REPUTATION AND COMMITMENT IN TWO-PERSON REPEATED GAMES WITHOUT DISCOUNTING

By Martin W. Cripps and Jonathan P. Thomas[1]

February 3, 1995

Abstract: Two-person repeated games with no discounting are considered where there is uncertainty about the type of the players. If there is a possibility that a player is an automaton committed to a particular pure or mixed stage-game action, then this provides a lower bound on the Nash equilibrium payoffs to a normal type of this player. The lower bound is the best available and is robust to the existence of other types. The results are extended to the case of two-sided uncertainty. This work extends Schmidt (1993) who analysed the restricted class of conflicting interest games.

Keywords: Reputation, Folk Theorem, repeated games, incomplete information.

JEL **Classification Nos.:** C73, D83, L14.

1. Introduction

"Reputation effects" arise when a player in a dynamic game is able to exploit some uncertainty that other players have concerning her preferences. There may be some probability that the player is of a type which would play in a particular way independently of the strategies of the other players. If, however, the player is not of this type, she might nevertheless wish to commit herself to playing in this way, because by mimicking the strategy of this type, the player can build up a "reputation" for following the type's strategy, even if the initial probability of the type is very small.

We shall consider two-player repeated games with no discounting, perturbed so that one of the players may be an automaton committed to a particular stage-game action. For all possible stage games and for each such automaton, we shall calculate a lower bound on the set of Nash equilibrium payoffs for this player, and thus we shall be able to describe the benefit of acquiring a reputation in any given stage game. Moreover we show that this bound is the best available. We allow players to acquire reputations for playing mixed actions in the stage game, and we generalise our results to account for games where there are non-negligible probabilities for the commitment type. The bounds we give are robust to the existence of other types. Finally the bound is shown to apply also in the case of two-sided incomplete information.

Our results extend those of Schmidt (1993), [2] who considers the effects of building a reputation for following the strategy of an automaton which plays the action that minmaxes the opponent. His analysis builds on that of Fudenberg & Levine (1989, 1992) who considered games where a single long-run player faces a sequence of short-lived (one period) opponents. Their key idea is that by mimicking the automaton action, the long-run player can eventually "convince" the short-run opponents that this action will be played in the next period, so the latter will play a best response. A patient long-run player will thus receive approximately the payoff she would get from public commitment to the automaton action. The payoffs from adopting such a strategy must provide a lower bound on any Nash equilibrium payoff. Schmidt demonstrates that these results can be extended to the case of a long-run opponent when the automaton plays the action which minmaxes the opponent.

Schmidt shows that if the automaton which always minmaxes the opponent has positive prior probability, then as the first player's discount factor converges to one (holding the opponent's discount factor fixed), so her Nash equilibrium average payoffs will be bounded below by an amount converging to the payoff she would get from committing to this action in the stage game. In games of "conflicting interests" this is shown to lead to a powerful lower bound: a game is of

---

[2] We originally obtained our results independently of this paper though the current version of our paper has benefitted considerably from our reading of Schmidt's paper.

conflicting interests if the action that one player, say player 1, would *most* like to commit to in the stage game[3], simultaneously holds the opponent, player 2, down to his minmax payoff. In such a game the lower bound associated with the minmaxing automaton will equal the best commitment payoff. Schmidt's results leave open the question of what lower bounds can be furnished from the possibility of mimicking automata playing non-minmax strategies, and whether these lower bounds may be better in games not of conflicting interests; or putting it differently, what is the best lower bound in an arbitrary repeated game? This is the question we address.

In Proposition 1 we give a lower bound derived from an *arbitrary* automaton action in *any* stage game, coinciding with Schmidt's bound when the automaton is a type playing the action which minmaxes the opponent. We also show how the bound is defined when the prior probability of the commitment type is non-negligible. In contrast to the analysis of Schmidt, who assumes discounted payoffs, we look at the no-discounting case.[4] Thus while Schmidt's result requires the first player to be potentially arbitrarily patient relative to its opponent, our analysis shows that the lower bound holds when both players are *equally* and infinitely patient (and in perturbed games with two-sided incomplete information we can therefore get a simultaneous lower bound for each player).[5]

Our lower bound is defined as the least payoff player 1 can get in the stage game when playing the automaton action, given that player 2 responds with a (possibly mixed) action which gives him at least his minmax payoff against the automaton action. (Clearly for a minmaxing automaton this coincides with the payoff from public commitment to this action since individually rational responses to the commitment strategy must also be best responses.) If the automaton is committed to playing an action other than the minmax action, then with a long-run opponent the reputation for playing that action is *not* sufficient to force the opponent to play a best response; the reputation is only for the *on-equilibrium path* behaviour of player 1, and the latter cannot convince player 2 that she will play the automaton action *off-equilibrium path*. Thus, rather than playing a best response to the automaton action each period, player 2 might play any individually rational response (on average) to the automaton action. Indeed, our tightness result (Section 5) shows that when the automaton is the "best" one, such an equilibrium can be constructed even when the game is perturbed to include other automata and other "rational" types, and so our bound is in this sense the best possible. The following example illustrates this lower bound.

---

[3]Assuming that the opponent plays the least favourable best response from player 1's point of view, so this can differ from the usual Stackelberg action when the best responses are not unique.

[4]In related work Watson (1992) studies reputation building in supergames where players do not discount payoffs and concentrates on equilibria where players use "forgiving strategies"; this restriction of the strategy space leads to stronger results than can be obtained in our framework.

[5]Time averaging is of course the limit of many sequences of pairs of discount factors, including sequences where player 1 becomes much more patient than player 2, the case studied by Schmidt. Proposition 1 below is extended in Cripps, Schmidt and Thomas (1993) to this latter case.

E<small>XAMPLE</small> 1:   Consider the stage game given below:

$$
\begin{array}{c c}
 & \begin{array}{c c} \text{L} & \text{R} \end{array} \\
\begin{array}{c} \text{T} \\ \text{B} \end{array} &
\left[\begin{array}{c c} (2,1) & (0,-1) \\ (0,0) & (0,-1) \end{array} \right]
\end{array}
$$

In an equilibrium of the supergame with no discounting and no uncertainty, player 1 (the row player) could get any payoff between 0 (her minmax payoff) and 2. Suppose the game is perturbed and that with probability p player 1 may be a commitment type always playing T. To calculate our lower bound on player 1's equilibrium payoff in the repeated game of incomplete information, consider in the stage game player 1 playing T and player 2 responding in such a way as to get at least his minmax payoff (0). This implies that player 2 must put at least a probability of one half on L; subject to this constraint the least payoff player 1 can get from T is 1. This is a lower bound for any p>0, and considerably restricts the set of equilibrium payoffs.[6] Notice that the bound attainable from a commitment type playing B (the action which minmaxes player 2) is zero, so in this example Schmidt's bound has no force.

In some games other than conflicting interest games, however, the best type to mimic is the one which minmaxes the opponent (best in the sense of providing the greatest lower bound), and, as already mentioned, the arguments in Schmidt (1993) can be used to calculate the bound for such types.   Whenever the minmax action is mixed, however, the Schmidt bound[7] is weak and there will be some other type which provides a better bound. This is illustrated in the next example.

E<small>XAMPLE</small> 2:   The "Battle of the Sexes" game:

$$
\begin{array}{c c}
 & \begin{array}{c c} \text{L} & \text{R} \end{array} \\
\begin{array}{c} \text{T} \\ \text{B} \end{array} &
\left[\begin{array}{c c} (3,1) & (0,0) \\ (0,0) & (1,3) \end{array} \right]
\end{array}
$$

To minmax the opponent, the row player (player 1) needs to play T with probability 3/4. Suppose that there is a positive probability automaton which plays this mixed action; then the lower bound derived by Schmidt is of no value because a best response to a minmaxing automaton by player 2 is any mixed action, including playing R which gives player 1 a payoff of 1/4; this is less than player 1's minmax payoff of 3/4. Our results show that a type which always plays T gives a bound of 9/4, a considerable restriction on the possible equilibrium payoffs. The general superiority of a bound

---

[6]Our more general statement of the bound in this example is 1+p, implying that if the prior probability of the automaton is non-negligible, a higher value for the bound is attained.

[7]His results are proved only when the minmax action is pure; that his argument goes through when it is mixed is established in Cripps, Schmidt and Thomas (1993).

based on a non-minmaxing automaton in the case where the minmax action is mixed is discussed in Section 5.

An outline of the paper is as follows. In Section 2 a complete information repeated game is described; in Section 3 a perturbed version of this game is given with player 1 being a number of different types; in Section 4 the lower bound on player 1's payoff is established; in Section 5 the value and tightness of the bound is analysed; in Section 6 the two-sided reputation problem is analysed.

