

Maintaining a Reputation Against a Patient Opponent¹

July 31, 2006

Marco Celentani

Drew Fudenberg

David K. Levine

Wolfgang Pesendorfer²

ABSTRACT: We analyze reputation in a game between a patient player 1 and a non-myopic but less patient opponent, player 2. Player 1's type is private information and he may be a "commitment type" who is locked into playing a particular strategy. We assume that players do not directly observe each other's action but rather see an imperfect signal of it. In particular, we assume that the support of the distribution of signals is independent of how player 2 plays. We show that in any Nash equilibrium of the game player 1 will get a payoff close to the largest payoff consistent with player 2 choosing a best response in a finite truncation of the game. Moreover, we show that if the discount factor of player 2 is sufficiently large then player 1 will essentially get the maximum payoff consistent with player 2 getting at least his pure strategy minmax payoff in any Nash equilibrium.

¹ The authors are grateful for financial support from NSF grants SBR-9223320 and SBR-9223175, DGICYT and the UCLA Academic Senate.

²Departments of Economics Universidad Carlos III de Madrid, Harvard, UCLA and Northwestern.

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 in to 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]).³ 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 consider the patient player 1 has discount factor δ_1 near 1, while his opponent player 2 is an infinite-lived player who discounts future payoffs with a smaller discount factor δ_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 different but identical player 2's, each of whom observe all previous play and who either alternate play, or play consecutively. 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.⁴

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

³Celentani and Pesendorfer (1992) 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.

⁴ Thus the intuition for our results is similar to that for the "sequential contests" model of Fudenberg and Kreps [1987]

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]).⁵

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. Indeed, we assume that the support of the distribution of signals is independent of how player 2 plays. Intuitively, this ensures that player 1 will be called on to periodically use all rewards or punishments, thus eliminating the problem of player 2 misperceiving how player 1 would respond to a deviation. 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 the 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.”⁶

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 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 [1993] provided tight bounds in the case of perfect observability; Cripps and Thomas [1994] provide the parallel result when both players have time-average payoffs.

⁵Celentani (1991) uses multiple types of short-run players to get around the problem of unobservable off-path behavior of the long-run player in non-degenerate extensive form stage games. His approach can be extended to short run players who live for more than one period.

⁶ In contrast, with a myopic opponent and simultaneous-move stage games, as in Fudenberg-Levine [1989], reputation effects often imply strong bounds for reasonable discount factors and prior beliefs.

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 example is based on the presence of a "perverse" type who uses a history-dependent strategy, and is indistinguishable from the "good" commitment type along the equilibrium path. Cripps, Schmidt, and Thomas [1993] show that the perverse type is not required if we consider only Nash equilibria. Their Theorem 3 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.

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 more 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, 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 [1993] 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.

| | | |
|---|------|------|
| | C | D |
| C | 1,1 | -1,2 |
| D | 2,-1 | 0,0 |

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 ε . 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 the short run player is patient, then a patient long run player will receive a payoff close to $3/2$ in every Nash equilibrium of this prisoner's dilemma. (Note that this payoff can be achieved if player 2 always chooses C and player 1 alternates between C and D.)

2. The Model

We consider a repeated game between two players, player 1 (the patient player) and player 2 (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 each period $t = 1, 2, \dots$ players commonly observe a stochastic outcome drawn from a finite set, $y \in Y$. The probability distribution over outcomes depends on the action profile and is given as $\rho(y|a)$; for mixed actions $\rho(y|\alpha)$ is defined in the obvious way.

Player 1 can be one of countably many types $\omega \in \Omega$. These types are drawn from a common knowledge prior μ assigning positive probability to all points in Ω . Player 1 is informed of his type before play begins, but this is purely private knowledge and is not revealed to player 2. We focus on a particular type $\omega_0 \in \Omega$, which we refer to as the rational type.

