

Discussion Paper No. 1075R

**MAINTAINING A REPUTATION AGAINST  
A LONG-LIVED OPPONENT**

by

Marco Celentani

Drew Fudenberg

David K. Levine

Wolfgang Pesendorfer

December 1993, revised April 1995

# MAINTAINING A REPUTATION AGAINST A LONG-LIVED OPPONENT

(forthcoming in: *Econometrica*)

We analyze reputation in a game between a patient player 1 and a non-myopic but less patient opponent, player 2. We assume that Player 1's type is private information and that players do not directly observe each other's action but rather see an imperfect signal of it. We show that in any Nash equilibrium of the game player 1 will get almost the largest payoff consistent with player 2 choosing a best response in a finite truncation of the game. If the discount factor of player 2 is sufficiently large, then player 1 will get approximately the maximum payoff consistent with player 2 getting at least his pure strategy minmax payoff .

KEYWORDS: Repeated games, commitment, reputation, patience.

## AUTHORS' ADDRESSES

Marco Celentani  
Departamento de Economía  
Universidad Carlos III de Madrid  
Getafe (Madrid) 28903  
Spain

Drew Fudenberg  
Department of Economics  
Harvard University  
Cambridge, MA 02138  
USA

David K. Levine  
Department of Economics  
University of California, Los Angeles  
Los Angeles, CA 90024  
USA

Wolfgang Pesendorfer  
Department of Economics  
Northwestern University  
Evanston, IL 60208  
USA

# MAINTAINING A REPUTATION AGAINST A LONG-LIVED OPPONENT

BY MARCO CELENTANI, DREW FUDENBERG, DAVID K. LEVINE, AND WOLFGANG

PESENDORFER<sup>1</sup>

## 1. INTRODUCTION

We consider a game between a patient player 1 and a non-myopic but less patient opponent, player 2. As usual in reputation models, we suppose that the patient player's type is private information, and that he may be a "commitment type" who is locked into playing a particular strategy. We investigate the extent to which an uncommitted or "normal" type of patient player can exploit his less patient opponent's uncertainty to maintain a reputation for playing like a commitment type.

Most previous work on reputation effects has supposed that player 2 is in fact completely myopic, or equivalently that player 2 corresponds to a sequence of short run players (see Kreps and Wilson (1982), Milgrom and Roberts (1982), Fudenberg and Levine (1989, 1992)).<sup>2</sup> Since a myopic player 2 will play a short-run best response in each period to that period's expected play, the best possible commitment for the long run player is to the Stackelberg strategy for the corresponding static game.

This paper considers the case in which the patient player, player 1, has discount factor  $\delta_1$  near 1, while his opponent, player 2, is an infinite-lived player who discounts future payoffs with a smaller discount factor  $\delta_2$ . Perhaps the best way to interpret this assumption of unequal discount factors is to interpret the model as a shorthand for a situation where player 1 faces a large number of identical player 2's, each of whom observe all previous play. Under one interpretation the player 2's alternate play so that for example

the first player 2 moves in periods 1,  $N+1$ , etc. Alternatively, the first player 2 might move in periods 1 to  $N$  the second in  $N+1$  to  $2N$  and so on. Under either interpretation, the key is that player 1 cares more about future payoffs of this game than player 2 does, because he will be playing in more future periods.<sup>3</sup>

A game with a non-myopic opponent differs from one with a myopic opponent in two main ways. First, because a non-myopic opponent cares about future payoffs, the static Stackelberg strategy is no longer necessarily the best possible commitment: Higher payoffs can sometimes be attained by the use of rewards and punishments. In the prisoner's dilemma, for example, "tit-for-tat" is a better commitment against a non-myopic opponent than the Stackelberg strategy of defecting. Second, it may be difficult to demonstrate that one is using a strategy with rewards and punishments unless these rewards and punishments are occasionally carried out. This is similar to the way in which incorrect off-path beliefs can weaken reputation effects in the play of extensive-form games against myopic opponents (Fudenberg and Levine (1989)).<sup>4</sup>

Our main assumption is that player 1 does not observe player 2's intended action, but only sees an imperfect signal of it, as in a model of moral hazard. We assume that the support of the distribution of signals is independent of how player 2 plays. Intuitively, this ensures that every reward and punishment in player 1's strategy will occasionally be triggered, so that player 2 will learn how player 1 responds to deviations. As a result, player 1's equilibrium payoff is bounded below by what he could get through commitment in the repeated game. In particular, if player 2 is sufficiently patient, player 1 gets approximately the greatest feasible payoff consistent with individual rationality of player 2.

This conclusion holds with an arbitrarily small amount of noise. However, as the amount of noise shrinks, the patient player's discount factor must be increasingly close to one to ensure that its Nash equilibrium payoff is close to its limit value. Consequently, our result is of the most relevance when the amount of noise is "significant."

The first general study of reputation with two non-myopic players is Schmidt (1993), who studied perfect observability. He showed that the long-run player can guarantee at least the payoff he would get from precommitment to a static strategy that minmaxes his opponent. This is a good bound in some games, but in others, such as the prisoner's dilemma, it imposes no restrictions beyond those implied by individual rationality. Subsequently Cripps, Schmidt and Thomas (1995) provided tight bounds in the case of perfect observability; Cripps and Thomas (1994) provide the parallel result when both players have time-average payoffs.

The key point is that with perfectly observed actions, the problem of off-path beliefs can prevent player 1 from obtaining the payoff he would most prefer. Schmidt (1993) gives an example of a perfect Bayesian equilibrium in which player 2's inability to learn the strategy played off the equilibrium path prevents player 1 from achieving the payoff he would get with a public commitment. Schmidt's (1993) example is based on the presence of a "perverse" type who plays like the "good" commitment type on the equilibrium path, but responds to deviations in a history dependent way. Cripps, Schmidt, and Thomas (1995) show that the perverse type is not required if we consider only Nash equilibria. Their Theorem 2 applies to games with observed actions where the patient player is either "normal" or plays an arbitrary finitely-complex strategy. It shows that there is a Nash

equilibrium where player 1's payoff is not substantially above the most he could obtain by playing a constant action, with player 2 choosing the individually rational response to this action that player 1 likes least.

In Section 5 we provide an example that shows the assumption of imperfect observability can in general not be dispensed with even if we restrict to Perfect Bayesian equilibria. The example shows that with perfect observability of actions there are Perfect Bayesian equilibria of the infinitely repeated game in which player 1 gets strictly less than his Stackelberg payoff.

