Gradualism and Irreversibility[1]

Ben Lockwood (University of Warwick)

Jonathan P. Thomas (University of St. Andrews)

Preliminary version, September 1998
This version, April 1999

Abstract: This paper considers a class of two-player dynamic games in which each player controls a one-dimensional variable which we interpret as a level of cooperation. In the base model, there is an irreversibility constraint stating that this variable can never be reduced, only increased. It otherwise satisfies the usual discounted repeated game assumptions. Under certain restrictions on the payoff function, which make the stage game resemble a continuous version of the Prisoners' Dilemma, we characterize efficient symmetric equilibria, and show that cooperation levels exhibit gradualism and converge, when payoffs are smooth, to a level strictly below the one-shot efficient level: the irreversibility induces a steady-state as well as a dynamic inefficiency. As players become very patient, however, payoffs converge to (though never attain) the efficient level. We also show that a related model in which an irreversibility arises through players choosing an incremental variable, such as investment, can be transformed into the base model with similar results. Applications to a public goods sequential contribution model and a model of capacity reduction in a declining industry are discussed. The analysis is extended to incorporate partial reversibility, asymmetric equilibria, and sequential moves.

Keywords: Cooperation, repeated games, gradualism, irreversibility, public goods.

JEL Classification Numbers: C73, H41.

---

[1] This paper was prepared for the ESRC Game Theory meeting in Kenilworth, September 1998. We are grateful for comments from participants at this seminar, and also at presentations at Royal Holloway College, St. Andrews University, Southampton University and the Centre for Globalisation and Regionalisation, University of Warwick. We are also particularly grateful for many helpful discussions with Carlo Perroni, for valuable comments and suggestions from Martin Cripps, and also to Daniel Seidmann, Norman Ireland, Steve Matthews and William Walker. Both authors gratefully acknowledge the financial support of the ESRC Centre for Globalisation and Regionalisation, University of Warwick.

1. Introduction

We consider a model in which in every period, there is a Prisoner's Dilemma structure; agents have some mutual interest in cooperating, despite the fact that it is not in any agent's individual interest to cooperate. We suppose that this situation is repeated over time, and, crucially, subject to irreversibility, in the sense that an agent cannot reduce her level of cooperation once increased. In this setting, irreversibility has two opposing e¤ects. First, it aids cooperation, through making deviations in the form of reduced cooperation impossible. Second, it limits the ability of agents to punish a deviator. We consider the complex interplay of these two forces.

The key role of irreversibility in a¤ecting cooperation can be explained more precisely as follows. In the above model, suppose that every player has a (continuous) scalar action variable, which we interpret as a level of cooperation. We say that partial cooperation occurs in some time period if some player chooses a level of this action variable higher than the stage-game Nash equilibrium level, where the latter is the smallest feasible value of the action variable. Full cooperation is a level of this action variable that maximizes the joint payo¤ of the players[2]. In general, partial cooperation in any time-period can only be achieved if deviation by any agent can be punished by the other agents in some way.

Now the above model without reversibility is just a repeated Prisoner's Dilemma, and in that case, it is well-known that the most e¤ective (and credible) punishments take the form of "sticks", i.e., threats to reduce cooperation back to the stage-game Nash equilibrium. With irreversibility, such punishments are no longer feasible; instead, deviators can only be punished by withdrawal of "carrots", that is, threats take the form of withdrawal of promised higher levels of cooperation in future. It follows immediately from this that irreversibility causes gradualism, i.e., any (subgame-perfect) sequence of actions involving partial cooperation cannot involve an immediate move to full cooperation[3].

[2] The model is symmetric, i.e., players have identical per-period payo¤s given a permutation of their actions. So, the full cooperation level is the same for each player.

[3] This observation is not entirely new; for example, Schelling (1960, p45) makes a similar point. Admati and Perry (1991) and Marx and Matthews (1998) present equilibria of a dynamic voluntary contribution game which exhibit gradualism. However, to the best of our knowledge, our paper provides the …rst general characterization of gradualism in cooperation due to irreversibility.

Our first contribution is to refine and extend this basic insight. First, we show that any (subgame-perfect) equilibrium sequence of actions involving cooperation must have the level of cooperation rising in every period, but that full cooperation is never reached in finite time. So, as the level of cooperation in any period is bounded above by the full cooperation level, all equilibrium sequences will converge. We focus on the (symmetric) efficient equilibrium sequence i.e. the one that maximises the present value of payoffs of either player. A key question then is: to what value does this efficient equilibrium sequence converge? It turns out that if payoffs are smooth (differentiable) functions of actions, convergence will be to a level strictly below the full cooperation level, no matter how patient agents are. For the case where payoffs are linear up to some joint cooperation level, and constant or decreasing thereafter (the linear kinked case), the results are different — above some critical discount factor equilibrium cooperation can converge asymptotically to the fully efficient level. Below this critical discount factor, no cooperation at all is possible.

