

# Cooperation and Punishment\*

Robert Evans

*St. John's College, University of Cambridge*

Jonathan P. Thomas

*University of St. Andrews*

January 2000

**ABSTRACT:** We show that, in repeated common interest games without discounting, strong ‘perturbation implies efficiency’ results require that the perturbations must include strategies which are ‘draconian’ in the sense that they are prepared to punish to the maximum extent possible. Moreover, there is a draconian strategy whose presence in the perturbations guarantees that any equilibrium is efficient. We also argue that the results of Anderlini and Sabourian (1995) using perturbation strategies which are cooperative (and hence non-draconian) are not due to computability *per se* but to the further restrictions they impose on allowable beliefs.

**KEYWORDS:** Common interests, repeated games, cooperation, computability, reputation.

*JEL CLASSIFICATION NUMBERS:* C73, D83, L14.

\* We are grateful to an editor and three referees for a number of helpful comments and suggestions, and to seminar participants at University College London and the Institute for Economic Analysis at the University of Pompeu Fabra.

# 1 Introduction

Suppose that two players repeatedly play a game of common interests (i.e., one with a strongly Pareto-dominant payoff vector). If there is complete information, there are equilibria in which both get very low payoffs, despite the obvious scope for cooperation. Imagine, then, that the game is perturbed in such a way that, for each player, there is a small positive probability  $\xi$  that that player, rather than being a rational agent with payoffs as specified in the game, will play a repeated game strategy drawn from some probability distribution  $\mu$ . It would be natural to conjecture that there are strategies, or sets of strategies, which have the property that their presence in the perturbations (i.e., in the support of  $\mu$ ) guarantees that both parties cooperate in equilibrium. We refer to such a set of strategies as *cooperation-inducing*. A ‘perturbation implies efficiency’ result of this kind would be *strong* in the sense that it would not depend on any property of the probability distribution  $\mu$  except that  $\mu$  gives strictly positive probability to each strategy in the specified set.<sup>1</sup> The logic would be that the ‘normal’ type (i.e., the one which is present with probability  $1 - \xi$ ) of, say, player 1 can, by imitating a strategy in the cooperation-inducing set, eventually persuade player 2 that 1 will always cooperate in the future if player 2 does likewise. Player 2 will then, rationally, respond by always cooperating himself. Thus, both players must, in equilibrium, get approximately the maximum available payoff (if there is little or no discounting). Natural candidates for cooperation-inducing strategies are strategies which are ‘nice’ in the sense that, against any opponent, they will eventually always cooperate a significant fraction of the time.

This kind of argument was first formalized by Aumann and Sorin (1989), for pure-strategy equilibria only, under a restriction of the perturbation to bounded recall strategies<sup>2</sup>.

---

<sup>1</sup>Following Fudenberg and Levine (1989), there is a literature which identifies ‘strong’ reputation effects in general discounted games where a long-run player is faced by a sequence of short-lived opponents, with the latter believing that the former might be irrational, as described above, and where no restrictions are placed on the perturbation. These ideas have been extended to repeated games where the short-run players are replaced by a relatively impatient long-run player; see Sorin (1999) for a synthesis of the literature.

<sup>2</sup>Their argument has been extended to general (undiscounted) games by Watson (1996), who stresses that the perturbation strategies must be “forgiving” if the approach is to work.

In another important paper Anderlini and Sabourian (1995) established a strong result of this kind, based on arguments from computability theory. They assume that the repeated game (undiscounted) is played by Turing machines and show that, under some further computability assumptions, there exists a cooperation-inducing set of Turing machines. The strategies implemented by the machines in this set are indeed ‘nice’: in fact they have the property that they will eventually always cooperate when paired with any strategy (such machines are referred to as ‘cooperative’).

In Theorem 1 below we show that if allowable strategies are unrestricted, the conjecture that a set of ‘nice’ strategies can induce cooperation in the sense described above is false: in other words, for any such set there will always exist perturbations, containing the set, which give rise to bad equilibria. In fact, a stronger statement is true: no set of strategies can induce cooperation merely by its presence in the support of the perturbation if it does not contain strategies which are ‘draconian’, that is, prepared to min-max some opponents almost all of the time (by contrast, a ‘cooperative’ strategy is prepared to min-max almost none of the time against any opponent, since it will eventually always cooperate). The intuition for this result is simple. Let  $\mu$  be the perturbation of player 1’s strategies, and suppose that  $\text{supp}(\mu)$ , the support of  $\mu$ , contains only cooperative strategies. Let  $\{a_{2t}\}$  be some infinite sequence of non-cooperating actions for player 2. For each strategy  $s_1$  in the support of  $\mu$  there is another strategy  $m'_1(s_1, \{a_{2t}\})$ , a kind of “nasty” alter ego of  $s_1$ , which, as long as 2 follows  $\{a_{2t}\}$ , plays exactly as  $s_1$  does, and, as soon as 2 deviates from  $\{a_{2t}\}$ , min-maxes 2 forever. Define  $\tilde{\mu}$  as the distribution which, for each  $s_1 \in \text{supp}(\mu)$ , gives probability  $\theta\mu(s_1)$  to  $s_1$  and  $(1 - \theta)\sum \mu(\tilde{s}_1)$  to  $m'_1(s_1, \{a_{2t}\})$ , where the sum is over  $\tilde{s}_1 \in \text{supp}(\mu)$  such that  $m'_1(\tilde{s}_1, \{a_{2t}\}) = m'_1(s_1, \{a_{2t}\})$ . Here  $\theta$  is a small positive number. This is a straightforward transformation of  $\mu$  with an expanded support. If player 1’s perturbation is  $\tilde{\mu}$  and 2 plays  $\{a_{2t}\}$  then no history with  $\tilde{\mu}$ -positive probability can persuade 2 that 1 is likely to be cooperative: for every  $s_1$ ,  $m'_1(s_1, \{a_{2t}\})$  will have a higher posterior probability than  $s_1$ . Thus, notwithstanding the presence in  $\text{supp}(\tilde{\mu})$  of the given arbitrary set of cooperative strategies,  $\text{supp}(\mu)$ , one can construct equilibria with low payoffs.

Theorem 1 shows that a perturbation containing only non-draconian strategies can never guarantee that every equilibrium is cooperative.<sup>3</sup> This raises the question whether draconian strategies can induce cooperation. The answer turns out, at least in the undiscounted case considered here, to be yes (see Theorem 2 below). We construct a strategy which, if present in the support of  $\mu$ , ensures that every Nash equilibrium is approximately efficient. Rather than sending a signal, through its actions, that it has cooperative intentions and then cooperating, this strategy takes a punitive approach: it punishes, with ever-increasing severity, an opponent who persistently fails to cooperate.

The conclusion is that the presence of certain types in the support of a player's prior does not in general imply that there is some signal (i.e., sequence of play) which will persuade him that the other player is cooperative with high probability; nevertheless, even if he is pessimistic about the benefits of cooperating, he can be persuaded, through punishment, that non-cooperative behaviour is even less appealing. In the context of this paper, that is all that is necessary to induce cooperation in the long run.

Section 2 below sets out the framework and definitions which we use. The results outlined above are contained in Section 3. Section 4 discusses the relation of our results to those of Anderlini and Sabourian (1995). Section 5 concludes.