2. The unperturbed game

We begin with a standard complete information infinitely repeated game Γ with two players: 1 and 2. Each period t=1,2,... player 1 selects an action from a finite set I={1, 2, ..., I} and player 2 simultaneously selects an action from a finite set J={1, 2, ..., J}. Payoffs from the stage game are given by a pair of payoff matrices (A,B), so from actions (i,j) player 1 receives A(i,j) and player 2 receives B(i,j).

We assume players can observe all previous moves. Let H_t, t=1,2,..., be the set of histories h_t up to, but not including, stage t: $H_t = (I \times J)^{t-1}$, and we define H_1 to consist of a single element. By Kuhn's Theorem we can restrict attention to behaviour strategies. A behaviour strategy for player 1 is a sequence of maps $\{\sigma_t\}_{t=1}^{\cdot}$ where $\sigma_t : H_t \oslash \Delta^I$, t=1,2,...(denoting by Δ^n the unit simplex in R^n). Likewise for player 2, a behaviour strategy is $\{\tau_t\}_{t=1}^{\cdot}$ where $\tau_t : H_t \oslash \Delta^J$, t=1,2,... . Payoffs in the repeated game are defined as the (Banach) limit of expected average stage game payoffs (it will be convenient to delay formal definitions until the next section), with Nash equilibria defined as usual. Denote by minmax[1](A) and minmax[2](B) the respective minmax payoffs of players 1 and 2. The following notation will also be needed. We start with an abuse: given u Δ^I and v Δ^J we let A(u,v) = Σ_i $_{I,j}$ $_Ju_iv_jA(i,j)$ be player 1's expected payoff when mixed stage-game strategies (hereafter *mixed actions)* u and v are selected. Define B(u,v) analogously. Now define player 2's *best response payoff* against u Δ^I by BR[2](u) := max$_v$ $_{\Delta^J}$ B(u,v).

3. The perturbed game of incomplete information

This section introduces a game $\Gamma(p)$ which may be considered as a perturbed version of the original game. In the new game, player 1 may be one of a number of types, including the type previously described, and player 2 does not know what type of opponent he is playing against (although he knows his own payoff matrix which is fixed). Using Harsanyi's (1967) notion of a game of incomplete information, we identify player 1 with a type k K, where K is a countable set. It is

assumed to be common knowledge that a type $\kappa \in K$ is selected at the beginning of the game according to a probability measure $p=(p^1,p^2,p^3,\dots)$ on K.[8] We identify the type described in Section 2 with $k=1$, and refer to this type as the 'normal type'. While we wish to allow for very general types of player 1, including automata, at least some of the types may be similar to type $k=1$ in that they have preferences defined by average stage-game payoffs. Some of the other types may be automata, by which we mean types k with a fixed strategy $\{\sigma_t^k\}_{t=1}^{\bullet}$. Of particular interest will be commitment types playing the same pure, or mixed, stage-game action each period independently of history. There may also be other types, for example, with discounted payoffs, but since we are only interested in necessary conditions, explicit description of such types is superfluous.[9]

Type k of player 1 plays a strategy $\sigma^k=\{\sigma_t^k\}_{t=1}^{\bullet}$, and we define $\sigma = (\sigma^k)_{k \in K}$, while player 2 plays $\tau = \{\tau_t\}_{t=1}^{\bullet}$; hence we can define a probability space as follows. Let $H_8 =\Pi_{t=1}^{\bullet}(I\infty J)$ be the set of infinite histories. For each $t=1, 2, \dots$, we define h_t to be the σ-field generated on H_8 by H_t and let h_8 be the σ-field generated by the union of the h_t's. Let $\Omega = H_8 \infty K$ be the set of states of nature and endow this with the σ-field $h_8 \; 2^K$. Strategies (σ,τ) and probabilities p determine a probability measure P on Ω and let P^k be the conditional probability measure on Ω given player 1 is type k (for k such that $p^k>0$). We will write $E[.]$ for expectation with respect to the measure P, and $E^k[.]$ for conditional expectation given player 1 is type k. Where it is necessary to emphasise the dependence of the expectation on the strategy played we shall also write $E_{\sigma,\tau,p}$ for E and $E^k_{\sigma,\tau}$ for E^k.

Long-run averages need not converge, so we use some Banach Limit L to define payoffs.[10] The bounds on equilibrium payoffs we obtain will always be independent of the particular form of Banach limit taken. The average payoffs for type 1 of player 1 and for player 2 up to period T are respectively

$$a_T = \frac{1}{T}\sum_{t=1}^{T} A(i_t,j_t), \qquad b_T = \frac{1}{T}\sum_{t=1}^{T} B(i_t,j_t),$$

and repeated game payoffs are respectively

$$a = L[\{E^1(a_T)\}_{T=1}^{\infty}], \qquad b = L[\{E^1(b_T)\}_{T=1}^{\infty}].$$

---

[8] We use κ to denote the random variable and k for a particular value.

[9] That is to say, the lower bounds we obtain on payoffs arise solely from the consideration that type $k=1$ optimises.

[10] L is a real linear operator on the space of all real bounded sequences, and for any $x=\{x_T\}$, $y=\{y_T\}$, satisfies (i) $L[\lambda x + \mu y] = \lambda L[x] + \mu L[y]$ for $\lambda, \mu \in R$, (ii) $\limsup_{T\varnothing} x_T = L[x] = \liminf_{T\varnothing} x_T$; it follows from (ii) that $L[x] = \lim_{T\varnothing} x_T$ whenever the latter exists. See Dunford & Schwartz (1988) and Myerson (1991, Chap. 7).

Henceforth $L[\{x_T\}]$ will be abbreviated to $L[x_T]$ and it will be assumed that the limits are taken with respect to the index T.[11] (Notice that to calculate type 1's equilibrium payoff the expectations are conditional on $\kappa=1$.) Necessary conditions which must be satisfied by any Nash equilibrium (σ,τ) are [12]

$$L[E^1{}_{\sigma,\tau}(a_T)] \;=\; L[E^1{}_{\sigma,\tau}(a_T)] \qquad \forall\sigma',$$
$$L[E_{\sigma,\tau,p}(b_T)] \;=\; L[E_{\sigma,\tau',p}(b_T)] \qquad \forall\tau'.$$

The lower bound we develop exploits *necessary* conditions of a Nash equilibrium. This is only a meaningful exerciseif there are equilibria which exist in this environment. The existence question will be discussed in Section 5.

We shall also need notation describing player 2's priors at the beginning of period s. For a fixed equilibrium (σ,τ), let $p_s^k = p_s^k(h_s)$ be the conditional probability of the true type of player 1, κ, being type k, given h_s. The value p_s^k is a random variable measurable with respect to the sigma-field h_s, and if $1_{\kappa=k}$ is the indicator function for the event $\kappa=k$, then the priors can be written $p_s^k = E[\,1_{\kappa=k} \mid h_s\,]$. It follows from this that p_s^k is a martingale with respect to the filtration $\{h_t\}$, with $p_1^k=p^k$.

4. The value of a reputation to player 1

We shall now construct a lower bound on player 1's equilibrium payoff using simple martingale arguments—because the rate of convergence does not matter to time-averaged payoffs such methods are sufficient, in contrast to the discounted case as in Schmidt (1993) and Fudenberg and Levine (1989, 1992). To do this we first estimate a lower bound on player 2's equilibrium payoff after each history. We then show that if player 1 mimics a commitment type, the bound on player 2's payoff implies that player 1 must receive at least a certain payoff.

Below, we introduce a second type of player 1 which corresponds to the event $\kappa=\hat{k}$. Type $\hat{k}$ plays the fixed stage game (possibly mixed) action $\hat{u}\ \Delta^I$ in every period of play independently of the history. Thus we think of type $\hat{k}$ as a commitment type.[13] We henceforth assume that type $\hat{k}$ occurs with strictly positive probability: $p^{\hat{k}}>0$.

---

[11] Where there is no ambiguity we shall write LE[.] for L[E(.)].

[12] We present our results as restrictions on *Nash* equilibrium payoffs but they only depend on the following inequalities; hence our results will also apply to some generalizations of Nash equilibrium.

[13] This is not a type in the sense of Harsanyi (1967-68) which maximises a payoff function; rather it adheres to a fixed strategy. For an analysis of one-sided incomplete information when K consists of a finite number of types with stage game payoff matrices see Shalev (1988); also see Forges(1992). Shalev shows that all equilibria are payoff equivalent to equilibria in which each type completely reveals itself at the beginning of the game. It can be shown that this is not true of the class of games considered here. Fudenberg and Levine (1992) show it is possible to derive mixed strategy commitment types from a belief distribution over pure strategy types.