Stage game payoffs are $u_1(a_1, y)$ for the type ω_0 patient player and $u_2(a_2, y)$ for player 2. It is also useful to define normal form payoffs by

$$u_i(a) \equiv \sum_y u_i(a_i, y)\rho(y|a)$$

with $u_i(\alpha)$ defined in the obvious way.

In the repeated game the type ω_0 patient player seeks to maximize the average present value of utility using the discount factor δ_1 , while player 2 uses the discount factor δ_2 . Types of player 1 other than type ω_0 have von Neumann-Morgenstern

preferences over sequences of player 1 actions and public outcomes, but these are not necessarily representable in a time separable form.

A *type behavior strategy* for 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 σ_1 and σ_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* specifies a map from types to type behavior strategies. 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 actions (for that player) 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.⁷ Let $N_1(\delta_1, \delta_2)$ denote the least (inf) expected payoff to player 1 conditional on type ω_0 in any Nash equilibrium.

Say that a type 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 σ_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 σ_1 is pure 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 α_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

⁷ See, for example, Fudenberg and Levine [1983].

number of types.) Assumption 2 is from Fudenberg and Levine [1992] who call it identifiability: it means that regardless of the play of player 2 there is enough statistical information revealed by the outcomes to determine 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 if 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. If he could, then there are many counterexamples to the theorems below. 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 minmax in order to avoid the complications involved in maintaining a reputation for playing a mixed strategy.⁸ The existence of a profile better for player 2 than the pure strategy minmax does rule 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..

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. This set has the feature that these payoffs can be approximated by profiles in a finitely repeated version of the game, in which player two's repeated game strategy is a best response to player 1's in a finite truncation of the game.

⁸ These complications were addressed in the context of a myopic player 2 in Fudenberg-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

Definition 1. $v_1 \in V_1(\delta_2)$ iff for every $\varepsilon > 0$ there is an N and a pure strategy σ_1^N in the N -fold repeated game such that if σ_2^N is a best response of player 2 with discount factor δ_2 in the N -fold repeated game then $(1/N) \sum_{t=1}^N u_1^t(\sigma_1^N, \sigma_2^N) \geq v_1 - \varepsilon$.

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

3. An Impatient Less Patient Player

Our main result is

Theorem 1: *Suppose that Assumptions 1-3 hold. Then $\liminf_{\delta_1 \rightarrow 1} N_1(\delta_1, \delta_2) \geq \bar{v}_1(\delta_2)$.*

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 are types of patient player who play these strategies regardless of the history. In the usual reputational story, this would mean that if player 1 plays one of these review strategies, 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 these strategies in the N -fold repeated game.

Lemma 1: *For every $\eta > 0$ there exists an N , $\underline{\delta}_1 < 1$, $\varepsilon > 0$ and pure type strategy for player 1 in the N -fold repeated game σ_1^N such that if $1 > \delta_1 \geq \underline{\delta}_1$ and player 2 plays a ε -best response to σ_1^N the payoff to type ω_0 player 1 is at least $\bar{v}_1(\delta_2) - \eta$.*

Proof: By definition, we can choose an N so that for some σ_1^N and for all σ_2^N which are a best response to σ_1^N in the N -fold repeated game we have

$$(*) \quad (1/N) \sum_{t=1}^N u_1^t(\sigma_1^N, \sigma_2^N) \geq \bar{v}_1 - \eta/3$$

Moreover, since the ε -best response correspondence is upper hemi-continuous in ε , we may choose an ε so that $(1/N) \sum_{t=1}^N u_1^t(\sigma_1^N, \tilde{\sigma}_2^N) \geq \bar{v}_1 - 2\eta/3$ for all ε -best responses $\tilde{\sigma}_2^N$.

Moreover, we can choose $\delta_1 < 1$ so that for all σ_2^N (best response or not) $|(1/N) \sum_{t=1}^N u_1^t(\sigma_1^N, \sigma_2^N) - (1-\delta_1) \sum_{t=1}^N \delta_1^{t-1} u_1^t(\sigma_1^N, \sigma_2^N)| < \eta/3$. q

If σ^N is a profile consisting of a patient player type behavior strategy and a less patient player strategy in the N -fold repeated game, let $p(\sigma^N)$ be the probability distribution over N -length sequences of public outcomes induced by ρ . Notice that this is a finite vector.