## 2 The Framework

Let  $G$  be a game of common interests between two players, 1 and 2. The (finite) pure action sets are  $A_1$  and  $A_2$ , with generic elements  $a_1$  and  $a_2$  respectively;  $\mathcal{A}_1$  and  $\mathcal{A}_2$  are the respective mixed action sets. The set of pure action pairs is denoted  $A$  and its generic element is  $a$ . The payoffs of  $G$ , given actions  $a \equiv (a_1, a_2)$ , are  $g(a) \equiv (g_1(a_1, a_2), g_2(a_1, a_2))$ , which we assume to be rational numbers. Abusing notation slightly, we let  $g_1(a_1, \alpha_2)$  be 1's expected payoff when she plays  $a_1$  and 2 plays the mixed action  $\alpha_2 \in \mathcal{A}_2$ .  $g_1(\alpha_1, a_2)$ ,

---

<sup>3</sup>Of course there will be non-draconian perturbations for which cooperative equilibria exist and some for which all equilibria are cooperative. What our result shows is that non-draconian strategies cannot ensure cooperation in equilibrium merely by their presence in the support, regardless of the extent of the support and the details of the distribution.

$g_2(a_1, \alpha_2)$ , etc., are defined similarly. The payoffs are normalized so that (for  $i, j = 1, 2; j \neq i$ ) player  $i$ 's pure min-max payoff,  $\min_{a_j} \max_{a_i} g_i(a_1, a_2)$ , is zero. Non-negative payoffs are referred to as individually rational. Let  $A_i^m$  denote the set of  $i$ 's pure min-max actions, i.e.,  $\hat{a}_1 \in A_1^m$  if  $\max_{a_2} g_2(\hat{a}_1, a_2) = 0$  and  $\hat{a}_2 \in A_2^m$  if  $\max_{a_1} g_1(a_1, \hat{a}_2) = 0$ . By the assumption of common interests, there exists a strongly Pareto-dominant payoff vector  $(v_1^*, v_2^*)$ ; we assume that  $v_i^* > 0$ ,  $i = 1, 2$ . Let  $a^* \equiv (a_1^*, a_2^*)$  be a fixed arbitrary "efficient" action pair which attains this payoff vector. Finally, let  $\underline{v}_i := \min_{a \in A} g_i(a)$ .

We are concerned with an infinite repetition of  $G$ . The *history* at stage  $t$ , denoted  $h_t$ , is the sequence of pure action pairs played up to and including stage  $t$ . At each stage  $t = 1, 2, 3, \dots$ , both players know all the pure actions taken so far, i.e., they know the history  $h_{t-1}$ . A *realised play path* is given by the infinite history  $h_\infty = ((a_1^1, a_2^1), (a_1^2, a_2^2), \dots)$ , where  $a_i^t$  is the action taken by player  $i$  at  $t$ . A pure strategy for player  $i$  specifies for each date  $t + 1$  and each history  $h_t$  an action  $a_i \in A_i$ ; the set of all pure strategies for player  $i$  is denoted  $\Sigma_i$ . A mixed strategy  $\sigma_i$  for  $i$  is a probability distribution over  $\Sigma_i$ .

With probability  $1 - \xi$  player  $i$  ( $i = 1, 2$ ) plays a strategy which maximizes expected long-run average payoff, defined as the expectation of  $\pi_i(h_\infty) := \liminf_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T g_i(a_1^t, a_2^t)$ . This probability will be close to 1. With probability  $\xi > 0$  player  $i$ 's strategy is drawn from some distribution  $\mu_i$  over  $i$ 's pure strategies, the support of which is denoted  $\text{supp}(\mu_i)$  and is assumed to be countable. We denote this perturbed game as  $G^\infty(\mu_1, \mu_2; \xi)$ . (We shall also use  $\mu_i$  to denote a distribution over Turing machines; the use will be apparent from the context.) A Nash equilibrium of  $G^\infty(\mu_1, \mu_2; \xi)$  is a pair of mixed strategies  $(\sigma_1, \sigma_2)$  such that, for each  $i$ ,  $\sigma_i$  is a best reply to  $(1 - \xi)\sigma_j + \xi\mu_j$ , ( $j \neq i$ ).

Let  $IR_1(a_2) := \{\alpha_1 \in \mathcal{A}_1 \mid g_1(\alpha_1, a_2) \geq 0\}$  be player 1's mixed action responses to  $a_2$  which give 1 at least 0 (i.e., an individually rational payoff in the unperturbed game), and define player 2's "reservation payoff" to be

$$\hat{v}_2 := \max\left\{\max_{a_2 \in A_2} \min_{\alpha_1 \in IR_1(a_2)} g_2(\alpha_1, a_2), 0\right\}. \quad (1)$$

The first expression in the brackets is the payoff which player 2 can guarantee (as a minimum) by playing a pure strategy in the stage game assuming that player 1 responds

with the least favourable (to 2) mixed action guaranteeing herself (1) at least her pure min-max payoff (if this expression is negative, we take zero to be the reservation payoff). Define  $\hat{v}_1$  symmetrically. There exist strategies which ensure that, if they have positive probability according to  $\mu_i$ , player  $i$  will obtain at least  $(1 - \xi)\hat{v}_i + \xi v_i$  in any pure-strategy Nash equilibrium of  $G^\infty(\mu_1, \mu_2; \xi)$ .<sup>4</sup> Also notice that  $\hat{v}_i < v_i^*$  since if player  $i$  selects an efficient action,  $j$  can always put positive probability on an inefficient reply (which exists by the assumption that  $v_i^* > 0$ ) while receiving a nonnegative payoff. As an illustration of these reservation payoffs, consider the following example.

	$L$	$R$
$T$	1, 8	-2, -1
$B$	-1, -6	0, 0

In this game,  $IR_1(L)$  consists of those mixed strategies which give  $T$  at least probability  $\frac{1}{2}$ . Therefore  $\min_{\alpha_1 \in IR_1(L)} g_2(\alpha_1, L) = 1$ .  $IR_1(R)$  is the singleton strategy which gives probability zero to  $T$  and so  $\min_{\alpha_1 \in IR_1(R)} g_2(\alpha_1, R) = 0$ . Hence  $\hat{v}_2 = 1$ . (Similarly  $\hat{v}_1$  in this game is zero.) Suppose that  $\mu_2$  gives positive probability to the strategy “always  $L$ ”. If player 2 plays “always  $L$ ” a rational player 1 must eventually play  $T$  at least half of the time in a pure-strategy equilibrium. Therefore, in any pure-strategy equilibrium, player 2 must get at least 1 against a rational opponent.

Next we define the “draconian” strategies mentioned in the Introduction. For any pair of pure strategies,  $(s_1, s_2)$ , we let  $h_t(s_1, s_2)$  be the play induced in the first  $t$  periods by this strategy combination, we define  $\beta_i^t(s_1, s_2)$  to be the number of times in  $h_t(s_1, s_2)$  that player  $i$  min-maxes the other player (plays an action in  $A_i^m$ ), and, for  $i, j = 1, 2$ ,  $i \neq j$ , we let

$$\beta_i(s_1, s_2) := \limsup_{t \rightarrow \infty} \frac{1}{t} \beta_i^t(s_1, s_2);$$

$$\beta_i(s_i) := \sup_{s_j \in \Sigma_j} \beta_i(s_1, s_2).$$

---

<sup>4</sup>If, say,  $\hat{v}_2 > 0$ , the strategy for 2 simply involves always playing an action which attains the maximum of the first term inside the curly bracket of (1); for a proof with a somewhat different payoff criterion see Cripps and Thomas (1995). As we do not restrict attention to pure-strategy equilibria, it is possible that payoffs below this level can be received in mixed-strategy equilibrium.