The bound we calculate in Proposition 1 can be explained with the following intuition. Suppose we have found a Nash equilibrium of the perturbed game $\Gamma(p)$ and, after some history h_s, the normal type of player 2 has a prior $p_s^{\hat{k}}$ that his opponent is the commitment type playing the action $\hat{u}$ in every period. From this point onwards player 2 *could* use the following review strategy: play a best response to $\hat{u}$ for a finite time period, while simultaneously watching the opponent's actions to find out whether player 1 really is playing $\hat{u}$ on average. As player 2 is long-sighted the costs associated with any finite experiment of this sort are negligible, whilst from the laws of large numbers he will be able to verify whether his opponent is playing close to $\hat{u}$ on average. With probability $p_s^{\hat{k}}$, player 2 faces type $\hat{k}$ and therefore receives approximately the payoff $BR2(\hat{u})$. If player 2 suspects a deviation from $\hat{u}$, then he must always be able to receive a payoff of at least $minmax2(B)$. At any stage of play in an equilibrium of the perturbed game, player 2's expected payoff must therefore be no less than $p_s^{\hat{k}} BR2(\hat{u}) + (1 - p_s^{\hat{k}})\, minmax2(B)$; see Lemma 1.

Now consider player 1's payoff in this Nash equilibrium of $\Gamma(p)$. Player 1 always has the option of mimicking the commitment type, that is, playing $\hat{u}$ each period. If she does this, player 2's beliefs $p_s^{\hat{k}}$ will converge (almost surely) to some positive value. This is only possible if player 2's beliefs only assign positive probability to types whose actions are near to $\hat{u}$ each period, as otherwise player 2's beliefs would continue to change. Hence "in the limit" player 2 expects $\hat{u}$ to be followed on the equilibrium path, and from above he must receive at least $p_s^{\hat{k}} BR2(\hat{u}) + (1 - p_s^{\hat{k}})minmax2(B)$ against this strategy. Taking expectations without conditioning on a particular history, h_s, gives a lower bound on player 2's expected payoff against the strategy of $\hat{u}$ each period. This restricts, by feasibility, player 1's expected payoff, and hence gives us our bound.

For any history h_s, player 2's expected payoffs for the future are $L[E(b_T \mid h_s)]$. Following the above argument we shall give a lower bound for player 2's expected payoffs, after any positive probability history at a Nash equilibrium, which depends upon $p_s^{\hat{k}}$, $minmax2(B)$ and player 2's best response payoff to the type $\hat{k}$, $BR2(\hat{u})$.

Lemma 1: *Let* s=1 *be given; then at any Nash equilibrium of* $\Gamma(p)$

$$LE[\, b_T \mid h_s\,] \;=\; p_s^{\hat{k}}\, BR2(\hat{u}) \;+\; (\,1 - p_s^{\hat{k}}\,)minmax2(B) \qquad\qquad a.s.$$

(Proofs of all lemmas are contained in Appendix A.) We shall use this result to construct a bound on player 1's payoffs. Before the main result is proved, two technical lemmas are needed. The first shows that conditional on type k's strategy being followed, p_t^k is a submartingale and it converges

to a positive number with probability one. The intuition for this is straightforward: if player 1 is type k, then the priors that player 2 attaches to this type on average cannot fall. (Recall that the expectation operator $E^k[.]$ denotes expectations conditional on player 1 playing as type k.)

    Lemma 2: *Suppose that* $p^k > 0$. *Then at any Nash equilibrium of* $\Gamma(p)$
      (i) $p_t^k = E^k[\,p_{t+1}^k \mid h_t\,]$ P^k-a.s.,
      (ii) p_t^k *converges* P^k-a.s. *to a strictly positive random variable.*

Since p_t^k is a martingale, at a Nash equilibrium the priors converge almost surely by martingale convergence. A consequence of this must be that player 1 ultimately reveals less and less information about her type. As the priors converge, so the strategies of positive probability types must also converge as otherwise the priors could not be close to their limits. Thus player 2 cares less about what he will learn about the true type of his opponent, in the sense that payoffs conditional upon a positive probability type k will be close to unconditional payoffs:

    Lemma 3: *At any Nash equilibrium of* $\Gamma(p)$

$$p_s^k \mid LE^k[\,b_T \mid h_s] - LE[\,b_T \mid h_s]\mid \varnothing 0, \text{ a.s.,} \qquad \text{as } s\varnothing 8 \,.$$

In the light of Lemma 2(ii), Lemma 3 provides a way of approximating the unconditional payoff of player 2 by its conditional payoff against type $\hat{k}$. Lemma 3 is essential because it allows us to use the bound on player 2's unconditional payoffs to make statements about player 2's conditional payoffs.[14] Next, define the function

(1) $a^*(p^k, u) := \displaystyle\min_{v \in \Delta^J} \left\{ A(u,v) \mid B(u,v) \bullet p^k BR^2(u) + (1-p^k)minmax^2(B) \right\}$.

We can now state our main result.

---

[14] An example might make this clearer. Suppose that player 1 is either the normal type or the automaton with equal probability, and in equilibrium player 1 will reveal her type in period 1, with player 2 getting $BR^2(\hat{u})$ against the automaton and $minmax^2(B)$ against the normal type. In period 2, the conditional and unconditional payoffs converge to $BR^2(\hat{u})$ if $p_2^{\hat{k}} > 0$ (i.e. if player 1 is the automaton), so the convergence of Lemma 3 takes place within a period. Since the inequality of Lemma 1 must be satisfied for player 2's unconditional payoff, it will be satisfied here also for player 2's conditional payoff, strictly in this example. If full revelation does not, even in the limit, take place in equilibrium, then the conditional payoff may be closer, or equal, to the right hand side of the inequality of Lemma 1, but the inequality cannot be violated.

Proposition 1: *At a Nash equilibrium of* $\Gamma(p)$, *where a positive probability type* $\hat{k}$ *plays mixed action* $\hat{u}$ *at every stage of the game, the normal type* (k=1) *of player 1 receives at least* $a^*(p^{\hat{k}},\hat{u})$.

Proof: Let $\delta>0$ and $\phi>0$ be given. If player 1 mimics type $\hat{k}$, then $p_s^{\hat{k}}$ converges a.s. to some $p_\bullet^{\hat{k}}>0$ by Lemma 2(ii). Hence there exists an $\eta>0$ and an S such that

(2) $\qquad$ $\Pr[\ \exists s=S \text{ such that } p_s^{\hat{k}} < \eta \ |\ \kappa=\hat{k}\ ] < \delta$.

Next, because $P^{\hat{k}}$ is absolutely continuous with respect to P, the convergence of Lemma 3 holds $P^{\hat{k}}$-almost surely. Hence for any $\xi>0$, there exists an S' such that

$$\Pr[\ \exists s=S': p_s^{\hat{k}} \ | L\,E^{\hat{k}}[\ b_T\ |\ h_s] - LE[\ b_T\ |\ h_s]\ | > \xi\ |\ \kappa=\ \hat{k}\ ] < \delta.$$

Hence using (2) and defining S*=max{S,S'},

$$\Pr[\ \exists s=S^*: |\ L\,E^{\hat{k}}[\ b_T\ |\ h_s] - LE[b_T\ |\ h_s]\ | > \xi/\eta\ |\ \kappa=\ \hat{k}\ ] < 2\delta.$$

Setting $\xi=\phi\eta$ and using Lemma 1, we therefore have that there exists an S* such that

$$\Pr[\ \exists s=S^*: L\,E^{\hat{k}}[\ b_T\ |\ h_s] < p_s^{\hat{k}}\,BR2(\hat{u})+(\ 1- p_s^{\hat{k}}\ )minmax2(B)-\phi\ |\ \kappa=\hat{k}\ ] < 2\delta.$$

This implies

$$\Pr[\ L\,E^{\hat{k}}[b_T\ |\ h_{S^*}]\ \varepsilon\ p_{S^*}^{\hat{k}}\,BR2(\hat{u})+(\ 1- p_{S^*}^{\hat{k}}\ )minmax2(B)-\phi\ |\ \kappa=\hat{k}\ ] = 1-2\delta\ .$$

Now taking expectations of $L\,E^{\hat{k}}[\ b_T\ |\ h_{S^*}]$ without conditioning on history (but conditioning on $\kappa=\hat{k}$) and using Lemma 2(i) on $p_{S^*}^{\hat{k}}$ (the operators $E^{\hat{k}}[.]$ and L[.] can be transposed as h_s is a finite sigma field: see Hart (1985, Lemma 4.6)):

$$L\,E^{\hat{k}}[\ b_T\ ]\ \varepsilon\ (1-2\delta)(\ p^{\hat{k}}BR2(\hat{u})+(\ 1- p^{\hat{k}}\ )minmax2(B)-\phi)-2\delta M_B,$$

where $M_B:=max_{i,j}|B(i,j)|$. Since $\delta>0$ and $\phi>0$ are arbitrary, we have

(3) $\qquad$ $L\,E^{\hat{k}}[\ b_T\ ]\ \varepsilon\ p^{\hat{k}}BR2(\hat{u})+(\ 1- p^{\hat{k}}\ )minmax2(B)$.