The reason for the asymptotic inefficiency in the smooth payoff case is that close to full cooperation, returns from additional mutual cooperation are second-order, whereas the benefits to deviation (not increasing cooperation when the equilibrium path calls for it) remain first-order. The future gains from sticking to an increasing mutually cooperative path will be insufficient to offset the temptation to deviate. It follows that it will be impossible to sustain equilibrium paths close to full cooperation.

Despite this result, inefficiency disappears in the limit as players become patient in the sense that the limit value of the sequence, and player payoffs, both converge to fully efficient levels as discounting goes to zero. However, the asymptotically efficient path of actions in our model is quite different that in the standard "folk theorem" for repeated games: that in the latter case, (without irreversibility) above some critical discount factor the efficient cooperation level can be attained exactly and immediately.

Later sections of the paper then extend the basic model in several directions. First, we recognize that our basic model is very stylized. In many economic applications, irreversibility arises more naturally when the level of "cooperation" is a stock variable which may benefit both players, and it is incremental investment in cooperation that is costly and non-negative, implying the stock variable is irreversible. Therefore, in Section 4, we present an “adjustment cost” model with these features, and show that it can be reformulated so that it is a special case of our base model. We then apply the adjustment cost model to study sequential public good contribution games (Admati and Perry (1991), Marx and Matthews (1998)) and capacity reduction in a declining industry (Ghemawat and Nalebu¤(1990)). These applications illustrate the extent to which our results are applicable to variety of disparate areas of economics.

A second key extension is to allow a small amount of irreversibility, so that any player can reduce his cooperation level by some (small) …xed percentage. This has two countervailing e¤ects. The …rst is to make deviation more pro…table; the deviator at t can lower his cooperation level below last period’s, rather than just keeping it constant. The second e¤ect is to make punishment more severe; the worst possible perfect equilibrium punishment of the deviator is for the other player to reduce his cooperation over time, rather than just not increase it. A priori, it is not clear which e¤ect will dominate. Nevertheless, we are able to show that for a small amount of reversibility the second e¤ect dominates, and in the linear kinked case it dominates for any degree of reversibility. In our model, then, reversibility is desirable in that it allows more cooperative equilibria to be sustained.

The base model also assumes that (two) players move simultaneously, and that they both choose the same[4] path of actions (the symmetric path). In Section 6 we allow players to choose di¤erent action paths, and in this Section, we obtain a (partial) characterization of the Pareto-frontier of the set of equilibrium payo¤s, and how it changes with the discount factor. In Section 7, we allow payers to move sequentially. We show that the equilibrium payo¤s in this game are a subset of those in the simultaneous move game, but that as discounting goes to zero, the e¢cient symmetric payo¤ in the symmetric move game can be arbitrarily closely approximated by equilibrium payo¤s in the sequential game, so that asymptotically, the order of moves has little e¤ect on achievable payo¤s.

There is a small literature on games with the features we consider here. Admati and Perry (1991) and Marx and Matthews (1998) in particular have considered sequential public good contribution games in a formally similar context. Cooperation in such models

[4] As the model is symmetric, i.e. players have identical per-period payo¤s given a permutation of their actions, this is a natural base case.

is the sum of an individual's contributions, and this is irreversible. Gale (1997) has considered a class of sequential move games which he dubs monotone games. For games with "positive spillovers", which include the class of games considered here, he characterizes long-run efficient outcomes when there is no discounting. In particular, his results imply that in a sequential-move version of our model without discounting, first-best outcomes are attainable.[5]

Of these papers, possibly the closest is Marx and Matthews (1998). The relationship between the two papers is as follows. First, the two papers consider quite different models, although there is some overlap. Marx and Matthews(1998) consider a number of different voluntary contribution games, where a number of players simultaneously make contributions to a public project over T periods, and where T may be finite or infinite. Each player gets a payoff that is linear in the sum of cumulative contributions, plus possibly a "bonus" when the project is completed. One case of their model (T infinite, two players, no bonus) can be reformulated as an "adjustment cost" variant of our model with linear kinked payoffs (as argued in detail in Section 4.1).

In this version of their model, Marx and Matthews (1998) construct a subgame-perfect equilibrium which is approximately efficient when discounting is negligible[6], whereas we are able to characterise efficient subgame-perfect equilibria for any fixed value of the discount factor. Specifically, our results show[7] that in their model, the equilibrium with completion which they construct is in fact efficient for any discount factor above a critical value, and conversely when the discount factor is below the critical value, there are no contributions made in the efficient equilibrium (see Section 4.1 for more details).

We see our model as being applicable to a wide variety of situations in addition to those already mentioned above. Nuclear disarmament between two countries is one example— here cooperation would be measured by the extent of disarmament. While it

[5] The games considered in this literature allow for the possibility that a player's payoff may be increasing in his or her own cooperation level (on completion of the project in the public good model). The lack of this feature here allows us to obtain results without needing to impose linearity or no discounting.