So  $\beta_i(s_i)$  is the maximum proportion of times that the strategy  $s_i$  is prepared to punish. For example, a ‘grim’ strategy which punishes forever after one period of non-cooperation by the other player has  $\beta_i(s_i) = 1$ . A strategy which min-maxes a recalcitrant opponent but relents every tenth period to see if the other player will cooperate will have  $\beta_i(s_i) = 0.9$ .

**Definition 1** *A pure strategy  $s_i \in \Sigma_i$  is draconian if  $\beta_i(s_i) = 1$ . A set  $S_i \subseteq \Sigma_i$  of pure strategies is draconian if  $\sup_{s_i \in S_i} \beta_i(s_i) = 1$ .*

The next two definitions make precise the notion of sets of strategies which induce cooperation.

**Definition 2** *Given  $(S_1, S_2)$ , where  $S_i \subseteq \Sigma_i$  ( $i = 1, 2$ ), a payoff pair  $(v_1, v_2)$  is consistent with  $(S_1, S_2)$  in Nash equilibrium if there exists  $(\mu_1, \mu_2)$  with  $S_i \subseteq \text{supp}(\mu_i)$  ( $i = 1, 2$ ) such that, for each  $\eta > 0$ , there exists  $\bar{\xi} > 0$  and, for each  $\xi \in (0, \bar{\xi})$ , a Nash equilibrium of  $G^\infty(\mu_1, \mu_2; \xi)$  with payoff within  $\eta$  of  $(v_1, v_2)$ .*

**Definition 3** *The pair  $(S_1, S_2)$ , where  $S_i \subseteq \Sigma_i$  ( $i = 1, 2$ ), induces cooperation in Nash equilibrium if the set of payoff pairs consistent with  $(S_1, S_2)$  in Nash equilibrium is the singleton set  $(v_1^*, v_2^*)$ .*

### 3 Draconian Strategies and Cooperation

The following shows that non-draconian strategies do not induce cooperation and indeed that their mere presence in the perturbations can do little to restrict equilibrium payoffs.

**Theorem 1** *If the countable set  $S_i \subseteq \Sigma_i$  ( $i = 1, 2$ ) is not draconian then any rational<sup>5</sup>, feasible payoff pair  $v = (v_1, v_2)$  with  $v_i > \hat{v}_i$  ( $i = 1, 2$ ) is consistent with  $(S_1, S_2)$  in Nash equilibrium.*

---

<sup>5</sup>That is,  $v_1$  and  $v_2$  are rational numbers.

The proof of the theorem is in the appendix, but we outline the essential ideas here. Take any  $(\mu_1, \mu_2)$  with  $S_i = \text{supp}(\mu_i)$  ( $i = 1, 2$ ) and any rational, feasible  $v = (v_1, v_2) \gg (\hat{v}_1, \hat{v}_2)$ . Clearly there exists a sequence of pure action pairs  $(\tilde{a}^n(v))_{n=1}^\infty$ , consisting of an infinite repetition of a finite sequence, such that  $\lim_{N \rightarrow \infty} N^{-1} \sum_{n=1}^N g(\tilde{a}^n(v)) = v$ . The following lemma (proof in appendix) establishes that if player  $j$  plays the same pure action in every period (not a min-max action) then player  $i \neq j$  has a strategy, consisting of the infinite repetition of a finite sequence of pure actions, which will on average give  $j$  strictly less than  $v_j$  and  $i$  strictly more than zero.

**Lemma:** *Given  $v_1 > \hat{v}_1$ , there exist  $\varepsilon_1 > 0$  and, for each  $a_1 \notin A_1^m$ , a sequence  $(\hat{a}_2^n(a_1, v_1))_{n=1}^\infty$ , consisting of an infinite repetition of a finite sequence of actions in  $A_2$ , such that  $\lim_{N \rightarrow \infty} \sum_{n=1}^N N^{-1} g_1(a_1, \hat{a}_2^n(a_1, v_1)) < v_1 - \varepsilon_1$  and  $\lim_{N \rightarrow \infty} \sum_{n=1}^N N^{-1} g_2(a_1, \hat{a}_2^n(a_1, v_1)) > \varepsilon_1$ . An equivalent statement holds for player 2.*

We define a pair of strategies  $(\hat{s}_1(v, \mu_2), \hat{s}_2(v, \mu_1))$  so that they have the following three properties. (i) When the two strategies play each other, the outcome path is  $(\tilde{a}^n(v))_{n=1}^\infty$ , giving payoff  $v$ . (ii) If, say, player 1 deviates from  $\hat{s}_1(v, \mu_2)$  and plays instead some strategy  $\tilde{s}_1$  in  $\text{supp}(\mu_1)$  then the strategy  $\hat{s}_2(v, \mu_1)$  eventually “identifies” how  $\tilde{s}_1$  will play and plays in such a way that in periods in which  $\tilde{s}_1$  does not play a min-max action 1 gets on average at most  $v_1 - \varepsilon_1$  (as defined in the lemma) and 2 gets on average at least  $\varepsilon_1$ ; that is, in periods when  $\tilde{s}_1$  specifies  $a_1 \notin A_1^m$ ,  $\hat{s}_2(v, \mu_1)$  follows the sequence  $\hat{a}_2^n(a_1, v_1)$ . Furthermore, if  $\tilde{s}_1$  specifies a min-max action then  $\hat{s}_2(v, \mu_1)$  plays a best reply which holds 1 to at most  $\hat{v}_1$ . (iii) If, say, player 1 deviates and plays some strategy  $\bar{s}_1$  inconsistent (against  $\hat{s}_2(v, \mu_1)$ ) with any  $\tilde{s}_1$  in  $\text{supp}(\mu_1)$ , then  $\hat{s}_2(v, \mu_1)$  min-maxes often enough to ensure that 1’s long-run payoff is zero.

Clearly, if  $\hat{s}_1(v, \mu_2)$  and  $\hat{s}_2(v, \mu_1)$  satisfy these three criteria then each will be a best reply to the other. It does not follow, however, that they form an equilibrium of  $G^\infty(\mu_1, \mu_2; \xi)$ . To establish that, one would have to show that  $\hat{s}_i(v, \mu_j)$  is optimal, given  $i$ ’s updated beliefs, after every  $\mu_j$ -positive probability history. As specified,  $\hat{s}_i(v, \mu_j)$