Letting $F^{\hat{u}}=co\{\ (A(\hat{u},v),B(\hat{u},v)\ )|\ v\ \Delta^J\ \}$ be the convex hull of the set of stage-game payoffs when player 1 plays $\hat{u}$, then $(L\,E^{\hat{k}}[a_T], L\,E^{\hat{k}}[b_T])\ F^{\hat{u}}$ (by Hart (1985, Lemma 4.7)). Inequality (3) gives a lower bound on $L\,E^{\hat{k}}[b_T]$, which thus implies that player 1 receives at least the amount

$a^*(p^{\hat{k}},\hat{u})$ from imitation of type $\hat{k}$. Player 1's payoff at a Nash equilibrium is therefore bounded below by this quantity. *Q.E.D.*

Remark 1: The bound $a^*(p^{\hat{k}},\hat{u})$ is non-decreasing in the initial belief $p^{\hat{k}}$, and, not surprisingly, converges as $p^{\hat{k}}$ converges to one to the (least) payoff that player 1 would get from committing to $\hat{u}$ in the stage game. Of particular interest is a bound that holds for all $p^{\hat{k}}>0$. This gives a bound valid for even very small perturbations of the complete information game. Such a bound is found by setting $p^{\hat{k}} = 0$ in the definition of $a^*(p^{\hat{k}},\hat{u})$, so that player 2 is being restricted only to play individually rational responses to $\hat{u}$. If $\hat{u}$ minmaxes player 2, then $a^*(0,\hat{u})$ equals the payoff player 1 would get from commitment to $\hat{u}$ (this is the main result of Schmidt (1993), established in the discounted case).

5. Value and tightness of the bound

Three questions arise about the value of the bound of Proposition 1. First, to what extent does it restrict the payoffs that player 1 can receive in equilibrium? Second, is it tight in the sense that for a given commitment automaton it is the highest possible lower bound? Finally, is the existence of Nash equilibrium guaranteed (otherwise the bound may be meaningless)? We address all of these questions in this section.

We shall restrict attention to the bound which applies for all $p^{\hat{k}} > 0$ (see Remark 1). Consider the "best" type to mimic, in the sense of a type which maximises the lower bound $a^*(0,\hat{u})$ of Proposition 1; this yields a value for the bound of

$$
(4) \qquad \sup_{\hat{u}\ \Delta^I}\ \min_{v\ \{v\ \Delta^J\ |\ B(\hat{u},v)=\text{minmax}_2(B)\}} A(\hat{u},v) = \sup_{\hat{u}\ \Delta^I}\ \min_{v\ \Delta^J}\ A(\hat{u},v),
$$

where the right hand side of the inequality is minmax^l(A) by the "minmax theorem". Hence the difference between the best lower bound and player 1's minmax payoff is that in the definition of the former, player 2 can only minimise over mixed actions which give him at least his own minmax payoff, whereas in the definition of the latter this is unrestricted. Our lower bound is valuable to player 1 whenever this restriction matters. Examples 1 and 2 in the Introduction are games where there is a strict inequality in (4). *Only* in games of conflicting interests (see the Introduction), however, is the bound equal to the best payoff from public commitment in the stage game. It should also be stressed that in many games commitment has no value. For example, in Prisoners' Dilemma games reputation for following a fixed action is worthless and our best bound equals the minmax payoff.

A further issue affecting the value of the bound derived in this paper concerns whether the best bound will be delivered by an automaton playing the action that minmaxes player 2; the bound from such a type has already been characterised by Schmidt (1993) in the discounted case. Again, Examples 1 and 2 illustrate cases where this is not true, and in generic games, if the action which minmaxes player 2 is mixed, there will be a type which will provide a better bound than the minmaxing automaton. This follows from

Lemma 4: *If* (A,B) *is a generic* $I \infty J$ (I, J = 2) *bi-matrix game and if the zero-sum game* (-B,B) *has a mixed strategy equilibrium* (u*,v*), *then there exists* $z \in \Delta^I$ *such that*

$$\min_{v \in \Delta^J} \{ z^t A v \mid z^t B v = minmax2(B) \} > \min_{v \in \Delta^J} \{ (u^*)^t A v \mid (u^*)^t B v = minmax2(B) \} .$$

Hence from equation (1), $a^*(0, z) > a^*(0, u^*)$, where z and u* are as in the statement of the lemma; $a^*(0, u^*)$ is the bound from a small perturbation involving a minmaxing automaton and $a^*(0, z)$ is a superior bound arising from a small perturbation involving an automaton playing z. (The intuition behind this result is demonstrated by the discussion of Example 2 in the Introduction.) When player 1's minmaxing action is pure, however, an automaton playing this action may or may not (cf. Example 1) deliver the best bound, even for generic games.

The second question concerned the tightness of the bound: is it is possible to find a better lower bound than that described in Proposition 1? We will be interested only in the best possible automaton (see above); this automaton plays $\hat{u}^{\hat{k}}$ such that $a^*(0, \hat{u}^{\hat{k}}) = a^*(0,u)$ for all $u \in \Delta^I$ and we shall identify type $\hat{k}$ with this strategy. It is clear that a tightness result cannot be established for other automata in games in which, for example, the best automaton is also present.[15] Proposition 2 considers perturbations involving a finite number of (fixed stage-game action) automata, including this best automaton, and also a possible finite number of other normal (i.e. payoff matrix, zero discounting) types.[16] Provided all other types ($k \eth 1$, $\hat{k}$) have sufficiently small probability, an equilibrium with payoffs to player 1, type 1, arbitrarily close to $a^*(p^{\hat{k}}, \hat{u}^{\hat{k}})$ can be constructed. In Cripps, Schmidt and Thomas (1993) perturbations involving a single *pure-strategy* automaton playing a history dependent strategy are considered for the discounted case. These automata do not provide a bound better than would the 'best' automaton playing a fixed, pure, stage-game action.

---

[15] If $a^*(p^k, \hat{u}^k) > a^*(0, \hat{u}^{\hat{k}})$ then the proof goes through for this type k *for such a value of* p^k. Tightness can also be established for any automaton (which gives a bound above player 1's minmax payoff) in a perturbation involving *only* this automaton; a simple variant of the proof of Proposition 2 establishes this.

[16] For games involving a finite number of normal types only, an equilibrium exists in which player 1 gets $a^*(0, \hat{u}^{\hat{k}})$ or less: see Israeli (1989) and also the discussion in Forges (1992).

Proposition 2 also offers a partial answer to the third question we raised, that of existence. It is closely related to the tightness issue, and a similar construction can be used to establish existence under the same assumption of finiteness of the sets of automata and normal types.

Proposition 2: *Consider the perturbed game* $\Gamma(p)$ *involving a finite number* K *types of player 1: a normal type* (k=1) *with preferences* A(i,j), $\hat{k}$-1 (=1) *automata* (k=2,..., $\hat{k}$) *playing the distinct actions* $\hat{u}^k$ Δ^I *in every period, and* K- $\hat{k}$ (=0) *further normal types* (k= $\hat{k}$+1,...,K) *with preferences* A^k(i,j). *(i) (Existence) There exists a Nash equilibrium of this game. (ii) (Tightness) Suppose* a*(0, $\hat{u}^{\hat{k}}$)=a*(0,u) *for all* u Δ^I. *Then given* η>0 *and* $p^{\hat{k}}$>0 *we can choose* ε>0 *such that provided* p^k<ε *for* k≠1, $\hat{k}$, *there exists an equilibrium of* $\Gamma(p)$ *in which type* k=1 *of player 1 receives a payoff that is at most* η *larger than* a*($p^{\hat{k}}$, $\hat{u}^{\hat{k}}$).

See Appendix B for the proof. The equilibrium of part (ii) can be described briefly as follows.

(The normal type, k=1) At the start of the game this type will randomly select and mimic one from a subset of the automata. The subset comprises the 'best' automaton, type $\hat{k}$, together with any automaton which with full reputation would give type k=1 a payoff greater than a*($p^{\hat{k}}$,$\hat{u}^{\hat{k}}$).

(Other normal types) Types k>$\hat{k}$ play out deterministic sequences which, when player 2 plays a corresponding sequence, replicate the average outcome (relative frequencies over action profiles) of a best response to the monitoring phase (see below). Their actions differ during the first K periods allowing player 2 to provisionally identify which of these normal types he is facing (with mixed strategy automata full revelation may not be possible since one of the automata may also follow such a sequence with positive probability).