Finally, we should acknowledge that Aoyagi (1994) independently obtains a result similar to ours for the case where player 1 maximizes his time-average payoff while player 2 discounts. Aoyagi's paper differs in interpreting the noise as "trembles," and, more significantly, in considering a complex class of "commitment types" that may be empty in some games, but the basic intuition for his results is the same.

To illustrate an application of our Theorems, as well as how they differ from the bounds of Cripps, Schmidt, and Thomas (1995), we examine a version of the Prisoner's Dilemma. It should be noted that the bound developed by Schmidt (1993) and by Cripps, Schmidt and Thomas (1995) does not imply any restriction on the Nash equilibria of this game: the best player 1 can get while minmaxing player 2 is his own minmax. We start with a traditional prisoner's dilemma (see Figure 1). We add incomplete observability, by supposing that if a player chooses a particular action then there is a small chance that the realized action will be different. In particular suppose that conditional on choosing action  $C$  (or  $D$ ) the realized action will be  $C$  ( $D$ ) with probability  $1 - \varepsilon$  and  $D$  ( $C$ ) with

probability  $\varepsilon$ . For this example a simple calculation shows that the greatest socially feasible payoff for player 1 that gives player 2 at least the minmax is  $3/2 - 2\varepsilon$ . Our results imply that if player 2 is patient, then a more patient player 1 will receive a payoff close to  $3/2$  in every Nash equilibrium of this prisoner's dilemma with commitment types. (Note that this payoff can be achieved if player 2 always chooses  $C$  and player 1 alternates between  $C$  and  $D$ .)

## 2. THE MODEL

Let  $g$  denote a stage game between two players, player 1 (the patient player) and player 2. We denote by  $A_1, A_2$  the finite (pure) action sets of the two players in the stage game with generic elements  $a_1, a_2$ , and use  $\alpha_1 \in \tilde{A}_1, \alpha_2 \in \tilde{A}_2$  for mixed actions. We denote by  $A, \tilde{A}$  the corresponding spaces of profiles. At the end of the stage game both players observe a stochastic outcome drawn from a finite set,  $y \in Y$ , but players do not observe each other's action. The probability distribution over outcomes depends on the action profile and is given as  $\rho(\cdot|a)$ ; for mixed actions  $\rho(\cdot|\alpha)$  is defined in the obvious way. Stage game payoffs, denoted  $u_i(\alpha)$ , take the form  $u_i(\alpha) \equiv \sum_y u_i(a_i, y)\rho(y|a)$  and therefore the realized payoffs do not convey additional information about the other player's action.

We consider an infinitely repeated game,  $G^\infty$ , constructed from  $g$  as follows. The public history at date  $t$ ,  $h^t$ , is the sequence of past realizations of the outcome  $y$ . Player  $i$ 's private history at  $t$ ,  $h_i^t$ , is the sequence of actions  $i$  chose in previous periods. Player 1 can be one of countably many types  $\omega \in \Omega$ . These types are drawn from a common

knowledge prior  $\mu$  assigning positive probability to all points in  $\Omega$ . Player 1's type is private information. We focus on a particular type  $\omega_0 \in \Omega$ , which we refer to as the normal type. Type  $\omega_0$  seeks to maximize the average present value of  $u_1$  using the discount factor  $\delta_1$ , while player 2 maximizes the average present value of  $u_2$  with the discount factor  $\delta_2$ . Types of player 1 other than type  $\omega_0$  have von Neumann-Morgenstern preferences over sequences of own actions and public outcomes, but these are not necessarily representable in a time separable form.

A behavior strategy for a type of player 1 or a behavior strategy for player 2 specifies a time indexed sequence of maps from private (for that player) and public histories to mixed actions (for that player). We denote these by  $\sigma_1$  and  $\sigma_2$  respectively. We also define  $u_t^i(\sigma_1, \sigma_2)$  to be the corresponding period  $t$  expected utility. Finally, a behavior strategy for player 1,  $\omega$ , specifies a behavior strategy for each type. A *Nash equilibrium* is a behavior strategy for each player such that given the opponent's behavior strategy, no other behavior strategy yields a distribution over time sequences of own actions and public outcomes that is preferred to that in the proposed equilibrium. It is quite easy to show by taking limits of finite truncations of this infinite game that Nash equilibria exist.<sup>5</sup> Let  $E_1(\delta_1, \delta_2)$  denote the least (inf) expected payoff to player 1 conditional on type  $\omega_0$  in any Nash equilibrium.

Say that a behavior strategy for player 1 has *bounded recall* if there exists a number  $N$  such that play at time  $t$  is entirely determined by the history between  $t - N$  and  $t - 1$ . A type of player 1 whose preferences make the type behavior strategy  $\sigma_1$  strictly dominant is called committed to that strategy, and we write the type as  $\omega(\sigma_1)$ .



We make four key assumptions:

ASSUMPTION 1: *If  $\sigma_1$  is a pure strategy of bounded recall, then  $\omega(\sigma_1) \in \Omega$ , that is, it has positive probability.*

ASSUMPTION 2: *If  $\rho(\cdot|\alpha_1, \alpha_2) = \rho(\cdot|\alpha'_1, \alpha_2)$ , then  $\alpha_1 = \alpha'_1$ .*

ASSUMPTION 3: *The support of  $\rho(\cdot|\alpha)$  is independent of  $\alpha_2$ .*

Define the (pure strategy) minmax for player 2:

$$\underline{u}_2 \equiv \min_{a_1} \max_{a_2} u_2(a).$$

ASSUMPTION 4: *There exists a pure profile  $a$  such that  $u_2(a) > \underline{u}_2$ .*

Assumption 1 ensures there are enough irrational types. (Since the set of bounded-recall strategies is countable this is consistent with our restriction to a countable number of types.) Assumption 2 requires that regardless of the play of player 2 the information revealed by the outcomes identifies the action of player 1. If it fails, player 2 may play an action that precludes him from learning what stage-game action player 1 is playing, preventing player 1 from developing a reputation, even when player 2 is myopic. Note that the assumption is satisfied if player 1's actions are perfectly observed, as in the previous papers on reputation effects with two long-run players.