has strictly positive continuation payoff after every such history because eventually  $i$  gets at least  $\varepsilon_j > 0$  in periods when  $i$  is not being min-maxed, while, since  $\text{supp}(\mu_j)$  is non-draconian, the latter happens a non-zero fraction of the time. Nevertheless this continuation payoff might be very low and if, for example,  $i$  comes to believe that  $j$  is probably playing a cooperative strategy, then  $i$  might well be able to improve on  $\widehat{s}_i(v, \mu_j)$  by playing cooperatively. However, we now define a pair of perturbation distributions  $(\mu'_1, \mu'_2)$  with wider supports than  $(\mu_1, \mu_2)$ , and argue that  $(\widehat{s}_1(v, \mu_2), \widehat{s}_2(v, \mu_1))$  is an equilibrium of  $G^\infty(\mu'_1, \mu'_2; \xi)$ . Given a pair of strategies  $(s_1, s_2) \in \Sigma_1 \times \Sigma_2$ , let  $m_1(s_1, s_2) \in \Sigma_1$  be a strategy which, after any history in which 2 has not deviated from  $s_2$ , plays the same action as  $s_1$ , and if 2 deviates from  $s_2$ , thereafter always plays a min-max action. Now, if  $s_1 \in \text{supp}(\mu_1)$ , let  $\mu'_1(s_1) = \theta\mu_1(s_1)$  (here  $\theta$  is a small positive number), and otherwise let  $\mu'_1(s_1) = (1 - \theta) \sum \mu_1(s'_1)$  where the sum is over  $s'_1$  such that  $s_1 = m_1(s'_1, \widehat{s}_2(v, \mu_1))$  (that is, a proportion  $(1 - \theta)$  of the probability assigned to a strategy in the original support is now assigned to its ‘alter ego’; the latter may also be an ‘alter ego’ to other strategies in the original support). Define  $\mu'_2$  similarly. Clearly  $\text{supp}(\mu_i) \subseteq \text{supp}(\mu'_i)$ . If 2 plays  $\widehat{s}_2(v, \mu_1)$  then the distributions  $\mu_1$  and  $\mu'_1$  have the same positive-probability histories: nevertheless, given prior  $\mu'_1$ , 2 will, after any such history, place much lower probability (in the ratio at most  $\theta : 1 - \theta$ ) on  $s_1$  than on the observationally indistinguishable strategy  $m_1(s_1, \widehat{s}_2(v, \mu_1))$ . The point is that there is no conceivable observation (as long as 2 plays  $\widehat{s}_2(v, \mu_1)$ ) which will persuade 2 that 1 is probably cooperative: she will place high probability on the event that, if she deviates from  $\widehat{s}_2(v, \mu_1)$ , she will be min-maxed forever. On the other hand,  $\widehat{s}_2(v, \mu_1)$  gives her a strictly positive expected continuation payoff, and so is optimal. This establishes that, as required,  $(\widehat{s}_1(v, \mu_2), \widehat{s}_2(v, \mu_1))$  is an equilibrium.

In view of this theorem a cooperation-inducing set must include strategies which are prepared to punish to the maximum extent. Does such a set exist? The next proposition shows that it does: we construct a set consisting of a single strategy  $\tilde{s}_i$  which induces cooperation. As long as this strategy has positive probability, however small, all Nash equilibria are approximately cooperative. Let  $\tilde{s}_1$  be defined as follows.

Start in the *normal* phase. In the normal phase, play  $a_1^*$ . In the *punishment* phase,

min-max player 2 by playing some  $a_1^m \in A_1^m$ . In a normal phase, so long as player 2 plays  $a_2^*$ , remain in the normal phase. If player 2 fails to play  $a_2^*$  in period  $t$ , then switch to the punishment phase for  $i$  periods, where  $i$  is the smallest positive integer satisfying  $tv_2^*/(t+i) < 1/t$  (ensuring that 2's average payoff after the punishment phase is below  $1/t$ ). After  $i$  periods, return to the beginning of the normal phase regardless of player 2's actions during the punishment phase.

**Theorem 2** *Suppose that in  $G$  each player's pure and mixed min-max payoffs coincide, and let  $\mu_1(\tilde{s}_1) > 0$ . Then in any Nash equilibrium of  $G^\infty(\mu_1, \mu_2; \xi)$  player 1 must receive at least  $(1 - \xi)v_1^* + \xi v_1$ . Therefore  $(\{\tilde{s}_1\}, S_2)$ , where  $S_2$  is arbitrary, induces cooperation in Nash equilibrium.*

The proof is in the Appendix. The point is simply that if player 2 does not eventually always cooperate against  $\tilde{s}_1$  then she will get a payoff of zero. As long as she plays a non-cooperative strategy she may attach no more than a very small probability to the event that 1 is cooperative; nevertheless the small positive expected payoff which she would get from experimenting with cooperation is enough to persuade her to do so. Theorems 1 and 2 together show that the ability of a set of perturbation strategies to restrict equilibrium payoffs is discontinuous, for the class of games of Theorem 2, in the degree to which they are prepared to punish (i.e., in  $\sup_{s_i \in S_i} \beta_i(s_i)$ ): if they are non-draconian then any payoffs above reservation payoffs are consistent with their presence in the perturbation; a draconian set of strategies can, however, induce cooperation.

**Remark.** To what extent do our results generalise to other environments? We can prove that, for a class of games including the example of Section 4 below, the conclusion of Theorem 1 holds for discounted games as the discount factors—which need not be symmetric—go to 1. This result requires a slightly different definition of a non-draconian set of strategies: there must exist an upper bound (uniform across the set) on the consecutive number of periods any strategy in the set is prepared to min-max. Then, given non-draconian perturbations  $(\mu_1, \mu_2)$ , it is possible to find perturbations  $(\mu'_1, \mu'_2)$  with

larger support, a perturbation probability  $\bar{\xi} > 0$  and a critical discount factor  $\bar{\delta} < 1$ , such that there are equilibria of  $G^\infty(\delta_1, \delta_2; \mu'_1, \mu'_2; \xi)$ —using an obvious notation—with payoffs arbitrarily close to some inefficient level, for all  $1 > \delta_1, \delta_2 > \bar{\delta}, 0 < \xi < \bar{\xi}$ . We conjecture that this holds for all common interest games. A similar result should hold for long but finite time horizons. The construction used in Theorem 2, on the other hand, does not work with symmetric discounting. A related version of this argument is used in Evans and Thomas (1997) to establish that with discounting a *relatively patient* player can achieve approximately the efficient payoff (though the other player may get less). Whether a version of the theorem can be established for symmetric discounting is an open question.

In contrast to Aumann and Sorin (1989) and (by the computability restriction) Anderlini and Sabourian (1995), we do not exclude mixed strategy equilibria in our analysis. Moreover, it is possible to show that if the mixed min-max payoff were strictly lower than the pure min-max payoff, then the conclusion of Theorem 2 would still hold provided that  $\text{supp}(\mu_1)$  contains a strategy, otherwise the same as  $\tilde{s}_1$ , which punishes using the mixed min-max punishment (although we have assumed here that perturbation strategies are *pure*).

## 4 Computability and Cooperation

As we noted in the introduction, Anderlini and Sabourian (1995) (henceforth AS) established a result which states that in a computability framework a cooperative, and hence non-draconian, set of strategies (more precisely, Turing machines) can, in our terminology, induce cooperation. In this section we explore the relation between our results and theirs, and we show that their result depends on excluding machines of the alter ego type (variations on the strategies  $m_1(s_1, \hat{s}_2)$ ).

AS impose a number of assumptions on the repeated game, all related to the notion of computability. Firstly, each strategy has to be computable in the sense that it can be implemented by a Turing machine (a program). This applies to the strategy chosen by

the normal type (present with probability  $1 - \xi$ ) and it also implies that  $supp(\mu_i)$  cannot contain strategies which are not Turing-implementable. The further assumptions are as follows (using our notation).

(AS1) The distributions  $(\mu_1, \mu_2)$  associated with the perturbations must be computable. In other words, for each  $\mu_i$ , which can now be regarded as a probability distribution over Turing machines (programs), there exists a Turing machine which enumerates a list  $(X_i^n, \mu_i(X_i^n))_{n=1}^\infty$  of pairs of coded descriptions of programs and (positive) probabilities. This list is exhaustive in the sense that any machine in  $supp(\mu_i)$  will appear in the list in finite time.

(AS2) For each  $\mu_i$ , the total probability,  $\mu_i(\overline{Q}_i)$ , of machines which are not quasi-cooperative is approximately computable: i.e., there is a Turing machine which, given any small rational number  $\varepsilon > 0$ , will compute  $\mu_i(\overline{Q}_i)$  to within  $\varepsilon$ . A machine is said to be *quasi-cooperative* (to belong to  $Q_i$ ) if, having once played the cooperative action  $a_i^*$ , it will always subsequently play  $a_i^*$ , regardless of the history.