(Player 2) So long as player 1 follows one of the above deterministic sequences then player 2 follows a corresponding deterministic sequence. If player 1 plays an action inconsistent with any such sequence, play enters the *monitoring phase* in which player 2 eventually verifies, through a sequence of review phases, which of the automaton strategies is being followed. If some other strategy is being followed then this will repeatedly trigger a punishment strategy which eventually holds type k=1 down to a*(0,$\hat{u}^{\hat{k}}$), and at the same time guarantees player 2 at least minmax2(B). In the review phases for the automaton strategies which type k=1 might follow, player 2 plays to give type k=1 the same payoff (of a*) that is within η of a*($p^{\hat{k}}$,$\hat{u}^{\hat{k}}$), and against the other automaton strategies player 2 plays a best response.

(Equilibrium) Type 1 of player 1 is indifferent about mimicking one of the subset of automata, and prefers this to facing repeated punishment phases. By construction none of the other normal types

can benefit by deviations. Likewise, player 2's strategy is constructed to be deterministic in all but punishment phases, and deviations are minmaxed by normal types. (Deviations by player 2 during punishment phases do not affect his long-run payoff.) Against the other normal types player 2 receives at least his minmax payoff since these outcomes correspond to monitoring phase outcomes, and player 2 always gets his minmax payoff in that case. His strategy against the automaton strategies together with the initial randomization of type k=1 are chosen to ensure that the threat of being minmaxed by normal types is sufficient to prevent player 2 from deviating.

6. Two-sided uncertainty

If player 2 is one of many possible types, the arguments above will still apply. Let player 2's types g belong to a countable set G and let q be an independent probability measure on this set describing how player 2's type is selected at the beginning of play. Identify the type described in Section 2 as g=1. To denote the game of two-sided uncertainty we shall use $\Gamma(p,q)$.

As player 1 mimics the commitment type by always playing $\hat{u}$, so player 2's learning will still settle down, as in Lemmas 2 and 3, and type g=1's expected payoff will still be bounded below by the deviation payoff as in Proposition 1. The only problem comes in evaluating player 1's payoff to mimicking type $\hat{k}$, because sometimes she will be playing against type g=1 of player 2 and sometimes she will be playing against other types of player 2. We therefore get a revised statement of the lower bound a* on player 1's payoff in $\Gamma(p,q)$, where we define $M_A := \max_{i,j} |A(i,j)|$:

$$a^* = q^1(\min_{v \ \Delta^J} \ \{ A(\hat{u},v) \mid B(\hat{u},v) = p^{\hat{k}}BR2(\hat{u}) + (1- p^{\hat{k}})\text{minmax2}(B)\}) + (1-q^1)(-M_A).$$

With probability q^1 player 2 faces the same incentive constraints as in the one-sided case; with probability $(1-q^1)$ however he does not.[17] In this paper we are interested in small perturbations of repeated games of complete information. Thus we are particularly interested in the lower bound on Nash equilibrium payoffs to both players as the probabilities of types k, g > 1 become small. As p^1 and q^1 tend to unity the lower bound for player 1's Nash equilibrium payoff becomes $a^*= a^*(0, \hat{u})$ (i.e. the one-sided bound calculated for $p^{\hat{k}}$ near zero); and likewise, a symmetric expression gives a lower bound for player 2 if he can mimic an automaton (which has positive probability) playing a fixed stage-game mixed action $\hat{v}$.

The possibility of two-sided uncertainty also implies that no Nash equilibrium may exist in the perturbed game. For example, if the Battle of the Sexes game (Example 2 in the Introduction) is the stage game and player 1 can mimic a commitment type playing T with probability one, then a*–

---

[17] Here it is essential that p and q are independent probability distributions.

2.25 if p^1 and q^1 are close to one. Similarly, if player 2 can mimic a commitment type playing R with probability one at every stage game, then 2.25 is approximately a lower bound on his equilibrium payoff. But there is no feasible payoff pair giving both players a payoff of 2.25, so no equilibrium can exist. The possibility of non-existence in undiscounted repeated games of incomplete information is well known: see Mertens, Sorin & Zamir (1991). For a similar example in a standard repeated game with "known own-payoffs" see Koren (1988).

7. Conclusion

In this note we have established the extent to which a player can guarantee herself a certain level of payoffs by exploiting reputation effects when she is playing against a long-run opponent. Even a small amount of uncertainty on the part of an opponent about a player's type can lead, in a class of games, to a large reduction in the set of possible equilibrium payoffs. Moreover the lower bound we derive is tight, and is robust to the existence of other possible types.

Department of Economics, University of Warwick, Coventry CV4 7AL, U.K.

APPENDIX A: PROOFS OF LEMMAS

Proof of Lemma 1: Without loss of generality, fix any h_s which has positive probability under type $\hat{k}$ (player 2 must optimise against player 1's strategy thereafter). If $BR^2(\hat{u}) = \text{minmax}^2(B)$ the result is true, because player 2 can always obtain his minmax payoff by playing independently of his opponent's type. If $BR^2(\hat{u}) > \text{minmax}^2(B)$ we shall construct a strategy for player 2 that ensures he receives a payoff arbitrarily close to $p_s^{\hat{k}} BR^2(\hat{u}) + (1 - p_s^{\hat{k}})\text{minmax}^2(B)$. Define f(i,s,t) to be the random variable which counts the number of times action i is played by player 1 in the periods s, ...,t. Suppose player 1 is truly playing $\hat{u} = (\hat{u}_1, \hat{u}_2, ..., \hat{u}_I)$; then, by the strong law of large numbers, $f(i,s,t)/(t-s-1) \varnothing \hat{u}_i$, as $t \varnothing$, for all i, almost surely conditional on h_s and $\hat{k}$. Hence for any $\varepsilon > 0$, $\delta > 0$ there exists by Egoroff's Theorem a $\tau(\varepsilon, \delta)$ such that

$$\Pr\left[\ \left|\ \frac{f(i,s,t)}{t-s+1} - \hat{u}_i\ \right| < \varepsilon, \ \ \text{for all } t > \tau(\varepsilon, \delta) \text{ and all } i\ \middle|\ \hat{k}, h_s\right] > 1-\delta\ .$$

Player 2 can be almost certain that if he faces type $\hat{k}$, and waits until after period $\tau(\varepsilon, \delta)$, and then calculates the relative frequency distribution of player 1's actions, this will be (and will remain) within ε of $\hat{u}$.

Let player 2 play a fixed pure strategy best response to $\hat{u}$ in periods s, s+1, ..., $\tau(\varepsilon, \delta)$, and once this review session is over continue to play the best response to $\hat{u}$ until there exists an i such that $|(f(i,s,t)/(t-s-1)) - \hat{u}_i| = \varepsilon$. If this condition is satisfied for the first time in some period t, player

2 plays to achieve his minmax payoff in all future periods. The condition is violated with probability $1-\delta$ if player 1 is truly playing $\hat{u}$. Player 1 is a commitment type playing $\hat{u}$ with probability $p_s^{\hat{k}}$, so with probability at least $(1-\delta)\, p_s^{\hat{k}}$ player 2 gets from this strategy $BR_2(\hat{u})$ and with probability at most $\delta\, p_s^{\hat{k}}$ he gets $minmax_2(B)$. In all other cases this strategy will get player 2 at least his minmax payoff provided ε is chosen sufficiently small. (Continuity ensures that playing a constant pure best response to $\hat{u}$ when player 1 actually plays within ε of $\hat{u}$ will give player 2 a payoff greater than $minmax_2(B)$, provided $BR_2(\hat{u}) > minmax_2(B)$ and ε is small.) Thus, player 2's expected payoff from the strategy must be at least $(1-\delta)\, p_s^{\hat{k}} BR_2(\hat{u}) + (1-(1-\delta)\, p_s^{\hat{k}})minmax_2(B)$. This completes the proof because at a Nash equilibrium a deviation by player 2 to this strategy must be unprofitable and δ can be chosen arbitrarily close to zero. *Q.E.D.*

Proof of Lemma 2: It is a standard result that the odds ratio is a supermartingale conditional on $\kappa = k$ (see, e.g., Fudenberg & Levine (1992)):

$$\frac{1-p_t^k}{p_t^k} \quad \bullet \quad E^k \left[\; \frac{1-p_{t+1}^k}{p_{t+1}^k} \; \Big| \; h_t \; \right] \qquad \qquad P^k\text{-a.s.,}$$

or, subtracting one and using Jensen's inequality,

(A1) $\qquad (p_t^k)^{-1} = E^k[\,(p_{t+1}^k)^{-1} \mid h_t\,] = (E^k[\,p_{t+1}^k \mid h_t\,])^{-1} \qquad P^k\text{-a.s.}$