[6] Corollary 3(ii), Marx and Matthews(1998). Note that their results are stated for $n > 2$ players also.

[7] We are also able to characterise equilibrium in the case of linear kinked payoffs (which includes the infinite-horizon contribution game without a bonus as a special case) when the two players contribute asymmetrically, whereas Marx and Matthews study only the symmetric equilibrium in this version of their model (although in their paper, they study other versions of their model where players behave asymmetrically).

may be desirable to move immediately to total disarmament, this is not an equilibrium because either country would prefer to have the other destroy its stockpile while retaining its own. Disarmament must proceed gradually, and our results give conditions under which the limit of the process is complete or only partial disarmament.

Another example would be in trade negotiations. For example, GATT negotiations are known for their gradualism, although there has been little theoretical work on this (see Bagwell and Staiger, 1997). If concessions are irreversible, or if irreversibilities arise in investment such that shifting capital away from import competing technologies cannot easily be reversed, then a similar story to the one we give can be told to explain gradualism. A formal treatment of a related idea in the negotiation context is in Comte and Jehiel (1998) who consider the impact of outside options in a negotiation model where concessions by one party increase the payoff the other party gets in a dispute resolution phase.

A further fruitful application is to environmental problems. For example, environmental cooperation may take the form of installation of costly abatement technology. Once installed, this technology may be very expensive to replace with a "dirtier" technology, e.g., conversion of automobiles to unleaded petrol would be expensive to reverse. Consequently it will again be difficult to punish deviants by reversing the investment.[8] Similarly, destruction of capital which leads to over-exploitation of a common property resource (e.g., fishing boats) will also fit into the general framework of the paper if it is difficult to reverse.

2. The Model and Preliminary Results

There are two players[9] $i = 1, 2$. In each period, $t = 1, 2, \ldots,$ each player i simultaneously chooses an action variable $c_i \in \Re_+$, measuring i's level of cooperation[10]. The per-period payoff to player 1 is $\pi(c_1, c_2)$ with that of player 2 being $\pi(c_2, c_1)$. So, payoffs of the two players are identical following a permutation of the pair of actions. Also, we assume that π is continuous, strictly decreasing in c_1 and strictly increasing in c_2. Payoffs over the

[8] We are grateful to Anthony Heyes for suggesting this application.

[9] Our main results generalise straightforwardly to more than two players.

[10] The action spaces can also be bounded, i.e., $c_i \in [0, \bar{c}]$, as long as $\bar{c} \geq c^*$.

infinite horizon are discounted by common discount factor δ; $0 < \delta < 1$:

In this setting, we shall initially be restricting attention to symmetric equilibria. So, we can define the first-best efficient level(s) of cooperation as the value(s) of c, that maximise $w(c) := \pi(c; c)$: We assume the following weak property of $w(c)$:

A1. There exists a $c^* > 0$ such that $w(c)$ is strictly increasing in c for all $0 \leq c < c^*$, and $w(c) \leq w(c^*)$ for all $c \in \Re_+$.

This is satisfied if $w(c)$ is concave with a finite maximum or even single-peaked: Note that c^* is the smallest first-best efficient level of cooperation: We assume that the choice of action is irreversible in every period, i.e.,

$$c_{i;t} \geq c_{i;t-1},\ i = 1; 2,\ t = 1; 2; \ldots; \tag{2.1}$$

where $c_{i;t}$ is i's action in period t; and, without loss of generality, we set $c_{1;0} = c_{2;0} = 0$.

A game history at time t is defined in the usual way as $\{(c_{1;\tau}; c_{2;\tau})\}_{\tau=1}^{t-1}$. Both players can observe game histories. A pure strategy for player $i = 1; 2$ is defined in the usual way as a sequence of mappings from game histories in periods $t = 1; 2 \ldots$ to values of $c_{i;t}$ in $\Re_+$, and where every pair $(c_{i;t-1}; c_{i;t})$ satisfies (2.1). An outcome path of the game is a sequence of actions $\{c_{1;t}; c_{2;t}\}_{t=1}^{\infty}$ that is generated by a pair of pure strategies. We are interested in characterizing subgame perfect Nash equilibrium outcome paths. For the moment, we restrict our attention to symmetric equilibrium[11] outcome paths where $c_{1;t} = c_{2;t} = c_t$, $t = 1; 2; \ldots$; and we denote such paths by the sequence $\{c_t\}_{t=1}^{\infty}$.