(AS3) The only machines allowable in the supports of the perturbations are those that always halt (after any history) and those that never halt.

Under these assumptions, AS establish the existence of a pair of sets of machines, one for each player, which induces cooperation in Nash equilibrium. *These sets are non-draconian*—in fact the machines which represent valid strategies play only the cooperative action beyond some point. Essentially the argument is as follows. Assumptions (AS1)-(AS3) ensure that there exists a machine (call it  $Z_i$ ) which will begin by, each period, enumerating a new non-quasi-cooperative machine in  $supp(\mu_i)$ , simulating the action which that machine would take after the current history, and then taking a different action. It does this until it has distinguished itself from a set of non-quasi-cooperative machines with prior probability of almost  $\mu_i(\overline{Q}_i)$ . Then it takes action  $a_i^*$  forever. In effect, it first signals to the other player, in a highly complicated way which depends on the other player's prior, that it is quasi-cooperative with high probability. Once it starts cooperating, the optimal behaviour of the other player must then be to do the same. Player  $i$  therefore

has a strategy available which will oblige the normal type of player  $j$  to cooperate forever after some point and so the equilibrium must be approximately cooperative. Clearly, for this argument to work,  $\mu_i(Z_i)$  must be (a) strictly positive and, more problematically, (b) sufficiently large relative to the total probability of non-quasi-cooperative machines which  $Z_i$  does not distinguish itself from. By an ingenious application of a fixed-point argument from computability theory AS show that there is a set  $R$  of machines which guarantees that if  $\mu_i(X_i) > 0$  for all  $X_i \in R$ , then (a) and (b) must be true for some  $Z_i \in R$ .

How does this result square with Theorem 1 above<sup>6</sup>? Suppose that only (AS1) is imposed, and that the allowable machines are those that, given any history of play as input, will halt and produce an action as output, i.e., those which implement strategies of the repeated game. Given this assumption, Theorem 1 still applies in this new setting. In other words, given any pair  $(\mu_1, \mu_2)$  of distributions over non-draconian Turing-implementable strategies which satisfy assumption (AS1), one can construct distributions  $(\mu'_1, \mu'_2)$  as in the proof of Theorem 1, with wider support, such that  $G^\infty(\mu'_1, \mu'_2, \xi)$  has computable equilibrium strategies giving low payoffs;  $(\mu'_1, \mu'_2)$  are also distributions over Turing-implementable strategies and satisfy (AS1). To establish this, note first that if  $s_1$  and  $s_2$  are computable strategies, implemented by machines  $X_1$  and  $X_2$  respectively, then there is a machine  $m_1(X_1, X_2)$  which implements the strategy  $m_1(s_1, s_2)$ . Secondly, the strategy  $\hat{s}_2(v, \mu_1)$  is defined in an algorithmic way and can therefore be implemented by some machine  $X_2(v, \mu_1)$  as long as the support,  $(X_1^n)_{n=1}^\infty$ , of  $\mu_1$  can be enumerated algorithmically and each of its elements  $X_1^n$  can be simulated. Both of these conditions apply because, by assumption,  $\mu_1$  is computable and all strategies in its support are computable. Thirdly, to see that  $\mu'_1$  is a computable distribution, note that, by assumption, there is an algorithm which computes  $\mu_1$ , and also that one can construct an algorithm which, given the code of any machine  $X_1$ , can produce the code of  $m_1(X_1, X_2(v, \mu_1))$ . Therefore there is an algorithm which enumerates the list  $(X_1^n, \theta\mu_1(X_1^n); m_1(X_1^n, X_2(v, \mu_1)), (1 - \theta)\mu_1(X_1^n))_{n=1}^\infty$ . This algorithm clearly computes  $\mu'_1$ . (The same is true for  $\mu'_2$ .)<sup>7</sup>

---

<sup>6</sup>Note that AS use the same payoff criterion as we do here.

<sup>7</sup>For  $X_1 \neq X'_1$ , the new machines can be constructed so that  $m_1(X_1, X_2(v, \mu_1)) \neq m_1(X'_1, X_2(v, \mu_1))$ , even if they represent the same strategy, so that the probability assignment is straightforward. A more

We conclude from this that restricting attention to computable strategies and distributions cannot alter the conclusion of Theorem 1, namely that non-draconian strategies cannot induce cooperation. AS indeed note that for their result to hold, one must include in the cooperation-inducing set which we refer to above as  $R$  some machines which never halt: given any history as input, they compute, in principle, for ever and never produce an output. How does the introduction of non-halting machines affect the logic of Theorem 1?

Consider the stage game  $\Gamma$  below<sup>8</sup> and suppose that the set of allowable machines is the set of all Turing machines, not just those that always halt and play a legal action.

	$\gamma_1$	$\gamma_2$	$\gamma_3$
$\gamma_1$	3, 3	0, 0	0, 1
$\gamma_2$	0, 0	2, 2	0, 1
$\gamma_3$	1, 0	1, 0	0, 0

It is necessary to make some assumption about payoffs when a non-halting strategy is used. We assume that if  $X_i$  fails to halt when played against  $X_j$ , then player  $i$  receives 0 (as a long-run payoff) and  $j$  receives  $K > 0$ ; both receive 0 if, after some history, they both fail to halt.<sup>9</sup> The definition of a draconian strategy (machine) is extended to this case by defining the long-run proportion of periods in which  $X_i$  min-maxes  $X_j$  to be zero whenever either machine fails to halt along the play path they induce; thus any machine which never halts is non-draconian.

**Claim:** *Suppose that all Turing machines are allowed in the supports of the perturbations. If  $\mu_1$  and  $\mu_2$  have non-draconian supports and satisfy (AS1) then there exist distributions over Turing machines  $\mu'_1$  and  $\mu'_2$  such that*

(a)  $\text{supp}(\mu_i) \subseteq \text{supp}(\mu'_i)$  ( $i = 1, 2$ );

(b)  $\mu'_1$  and  $\mu'_2$  satisfy (AS1); and

---

detailed account is provided in an earlier version of the paper.

<sup>8</sup>This example is extended to a wide class of common interest games in an earlier version of the paper.

<sup>9</sup>This is in the spirit of assumptions made in AS; AS justify such assumptions on the grounds that employing a non-halting strategy is equivalent to making an illegal move, and the opponent wins the game provided only that his strategy is legal.

(c)  $\Gamma^\infty(\mu'_1, \mu'_2, \xi)$  has an equilibrium in computable strategies, independent of  $\xi$ , with payoffs which converge to  $(2, 2)$  as  $\xi \rightarrow 0$ .