Assumption 3 is the substantive assumption: It says that player 2 cannot determine the set of possible outcomes through his own action. Note that this assumption does not require that player 1's action be imperfectly observed, and indeed, whether player 1's action is observed or not is irrelevant.

Assumption 4 says that there is a profile that is better for player 2 than the pure strategy minmax payoff. If we used mixed strategies in place of pure, this would be a mild

non-degeneracy condition: failure would mean that the indifference of player 2 might well make him immune to threats by player 1. We restrict attention to the pure strategy min-max in order to avoid the complications involved in maintaining a reputation for playing a mixed strategy.<sup>6</sup> The existence of a profile better for player 2 than the pure strategy min-max rules out some interesting games, but the assumption is satisfied in other games of interest, such as the prisoner's dilemma and the battle of the sexes.

Let  $G_0^N$  be the  $N$  period repetition of  $g$ ; note that this is a game of complete information. Let  $G^N$  be the  $N$  period game with type space  $\Omega$  and prior  $\mu$ .

Before analyzing reputation in our model, we calculate as a benchmark how much player 1 might hope to get by precommitting. First we define a set of payoffs for player 1.

DEFINITION 1:  $v_1 \in V_1(\delta_2)$  if and only if for every  $\varepsilon > 0$  there is an  $N$  and a pure strategy  $\sigma_1^N$  such that for every best response  $\sigma_2^N$  to  $\sigma_1^N$  in  $G_0^N$ ,

$$(1/N) \sum_{t=1}^N u_1^t(\sigma_1^N, \sigma_2^N) \geq v_1 - \varepsilon .$$

Through precommitment (to a pure strategy) a patient player can guarantee himself  $\bar{v}_1(\delta_2) = \sup V_1(\delta_2)$ .<sup>7</sup>

### 3. AN IMPATIENT PLAYER 2

Our main result is:

THEOREM 1: *Suppose that Assumptions 1-3 hold. Then*

$$\liminf_{\delta_1 \rightarrow 1} E_1(\delta_1, \delta_2) \geq \bar{v}_1(\delta_2) .$$

In other words, if player 1 is very patient, then he gets nearly as much in any Nash equilibrium as the greatest amount consistent with the short run player choosing a best response in a finite truncation of the game.

The idea is that if player 1 commits to an appropriate bounded recall strategy and player 2 plays a best response to it, then player 1 gets a payoff very close to the lower bound given above. Note that since the strategy has bounded recall there is a type who is committed to playing this strategy. In the usual reputational story, this would mean that if player 1 chooses this strategy, player 2 must either play a best response to it, or come to believe that he faces a committed type. The situation here is complicated by the need to show that player 2 can learn the punishment strategy of player 1 without deviating: this is where Assumption 3 comes in.

We proceed via several Lemmas. Our initial focus is on the response of player 2 to bounded recall strategies in  $G_0^N$ .

LEMMA 1: *For every  $\eta > 0$  there exists an  $N, \underline{\delta}_1 < 1, \varepsilon > 0$  and pure strategy  $\sigma_1^N$  for player 1 in  $G_0^N$ , such that if  $1 > \delta_1 \geq \underline{\delta}_1$  and player 2 plays an  $\varepsilon$ -best response to  $\sigma_1^N$  in  $G_0^N$ , then the payoff to player 1 is at least  $\bar{v}_1(\delta_2) - \eta$ .*

PROOF: Omitted, see Celentani et al. (1993).

If  $\sigma^N$  is a profile in  $G_0^N$ , let  $p(\sigma^N)$  be the probability distribution over  $N$ -length sequences of public outcomes induced by  $\rho$ . Notice that this is a finite vector.

LEMMA 2: For any  $\varepsilon > 0, N$  there exists a  $\gamma > 0$  such that in  $G_0^N$  if  $\|p(\sigma_1^N, \sigma_2^N) - p(\tilde{\sigma}_1^N, \sigma_2^N)\| < \gamma$  and  $\sigma_2^N$  is an  $\varepsilon$ -best response by player 2 to  $\sigma_1^N$ , then it is a  $2\varepsilon$ -best response to  $\tilde{\sigma}_1^N$ .<sup>8</sup>

PROOF: We identify type behavior strategies by player 1 that differ only at information sets that are unclosed set is closed, so it suffices to show that  $p(\cdot, \sigma_2^N)$  is continuous and 1-1. Continuity is obvious. Suppose in fact that  $p(\tilde{\sigma}_1^N, \sigma_2^N) = p(\sigma_1^N, \sigma_2^N)$ , but that  $\sigma_1^N, \tilde{\sigma}_1^N$  are not equivalent. Let  $(h', h'_1, h'_2)$  be a triple consisting of a public and private histories (of the same length) possible under  $\sigma_1^N$  such that  $\tilde{\sigma}_1^N(h', h'_1) \neq \sigma_1^N(h', h'_1)$ . By Assumption 2 it follows that  $\rho(\tilde{\sigma}_1^N(h', h'_1), \sigma_2^N(h', h'_2)) \neq \rho(\sigma_1^N(h', h'_1), \sigma_2^N(h', h'_2))$ . Since  $h$  has positive probability under  $\sigma_1^N$  for some  $\tilde{\sigma}_2^N$  by Assumption 3 it has positive probability under  $\sigma_1^N, \sigma_2^N$  (and by hypothesis the same probability under  $(\tilde{\sigma}_1^N, \sigma_2^N)$ ). This contradicts  $p(\tilde{\sigma}_1^N, \sigma_2^N) = p(\sigma_1^N, \sigma_2^N)$ . Q.E.D.