We now derive necessary and sufficient conditions for some fixed symmetric outcome path $\{c_t\}_{t=1}^{\infty}$ to be an equilibrium. Note that the worst punishment that j could impose on i for deviating at date t from such a path is for j to set c_j as low as possible. So, if i deviates at t, the worst punishment is for j to set $c_{j;\tau} = c_{j;t}$, all $\tau > t$: Also whatever action is chosen by j, it is always a best response for i to set c_i as low as possible. It follows that this punishment is credible, and, given the punishment, i's optimal deviation at t from the symmetric path $\{c_t\}_{t=1}^{\infty}$ is to set $c_{i;\tau} = c_{t-1}$ for all $\tau \geq t$. Consequently, for a non-decreasing sequence $\{c_t\}_{t=1}^{\infty}$ to be an equilibrium outcome path it is necessary and

[11] In the sequel, it is understood that "equilibrium" refers to subgame-perfect Nash equilibrium.

sufficient that $\{c_t\}_{t=1}^{\infty}$ satisfies, for all $t \geq 1$, the inequalities

$$\frac{\pi(c_{t-1}, c_t)}{1-\delta} \leq \pi(c_t, c_t) + \delta\pi(c_{t+1}, c_{t+1}) + \ldots \quad . \tag{2.2}$$

So, as $c_t \geq c_{t-1}$ from the irreversibility constraint (2.1), the interpretation of (2.2) is that in the event of defection, both players stop increasing their levels of cooperation.

Let C_{SE} be the set of non-decreasing paths $\{c_t\}_{t=1}^{\infty}$ that satisfy (2.2), and we refer to any path in C_{SE} as a (symmetric) equilibrium path. We now note two basic properties of sequences in C_{SE}:

Lemma 2.1. If $\{c_t\}_{t=1}^{\infty}$ is an equilibrium path, then (i) $c_t < c^*$, for all $t \geq 1$; and (ii) if $c_t > c_{t-1}$ for some $t > 0$, then for all $\tau \geq 0$, there exists a $\tau' > \tau$ such that $c_{\tau'} > c_\tau$ (i.e., the sequence never attains its limit).

Proof. (i) Suppose to the contrary that $c_t \geq c^*$ for some $t > 0$, with $c_{t-1} < c^*$. From the definition of c^*, and A1, we must have

$$\pi(c_t, c_t) \geq \pi(c_{t+1}, c_{t+1}), \ \tau \geq 1$$

Consequently,

$$\pi(c_t, c_t) + \delta\pi(c_{t+1}, c_{t+1}) + \ldots < \frac{\pi(c_t, c_t)}{1-\delta}.$$

Then, by (2.2), we have

$$\frac{\pi(c_{t-1}, c_t)}{1-\delta} < \frac{\pi(c_t, c_t)}{1-\delta}.$$

But as $c_{t-1} < c_t$, and π decreasing in its first argument, $\pi(c_{t-1}, c_t) > \pi(c_t, c_t)$, a contradiction.

(ii) If this is not the case, then $c_t > c_{t-1}$ for some $t > 0$, and there exists a $T \geq t$ with $c_\tau = \mathbf{e}$ for all $\tau \geq T$ and $c_\tau < \mathbf{e}$ for $\tau < T$. Player 1, by deviating at T, would receive

$$\frac{\pi(c_{T-1}, \mathbf{e})}{1-\delta} > \frac{\pi(\mathbf{e}, \mathbf{e})}{1-\delta},$$

where the inequality follows from π decreasing in its first argument. Thus the deviation is profitable, contradicting the equilibrium assumption. ■

Say that the path $\{\hat{c}_t\}_{t=1}^{\infty} \in C_{SE}$ is efficient[12] (i.e., among symmetric equilibrium paths) if there does not exist another sequence $\{c_t'\}_{t=1}^{\infty} \in C_{SE}$ such that

$$\sum_{t=1}^{\infty} \delta^{t-1}\pi(c_t', c_t') > \sum_{t=1}^{\infty} \delta^{t-1}\pi(\hat{c}_t, \hat{c}_t):$$

We now have:

Lemma 2.2. An efficient sequence $\{\hat{c}_t\}_{t=1}^{\infty}$ exists, and this sequence satisfies inequalities (2.2) with equality, i.e., for all $t \geq 1$;

$$\frac{\pi(\hat{c}_{t-1}, \hat{c}_t)}{1-\delta} = \pi(\hat{c}_t, \hat{c}_t) + \delta\pi(\hat{c}_{t+1}, \hat{c}_{t+1}) + \cdots \quad : \tag{2.3}$$

Proof. As all the inequalities in (2.2) are weak, existence follows from standard arguments. We refer to (2.2) holding at t the t-constraint. To show that all the $t-$constraints hold with equality, suppose to the contrary that for some t,

$$\frac{\pi(\hat{c}_{t-1}, \hat{c}_t)}{1-\delta} < \pi(\hat{c}_t, \hat{c}_t) + \delta\pi(\hat{c}_{t+1}, \hat{c}_{t+1}) + \cdots \quad :$$