**Proof.** Let  $\widehat{s}_i$  be the following strategy: ‘Play  $\gamma_2$  as long as  $j \neq i$  has always played  $\gamma_2$ ; if  $j$  does not play  $\gamma_2$ , play  $\gamma_3$  forever’. This is clearly computable by some machine  $\widehat{X}_i$  and  $(\widehat{s}_1, \widehat{s}_2)$  gives each player a long-run payoff of approximately 2 for small  $\xi$ . New perturbations  $(\mu'_1, \mu'_2)$  are defined as in the proof of Theorem 1 (adapted to the computable case as discussed above), where now if  $X_1 \in \text{supp}(\mu_1)$  fails to halt after some history in which 2 has not deviated from  $\widehat{X}_2$ ,  $m_1(X_1, \widehat{X}_2)$  will also fail to halt (and likewise for  $m_2(\widehat{X}_1, X_2)$ ). Given these definitions,  $(\widehat{s}_1, \widehat{s}_2)$  is an equilibrium, for  $\theta$  small enough, of  $\Gamma^\infty(\mu'_1, \mu'_2, \xi)$ , and yields a long-run payoff of approximately 2 for small  $\xi$ , which is inefficient. To see that this is so, note first that  $\widehat{s}_1$  and  $\widehat{s}_2$  are best replies to each other. Is  $\widehat{s}_2$  (or, equivalently,  $\widehat{X}_2$ ) a best reply to  $\theta X_1 + (1 - \theta)m_1(X_1, \widehat{X}_2)$ , where  $X_1 \in \text{supp}(\mu_1)$ ?  $\widehat{X}_2$  gives a strictly positive payoff against any non-draconian strategy (hence against  $X_1$ , which is non-draconian by assumption, and also against  $m_1(X_1, \widehat{X}_2)$ , which is indistinguishable from  $X_1$  when paired with  $\widehat{X}_2$ ). If, along the play path induced by  $X_1 \in \text{supp}(\mu_1)$  and  $\widehat{X}_2$ ,  $X_1$  always halts then so does  $m_1(X_1, \widehat{X}_2)$  and the argument is as in the proof of Theorem 1 (for small  $\theta$ , a deviation by player 2 from this path will yield almost zero, whether or not both  $X_1$  and  $X_2$  always halt following the deviation). Suppose that, along this path,  $X_1$  fails to halt at time  $t$ . Then  $\widehat{X}_2$  gives player 2 a payoff of  $K > 0$  against both  $X_1$  and  $m_1(X_1, \widehat{X}_2)$ . Any deviation from  $\widehat{X}_2$  after  $t$  will also (trivially) give  $K$  while any deviation before  $t$  will be min-maxed forever by  $m_1(X_1, \widehat{X}_2)$ , giving less than  $K$ , for small  $\theta$ . Therefore, for small  $\theta$ ,  $\widehat{X}_2$  is optimal against  $\theta X_1 + (1 - \theta)m_1(X_1, \widehat{X}_2)$ . (A symmetric argument applies to  $\widehat{X}_1$ .) This shows that  $(\widehat{s}_1, \widehat{s}_2)$  is an equilibrium. The argument that  $\mu'_i$  satisfies (AS1) is the same as in the case where only halting machines were allowed. ■

This example shows that the logic of Theorem 1 extends to the case where non-halting machines are allowed. AS’s result therefore depends on the two further assumptions (AS2) and (AS3). On its own, (AS2) does not affect the logic of the above constructions. Suppose that  $(\mu_1, \mu_2)$  satisfies (AS2). As defined,  $(\mu'_1, \mu'_2)$  may not do so and our construction may

therefore violate the assumptions of AS's theorem. However, with slight modifications of the definitions, (AS2) will carry over from  $(\mu_1, \mu_2)$  to  $(\mu'_1, \mu'_2)$ . Taking first the case in which machines must always halt, define a machine  $n_i(X_1, X_2)$  which is the same as  $m_i(X_1, X_2)$  except that if player  $j \neq i$  deviates from  $X_j$ ,  $i$  plays  $a_i^*$  once and thereafter always min-maxes (as before). Let  $n_i$  take the place of  $m_i$  in the preceding arguments (i.e., in the definition of  $\mu'_i$ ). Any strategy of the form  $n_i(X_1, X_2)$  is necessarily non-quasi-cooperative because there are paths along which it cooperates and thereafter always min-maxes (hence does not cooperate). Any such strategy is also draconian and so not an element of  $\text{supp}(\mu_i)$ ; hence  $\mu'_i(\overline{Q}_i) = \theta\mu_i(\overline{Q}_i) + (1-\theta)$ , which is approximately computable. Therefore  $(\mu'_1, \mu'_2)$  also satisfies (AS2). Next, take our example  $\Gamma$  above in which all Turing machines are allowed (including those which halt after some histories but not others). The problem here is that if  $X_1$  fails to halt in period 1 then  $n_1(X_1, \widehat{X}_2)$  will also fail to halt and will therefore be quasi-cooperative (there is no path on which it cooperates and subsequently fails to cooperate since it does not generate any path). Therefore we cannot conclude, as we did above, that  $\mu'_1(\overline{Q}_1) = \theta\mu_1(\overline{Q}_1) + (1-\theta)$ , an approximately computable number. However, a further modification of  $(\mu'_1, \mu'_2)$  avoids this difficulty too: for each  $X_1 \in \text{supp}(\mu_1)$ , instead of adding  $n_1(X_1, \widehat{X}_2)$  to the support, add three machines, giving each the same probability, all of which are identical to  $n_1(X_1, \widehat{X}_2)$  except that each plays a different action in period 1 (one machine for each of the three actions). Call these  $n_1(X_1, \widehat{X}_2, \gamma_1)$  etc. Amend  $\mu'_2$  similarly. Along any path, consistent with  $\widehat{X}_2$ , on which  $X_1$  has always played an action, one of the three will be indistinguishable from  $X_1$  (and the probabilities of the new machines can be chosen so that the posterior probabilities for  $X_1$  and this indistinguishable machine will again be in the ratio  $\theta : 1 - \theta$ ).<sup>10</sup> The three machines  $n_1(X_1, \widehat{X}_2, \gamma_i)$ ,  $i = 1, 2, 3$ , are all non-quasi-cooperative (and draconian, hence disjoint from  $\text{supp}(\mu_1)$ ). Again, therefore, we conclude that  $(\mu'_1, \mu'_2)$  satisfies (AS2). The reason that  $n_1(X_1, \widehat{X}_2, \gamma_i)$  is non-quasi-cooperative is that, if player 2 deviates from  $\widehat{X}_2$  in period 1,  $n_1(X_1, \widehat{X}_2, \gamma_i)$  plays the sequence  $(\gamma_i, \gamma_1 = a_1^*, \gamma_3, \gamma_3, \dots)$ ; therefore there is a path along which it cooperates and subsequently fails to cooperate.

---

<sup>10</sup>  $\widehat{X}_2$  is a best response against the other two new machines, and so is a best response against player 1's perturbed strategy for small  $\theta$ .

AS’s final assumption is (AS3), namely that the only machines allowable in the supports of the perturbations are those that always halt and those that never halt. In particular, if  $X_1$  is one of the latter type, then  $n_1(X_1, \widehat{X}_2, \gamma_i)$  is not allowed because it will halt in every period if player 2 fails to play  $\gamma_2$  in period 1, but will never halt after period 1 if player 2 plays  $\gamma_2$  (because it is mimicking  $X_1$ ). The discussion above shows that in the Turing machine framework, given assumptions (AS1) and (AS2), AS’s result requires that machines of the form  $n_1(X_1, \widehat{X}_2, \gamma_i)$  are not all allowed. An intuition for why AS’s result does not work if all Turing machines are allowed is as follows. Suppose that a machine attempts to enumerate recursively, and distinguish itself from, almost all of the non-quasi-cooperative machines which will, on some paths, play in every period. There is nothing to guarantee that it will not, at some stage, get stuck trying to simulate a machine which will fail to halt on the current history.

