_\varepsilon$ be the smallest integer greater or equal to θ . Then, define the loop in the following way: let $t_1 = 1$ and $t_j = t_{j-1} + m + 1$ for all $j = 1, \dots, p_4$ and let $\pi^t = (C, C)$ for all $t \in \{t_1, \dots, t_{p_4}\}$. Finally, insert the p_1 (D, D) s, the p_2 (D, C) s and the p_3 (C, D) s arbitrarily in periods $t \in \{1, \dots, q\} \setminus \{t_1, \dots, t_{p_4}\}$. Clearly, the number of consecutive D s

for any one of the players is bounded above by $m \leq M_\varepsilon$.

Since in both cases the number of consecutive D s for any one of the players is bounded above by M_ε , the outcome π can be supported by strategy with M_ε -recall as in the proof of Theorem 1 ■

5 Concluding Remarks

This paper studies a notion of limited memory in the repeated discounted prisoners' dilemma with the additional feature that agents may use time in their behavior.

Implementing equilibrium outcomes of any discounted repeated game by finite recall strategies involves the following essential aspects: First, the equilibrium outcome and its associated punishment paths have to be such that players can distinguish them by employing finite recall strategies. Second, those paths must obey the property that players can precisely identify who has deviated.

In the case of the prisoners' dilemma, the first difficulty was solved by Lemmas 1 and 2, whereas Lemma 3 deals with the second. The fact that in the prisoners' dilemma the most severe punishment path is the same for both players and has a simple structure enables us to overcome the above difficulties.⁶

The following example demonstrates a discounted infinitely repeated game in which time dependent finite recall is not enough to support all equilibrium outcomes. More precisely, in this game for a carefully chosen value of the discount factor there exists an equilibrium path which cannot be obtained by any finite recall equilibrium strategy.

$1 \setminus 2$	C	D
C	$\frac{4}{3}, 0$	$-1, -1$
D	$2, 2$	$0, 3$

Let $\bar{\delta} \in (1/3, 1/2)$ be such that

$$(1 - \bar{\delta}) \left[-1 + 2 \sum_{k=1}^{\infty} \bar{\delta}^{2^k - 1} \right] = 0. \quad (4)$$

The existence of such a $\bar{\delta}$ can be established as follows: let $h : [0, 1) \rightarrow \mathbb{R}$ be defined by $h(\delta) = (1 - \delta)(-1 + 2 \sum_{k=1}^{\infty} \delta^{2^k - 1})$; then h is continuous, $h(1/3) < 0$ and $h(1/2) > 0$ and so, by the intermediate value theorem, there exists $\bar{\delta} \in (1/3, 1/2)$ such that $h(\bar{\delta}) = 0$.

⁶That is, the most severe punishment path of each agent consists of repetitions of the unique Nash equilibrium of the stage game.

Consider the following outcome π : let $\pi^1 = (C, D)$ and for $l > 1$, let $\pi^l = (D, C)$ if $l = 2^k$, $k \in \mathbb{N}$ and $\pi^l = (D, D)$ otherwise. Player 1 and player 2 are to be punished by repeating (D, D) , and (C, C) respectively, forever. In this outcome, for any $k \geq 2$, player 1 is playing a best-reply, hence, he has no incentives to deviate from π . Moreover, in the light of equation (4) the discount factor is such that in period 1 player 1 is indifferent between conforming to π and deviating to D and then facing (D, D) forever. Moreover, it follows from equation (4) that player 2 prefers to conform to π in period 1, obtaining a payoff strictly higher than $(1 - \bar{\delta})(-1 + 2 \sum_{k=1}^{\infty} \bar{\delta}^{2^k-1}) = 0$, rather than deviating to C and then facing (C, C) forever, which would yield him a payoff of 0. Furthermore, in period 2^k , $k \in \mathbb{N}$, player 2 again prefers to conform to π rather than deviating to D and then facing (C, C) forever, since the gain of conforming is greater than $(1 - \bar{\delta})(-1 + 2 \sum_{l=1}^{\infty} \bar{\delta}^{2^l-1}) = 0$. The punishment of player 1, repeating (D, D) forever, is a subgame perfect outcome, because (D, D) is a Nash action profile in the stage game. Finally, what remains to be shown is that $\bar{\pi}$ defined by $\bar{\pi}^l = (C, C)$ for all $l \in \mathbb{N}$ is subgame perfect. In fact, player 2 is playing a best-reply in every period; by playing $\bar{\pi}$ player 1 receives $4/3$, and if he were to deviate to D and then face (D, D) forever, he obtains $2(1 - \bar{\delta})$ strictly smaller than $4/3$ due to $\bar{\delta} \in (1/3, 1/2)$. This shows that π is a subgame perfect outcome.