Inverting the extremes of (A1) gives (i). By (i) conditional on player 1 being type k, p_t^k is a bounded submartingale and it therefore converges almost surely to some random variable $p_\bullet^k$. Suppose $p_\bullet^k$ were zero on a set of positive measure; then $(1-p_\bullet^k)/p_\bullet^k$ would be infinite on a set of non-zero measure, which implies $E^k[(1-p_t^k)/p_t^k]\varnothing 8$ as $t\varnothing 8$. This leads to a contradiction since the supermartingale property of the odds ratio implies $> (1-p_t^k)/p_t^k \; \varepsilon \; E^k[(1-p_t^k)/p_t^k]$, for all t. Hence (ii) is established. *Q.E.D.*

Proof of Lemma 3: Fix a positive probability h_s. Consider how player 2 updates his beliefs after a history h_t where $t>s$ (and h_t has positive probability conditional on h_s). From Bayes' Theorem we have

$$p_t^k = \frac{p_s^k \, \Pr[h_t \mid k, h_s]}{\sum_{k'} p_s^{k'} \, \Pr[h_t \mid k', h_s]} \quad .$$

This gives a relation between the conditional and unconditional distributions in the expectations $E[\,.\mid h_s\,]$ and $E^k[\,.\mid h_s\,]$. If we substitute for the conditional distribution we get that for any h_t-measurable random variable x, where $t>s$,

$$p_s^k \, Ek[\, x \,|h_s] = \sum_{h_t \in H_t} p_s^k \, Pr[h_t|k,h_s]x(h_t) = \sum_{h_t \in H_t} p_t^k(h_t)\sum_{k'} p_s^{k'} \, Pr[h_t|k',h_s]x(h_t) = E[\, p_t^k \, x \,|h_s];$$

thus it is possible to write the conditional expectation of the random variable x as an unconditional expectation of p_t^k x. Notice that player 2's payoff in period t, $B(i_t,j_t)$, is an h_{t+1}-measurable random variable, and hence the following holds almost surely:

$$p_s^k \, |\, LEk[\, b_T \,|\, h_s \,] - LE[\, b_T \,|\, h_s \,] \,|$$

$$= \& LEk[\, T^{-1}\Sigma_{t=1}^T \, p_s^k B(i_t,j_t) \,|\, h_s \,] - LE[\, T^{-1}\Sigma_{t=1}^T \, p_s^k B(i_t,j_t) \,|\, h_s \,] \,\&$$

$$= 2L[T^{-1}(s-1)M_B] + \& L[\, p_s^k Ek(\, T^{-1}\Sigma_{t=s}^T B(i_t,j_t)|\, h_s \,)] - LE[\, T^{-1} \, \Sigma_{t=s}^T \, p_s^k B(i_t,j_t)|\, h_s \,] \,\&$$

$$= 0 + \& LE[\, T^{-1}\Sigma_{t=s}^T \, p_{t+1}^k B(i_t,j_t) \,|\, h_s \,] - LE[\, T^{-1}\Sigma_{t=s}^T \, p_s^k B(i_t,j_t) \,|\, h_s \,] \,\&$$

$$= M_B LE[\, T^{-1}\Sigma_{t=s}^T \, \& \, p_{t+1}^k - p_s^k \, \& \,|\, h_s \,],$$

where both inequalities use the fact that L is a linear operator (recall that M_B bounds payoffs) and in the third line we take p_s^k outside $Ek[\, \cdot \,|\, h_s \,]$ in order to use the above result on conditional expectations. Following Hart (1985) we now define a non-increasing sequence of random variables $z_s := \sup_{t=s} \& \, p_{t+1}^k - p_\bullet^k \, \&$ which converges a.s. to zero. It follows that $E[\, z_s \,|\, h_s \,]$ is a bounded supermartingale w.r.t. $\{h_s\}$ and $E[\, z_s \,|\, h_s \,] \oslash \hat{z}_8$ a.s. for some random variable $\hat{z}_8$. But $E[E[\, z_s \,|\, h_s \,]] = E[z_s]\oslash 0$, so $E[\, \hat{z}_8 \,] = 0$ and $\hat{z}_8 = 0$ a.s. We then have $LE[\, T^{-1}\Sigma_{t=s}^T \, \& \, p_{t+1}^k - p_s^k \, \& \,|\, h_s \,] = E[\sup_{t=s} \& \, p_{t+1}^k - p_s^k \, \& \,|\, h_s \,] = E[\, 2z_s \,|\, h_s \,] \oslash 0$ a.s.

Q.E.D.

Proof of Lemma 4: A generic zero-sum game (-B,B) has a unique equilibrium (u*,v*): this equilibrium is regular and quasi-strict (see Van Damme (1987, p.56)), and we are assuming h = #supp(u*)=#supp(v*)>1. The equilibrium is regular so there is a non-singular hxh submatrix **B** of B such that $(\mathbf{u}^*)^t\mathbf{B} = \mathbf{e}^t(minmaxB)$ and $\mathbf{B}\mathbf{v}^* = \mathbf{e}(minmaxB)$, where $\mathbf{u}^*$ ($\mathbf{v}^*$) is u* (v*)with all zeros deleted and $\mathbf{e}$ is a vector of ones.

Let $\mathbf{z}(\mathbf{p}) := \alpha^{-1} \{\, \mathbf{u}^* + \delta(\mathbf{p}^t\mathbf{B}^{-1})^t \,\}$ for $\mathbf{p}$ R^h, where $\alpha,\delta>0$. Choose δ small enough so that $\mathbf{z}(\mathbf{p}) >> 0$ and α so that $\mathbf{z}(\mathbf{p})^t\mathbf{e}=1$; thus $\mathbf{z}(\mathbf{p})$ is a potential mixed strategy for player 1. Notice that if $\mathbf{v}$ Δ^h, then

$$\mathbf{z}(\mathbf{p})^t\mathbf{B}\mathbf{v} \;=\; \alpha^{-1} \{\, \mathbf{e}^t(minmax2(B)) + \delta\mathbf{p}^t \,\}\mathbf{v} \;=\; \alpha^{-1}\{\, (minmax2(B)) + \delta\mathbf{p}^t\mathbf{v} \,\}.$$

Thus, $\mathbf{z}(\mathbf{p})^t\mathbf{B}\mathbf{v} = minmax2(B)$ $\delta\mathbf{p}^t\mathbf{v} = (\alpha-1)minmax2(B)$ $\mathbf{p}^t\mathbf{v} = \mathbf{p}^t\mathbf{v}^*$. ($\mathbf{z}(\mathbf{p})^t\mathbf{e}=1$ together with $\mathbf{v}^* = \mathbf{B}^{-1}\mathbf{e}(minmax2(B))$ implies $\alpha-1 = \delta\mathbf{p}^t\mathbf{v}^*minmax2(B)^{-1}$.) The equilibrium (u*,v*) is quasi-strict so $V:=\{v$ $\Delta^J \,|\, (u^*)^t Bv = minmax2(B) \}$ consists of all v's satisfying supp(v)$\subseteq$ supp(v*). Let

$z(\mathbf{p})$ Δ^m be $\mathbf{z(p)}$ with zeros in appropriate locations and let p Δ^J be $\mathbf{p}$ with appropriate extra zeros. As $\delta\varnothing 0$ the set $V(\mathbf{p}) := \{ v \ \Delta^J \mid z(\mathbf{p})^t Bv = \text{minmax}2(B) \}$ converges to $V^*(\mathbf{p}):=V\leftrightarrow\{v \ \Delta^J\mid p^t v=p^t v^* \}$ in the Hausdorff topology. $(V^*(\mathbf{p}) \subseteq V(\mathbf{p})$ and if v $V(\mathbf{p})$, but supp(v) $\not\subset$ supp(v*), then v plays an action which is not a best response to u*; hence $z(\mathbf{p})^t Bv < \text{minmax}2(B)$ as $\delta\varnothing 0$. If v V but v $V^*(\mathbf{p})$ then $\mathbf{p}^t\mathbf{v} < \mathbf{p}^t\mathbf{v^*}$ (where $\mathbf{v}$ is v with appropriate zeros deleted), and this implies $\mathbf{z(p)}^t\mathbf{Bv} < \text{minmax}2(B)$: a contradiction.)