Lemma 2: *For any $\varepsilon > 0$, N there exists a $\gamma > 0$ such that in the N -fold repeated game if $\|p(\sigma_1^N, \sigma_2^N) - p(\tilde{\sigma}_1^N, \sigma_2^N)\| < \gamma$ and σ_2^N is a ε -best response by player 2 to σ_1^N then it is a 2ε -best response to $\tilde{\sigma}_1^N$.⁹*

Proof: We identify type behavior strategies by player 1 that differ only at information sets that are unreachable given that strategy. It is sufficient to show that $p(\tilde{\sigma}_1^N, \sigma_2^N) \rightarrow p(\sigma_1^N, \sigma_2^N)$ implies $\tilde{\sigma}_1^N \rightarrow \sigma_1^N$. This in turn will follow if $p(\cdot, \sigma_2^N)$ has a continuous inverse. Since the domain of $p(\cdot, \sigma_2^N)$ is compact, the image of a closed 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 σ_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 σ_1^N for some $\tilde{\sigma}_2^N$ by Assumption 3 it has positive probability under σ_1^N, σ_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

⁹ 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.

We will refer to the N -fold repeated game as the *superstage* game. Fixing a strategy for all types of more patient player, let $M(K)$ be the finite set of probability distributions over types of more patient player that can be generated by Bayesian updating with no more than K observations of more patient player play.

Lemma 3: *Suppose the more patient player strategy in the infinitely repeated superstage game is such that type ω_0 plays σ_1^N in each superstage game. In any of the superstage games, let $\tilde{\sigma}_1^N(\mu')$ be the conditional probability of different pure strategies according to the beliefs of player 2, when his prior beliefs over more patient player types is μ' . For every $\lambda > 0, \gamma > 0, K$ there is an L such that if $\mu' \in M(K)$ the probability is less than λ that there are more than L superstage games with $\|p(\sigma_1^N, \sigma_2^N) - p(\tilde{\sigma}_1^N(\mu'), \sigma_2^N)\| \geq \gamma$.*

Proof: This is a restatement of Theorem 4.1 in Fudenberg and Levine [1992]. q

Proof of Theorem 1: 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 $\eta/4$.

The idea is to consider what happens when the more patient player repeatedly plays σ_1^N and player 2 plays a best response. Our conclusion follows by demonstrating that the more patient player 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 number μ so that $\delta_2^{\mu N}(\max u_2 - \min u_2) < \varepsilon(1 - \delta_2)/4$. We refer to the game consisting of μ repetitions of the superstage game (and μN repetitions of the stage game) as the *superduperstage* game. Apply Lemma 2 to any superduperstage game where the tolerance is $\varepsilon(1 - \delta_2)/2$ to find a value for γ . Apply Lemma 3 to the repeated superduperstage game using this value of γ as the tolerance level for beliefs, choosing the probability λ that this tolerance level is exceeded to be such that $\lambda(\max u_1 - \min u_1) < \eta/4$, and choosing $K = \mu N$. We refer to superduperstage games in which the tolerance is exceeded as anomalous.

Player 2 is playing at worst an $\varepsilon(1 - \delta_2)/2$ best response to his beliefs in the superduperstage game, since $\delta_2^{\mu N}(\max u_2 - \min u_2) < \varepsilon(1 - \delta_2)/4$. By Lemma 2 (except in the anomalous case) player 2 is playing an $\varepsilon/(1 - \delta_2)$ best response to the μ -fold repetition of σ_1^N . Consequently, player 2 is playing an ε -best response to σ_1^N in the first of the superstage games of the superduperstage game. By Lemma 1, in the first of the

superstage games of a non-anomalous superduperstage game player 1 gets at least $\bar{v}_1(\delta_2) - \eta/2$.