For a given  $\delta_1$  let  $M(K, \delta_1)$  be the finite set of probability distributions over types of player 1 that can be generated by Bayesian updating with no more than  $K$  observations of player 1's play.<sup>9</sup>

We can partition  $G^\infty$  into a sequence of blocks of length  $N$ : each of these blocks is a game of the same form as  $G^N$  with the prior beliefs given by Bayes rule. By  $G^{N,k}$  we denote the  $k$ -th block of length  $N$  which starts in period  $N(k-1)+1$ . For a given  $\sigma_2$ , let  $\sigma_2^N(h')$  denote the strategy in  $G_0^N$  corresponding to the play of  $\sigma_2$  in periods

$t=1, \dots, t=N$ . For any  $\omega_0$ ,  $k$ , and for any distribution  $\mu'$  over types let  $\phi_1^k(h^{N(k-1)-1}, \mu', \omega_0)$  be the probability distribution over 1's strategies in  $G_0^N$  defined by

$$\phi_1^k(h^{N(k-1)-1}, \mu', \omega_0)(\sigma_1^N) = \sum_{\omega} \mu'(\omega) \omega_0(\sigma_1^N | \omega, h^{N(k-1)-1}).$$

LEMMA 3: Fix a Nash equilibrium  $(\omega_0, \sigma_2)$  and suppose for some  $N$  and  $\sigma_1^N$ ,  $\omega_0$  is such that type  $\omega_0$  plays  $\sigma_1^N$  in each  $G^{N,k}$ . For every  $\lambda > 0, \gamma > 0, K$  there is an  $L$  such that if  $\mu' \in M(K)$  the probability is less than  $\lambda$  that there are more than  $L$  stages  $G^{N,k}$  with

$$\left\| p(\sigma_1^N, \sigma_2^N(h^{N(k-1)-1})) - p(\phi_1^k(h^{N(k-1)-1}, \mu', \omega_0), \sigma_2^N(h^{N(k-1)-1})) \right\| \geq \gamma.$$

PROOF: This is a restatement of Theorem 4.1 in Fudenberg and Levine (1992).

PROOF OF THEOREM 1: Fix a Nash equilibrium  $(\omega_0, \sigma_2)$ . For any number  $\eta > 0$  we may choose  $N, \underline{\delta}_1 < 1, \varepsilon > 0, \sigma_1^N$  so that Lemma 1 is satisfied for the tolerance  $\eta/4$  and hence if player 2 plays an  $\varepsilon$ -best response to  $\sigma_1^N$  in  $G_0^N$  then the payoff to player 1 in  $G_0^N$  is at least  $\bar{v}_1(\delta_2) - \eta/4$ . The idea is to consider what happens when player 1 of type  $\omega_0$  repeatedly plays  $\sigma_1^N$  and player 2 plays a best response. Our conclusion follows by demonstrating that if player 1 is more patient than  $\underline{\delta}_1$  he gets at least  $\bar{v}_1(\delta_2) - \eta$ .

To analyze the best response of player 2, we fix any integer  $\kappa$  so that  $\delta_2^{\kappa N} (\max u_2 - \min u_2) < \varepsilon(1 - \delta_2)/4$ . Set  $N' = \kappa N$ . Apply Lemma 2 to  $G_0^{N'}$  where the tolerance is  $\varepsilon(1 - \delta_2)/2$  to find a value for  $\gamma$ . Apply Lemma 3 to blocks  $G^{N'}$  using this

same value of  $\gamma$ , and choosing  $\lambda$  such that  $(1 - (1 - \lambda)^\kappa)(\max u_1 - \min u_1) < \eta / 4$ . We refer to blocks  $G^{N',k}$  in which

$$\left\| p(\sigma_1^{N'}, \sigma_2^{N'}(h^{N'(k-1)-1})) - p(\phi_1^k(h^{N'(k-1)-1}, \mu^1, \delta_2), \sigma_2^{N'}(h^{N'(k-1)-1})) \right\| \geq \gamma$$

as anomalous.

Strategy  $\sigma_2$  is at worst an  $\varepsilon(1 - \delta_2) / 2$  best response to  $\phi_1^k(h^{N'(k-1)-1}, \mu(h^{N'(k-1)-1}), \delta_2)$  in  $G^{N',k}$ , where  $\mu(h^{N'(k-1)-1})$  denotes 2's posterior beliefs given  $h^{N'(k-1)-1}$  (this is the case since  $\delta_2^{\kappa N'}(\max u_2 - \min u_2) < \varepsilon(1 - \delta_2) / 4$ ). By Lemma 2 player 2 is playing an  $\varepsilon(1 - \delta_2)$  best response to the  $\kappa$ -fold repetition of  $\sigma_1^{N'}$  except in the anomalous blocks. Consequently, except in anomalous blocks, player 2 is playing an  $\varepsilon$ -best response to  $\sigma_1^{N'}$  in the first  $N$  periods of each  $N'$  block. Lemma 1 then shows that player 1 gets at least  $\bar{v}_1(\delta_2) - \eta / 2$  in the first  $N$  periods of each non-anomalous block. Furthermore, with probability at least  $1 - \lambda$  there are at most  $L$  anomalous blocks.

Now consider the infinitely repeated game beginning in any period  $kN + 1$ ,  $1 \leq k \leq \kappa - 1$ . This is identical to the game that begins in period 1, except that the prior of player 2 may have changed. We may again organize this game into blocks of length  $N'$  and, since Lemma 3 applies to all priors reachable during periods up to  $N'$ , the previous argument shows that for every  $k$  there are at most  $L$  anomalous blocks with probability at least  $1 - \lambda$ . Thus with probability at least  $(1 - \lambda)^\kappa$  in the overall game there are at most  $\kappa L$  anomalous blocks. The payoff of player 1 is therefore at least

$$\delta_1^{\kappa L \kappa N'} \bar{v}_1(\delta_2) - 3\eta / 4 + (1 - \delta_1^{\kappa L \kappa N'}) \min u_1$$

Letting  $\delta_1 \rightarrow 1$  gives the desired result.

*Q.E.D.*

## 4. A PATIENT PLAYER 2

Our second Theorem shows that as  $\delta_2 \rightarrow 1$  player 1 can approximately obtain the maximum feasible payoff that gives player 2 his pure strategy minmax payoff. Note that this bound is derived by taking a particular order of limits. First, we derive a lower bound on player 1's payoff as  $\delta_1 \rightarrow 1$ , and then we ask how this bound behaves as  $\delta_2 \rightarrow 1$ . Let  $V^*$  denote the convex hull of feasible payoffs that are at least as great as the pure strategy minmax. By  $V_1^*$  we denote the projection of  $V^*$  onto the payoffs of player 1.

THEOREM 2: *Suppose Assumptions 1-4 hold. Then*

$$\liminf_{\delta_2 \rightarrow 1} \liminf_{\delta_1 \rightarrow 1} E_1(\delta_1, \delta_2) \geq \max V_1^*$$

It is worth noting that if we allowed types of player 2, this result would remain valid, and the proof (and Theorem 1 from which it follows) would involve only notational changes. However, we cannot turn the Theorem around and use the fact that there are types player 2 to find a bound on his payoffs similar to that for player 1: The validity of Theorem 2 depends crucially on the order of limits.