Then, by continuity, we can increase $\hat{c}_t$, holding $\hat{c}_{t+1}, \hat{c}_{t+2}, \ldots,$ fixed; without violating the $t-$constraint. Moreover, the $t+1$-constraint is relaxed by an increase in $\hat{c}_t$, holding $\hat{c}_{t+1}, \hat{c}_{t+2}, \ldots$ fixed, as π is decreasing in its first argument. Finally, we can hold $\hat{c}_{t-1}, \hat{c}_{t-2}, \ldots, \hat{c}_1$ fixed since the only effect of an increase in $\hat{c}_t$ is to relax the τ-constraints, for $\tau < t$: $\square$

It now follows quite straightforwardly from Lemmas 1 and 2 that the efficient path must satisfy a second-order difference equation. First note that the efficient path must solve the sequence of equations (2.3). Let the sequence $\{c_t(c_1; \delta)\}_{t=1}^{\infty}$ solve the second-order difference equation

$$\pi(c_t, c_{t+1}) = \frac{1}{\delta}\left[\pi(c_{t-1}, c_t) - \pi(c_t, c_t)\right] + \pi(c_t, c_t); \; t > 1 \tag{2.4}$$

with initial conditions $c_0 = 0; c_1 \geq 0$: It is easily checked[13] that any solution to this difference equation is non-decreasing, so the sequence $\{c_t(c_1; \delta)\}_{t=1}^{\infty}$ has a limit $c_\infty(c_1; \delta)$ which is finite or $+\infty$. Then we have:

[12] We use the term 'first-best' to refer to unconstrained efficient outcomes.
[13] This fact follows directly from the proof of Lemma 2.3 below.

Lemma 2.3. Any sequence $\{c_t\}_{t=1}^{\infty}$ solves (2.3) if and only if it solves (2.4) with initial conditions $c_0 = 0; c_1 \geq 0$, and $c_\infty := \lim_{t\to\infty} c_t < +\infty$.

Proof. Necessity. From the irreversibility constraint, $\{c_t\}_{t=1}^{\infty}$ is a non-decreasing sequence, so it converges to some finite limit c_∞ or diverges to $+\infty$. Since (2.3) implies (2.2), $\{c_t\}_{t=1}^{\infty}$ is an equilibrium sequence and by Lemma 2.1, $\{c_t\}_{t=1}^{\infty}$ must converge to $c_\infty \leq c^*$. Now, (2.3) can be written

$$\frac{\pi(c_{t-1}; c_t)}{1-\delta} = S_t;$$

where we again write $S_t := \pi(c_t; c_t) + \delta\pi(c_{t+1}; c_{t+1}) + \cdots$. Advancing by one period, we get

$$\frac{\pi(c_t; c_{t+1})}{1-\delta} = S_{t+1}.$$

Also,

$$S_t = \pi(c_t; c_t) + \delta S_{t+1}.$$

So,

$$\frac{\pi(c_{t-1}; c_t)}{1-\delta} = \pi(c_t; c_t) + \frac{\delta\pi(c_t; c_{t+1})}{1-\delta}. \tag{2.5}$$

Rearrangement of (2.5) gives (2.4).

Sufficiency. As just shown above, (2.4) is equivalent to (2.5). By successive substitution using (2.5), we get

$$\frac{\pi(c_{t-1}; c_t)}{1-\delta} = \pi(c_t; c_t) + \cdots + \delta^{n-1}\pi(c_{t+n-1}; c_{t+n-1}) + \frac{\delta^n\pi(c_{t+n-1}; c_{t+n})}{1-\delta} \tag{2.6}$$

Now, as $\{c_t\}_{t=1}^{\infty}$ converges by assumption, we must have

$$\lim_{n\to\infty} \frac{\delta^n\pi(c_{t+n-1}; c_{t+n})}{1-\delta} = 0$$

So, taking the limit in (2.6), we recover (2.3). ∎

We now know that the efficient path solves the difference equation (2.4) with initial conditions $c_0 = 0$ and c_1 yet to be determined. The following lemma allows us to determine c_1 and hence the efficient path itself. This lemma shows that the efficient path is the upper envelope of all equilibrium paths (and hence it is unique). It then follows from Lemma 2.5 (ii) below that c_1 is simply the highest value consistent with convergence of the solution to the difference equation.