A referee has suggested that AS’s result would still apply if, in addition to the machines permitted by (AS3), any machine is permitted which, for some  $t \geq 1$ , halts after all histories of length  $t$  or less and thereafter never halts. ((AS2) would also have to be adapted to this context and one would need the further assumption that the set of machines in the support of the perturbation which always halt and are non-quasi-cooperative is recursively enumerable). Note, however, that this would still disallow machines of the form  $n_1(X_1, \widehat{X}_2, \gamma_i)$ . Our view is that, for the purposes of establishing a strong “perturbation implies efficiency” result, if a player is allowed to believe that the other player might play  $X_1$  then she should also be allowed to believe that he might play ‘alter ego’ strategies of the form  $m_1(X_1, \widehat{X}_2)$  and  $n_1(X_1, \widehat{X}_2, \gamma_i)$ . After all, the program which defines  $n_1(X_1, \widehat{X}_2, \gamma_i)$  is only a small modification of  $X_1$ ’s program.<sup>11</sup>

Is it possible that a version of AS’s result is obtainable without restricting the set of allowable machines? Suppose that some restriction on allowable perturbations could be

---

<sup>11</sup>We do not argue that if player 2 puts strictly positive probability on 1 playing  $X_1$  then she *must* also put strictly positive probability on, say,  $n_1(X_1, \widehat{X}_2, \gamma_i)$ ; only that, in the context of a strong ‘perturbation implies efficiency’ theory, the analyst should not forbid her to do so. The former viewpoint—that if a player’s prior gives strictly positive probability to strategy  $s$  then it must give strictly positive probability to strategies similar to  $s$ —is reminiscent of the assumption of *weak caution* found in Nachbar (1999) although he uses a different notion of similarity to the one implied here.

found (call it  $P$ ) which delivers the result even when all Turing machines are allowable: either a variant of AS2 or something unrelated to it. Then our arguments above show that this restriction, though satisfied by  $\mu_i$ , must be unsatisfied by  $\mu'_i$ . We would argue that the set of legitimate perturbations should satisfy the property that if  $\mu_i$  is a legitimate perturbation then  $\mu'_i$  is also legitimate because  $\mu'_i$  is obtained from  $\mu_i$  by adding legitimate machines via a straightforward algorithmic process. If this is so, however, it follows that property  $P$  cannot be satisfied by all legitimate perturbations.

## 5 Concluding Comments

A number of papers have explored the possibility that equilibria of perturbed repeated common interest games must, if the perturbation is sufficiently inclusive, be efficient because a player is able to ‘signal’ to the other player, through his actions, that he is likely to behave cooperatively if the other player does likewise. The logic of Theorem 1 is that whatever the signal might be there is always a strategy which sends exactly that signal on the (bad) equilibrium path and enforces the equilibrium with non-cooperative punitive actions off that path. Therefore it is not necessarily possible for one player to persuade the other that he is ‘nice’. Rather than attempting this, he might try to persuade the other player to cooperate by punishing him severely if he does not. Theorem 2 shows that in the undiscounted case this approach works and therefore that if the perturbation has a wide enough support any equilibrium must be efficient.

### References

- Anderlini, L., and H. Sabourian (1995): “Cooperation and Effective Computability,” *Econometrica*, 63, 1337-1369.
- Aumann, R., and S. Sorin (1989): “Cooperation and Bounded Recall,” *Games and Economic Behavior*, 1, 5-39.
- Cripps, M., and J.P. Thomas (1995): “Reputation and Commitment in Two-Person Repeated Games without Discounting,” *Econometrica*, 63, 1401-1419.

Evans, R., and J. P. Thomas (1997): “Reputation and Experimentation in Two-Player Repeated Games with Discounting,” *Econometrica*, 65, 1153-1173.

Fudenberg, D., and D. K. Levine (1989): “Reputation and Equilibrium Selection in Games with a Patient Player,” *Econometrica*, 57, 759–778.

Nachbar, J. H. (1999): “Rational Bayesian Learning in Repeated Games,” Mimeo, Washington University.

Sorin, S. (1999): “Merging, Reputation and Repeated Games with Incomplete Information,” *Games and Economic Behavior*, 29, 274-308.

Watson, J. (1996): “Reputation in Repeated Games with No Discounting,” *Games and Economic Behavior*, 15, 82-109.

## 6 Appendix

PROOF OF LEMMA: Take  $v_1$  as in the statement of the lemma and  $a_1 \notin A_1^m$ . Since  $a_1 \notin A_1^m$  player 2 has an action  $a_2 \in A_2$  such that  $g_2(a_1, a_2) > 0$ . By the definition of  $\hat{v}_1$ , 2 has a mixed action  $\tilde{\alpha}_2$  such that  $g_1(a_1, \tilde{\alpha}_2) \leq \hat{v}_1$  and  $g_2(a_1, \tilde{\alpha}_2) \geq 0$ . By continuity, there exist  $\hat{\alpha}_2(a_1)$ , with rational weights, and  $\varepsilon_1(a_1)$  such that  $g_1(a_1, \hat{\alpha}_2(a_1)) < v_1 - \varepsilon_1(a_1)$  and  $g_2(a_1, \hat{\alpha}_2(a_1)) > \varepsilon_1(a_1)$ . Let  $\varepsilon_1 := \min_{a_1 \notin A_1^m} \varepsilon_1(a_1)$ . Let  $(\hat{a}_2^n(a_1, v_1))_{n=1}^\infty$  consist of the infinite repetition of a finite sequence in which the frequency of each action  $a_2$  equals  $\hat{\alpha}_2(a_1)(a_2)$ . ■

PROOF OF THEOREM 1: Take any  $(\mu_1, \mu_2)$  with  $S_i = \text{supp}(\mu_i)$  ( $i = 1, 2$ ) and any  $v = (v_1, v_2)$  as in the statement of the theorem. Let  $\{s_i^n\}_{n=1}^\infty$  be an enumeration of  $S_i$ . Now define a pair of strategies  $(\hat{s}_1(v, \mu_2), \hat{s}_2(v, \mu_1))$  as follows.  $\hat{s}_1(v, \mu_2)$  is defined in the following algorithmic way ( $\hat{s}_2(v, \mu_1)$  is defined symmetrically):

Begin in Phase I.

Phase I: follow the path  $(\tilde{a}^t(v))_{t=1}^\infty$ . If player 2 deviates from  $(\tilde{a}^t(v))_{t=1}^\infty$  switch to Phase II(1).

Phase II( $n$ ) ( $n \geq 1$ ): check if the history so far,  $h_t$ , is consistent with  $s_2^n$ , the  $n$ th strategy in the enumeration of  $\text{supp}(\mu_2)$ . If it is consistent ‘predict’ 2’s action to be  $s_2^n(h_t)$  and

1. if  $s_2^n(h_t) \in A_2^m$  then play a best response which holds player 2 to at most  $\hat{v}_2$ ;
2. if  $s_2^n(h_t) \notin A_2^m$  then play  $\hat{a}_1^N(s_2^n(h_t), v_2)$  where this is the  $N$ th time during this phase that  $s_2^n(h_t)$  has been  $s_2^n$ ’s action.

If  $h_t$  is inconsistent with  $s_2^n$ , switch to Phase III( $n$ ).

Phase III( $n$ ): min-max 2 until her average payoff computed to the previous period is below  $n^{-1}$ . Then switch to Phase II( $n+1$ ) (this could be immediate) unless (in the case of a finite perturbation)  $n = \#S_2$ , in which case min-max forever.