Note also that the definition of  $V^*$  makes use of the *pure* strategy minmax. As was argued in Section 2, we allow player 1 to establish a reputation only for a pure strategy, since allowing him to establish a reputation for mixed strategy punishments would require the existence of a continuum of types that would make the notation significantly heavier. If we allowed a reputation for mixed strategies, we could use the usual feasible individually rational set, and replace the inequality in Theorem 2 with an equality.

The proof of Theorem 2 is an immediate consequence of Theorem 1 and the following Lemma.

LEMMA 4: *For any  $v = (v_1, v_2)$  with  $v \in \text{int}V^*$  there is a  $\bar{\delta}_2$  such that, for  $\delta_2 > \bar{\delta}_2$ ,  $v_1 \in V_1(\delta_2)$ .*

*Remark:* We should emphasize that this Lemma concerns the complete information game, where reputation plays no role. The Lemma is thus more closely related to the literature on repeated games than to that on reputation effects, and indeed our proof uses “review strategies” of the sort introduced in Radner’s (1981, 1985) study of repeated agency games, and subsequently used in a number of papers on the folk Theorem in repeated games. Despite this close link to the repeated game literature, the Lemma we need does not seem to be a direct consequence of previous work, so we give a complete proof in the Appendix.

## 5. EXAMPLE

In this Section we give an example that shows that the conclusions of Theorems 1 and 2 fail with perfectly observable actions even if the equilibrium concept is strengthened to perfect Bayesian equilibrium.<sup>10</sup> Suppose that the following stage game (see Figure 2) is played repeatedly with perfect observability. Player 1 is the row player and player 2 the column player. Notice that the Stackelberg payoff here is 2, with player 1 playing  $C$  and player 2 playing  $c$ . Suppose that there are two types of patient player: type  $\omega_0$  is the normal type and type  $\omega_1$  always plays  $C$ . (Note that there are no perverse types.) The



prior probability of the commitment type is  $\mu_1 = .001$ . For simplicity we assume that there is a public randomization device.

We claim that for all  $\delta_1 \geq \delta_2 \geq .99$  there is a perfect Bayesian equilibrium<sup>11</sup> in which type  $\omega_0$ 's payoff is no more than 1; in particular this is the case if  $\delta_1 = \delta_2 \geq .99$ .<sup>12</sup> The equilibrium strategies we describe below can be adapted to support any payoff of type  $\omega_0$  player 1 strictly above his minmax payoff of  $2/3$  if  $\delta_1 > \delta_2 > \bar{\delta}$  for some  $\bar{\delta} < 1$ .

Notice that if player 1 ever plays  $D$ , player 2 learns for certain that 1 is the normal type; at this point every equilibrium of the game where 1 is known to be normal (the “complete-information game”) is a continuation equilibrium of the original game. For this reason, it is useful to understand the perfect equilibria of the complete-information repeated game. First, there are three static equilibria:  $(2,2)$  and  $(1,-1)$  are pure strategy equilibria, and there is a mixed strategy equilibrium in which player 1 plays  $C$  with probability  $1/3$  and player 2 plays  $c$  with probability  $1/3$ . The payoff vector to this mixed equilibrium is  $(2/3, -2/3)$ . Naturally, the infinite repetition of any of these static equilibria is a subgame-perfect equilibrium of the repeated game. Moreover, any payoff vector  $(u_1, u_2)$  in the convex hull of the pure strategy payoffs that satisfies  $(u_1, u_2) \geq (.687, -.646)$  can also be obtained in a perfect equilibrium. These payoffs can be achieved by strategies that specify a repeated public randomization yielding  $(u_1, u_2)$  each period along the equilibrium path, with play switching to the static mixed strategy equilibrium if any player deviates.<sup>13</sup>

Finally, when the two players have different discount factors, there are socially feasible payoffs that are not in the convex hull of pure strategy payoffs. Call a sequence of

expected present values  $(v_1(t), v_2(t))$  socially feasible if there is a strategy profile that gives rise to this sequence. If in each period  $(v_1(t), v_2(t)) \geq (.687, -.646)$ , then each payoff vector in the sequence is a perfect equilibrium payoff: again the punishment is reversion to the static mixed strategy equilibrium. A particular example of such payoffs are any payoffs of the form  $(\delta_1^\tau u_1, \delta_2^\tau u_2) \geq (.687, 0) \geq (.687, -.646)$  where  $(u_1, u_2)$  is an element of the convex hull of pure strategy payoffs. (Note that payoffs of this form need not themselves be a convex combination of pure strategy payoffs.) Such payoff sequences can be realized by playing  $(C, d)$  starting in period 1 and for  $\tau$  periods, and then switching to the stationary equilibrium  $(u_1, u_2)$ . That this non-stationary sequence is an equilibrium follows from

$$(v_1(t), v_2(t)) = (\delta_1^{\tau-t-1} u_1, \delta_2^{\tau-t-1} u_2) \geq (\delta_1^\tau u_1, \delta_2^\tau u_2) \geq (.687, 0) \quad t \leq \tau + 1.$$

We can now describe a perfect Bayesian equilibrium of the incomplete information game in which type  $\omega_0$ 's payoff is exactly 1 in equilibrium, while player 2 also gets 1. This equilibrium consists of an equilibrium path and continuation equilibria in response to deviations. Following a play of  $D$  by player 1, the game becomes a complete information game, and the continuation equilibrium will always be one of the equilibria described above. There are two other continuation equilibria that will be needed: the *player 2 punishment equilibrium*, described below, and the *cooperative* equilibrium in which regardless of observed play both types of player 1 play  $C$  and player 2 plays  $c$ .

*The Equilibrium Path:* Along the equilibrium path player 1 always plays  $C$  while a 50-50 public randomization determines whether player 2 chooses  $d$  or  $c$ . This implies that both players get a present value of 1 as desired. No information about player 1's type is

revealed on the equilibrium path, so player 2 does not update his beliefs. If either player deviates in a period in which player 2 is supposed to play  $c$ , the deviation is ignored. This is consistent with equilibrium, since such deviations do not yield a short term gain. If both players deviate simultaneously both deviations are ignored.<sup>14</sup>

If player 1 is the only deviator in a period where 2 plays  $d$ , player 1 is revealed to be a normal type, and play switches to the repeated mixed strategy equilibrium of the complete information game. In this case, the deviation yields  $(1 - \delta_1) + \delta_1 \cdot 2/3$  which for any  $\delta_1$  is less than the equilibrium value of 1. Finally, if player 2 is the only deviator in a period in which 2 is supposed to play  $d$ , no information is revealed about player 1's type, and play switches to a second equilibrium of the incomplete information game that we refer to as the *player 2 punishment equilibrium*. We verify below that the expected present value to player 2 in this punishment equilibrium is constructed to be no greater than .81. Since  $1 \geq (1 - \delta_2) \cdot 2 + .81\delta_2$  player 2 indeed has no incentive to deviate from the equilibrium path.