Lemma 2.4. The efficient path $\{\hat{c}_t\}_{t=1}^{\infty}$ is the upper envelope of all equilibrium sequences, i.e., there does not exist a $\{c_t^0\}_{t=1}^{\infty} \in C_{SE}$ with $c_t^0 > \hat{c}_t$, for some t:

Proof. See Appendix. $\blacksquare$

As before, let the sequence $\{c_t(c_1;\delta)\}_{t=1}^{\infty}$ solve the difference equation (2.4), and consider the set of initial conditions c_1 such that $\{c_t(c_1;\delta)\}_{t=1}^{\infty}$ converges to a finite limit, i.e.,

$$C_1(\delta) = \{c_1 | c_\infty(c_1;\delta) < +\infty\}:$$

Then we have our final result of this section:

Lemma 2.5. (i) If, for any $c_1 \geq 0$; $\{c_t(c_1;\delta)\}_{t=1}^{\infty}$ is a convergent sequence, then it is also an equilibrium sequence; (ii) The efficient path satisfies $\{\hat{c}_t\}_{t=1}^{\infty} = \{c_t(\hat{c}_1;\delta)\}_{t=1}^{\infty}$, where $\hat{c}_1 = \max C_1(\delta)$, and $c_t(\hat{c}_1;\delta) \geq c_t(c_1^0;\delta)$; all $c_1^0 \in C_1(\delta)$; all $t \geq 0$.

Proof. (i) In view of the fact that (2.3) guarantees the sequence is equilibrium, sufficiency implies (i) of Lemma 2.3.

(ii) From Lemma 2.2 and Lemma 2.3, the efficient path exists, solves (2.4) with initial conditions $c_0 = 0; c_1 \geq 0$ and must also converge. Consequently, $\{\hat{c}_t\}_{t=1}^{\infty} = \{c_t(\hat{c}_1;\delta)\}_{t=1}^{\infty}$ for some $\hat{c}_1 \in C_1(\delta)$: Now suppose that there exists another $c_1^0 \in C_1(\delta)$ with $c_t(c_1^0;\delta) > c_t(\hat{c}_1;\delta)$ at some $t > 0$. In this case, $\{c_t(c_1^0;\delta)\}_{t=1}^{\infty}$ is an equilibrium (by part (i)) with $c_t(c_1^0;\delta) > c_t(\hat{c}_1;\delta)$ at some t, which contradicts Lemma 2.4. In particular this implies that $c_1^0 \in C_1(\delta)$ and $c_1^0 > \hat{c}_1$ is not possible. $\blacksquare$

3. Main Results

We know that the efficient path is the equilibrium path that is not crossed by any other, and which is the highest (at each point) of all convergent sequences that satisfy the difference equation (2.4). We now proceed to get an exact characterization of the limit $\hat{c}_\infty$. To do this, we consider two particular cases.

The Differentiable Case.

π is twice continuously differentiable, with $\pi_1 < 0; \pi_2 > 0; \ \pi_{11}; \pi_{22} < 0; \ \pi_{12} \leq 0.$

The Linear Kinked Case.

$$\pi = \begin{cases} \pi_1 c_1 + \pi_2 c_2 & \text{if } c_1 + c_2 \leq 2c^* \\ 2\pi_2 c^* - (\pi_2 - \pi_1)c_1 & \text{if } c_1 + c_2 > 2c^* \end{cases}$$

where $\pi_1 < 0; \pi_2 > 0$ are constants[14] with $\pi_1 + \pi_2 > 0$.

Note that both these cases satisfy our assumption A1 above on the shape of $w(c)$: In the differentiable case, $w(c)$ is strictly concave, as $w'' = \pi_{11} + \pi_{22} + 2\pi_{12} < 0$, with a unique maximum at c^*. In the linear kinked case, $w(c)$ is linear and increasing in c until c reaches the efficient level c^*, and after that, higher cooperation yields negative benefit.

Consider the differentiable case first. Define the function

$$\gamma(c) := \frac{-\pi_1(c; c)}{\pi_2(c; c)} > 0.$$

Note from the assumed properties of π; we have

$$\gamma'(c) = \frac{-1}{\pi_2}\left[\pi_{11} + \pi_{12} + \gamma(\pi_{22} + \pi_{12})\right] > 0;$$

and also that c^* solves $\gamma(c^*) = 1$. Consequently, provided $\gamma(0) \leq \delta$; there is a unique solution $\widehat{c}(\delta)$ to the equation

$$\gamma(\widehat{c}) = \delta; \tag{3.1}$$

and moreover, $\widehat{c}(.)$ is strictly increasing in δ. If $\gamma(0) > \delta$; we set $\widehat{c}(\delta) = 0$. Clearly $\widehat{c}(\delta) < c^*$, $\delta < 1$, with $\lim_{\delta \to 1} \widehat{c}(\delta) = c^*$. We can now state our first main result:

Proposition 3.1. Assume the differentiable case. Then the limit of the efficient symmetric path, $\widehat{c}_\infty$; is equal to $\widehat{c}(\delta)$. Consequently, for all $\delta < 1$, the efficient path is uniformly bounded below the first-best efficient level of cooperation; i.e., $\widehat{c}_t < \widehat{c}(\delta) < c^*$ for all t.