By linearity $\min_{v \ V}(u^*)^t Av$ occurs on a convex subset of the boundary of V; choose $\mathbf{p}$ so that $V^*(\mathbf{p})$ does not include this set (possible for generic A). By continuity, as $\delta\varnothing 0 \min_{v \ V(\mathbf{p})}(u^*)^t Av$ is strictly greater than $\min_{v \ V}(u^*)^t Av$. As $\delta\varnothing 0$ so $z(\mathbf{p})\varnothing u^*$ and thus we also have $\min_{v \ V(\mathbf{p})} z(\mathbf{p})^t Av > \min_{v \ V}(u^*)^t Av$.   *Q.E.D.*

APPENDIX B: TIGHTNESS OF THE BOUND

Proof of Proposition 2:   We prove part (ii) first.
(The strategy of the normal type, k=1) Type k=1 of player 1 randomly selects one of the automata k, $2=k=\hat{k}$, with probability $\theta^k =0$ and mimics its stage-game mixed action $\hat{u}^k$ in all future periods (where $\theta^2+...+\theta^{\hat{k}}=1$). Let $\pi^k = p^k(\theta^k p^1+p^k)^{-1}$; this will be player 2's conditional probability for type k given he knows $\hat{u}^k$ is being played (thus $\pi^k=p^k$). If type k=1 mimics $\hat{u}^k$, then by Proposition 1, once learning has settled down, this must guarantee her at least $a^*(\pi^k,\hat{u}^k)$. We shall construct an equilibrium which delivers exactly $a^*(\pi^k,\hat{u}^k)$ to type 1 when she mimics type k. Relabel the automata $k<\hat{k}$ so that $a^*(1,\hat{u}^k) > a^*(p^{\hat{k}},\hat{u}^{\hat{k}})$ if $k_m=k<\hat{k}$ (set $k_m=\hat{k}$ if there are no such automata). The θ^k's are chosen so that $\max_{2\check{s}k\check{s}\hat{k}} a^*(\pi^k,\hat{u}^k)$ is minimised: this will minimise the payoff to type 1 and simultaneously ensure that all types which are mimicked with positive probability offer the same (maximal) payoff, as $a^*(\pi^k,\hat{u}^k)$ is a decreasing continuous function of θ^k for $k_m=k<\hat{k}$ (clearly $\theta^k =0$ for $2=k<k_m$: mimicking one of these automata will give player 1 full reputation, $\pi^k=1$, but this is not desirable). Hence define $a^*:=\min_\theta \max_{2\check{s}k\check{s}\hat{k}} a^*(\pi^k,\hat{u}^k)$, where $\theta:=(\theta^2,...,\theta^{\hat{k}})$. We need to choose ε small so that a^* is close to $a^*(p^{\hat{k}},\hat{u}^{\hat{k}})$. For ε small, even a small increase in θ^k away from zero will lead to π^k falling from 1 to close to zero. Hence for ε sufficiently small, our choice of the θ^k's results in $a^*(\pi^k,\hat{u}^k)$ having the same value for all $k_m=k=\hat{k}$, that is $a^*=a^*(\pi^k,\hat{u}^k)$ for $k_m=k=\hat{k}$. Moreover as ε shrinks, $\theta^{\hat{k}}$ will go to 1 and p^1 to $1-p^{\hat{k}}$, so $\pi^{\hat{k}}$ (respectively a^*) can be made arbitrarily close to $p^{\hat{k}}$ (respectively $a^*(p^{\hat{k}},\hat{u}^{\hat{k}})$): we choose ε so that $a^*(p^{\hat{k}},\hat{u}^{\hat{k}})+\eta > a^*$.

Deviations by player 2 during the punishment phase are ignored. If player 2 deviates from his (deterministic) strategy at any other point of play, type k=1 minmaxes him thereafter.
(The strategy of player 2) Player 2 plays an arbitrary constant action for the first K periods and then follows a prescribed deterministic sequence $\{j_t^{ik} \}$ so long as the sequence $\{i_t^k \}$

(some k> $\hat{k}$) is observed–this corresponds to the normal type k being provisionally identified. Player 2 continues to follow { j_t^k } until an action i_t? i_t^k is observed (the sequences { i_t^k },{ j_t^k }, will be determined below). As soon as an action inconsistent with any sequence { i_t^k }, k> $\hat{k}$, is observed, play enters the monitoring phase in the following period–this means that player 2 believes he is not facing one of the normal types k> $\hat{k}$.

In the monitoring phase, player 2 reviews whether player 1 keeps to one of the automaton strategies ûk, k=2,..., $\hat{k}$, in a sequence of customised review phases of increasing length. Each time player 1 fails to play close to ûk on average during the review phase, she will face a finite period of punishment, or "punishment phase". Player 2's strategy will eventually identify the correct strategy ûk and play the correct review for this strategy.

Review phases are indexed by n, the number of the review in the natural order. Review n lasts for a fixed number of periods L(n); its (random) starting period is denoted by t(n). The automaton for which player 1 is being reviewed is denoted by k(n); this is an integer random variable taking values between 2 and $\hat{k}$. For k(n)=k, player 2 follows a fixed seque nce of actions { j_t^k } during the review. Let vk Δ^J be a (possibly mixed) action which attains the minimum in the definition of a*(πk,ûk) (see equation (1)) if k_m=k= $\hat{k}$, and which is a best response against ûk if 2=k<k_m. For each review n with k(n)=k, choose a fixed sequence[18] $\left\{ j_t^k \right\}_{t=t(n)}^{t(n)+L(n)-1}$ so that v_j(n,k):= #{t {t(n), ..., t(n)+L(n)+1} | j_t^k =j}/L(n), the relative frequency with which action j is chosen, converges to v_j^k as n$\emptyset$, with v_j(n,k)=0 if and only if v_j^k =0. Let f_{ij} be the number of times profile (i,j) is chosen during the review. Then player 1 fails review n with k(n)=k if the relative frequency of her actions against *each* of player 2's actions is not sufficiently close to ûk, specifically if

& f_{ij} (v_j(n,k) L(n))-1 - $\hat{u}_i^k$ &= $n^{-4/3}$ for any (i,j) such that v_j^k >0.

If the nth review is failed, play enters punishment phase n, which lasts for P(n) periods. To construct the punishment strategy, define IR(u):= {v Δ^J | B(u, v)=minmax2(B)}. Next, for λ [0,1],

$$\min_{v \in \Delta^J} (\lambda A(u,v) + (1-\lambda)(-B(u,v)) \leq \min_{v \in IR(u)} (\lambda A(u,v) + (1-\lambda)(-B(u,v))$$

$$\leq \min_{v \in IR(u)} \lambda A(u,v) + \max_{v \in IR(u)} (1-\lambda)(-B(u,v)) \leq \min_{v \in IR(u)} \lambda A(u,v) + (1-\lambda)(-\min \max{}^2(B)) \cdot$$

Taking the maximum over u $\in \Delta^I$,

$$\max_{u \in \Delta^I} \min_{v \in \Delta^J} (\lambda A(u,v) + (1-\lambda)(-B(u,v)) \leq \lambda a*(0, \hat{\hat{u}}^k) + (1-\lambda)(-\min\max{}^2(B)).$$

---

[18]The sequence depends only on n and k, but the starting period is random.

Hence by Blackwell's approachability theorem (see Myerson (1991, ch.7), Mertens *et al.* (1991)) the set of payoffs below $(a*(0, \hat{u^k}), -\text{minmax}_2(B))$ is *approachable* by player 2 when player 1 has vector payoffs $(A, -B)$. Define

$$S := \{(a, b) \quad R^2 \ \& \ a = a*(0, \hat{u^k}), b = \text{minmax}_2(B)\},$$

and for $t=0$, where $\| \cdot \|$ is used to denote the Euclidean norm,

$$\delta_t := \min_{(a,b) \in S} \left\| \left(\frac{L(n)(\bar{a}, \bar{b})}{(L(n) + t)} + \sum_{s=t(n)+L(n)}^{t(n)+L(n)+t} \frac{(A(i_s, j_s), B(i_s, j_s))}{(L(n) + t)} \right) - (a, b) \right\|,$$

where $\bar{a} = \max_{i,j} A(i,j)$ and $\bar{b} = \max_{i,j} B(i,j)$. Then by approachability there exists a continuation strategy $\tilde{\tau}$ for player 2 after the history $h_{t(n)+L(n)}$, and a finite integer R, depending only on n, such that for *any* continuation strategy σ followed by player 1 during the punishment phase $P_{\sigma, \tilde{\tau}}[$ $\sup_{t=t(n)+L(n)+R} \delta_t = n^{-1} \ \& \ h_{t(n)+L(n)}] = n^{-1}$. Define $P(n) = R+1$ and let $\tilde{\tau}$ be followed for the $P(n)$ periods of punishment phase n.

Review 1 with $k(1) = \hat{k}$ is always the first review phase played once the monitoring phase has begun. Review n with $k(n)=k$ is followed by review n+1 with $k(n+1)=k$ if review n with $k(n)=k$ is passed. If the review n with $k(n)=k$ is failed punishment phase n occurs. Player 1 could fail a review because she is following a different automaton strategy, so at the end of the punishment phase, review n+1 with $k(n+1)=k$ is played unless there exists a k', $1 < k' = \hat{k}$, such that $\& f_{ij}(v_j(n,k') L(n))^{-1} - \hat{u}_i^{k'} \& = n^{-4/3}$ for all (i,j) such that $v_j^{k'} > 0$, in which case review n+1 with $k(n+1)=k'$ is played (choose the smallest if there are multiple such k').