*Player 2 Punishment Equilibrium:* In the punishment equilibrium, play has two phases. The first phase lasts at most  $\tau$  periods, where  $\delta_2^\tau \geq 0.4 \geq \delta_2^{\tau-1}$ . In this phase, the normal type of player 1 always randomizes between  $C$ ,  $D$  as described below, while player 2 always chooses  $d$ . The phase ends when either time runs out (e.g. after  $\tau$  periods) or when player 1 plays  $D$ . If time runs out play reverts to the cooperative equilibrium  $(C, c)$ . If player 1 plays  $D$  in the  $t$ -th period of the punishment equilibrium, he reveals his type and in the following period play switches to an equilibrium of the complete information game

in which  $(C, d)$  is played up until and including period  $\tau$  and whose present value payoffs are  $(\delta_1^{\tau-t} u_1(t), \delta_2^{\tau-t} u_2(t))$ , where

$$(u_1(t), u_2(t)) = \left(2 - \frac{1 - \delta_1}{\delta_1^{\tau-t-1}}, 2 - \frac{1 - \delta_1}{\delta_1^{\tau-t-1}}\right).$$

It can easily be checked that the vector  $(u_1(t), u_2(t))$  is socially feasible for all  $1 \leq t \leq \tau$ , and that  $(\delta_1^{\tau-t} u_1(t), \delta_2^{\tau-t} u_2(t)) \geq (\delta_1^{\tau-1} u_1(1), \delta_2^{\tau-1} u_2(1)) \geq (.79, .79)$  so from our earlier analysis these are indeed perfect equilibrium payoffs of the complete information game.

Notice that the payoffs to player 1 following playing  $D$  are calibrated so that he is exactly indifferent between  $C$  and  $D$ , implying that his play during the initial phase is optimal. If player 2 deviates and player 1 chooses  $C$  in the same period then the deviation is ignored. If player 2 deviates in the  $t$ -th period of the phase and player 1 chooses  $D$  then player 1 has revealed his type. In this case we switch to the repeated mixed strategy equilibrium of the complete information game.<sup>15</sup> To show that the deviation is not optimal, we must now specify the randomization probabilities for player 1 and the corresponding beliefs for player 2.

To define the randomization probabilities, denote by  $\mu_t$  player 2's subjective probability that player 1 is a commitment type in the  $t$ -th period of the first phase. Let  $p = 1 - 6(1 - \delta_2)$ . In the  $t$ -th period of this phase type  $\omega_0$  chooses  $C$  with probability  $p_t = (p - \mu_t) / (1 - \mu_t)$ , so that the overall probability that player 1 chooses  $C$  equals  $p$ . The beliefs of player 2 are given by  $\mu_t = \mu_0 / p^{t-1}$ . Since  $\tau \leq \ln(.4) / \ln(\delta_2)$  we have

$$\begin{aligned}
\mu_t &\leq \frac{\mu_0}{p^\tau} \\
&\leq \frac{\mu_0}{(1-6(1-\delta_2))^{\ln(4) \cdot \ln(\delta_2)}} \\
&\leq \frac{\mu_0}{(1-6(1-.99))^{\ln(4) \cdot \ln(.99)}} = .282 \\
&< .94 = 1-6(1-.99) \leq 1-6(1-\delta_2) = p
\end{aligned}$$

for  $\delta_2 > .99$ .<sup>16</sup> Therefore, since  $p > \mu_t$ ,  $1 \leq t \leq \tau$ ,  $p_t$  is greater than 0.

Next we consider player 2's incentives to deviate in the various phases. Let  $V_2^t$  denote player 2's expected payoff in period  $t$  of the punishment equilibrium. First note that  $.81 \geq 2\delta_2^\tau \geq V_2^1$ , so that player 2's expected payoff in the initial period of the punishment equilibrium is indeed no greater than .81. As we remarked above, this shows that player 2 cannot gain by deviating along the equilibrium path.

Finally, we have to show that player 2 has no incentive to deviate in the punishment equilibrium. First we give a lower bound for  $V_2^t$ ,  $1 \leq t \leq \tau$ . Note that in the first  $\tau$  periods of the punishment equilibrium either  $(C,d)$  is played in every period or  $(D,d)$  is played exactly once. After the first  $\tau$  periods of the punishment equilibrium the continuation payoff of player 2 is either 2 or  $2 - \frac{1-\delta_1}{\delta_1^{\tau-t-1}}$  where  $1 \leq t \leq \tau$ . Thus  $V_2^t$  is bounded by

$$V_2^t \geq -(1-\delta_2) + 2\delta_2^{\tau-1-t} - \frac{1-\delta_1}{\delta_1^{\tau-1-t}} \cdot \delta_2^{\tau-1-t} \geq .78$$

Player 2 has no incentive to deviate in the punishment equilibrium if

$$2(1-\delta_2) + \delta_2[pV_2^{\tau-1} + (1-p)\delta_2(-2/3)] \leq (1-p)(1-\delta_2)(-1) + \delta_2V_2^{\tau-1}$$

which is equivalent to

$$2(1 - \delta_2) + \delta_2(6(1 - \delta_2))\delta_2(-2/3) \leq 6(1 - \delta_2)(1 - \delta_2)(-1) + \delta_2 6(1 - \delta_2)I_2^{t-1}$$

Since  $I_2^{t-1} \geq .78$  this inequality holds if

$$2 + \delta_2 \cdot 6 \cdot \delta_2(-2/3) \leq 6(1 - \delta_2)(-1) + \delta_2 \cdot 6 \cdot 0.78$$

which holds for  $\delta_2 \geq .99$ .