[14] An interpretation is that payoffs depend positively on $(c_1 + c_2)$ up to $2c^*$ with a coefficient of π_2, but there is a marginal utility cost of $(\pi_2 - \pi_1)$ to increasing one's own c_i. For $c_1 + c_2 > 2c^*$, there is no more benefit from joint contributions, only the cost remains, so that joint payoffs are declining in $(c_1 + c_2)$. For $c_1 + c_2 > 2c^*$; all that is needed for the results is that joint payoffs are nonincreasing in $(c_1 + c_2)$ and also own payoffs are declining in own c_i.

Proof. (a) By the Mean Value Theorem,

$$\begin{aligned}
\pi(c_{t-1}, c_t) - \pi(c_{t-1}, c_{t-1}) &= \pi_2(c_{t-1}, \theta_t)\Delta c_t, \ \theta_t \in [c_{t-1}, c_t] \\
\pi(c_{t-2}, c_{t-1}) - \pi(c_{t-1}, c_{t-1}) &= -\pi_1(\theta_{t-1}, c_{t-1})\Delta c_{t-1}, \ \theta_{t-1} \in [c_{t-2}, c_{t-1}];
\end{aligned}$$

where $\Delta c_t := c_t - c_{t-1}$. So, substituting in (2.4) and rearranging , we get

$$\begin{aligned}
\Delta c_t &= -\frac{\pi_1(\theta_{t-1}, c_{t-1})}{\delta\pi_2(c_{t-1}, \theta_t)}\Delta c_{t-1} \\
&\equiv a(c_{t-1}, c_t)\Delta c_{t-1}.
\end{aligned} \tag{3.2}$$

(b) Suppose that $\bar{c}_\infty > \bar{c}(\delta)$. There must, by $\pi(\cdot,\cdot)$ being twice continuously differentiable and $a(\bar{c}_\infty, \bar{c}_\infty) = \gamma(\bar{c}_\infty)/\delta > 1$, exist a T such that for $t > T$, $a(c_{t-1}, c_t) > 1$. But then from (3.2), for all $t > T$, $\Delta c_t > \Delta c_{t-1}$ whenever $\Delta c_{t-1} > 0$ and by Lemma 2.1 (ii), $\Delta c_{t-1} > 0$ for some $t-1 > T$; so c_t cannot converge, contrary to hypothesis. We conclude $\bar{c}_\infty \leq \bar{c}(\delta)$

(c) Suppose that $0 < \bar{c}_\infty < \bar{c}(\delta)$. We show that this is impossible. Find a neighborhood around $\bar{c}_\infty$, $(\bar{c}_\infty - \varepsilon, \bar{c}_\infty + \varepsilon)$, such that $a(c, c') < k < 1$ for all $c, c' \in (\bar{c}_\infty - \varepsilon, \bar{c}_\infty + \varepsilon)$. Define $\psi := (1-k)\varepsilon$, and consider T such that $c_T(\bar{c}_1, \delta) > \bar{c}_\infty - \psi$ (this must exist by definition of $\bar{c}_\infty$). Now, since $c_T(\bar{c}_1, \delta) < c_{T+1}(\bar{c}_1, \delta) < \bar{c}_\infty$, by $c_T(c_1, \delta)$ being continuous in c_1, we can find $c_1' > \bar{c}_1$ such that $c_T(c_1')$ and $c_{T+1}(c_1') \in (\bar{c}_\infty - \psi, \bar{c}_\infty)$, and moreover, since $0 < c_{T+1}(\bar{c}_1, \delta) - c_T(\bar{c}_1, \delta) < \psi$, c_1' can also be chosen so that $0 < c_{T+1}(c_1', \delta) - c_T(c_1', \delta) < \psi$. Hence for all $t > T$, $\Delta c_t < k\Delta c_{t-1}$ by (3.2), and consequently $\{c_t(c_1', \delta)\}_{t=1}^\infty$ must converge to some $c_\infty(c_1', \delta) < \bar{c}_\infty + \frac{\psi}{1-k}$ ($= \bar{c}_\infty + \varepsilon$). Since $\{c_t(c_1', \delta)\}_{t=1}^\infty$ is a convergent path it is also an equilibrium path (Lemma 2.5(i)) and $c_1' > \bar{c}_1$, which contradicts the envelope property of the efficient equilibrium (Lemma 2.4). Finally, a minor modification to this argument establishes that $\bar{c}_\infty = 0$ is impossible whenever $\bar{c}(\delta) > 0$. ∎

Next, consider the linear kinked case. Here, we have the following striking result.

Proposition 3.2. Assume the linear kinked case. If there is sufficiently little discounting ($\delta > -\pi_1/\pi_2$), then the limit of the efficient symmetric sequence, $\bar{c}_\infty$, equals c^*, i.e., first-best efficient cooperation can be asymptotically obtained. Otherwise, no cooperation can ever be obtained, i.e., $\bar{c}_t = 0$, all t.