(The strategies of the normal types $k > \hat{k}$) Consider type k playing a best response to that part of player 2's strategy which is played during the monitoring phase. Choose the deterministic sequences $\{i_t^k\}$, $\{j_t^k\}$, to replicate the best-response outcome, that is so that the limiting average relative frequency over stage-game action profiles is the same in either case. (If a best response does not exist, then replicate the limit relative frequencies from a convergent sequence of frequencies of $\hat{\varepsilon}$-best responses as $\hat{\varepsilon} \oslash 0$). The sequences $\{i_t^k\}$ for $k > \hat{k}$ are chosen additionally to ensure that they differ during the first K periods: for each k, $k' > \hat{k}$, there is some $t=K$ such that $i_t^k \neq i_t^{k'}$. Finally if any of the automata play deterministic strategies, ensure that each $\{i_t^k\}$, $k' > \hat{k}$, differs from all such automaton strategies at some point during the first K periods. If player 2 does not respond to $\{i_t^k\}$ with $\{j_t^k\}$ then type $k > \hat{k}$ minmaxes player 2 thereafter.

(Optimality of Players' Strategies) Suppose that type 1 chooses to mimic the automaton strategy of $\hat{u}^k$ each period. This implies that the monitoring phase begins in finite time with probability one. We choose $L(n)=n^4$. Consider review n with $k(n)=k$; by Chebychev's Inequality, for (i,j) such that $v_j^k > 0$,

$$\Pr\left[\left|\frac{f_{ij}}{v_j(n,k)L(n)} - \hat{u}_i^k\right| \geq n^{-4/3}\bigg| k\right] = \frac{1}{4v_j(n,k)L(n)n^{-8/3}} = \frac{n^{-4/3}}{4v_j(n,k)},$$

where we are using the fact that $f_{ij}\,(v_j(n,k)\,L(n))^{-1}$ has mean $\hat{u}_i^k$ and variance $\hat{u}_i^k\,(1-\hat{u}_i^k\,)(v_j(n,k)\,L(n))^{-1}=(4v_j(n,k)\,L(n))^{-1}$. Choose ϕ, $\tilde{n}$, such that $0<\phi<v_j(n,k)$ for all k, j, with $v_j^k>0$, and all $n=\tilde{n}$. Then the probability of failing review n with $k(n)=k$ for $n=\tilde{n}$ is no greater than $IJn^{-4/3}(4\phi)^{-1}$. Since $\Sigma_{n=1}^{\bullet}n^{-4/3}<8$, by the Borel-Cantelli Lemmas with probability one a player playing $\hat{u}k$ fails only a finite number of reviews n with $k(n)=k$ in equilibrium, and likewise (by essentially the same argument) only undertakes a finite number of reviews n with $k(n)=\tilde{k}\,\eth k$. Hence type 1 must receive the limit of her expected payoffs from review phases n with $k(n)=k$ which by construction equals a* for $k_m=k=\hat{k}$, and is no greater than a* for $2=k<k_m$.

If type 1 chooses to mimic one of the types $k>\hat{k}$ and play the sequence $\{i_t^k\}$ then by construction the outcome corresponds to some response to the monitoring phase; thus this cannot be a profitable deviation. Nor can type 1 profitably deviate by playing some other strategy during the monitoring phase: failing a review phase leads, with a probability approaching one as $n\varnothing 8$, to a payoff, for the entire review/punishment phase, no greater than a number approaching $a^*(0,\hat{u}^k)=a^*$, and since $L(n)/t(n)=n^4/\Sigma_{m=1}^{n-1}m^4\varnothing 0$ as $n\varnothing 8$, her possible gains obtained by deviating from $\hat{u}k$ in the current review phase become insignificant in the calculation of a_T for T large. It is then straightforward, but tedious, to show that $\limsup_{T\varnothing 8}E_{\sigma,\tau^*}^1[a_T]=a^*$ for any σ, where τ^* is player 2's equilibrium strategy, so $L^{E_{\sigma,\tau^*}^1}[a_T]=\limsup_{T\varnothing 8}E_{\sigma,\tau^*}^1[a_T]=a^*$. (Deviations during the punishment phase do not affect average payoffs.)

By construction the types $k>\hat{k}$ receive payoffs generated by the sequence $\{(i_t^k,j_t^k)\}$ which gives them their payoff from an optimal response to the monitoring phase strategy and so they cannot gain by deviation.

Since in equilibrium punishment phases only take place a finite number of times with probability one, deviations during such phases do not affect player 2's payoff. In all other phases player 2 plays deterministically. It is then sufficient to show that a deviation during a deterministic phase of play cannot improve player 2's payoff after any positive probability history h_t. Conditional upon $\hat{u}k$ being played, player 2's equilibrium payoff is at least $\pi^k BR2(\hat{u}k) + (1-\pi^k)minmax2(B)$ — note that $\pi^k=1$ for $2=k<k_m$. When $k>\hat{k}$ player 2's equilibrium payoff is at least $minmax2(B)$: the deterministic sequences $\{i_t^k\}$, $\{j_t^k\}$, $k>\hat{k}$, correspond to some response against player 2's monitoring phase strategy and from each review/punishment phase player 2 plays so as to guarantee himself a payoff converging, as n increases, to $minmax2(B)$ with probability approaching one. Player 2's expected payoff given his priors $(p_t^k)_{k\in K}$ is therefore at least

$$\sum_{k=2}^{\hat{k}} \frac{p_t^k}{\pi^k} \{ \pi^k BR2(\hat{u}k) + (1-\pi^k)minmax2(B) \} + \sum_{k>\hat{k}} p_t^k \, minmax2(B),$$

where p_t^k / π^k is his belief that $\hat{u}k$ is being followed. If player 2 deviates he gets at most $\Sigma_{k=2}^{\hat{k}} p_t^k BR2(\hat{u}k) + (1-\Sigma_{k=2}^{\hat{k}} p_t^k)minmax2(B)$, which equals the above expression. Therefore player 2 never benefits from deviating.

To prove part (i), a similar construction is used. Type k=1 can now be treated in the same way as the other normal types, and reveal its type at the beginning of the game, and play out a deterministic equilibrium corresponding to a best response against the monitoring phase strategy of player 2. Moreover player 2 no longer needs to punish type 1 during the monitoring phase; all that is necessary is that player 2 plays a prudent stage-game mixed action (i.e. to guarantee his minmax payoff) during the punishment phases. During the review phases for automaton k, player 2 plays a best response against $\hat{u}k$. The construction is otherwise identical to that of part (ii).   *Q.E.D.*

REFERENCES

CRIPPS, M. W., AND J. P. THOMAS (1992): "Reputation and Commitment in Two-Person Repeated Games," *Discussion Paper,* No.9210, CentER, Tilburg University.

CRIPPS, M. W., K. M. SCHMIDT, AND J. P. THOMAS (1993): "Reputation in Perturbed Repeated Games," *Discussion Paper,* No. A-410, University of Bonn.

DUNFORD, N., AND J. T. SCHWARTZ (1988): *Linear Operators: Part I General Theory,* Wiley Classics Library. New York: John Wiley.

FORGES, F. (1992): "Non-Zero-Sum Repeated Games of Incomplete Information," in *Handbook of Game Theory*, ed. by R.J.Aumann and S.Hart. Amsterdam: North Holland.

FUDENBERG, D., AND D. K. LEVINE (1989): "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, 57, 759-778.

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

HART, S. (1985): "Nonzero-Sum Two-Person Repeated Games with Incomplete Information," *Mathematics of Operations Research,* 10, 117-153.

HARSANYI, J. (1967-68): "Games with Incomplete Information Played by Bayesian Players," *Management Science,* 14, 159-182, 320-334.

ISRAELI, E. (1989): "Sowing Doubt Optimally in Two-Person Repeated Games," M.Sc. Thesis, Tel-Aviv University.

KOREN, G. (1988):"Two-Person Repeated Games with Incomplete Information and Observable Payoffs," M.Sc. Thesis, Tel-Aviv University.

MERTENS, J.-F., S. SORIN, AND S. ZAMIR (1991): *Repeated Games*, forthcoming.

MYERSON, R. B. (1991): *Game Theory.* Cambridge: Harvard University Press.

SCHMIDT, K.M. (1993): "Reputation and Equilibrium Characterization in Repeated Games of Conflicting Interests," *Econometrica,* 61, 325-352.

SHALEV, J. (1988): "Nonzero-Sum Two-Person Repeated Games with Incomplete Information and Observable Payoffs," *The Israel Institute of Business Research,* Working Paper 964/88, Tel-Aviv University.

VAN DAMME, E.E.C. (1987): *Stability and Perfection of Nash equilibria.* Berlin: Springer-Verlag.

WATSON, J. (1992): "Reputation and Outcome Selection in Perturbed Supergames: An Intuitive Behavioral Approach," Graduate School of Business, Stanford.