Taking account of the probability λ that there are more than L anomalous superduperstage games and the probability $1 - \lambda$ that there are less than or equal to L , we conclude the more patient player gets an expected average present value of at least $\delta_1^{\mu NL}(\bar{v}_1(\delta_2) - 3\eta/4) + (1 - \delta_1^{\mu NL}) \min u_1$ in all first stages of all superduperstage games combined.

Now consider the infinitely repeated game beginning in any period $\kappa N + 1$, $1 \leq \kappa \leq \mu - 1$. This is identical to the game that begins in period 1, except that the prior of player 2 may have changed. We may organize this game also into superstage and superduperstage games, and since Lemma 3 applies to all priors reachable during periods up to μN the previous argument shows that the more patient player receives an expected average present value of at least $\delta_1^{\mu NL}(\bar{v}_1(\delta_2) - 3\eta/4) + (1 - \delta_1^{\mu NL}) \min u_1$ in the first stages of these superduperstage games. However, the first stage of one of the superduperstage games in the game beginning in period $\kappa N + 1$ is stage κ of one of the superduperstage games in the game beginning in period 1, so we average over all stages of the superduperstage games beginning in period 1 to conclude that the more patient player receives an expected average present value of $\delta_1^{\mu NL}(\bar{v}_1(\delta_2) - 3\eta/4) + (1 - \delta_1^{\mu NL}) \min u_1$ for the entire game. Letting $\underline{\delta}_1 \rightarrow 1$ now gives the desired result. q

4. A Patient Less Patient Player

We now investigate what happens as $\delta_2 \rightarrow 1$. Our second theorem shows that if we consider a patient less patient player then player 1 can essentially achieve a payoff that is equal to the maximum he can get while giving player 2 his 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, when this player is arbitrarily patient. Then we ask how this bound behaves as also player 2 becomes very patient. Implicit in this construction is that player 1 is always infinitely more patient than player 2. (If we reverse the order of limits, then the result does not hold.)

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*

$$\liminf_{\delta_2 \rightarrow 1} \liminf_{\delta_1 \rightarrow 1} N_1(\delta_1, \delta_2) \geq \max V_1^*$$

Before turning to the proof, it is worth noting that if we allowed types of player 2 as well as types of patient player, this result would remain valid, and the proof of this result (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 of less patient player 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. As we make player 2 increasingly patient, Theorem 2 allows (and in fact requires) us to make player 1 even more patient.

Note also that the definition of V^* makes use of the *pure* strategy minmax. As was argued in Section 2, we only allow player 1 to establish a reputation for 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. Our pure strategy definition implies that $\max V_1^*$ is the largest socially feasible and individually rational payoff for player 1, if player 2's pure strategy minmax payoff is equal to his mixed strategy minmax payoff. If we allowed a reputation for mixed strategies, we could use the usual socially 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 here.

Proof: Given $v = (v_1, v_2)$ with $v \in \text{int} V^*$ let $\vec{a}_1 = (\vec{a}_1^1, \dots, \vec{a}_1^N)$ be a sequence of actions by player 1 such that there is a sequence of actions for player 2 such that in the N -times repeated game 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 in which player two has the time average payoff as a payoff function. In the following we call each MN -fold repetition of the game a superstage game, and hence we consider a K -fold repetition of the superstage game. Denote by $\vec{\vec{a}}_1 = (\vec{\vec{a}}_1, \dots, \vec{\vec{a}}_1)$ the sequence of M repetitions of \vec{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 the k -th superstage game.