*Universidad Carlos III de Madrid*

*Harvard University*

*University of California, Los Angeles*

*and*

*Northwestern University*

## APPENDIX

PROOF OF LEMMA 4 : Given  $v = (v_1, v_2)$  with  $v \in \text{int} V^*$  let  $\bar{a}_1 = (\bar{a}_1^1, \dots, \bar{a}_1^N)$  be a sequence of actions by player 1 such that there is a sequence of actions for player 2 such that in  $G_0^N$  the expected average payoff of player 1 is larger than  $v_1 - \varepsilon / 2$  and the expected average payoff of player 2 is larger than  $v_2$ . (Clearly for  $N$  sufficiently large such a sequence exists.)

Consider the  $KMN$ -fold repeated game  $G_0^{KMN}$  in which player 2 has the time average payoff as a payoff function. Again we can partition  $G_0^{KMN}$  into blocks of length  $MN$ ,  $G_0^{MN,k}$ ,  $k = 1, \dots, K$ . Denote by  $\bar{\bar{a}}_1 = (\bar{\bar{a}}_1^1, \dots, \bar{\bar{a}}_1^N)$  the sequence of  $M$  repetitions of  $\bar{a}_1$ . Let  $\underline{a}_1$  be a (pure) action of player 1 that minmaxes the payoff of player 2 in pure strategies. Further, let  $u_1^k$  denote player 1's average payoff in  $G_0^{MN,k}$ .

Let  $\hat{\sigma}_1$  be the following strategy: in  $G_0^{MN,1}$ , player 1 chooses  $\bar{\bar{a}}_1$ . In  $G_0^{MN,2}$  if  $v_1 - u_1^1 - \varepsilon / 2 < \eta$ , then player 1 again chooses  $\bar{\bar{a}}_1$  and so on. If for any  $k$ ,  $v_1 - u_1^k - \varepsilon / 2 > \eta$ , then player 1 plays action  $\underline{a}_1$  for the next  $P$  repetitions of  $G_0^{MN,k}$ .

CLAIM: Given  $\varepsilon > 0$  there are numbers  $\eta$ ,  $K$ ,  $N$ ,  $M$ , and  $P$  with  $P / K < \varepsilon$  such that for any (time average) best response  $\hat{\sigma}_2$  to  $\hat{\sigma}_1$  the following is satisfied:

(i) if player 1 chooses  $\bar{\bar{a}}_1$  in  $G_0^{MN,k}$ ,  $k = 1, \dots, K - P$ , then  $\Pr(v_1 - u_1^k < \varepsilon) > 1 - \varepsilon$ .

(ii) the fraction of stages  $k$  in which player 1 uses his punishment strategy is smaller than  $\varepsilon$  with probability  $(1 - \varepsilon)$ .

Assuming for the moment the truth of the claim, a straightforward upper hemi-continuity argument shows that the claim remains true if  $\hat{\sigma}_2$  is discounted best response with  $\delta_2$  sufficiently close to 1. Moreover, the claim implies that the average payoff to player 1 in  $G_0^{kM}$  is greater than  $v_1 - H\varepsilon$  where  $H$  is a positive constant independent of  $\varepsilon$ . This is the desired result.

To demonstrate the validity of the claim, first choose  $\eta < \varepsilon^2 / (\beta + 1)$ , where  $\beta$  is a fixed constant whose computation is described below.

Denote by  $Eu_1^k | h^{(k-1)MN-1}$  the expected payoff of player 1 in  $G_0^{MN,k}$  given the history  $h^{(k-1)MN-1}$ . Given  $K$  and  $\eta$  we can choose  $M$  sufficiently large, so that

- (a) if  $v_1 - Eu_1^k | h^{(k-1)MN-1} - \varepsilon / 2 > 2\eta$  then punishment occurs with probability greater than  $(1 - \eta)$  in  $G_0^{MN,k}$ ;
- (b) if  $v_1 - Eu_1^k | h^{(k-1)MN-1} - \varepsilon / 2 < \eta$  for all  $k$  and for all histories  $h^{(k-1)MN-1}$  such that player 1 chooses  $\bar{a}_1$  in  $G_0^{MN,k}$ , then the probability that no punishment occurs in any  $G_0^{MN,k}$  is larger than  $(1 - \eta)$ .

Note that the utility loss from a punishment is bounded below by  $(v_2 - \underline{u}_2)P$ , ( $v_2 - \underline{u}_2 > 0$ ), whereas the gain from a deviation is bounded above by  $\bar{u}_2 - \underline{u}_2$ , where  $\bar{u}_2$  is the largest attainable payoff for player 2 in the stage game. Thus for appropriate choice of  $P$  we have

$$(A1) \quad v_1 - Eu_1^k | h^{(k-1)MN-1} - \varepsilon / 2 < 2\eta$$

in all of the first  $K - P$  stages for any best response of player 2 and for all histories  $h^{(k-1)MN-1}$  such that player 1 chooses  $\bar{a}_1$  in  $G_0^{MN,k}$ . Hence part (i) of the claim follows.



Now we establish part (ii) of the claim. Suppose to the contrary that player 2's (optimal) strategy triggers a punishment in more than an  $\varepsilon$  fraction of the first  $K - P$  stages with probability greater than  $\varepsilon$ . We claim that this implies that a profitable deviation exists. Suppose player 2 deviates so that  $v_1 - Eu_1^k | h^{(k-1)M^{N-1}} - \varepsilon / 2 < \eta$  for all  $k=1, \dots, K - P$  and for all histories  $h^{(k-1)M^{N-1}}$  such that player 1 chooses  $\bar{a}_1$  in  $G_0^{M^N, k}$ . Since inequality (A1) has to be satisfied for all  $k = 1, \dots, K - P$ , this deviation can be chosen so that the loss in every stage is bounded above by  $\eta(\max u_2 - \min u_2)$ . Note that (after the deviation) the probability that a punishment occurs in any stage is smaller than  $\eta$ . Thus player 2 also improves his average payoff over the first  $K - P$  stages by at least  $(\varepsilon^2 - \eta)(v_2 - \underline{u}_2) \cdot P$  by reducing the probability of punishment. Consequently, if we choose

$$\beta = \frac{(\max u_2 - \min u_2)K}{(v_2 - \underline{u}_2) \cdot P}$$

player 2 gains from the deviation and part (ii) of the claim follows.