Since  $S_1$  and  $S_2$  are not draconian there exists  $k \in (0, 1)$  such that  $\sup_{s_i \in S_i} \beta_i(s_i) \leq k$  for  $i = 1, 2$ . Let  $\epsilon_i$  be as defined in the statement of the lemma. Since  $k < 1$  and each  $\epsilon_i > 0$ , we can find a rational  $\theta \in (0, 1)$  so that

$$v_i > \theta v_i^*, \quad i=1, 2, \quad (2)$$

and

$$(1 - k)\epsilon_j > \theta v_i^*, \quad i, j = 1, 2, i \neq j. \quad (3)$$

For convenience write  $\hat{s}_i$  in place of  $\hat{s}_i(v, \mu_j)$ . Now let  $\mu'_1$  be defined as follows: if  $s_1 \in S_1$ , let  $\mu'_1(s_1) = \theta \mu_1(s_1)$ , and otherwise let  $\mu'_1(s_1) = (1 - \theta) \sum \mu_1(s'_1)$  where the sum is over  $s'_1$  such that  $s_1 = m_1(s'_1, \hat{s}_2)$ . Since each  $s_1^n \in S_1$  is non-draconian while  $m_1(s_1^n, \hat{s}_2)$  is draconian, the latter strategies are disjoint from  $S_1$ ; consequently  $\mu'_1$  is well-defined. Define  $\mu'_2$  symmetrically.

We shall show that, for all  $\xi$ ,  $(\hat{s}_1, \hat{s}_2)$  is an equilibrium of the perturbed game  $G^\infty(\mu'_1, \mu'_2; \xi)$ ; for  $\xi$  small enough, expected payoffs will be within  $\eta$  of  $v$  as required. (a) Conditional on facing  $\hat{s}_2$ , if 1 plays  $\hat{s}_1$  play remains in Phase I and player 1 receives  $v_1$ ; if 1 deviates, then either (i) 1 plays a strategy which is consistent, against  $\hat{s}_2$ , with some  $s_1^n \in S_1$ , in which case play ends up in Phase II( $n$ ) of  $\hat{s}_2$  (or Phase II( $n'$ ) for some  $n' < n$  if  $s_1^{n'} \in S_1$  plays the same way against  $\hat{s}_2$  as does  $s_1^n$ ) with 2 correctly predicting 1’s actions,

and so 1 receives at most  $v_1 - \epsilon_1$ ; or (ii) 1 plays a strategy which is inconsistent, against  $\hat{s}_2$ , with any  $s_1 \in S_1$ , thus triggering an infinite number of punishment phases, so  $\pi_1 = 0$ . Hence  $\hat{s}_1$  is a best response against  $\hat{s}_2$ . (b) Conditional on facing  $\theta s_2^n + (1 - \theta)m_2(s_2^n, \hat{s}_1)$ , where  $s_2^n \in S_2$ , there are two possibilities. (i)  $s_2^n$  plays the same against  $\hat{s}_1$  as does  $\hat{s}_2$ , so that play remains in Phase I against both  $s_2^n$  and  $m_2(s_2^n, \hat{s}_1)$ . Hence  $\hat{s}_1$  yields  $v_1$  against  $\theta s_2^n + (1 - \theta)m_2(s_2^n, \hat{s}_1)$ , while a deviation from  $\hat{s}_1$  which leads to a different play path against  $s_2^n$  will be min-maxed with probability  $(1 - \theta)$ , and thus yields an expected payoff no greater than  $\theta v_1^*$ , which is smaller than  $v_1$  by (2). (ii)  $s_2^n$  plays differently against  $\hat{s}_1$  than does  $\hat{s}_2$ . In this case  $\hat{s}_1$  yields  $\pi_1 \geq (1 - k)\epsilon_2$  against  $\theta s_2^n + (1 - \theta)m_2(s_2^n, \hat{s}_1)$ , as after some finite time play will remain (against both  $s_2^n$  and  $m_2(s_2^n, \hat{s}_1)$ ) in Phase II( $n'$ ),  $n' \leq n$ , of  $\hat{s}_1$ , and 2 plays a min-max action no more than a fraction  $k$  of the time. Any deviation from  $\hat{s}_1$  which leads to a different play path against  $s_2^n$  will be min-maxed with probability  $(1 - \theta)$ , so the expected payoff is no greater than  $\theta v_1^*$ , which is smaller than  $(1 - k)\epsilon_2$  by (3). So  $\hat{s}_1$  is a best response against  $\theta s_2 + (1 - \theta)m_2(s_2, \hat{s}_1)$  for all  $s_2 \in S_2$ . It follows that  $\hat{s}_1$  must also be a best response against 2's unconditional strategy. The argument for the optimality of  $\hat{s}_2$  is symmetric. ■

**PROOF OF THEOREM 2:** Fix a Nash equilibrium of the game, and let  $s_2$  be an optimal pure strategy for player 2 against player 1's perturbed strategy  $\sigma_1$ . Now, suppose that  $s_2$  triggers an infinite number of punishment phases when played against  $\tilde{s}_1$ . We shall establish a contradiction. Let  $\tilde{h}_\infty$  be the infinite history resulting from the play of  $\tilde{s}_1$  against  $s_2$ , with, for any integer  $t$ ,  $\tilde{h}_t$  the corresponding  $t$ -period history. Player 2's payoff  $\pi_2(\tilde{h}_\infty) = 0$ . For any  $\epsilon \in (0, 1)$  there exists a  $t(\epsilon)$  such that  $\text{prob}_{\sigma_1, s_2}(\tilde{h}_\infty | \tilde{h}_{t(\epsilon)}) \geq 1 - \epsilon$ . (To see this, let  $\tilde{H}_t$  be the set of infinite histories consistent with  $\tilde{h}_t$  and note that  $\tilde{H}_t \supseteq \tilde{H}_{t+1} \supseteq \tilde{H}_{t+2} \dots$  and that  $\lim \tilde{H}_t = \{\tilde{h}_\infty\}$ . Therefore, by the monotone convergence theorem,  $\text{prob}(\tilde{H}_t) \rightarrow \text{prob}(\tilde{h}_\infty)$ . Thus,  $\text{prob}(\tilde{h}_\infty | \tilde{h}_t) \equiv \text{prob}(\tilde{h}_\infty | \tilde{H}_t) = \text{prob}(\tilde{h}_\infty) / \text{prob}(\tilde{H}_t) \rightarrow 1$ ). Choose  $\epsilon$  so that  $\epsilon < \mu_1(\tilde{s}_1)$ . Then  $E_{\sigma_1, s_2}[\pi_2 | \tilde{h}_{t(\epsilon)}] \leq \epsilon v_2^* < \mu_1(\tilde{s}_1)v_2^*$ . Suppose that player 2 changes her strategy after the history  $\tilde{h}_{t(\epsilon)}$  to the following: so long as player 1 adheres to  $\tilde{s}_1$ , then play  $a_2^*$ ; otherwise play prudently to achieve the min-max payoff (0). With probability at least  $\mu_1(\tilde{s}_1)$  (conditional on  $\tilde{h}_{t(\epsilon)}$ ) player 2 is facing  $\tilde{s}_1$  and this strategy will then lead to cooperation at the Pareto dominant payoff vector; otherwise the payoff is

at least zero. Consequently this strategy yields a conditional expected payoff of at least  $\mu_1(\tilde{s}_1)v_2^*$ , which is greater than  $E_{\sigma_1, s_2}[\pi_2 \mid \tilde{h}_{t(\epsilon)}]$ . Therefore the overall (unconditional) payoff can be increased, contradicting the assumption that  $s_2$  is optimal. Thus, an optimal  $s_2$  can only trigger a finite number of punishment phases against  $\tilde{s}_1$ , which implies that cooperation (at  $a^*$ ) occurs after some point. Finally, it follows that if player 1 plays the strategy  $\tilde{s}_1$ , she will receive a payoff of  $v_1^*$  against any optimal pure strategy  $s_2$  of player 2, and because the latter's equilibrium (unperturbed) strategy must put probability one on such strategies, the result follows. ■