This example is generic in terms of payoffs but non-generic in terms of the discount factor: For all stage games sufficiently close to our example, there exists a carefully chosen discount factor with which the above conclusion holds. On the other hand, the “strategic” choice of the discount factor suggests that such cases are exceptional. This, in turn, implies that the analysis of the set of games for which finite recall is sufficient is an appealing and open question.

Our approach allows us to introduce a new notion of complexity modeled as follows: Note that for any strategy f_i , there exists a function $g_i : \Omega_i \times \mathbb{N} \rightarrow A_i$ and a function $T_i : H \rightarrow \Omega_i$ such that $f_i(h) = g_i(T_i(h), n)$ for all $h \in H_n$, all $n \in \mathbb{N}$ and all $i = 1, 2$.⁷ We then say that a strategy f_i is of *at most K -time dependent complexity* if it can be represented as above by (Ω_i, T_i, g_i) with $|\Omega_i| \leq K$. In this representation, one can imagine that each player possesses an almanac that (1) has countably many pages, each page corresponding to a time period, and (2) cannot contain more than K entries on any page. The number K then provides us partial information about the complexity of the almanac.⁸ For the strategies establishing Theorems 1 and 3, we can let $\Omega_i = A^M$ and $T_i = T^M$. In

⁷Simply let $\Omega_i = H$, T_i be the identity function and $g_i(h, n) = f_i(h)$ for all $h \in H$ and $n \in \mathbb{N}$.

⁸More precisely, we could define the complexity $\text{comp}(f_i)$ of a strategy f_i by $\text{comp}(f_i) = \inf\{K \in \mathbb{N} : f_i \text{ is of at most } K\text{-time dependent complexity}\}$.

particular, Ω_i is finite; thus, the following corollary is just a reformulation of our Theorems:

Corollary 1 *For all $\delta \in [0, 1)$, there is an $M \in \mathbb{N}$ such that for any $\pi \in E(\delta)$, there is a subgame perfect equilibrium strategy f with at most M -time dependent complexity such that $\pi(f) = \pi$.*

Furthermore, for all $\varepsilon > 0$, there exists $M_\varepsilon \in \mathbb{N}$ such that for all $u \in \mathcal{U}^\varepsilon$, and all $\zeta > 0$, there is $\delta^ \in [0, 1)$ with the following property: For all $\delta > \delta^*$, there exists an equilibrium strategy f with at most M_ε -time dependent complexity, and $\|U(f) - u\| < \zeta$. Moreover, if $u \in \mathcal{U}_Q^\varepsilon$, the above conclusion holds with $\zeta = 0$.*

References

- AUMANN, R. (1981): “Survey of Repeated Games,” in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*. Bibliographisches Institut, Mannheim.
- AUMANN, R., AND S. SORIN (1989): “Cooperation and Bounded Recall,” *Games and Economic Behavior*, 1, 5–39.
- BARLO, M., AND G. CARMONA (2003): “Time Dependent Bounded Recall Equilibria in Discounted Repeated Games,” *Sabancı University and Universidade Nova de Lisboa*.
- BARLO, M., G. CARMONA, AND H. SABOURIAN (2006): “One Memory in Repeated Games,” *Sabancı University, Universidade Nova de Lisboa, and University of Cambridge*.
- COLE, H., AND N. KOCHERLAKOTA (2005): “Finite Memory and Imperfect Monitoring,” *Games and Economic Behavior*, 53, 59–72.
- FUDENBERG, D., AND E. MASKIN (1985): “The Folk Theorem in Repeated Games with Discounting and with Incomplete Information,” *Econometrica*, 54, 533–554.
- KALAI, E. (1990): “Bounded Rationality and Strategic Complexity in Repeated Games,” in *Game Theory and Applications*, ed. by A. N. Tatsuro Ichiishi, and Y. Tauman. Academic Press, New York.
- KALAI, E., AND W. STANFORD (1988): “Finite Rationality and Interpersonal Complexity in Repeated Games,” *Econometrica*, 56, 397–410.
- LEHRER, E. (1988): “Repeated Games with Stationary Bounded Recall Strategies,” *Journal of Economic Theory*, 46, 130–144.

- NEYMAN, A. (1985): “Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner’s Dilemma,” *Economic Letters*, 9, 227–229.
- RUBINSTEIN, A. (1986): “Finite Automata Play the Repeated Prisoner’s Dilemma,” *Journal of Economic Theory*, 39, 83–96.
- SABOURIAN, H. (1989): “The Folk Theorem of Repeated Games with Bounded Recall (One-Period) Memory,” *Economic Theory Discussion Paper 143, University of Cambridge*.
- (1998): “Repeated Games with M -period Bounded Memory (Pure Strategies),” *Journal of Mathematical Economics*, 30, 1–35.