*Q.E.D.*

## REFERENCES

- AOYAGI, M (1994).: "Reputation and Dynamic Stackelberg Leadership in Infinitely Repeated Games," mimeo.
- CELENTANI, M. (1995): "Reputation with Observable Actions," *Economic Theory*, forthcoming.
- CELENTANI, M., AND W. PESENDORFER (1995): "Reputation in Dynamic Games," *Journal of Economic Theory*, forthcoming.
- CELENTANI, M., D. FUDENBERG, D. K. LEVINE, AND W. PESENDORFER (1993): "Maintaining a Reputation Against a Patient Opponent," CMSEMS Discussion Paper No. 1075, Northwestern University.
- CRIPPS M., K. SCHMIDT, AND J. P. THOMAS (1995): "Reputation in Perturbed Repeated Games," *Journal of Economic Theory*, forthcoming.
- CRIPPS, M., AND J. P. THOMAS (1994): "Reputation and Equilibrium Selection in Two-Person Repeated Games without Discounting," mimeo.
- (1993): "Reputation and Perfection in a Repeated Common Interest Game," mimeo.
- FUDENBERG, D., AND D. M. KREPS (1987): "Reputation and Simultaneous Opponents," *Review of Economic Studies*, 54, 541-568.
- FUDENBERG, D., AND D. K. LEVINE (1983): "Subgame Perfect Equilibria of Finite and Infinite Horizon Games," *Journal of Economic Theory*, 31, 251-268.

————— (1989): “Reputation and Equilibrium Selection in Games with a Single Patient Player,” *Econometrica*, 57, 759-778.

————— (1992): “Maintaining a Reputation when Strategies are Imperfectly Observed,” *Review of Economic Studies*, 59, 561-579.

FUDENBERG, D., AND J. TIROLE (1991): “Perfect Bayesian Equilibrium and Sequential Equilibrium,” *Journal of Economic Theory*, 53, 236-260.

KREPS, D. M., AND R. WILSON (1982): “Reputation and Imperfect Information,” *Journal of Economic Theory*, 27, 253-279.

MILGROM, P., AND J. ROBERTS (1982): “Predation, Reputation and Entry Deterrence,” *Journal of Economic Theory*, 27, 280-312.

RADNER, R. (1981): “Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship,” *Econometrica*, 49, 1127-1148.

————— (1985): “Repeated Principal Agent Games with Discounting,” *Econometrica*, 53, 1173-1198.

SCHMIDT, K. (1993): “Reputation and Equilibrium Characterization in Repeated Games with Conflicting Interests,” *Econometrica*, 61, 325-351.

## FIGURES

	<i>C</i>	<i>D</i>
<i>C</i>	1,1	-1,2
<i>D</i>	2,-1	0,0

FIGURE 1

	<i>c</i>	<i>d</i>
<i>C</i>	2,2	0,0
<i>D</i>	0,-2	1,-1

FIGURE 2

## FOOTNOTES

---

<sup>1</sup> The authors are grateful for financial support from NSF grants SBR-9223320, SBR-9223175, SBR-9409180, DGICYT PB92-0245, Human Capital Mobility Programme ERBCHBICT940975, and the UCLA Academic Senate.

<sup>2</sup> Celentani and Pesendorfer (1995) consider dynamic games with one large player and a continuum of small but long lived players. Small players in this setting care about the future but cannot influence the relevant history of the game and hence are strategically myopic.

<sup>3</sup> Thus the situation is similar to the “sequential contests” model of Fudenberg and Kreps (1987).

<sup>4</sup> Celentani (1995) uses multiple types of short-run players to get around the problem of unobservable off-path behavior of the long-run player in repeated games with observable actions. His approach can be extended to short run players who live for more than one period.

<sup>5</sup> See, for example, Fudenberg and Levine (1983).

<sup>6</sup> These complications were addressed in the context of a myopic player 2 in Fudenberg and Levine (1992). The restriction is largely a matter of technical convenience: for

---

player 1 to develop a reputation for mixed strategy punishments it would be necessary to allow a continuum of types. Working with a continuum of types increases the complexity of the notation substantially.

<sup>7</sup> Because we restrict attention to pure strategy commitment types, the worst punishment that player 1 can hope to “teach” player 2 to fear is the pure-strategy minimax; this restriction on punishments can result in a lower maximum payoff for player 1 than if mixed punishments were considered .

<sup>8</sup> The norm  $\| \cdot \|$  may be taken to be ordinary Euclidean distance, where the probability distribution over a finite set  $p(\cdot)$  is viewed as a vector.

<sup>9</sup> Note that, although  $\sigma_2$  influences the probability of the various elements of  $M(K, \sigma_1)$ , it does not influence the set itself.

<sup>10</sup> Moreover this example is robust to small changes in payoffs, and player 1 cannot obtain the Stackelberg payoff by commitment to a mixed strategy. We thank a referee for finding an error in a previous example, and for encouraging us to find a robust example.

<sup>11</sup> Although sequential equilibrium has not been defined for infinite games, in the finitely repeated versions of the two-types example considered here the outcomes of perfect

---

Bayesian equilibria (PBE) and sequential equilibria coincide (Fudenberg and Tirole (1991)). We use the simpler PBE concept because its conditions on beliefs are easier to check.

<sup>12</sup> Cripps and Thomas (1993) give a related example for the case where both players have the same discount factor.

<sup>13</sup> The bound on player 1's payoff is derived as follows. If the public randomization prescribe the play of  $(D,c)$ , player 1 obtains  $2(1-\delta_1)+(2/3)\delta_1 v_1$  from deviating, and  $0+\delta_1 v_1$  from conforming, where  $v_1$  is player 1's equilibrium payoff; conforming is better than deviating if  $v_1$  is greater than  $.68/99 < .687$ . Similar calculations show that this bound on 1's equilibrium payoff implies that player 1 will not want to deviate when any of the other three pure strategy profiles are supposed to be played. The computations for player 2 are analogous.

<sup>14</sup> Play following simultaneous deviations is always irrelevant to the determination of equilibrium.

<sup>15</sup> The idea here is that player 2 is only punished if player 1 has revealed his type. This construction depends on the fact that the game is simultaneous move.

---

<sup>16</sup> Note that  $\frac{\mu_0}{(1 - 6(1 - \delta_2))^{\ln(4) \ln(\delta_2)}} \rightarrow \frac{\mu_0}{4^6} = .244$  as  $\delta_2 \rightarrow 1$ .