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

Claim: Given $\varepsilon > 0$ there are numbers η , K , N , M , and P with $P/K < \varepsilon$ such that for any (time average) best response $\vec{\vec{\sigma}}_2$ to $\vec{\vec{\sigma}}_1$

- (i) $\text{prob}\{v_1 - u_1^k < \varepsilon\} > 1 - \varepsilon$ for all superstage games $k = 1, \dots, K - P$ for which player 1 chooses $\vec{\vec{a}}_1$ (i.e., for all superstage games for which player 1 does not use his punishment strategy.)
- (ii) the fraction of superstage games in which player 1 uses his punishment strategy is smaller than ε 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 $\vec{\vec{\sigma}}_2$ is discounted best response with δ_2 sufficiently close to 1. Moreover, the claim implies that the average payoff to player 1 in the KMN -fold repeated game is greater than $v_1 - C\varepsilon$ where C is a positive constant independent of ε . This is the desired result.

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

Denote by Eu_k^1 the expected payoff of player 1 in superstagegame k . Given K and η we can choose M sufficiently large, so that

(a) if $v_1 - Eu_1^k - \varepsilon/2 > 2\eta$ in any superstage game then punishment occurs with probability greater than $(1 - \eta)$;

(b) if $v_1 - Eu_1^k - \varepsilon/2 < \eta$ for all k , then the probability that no punishment occurs in any superstagegame 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

$$(*) \quad v_1 - Eu_1^k - \varepsilon/2 < 2\eta$$

in all of the first $K - P$ superstage games for any best response of player 2 and 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 ε fraction of the first $K - P$ superstage games with probability greater than ε . We claim that this implies that a profitable deviation exists. Suppose player 2 deviates so that $v_1 - Eu_1^k - \varepsilon/2 < \eta$ for all $k=1, \dots, K - P$. Since (*) has to be satisfied in every superstagegame $k = 1, \dots, K - P$ this deviation can be chosen so that the loss in every superstagegame is bounded above by ηB (where B is a constant that depends only on the payoff matrix). Note that (after the deviation) the probability that a punishment occurs in any superstagegame is smaller than η . Thus player 2 also improves his average payoff over the first $K - P$ superstage games by at least $(\varepsilon^2 - \eta)(v_2 - \bar{u}_2) \cdot P$ by reducing the probability of punishment. Consequently, if we choose $\beta = B / (v_2 - \bar{u}_2) \cdot P$ player 2 gains from the deviation and part (ii) of the claim follows. \square

5. An Example

SAVING FN¹⁰

This example thus shows that the conclusion of theorem 1 below fails with perfectly observable actions, even if the equilibrium concept is strengthened.¹¹

SAVING FN¹²

¹⁰ Although sequential equilibrium has not been defined for infinite games, in the finitely repeated versions of the two-types example considered here perfect Bayesian equilibrium (PBE) and sequential equilibria coincide (Fudenberg-Tirole [1991]). We use the simpler PBE concept because its conditions on beliefs are easier to check.

¹¹ However, since for a fixed δ_2 the example requires an upper bound on the probability that 1 is the commitment type, it is not a counterexample to the strengthened version of theorem 2.

¹² Moreover this example is robust to small changes in payoffs, and 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.

REFERENCES

- Aoyagi, M. [1984] "Reputation and Dynamic Stackelberg Leadership in Infinitely Repeated Games," mimeo.
- Celentani, M (1991), "Reputation with Deterministic Stage Games," UCLA Working Paper 636S.
- Celentani, M. and W. Pesendorfer (1992), "Reputation in Dynamic Games," Northwestern University CMSEMS Discussion Paper 1009.
- Cripps M., K. Schmidt, and J. P. Thomas (1993), "Reputation in Perturbed Repeated Games", Discussion Paper No. A-410. Projektbereich A, Universitaet Bonn.
- Cripps, M. and J.P. Thomas [1994] "Reputation and Equilibrium Selection in Two-Person Repeated Games without Discounting," 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.
- Fudenberg, D. and D.K. Levine (1989), "Reputation and Equilibrium Selection in Games with a Single Patient Player," *Econometrica* 57,759-778.
- Fudenberg, D. and D.K. Levine (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. 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.
- K. Schmidt (1993) "Reputation and Equilibrium Characterization in Repeated Games with Conflicting Interests", *Econometrica*, 61, 325-351.