Proof. From Lemma 2.1, we can restrict attention to those paths with $c_t < c^*$, all t, as no other path can be an equilibrium one. Writing out (2.4) for this case, using the definition of π for the kinked linear case, we get:

$$\pi_1 c_t + \pi_2 c_{t+1} = \frac{1}{\delta}[\pi_1 c_{t-1} + \pi_2 c_t - \pi_1 c_t - \pi_2 c_t] + \pi_1 c_t + \pi_2 c_t,$$

which rearranges to

$$\Delta c_t = a \Delta c_{t-1}, \tag{3.3}$$

where $a := \left(-\frac{\pi_1}{\delta \pi_2}\right)$, $\Delta c_t := c_t - c_{t-1}$. Thus, $\Delta c_t = a^{t-1} \Delta c_1$ where $\Delta c_1 = c_1 - c_0 = c_1$, and c_1 can be chosen freely. So, we have

$$c_t = \sum_{\tau=1}^{t} \Delta c_\tau = (1 + a + \ldots a^{t-1}) c_1. \tag{3.4}$$

First suppose that $a \geq 1$. If $c_1 > 0$, then from (3.4), $c_t \to \infty$ as $t \to \infty$, contradicting the assumption that $c_t < c^*$, all t. So, we must have $c_1 = 0$, in which case $c_t = 0$, all t. Thus if $a \geq 1 \iff \delta \leq (-\pi_1/\pi_2)$, no cooperation is possible as claimed. Now suppose that $a < 1$. Then the series in (3.4) converges, so we get

$$c_\infty = \frac{1}{1-a} c_1 = \frac{1}{1 + \frac{\pi_1}{\delta \pi_2}} c_1.$$

So by appropriate choice of c_1, we can choose a path that converges to c^*, and this must be the efficient path by virtue of Lemma 2.4. $\square$

Note that in both cases, we have shown that as $\delta \to 1$, the limiting level of cooperation on the efficient equilibrium path, c_∞, tends to the first-best efficient level, c^*. It turns out that this fact implies that payoffs also converge to their efficient levels as $\delta \to 1$; i.e., there is no limiting inefficiency in this model.

Corollary 3.3. In either the differentiable or linear kinked cases, as $\delta \to 1$, the normalized discounted payoff from the efficient path, $\hat{\pi} = (1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} \pi(\hat{c}_t, \hat{c}_t)$, converges to the first-best payoff $\pi(c^*, c^*)$.

Proof. Consider, for some fixed δ, rewriting the equilibrium condition (2.2) as, for each t,

$$\pi(c_{t-1}, c_t) \leq (1-\delta) \sum_{\tau=t}^{\infty} \delta^{\tau-t} \pi(c_\tau, c_\tau). \tag{3.5}$$

Now, if $\{c_t\}_{t=1}^{\infty}$ is an equilibrium sequence at δ, then $\{c_t\}_{t=1}^{\infty}$ is also an equilibrium at any $\delta' > \delta$ since, as $\pi(c_t, c_t)$ is a non-decreasing sequence, the R.H.S. of (3.5) is non-decreasing in δ, and the L.H.S. is constant.

Now for the differentiable case, define $\hat{c}(\delta)$ as in (3.1), and in the linear kinked case, define

$$\hat{c}(\delta) = \begin{cases} c^* & \text{if } \delta > -\pi_1/\pi_2 \\ 0 & \text{otherwise} \end{cases}$$

So, for any $\varepsilon > 0$, find a $\bar{\delta}$ such that $\pi(\hat{c}(\bar{\delta}), \hat{c}(\bar{\delta})) > \pi(c^*, c^*) - \varepsilon$ (where in the differentiable case, we use the continuity of $\pi(\cdot,\cdot)$, and, as already remarked, $\lim_{\delta \to 1} \hat{c}(\delta) = c^*$). From Propositions 3.1 and 3.2, at $\bar{\delta}$, $\hat{c}_t \to \hat{c}(\bar{\delta})$, so holding $\{\hat{c}_t\}_{t=1}^{\infty}$ fixed, $\lim_{\delta \to 1}(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1}\pi(\hat{c}_t, \hat{c}_t) \to \pi(\hat{c}(\bar{\delta}), \hat{c}(\bar{\delta}))$; and hence there exists a $\delta' > \bar{\delta}$ such that for δ satisfying $\delta' < \delta < 1$, $(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1}\pi(\hat{c}_t, \hat{c}_t) > \pi(c^*, c^*) - \varepsilon$. Since $\{\hat{c}_t\}_{t=1}^{\infty}$ is an equilibrium sequence for such δ, the efficient path at such δ must also give a payoff greater than $\pi(c^*, c^*) - \varepsilon$. As ε is arbitrary, this completes the proof. $\square$

An alternative way of viewing this result is to note that if we shrink the period length, holding payoffs per unit of time constant, then inefficiency disappears as period length goes to zero.[15]

4. A Model with Adjustment Costs

The model studied above is very stylized. In many economic applications, irreversibility arises more naturally when there is a stock variable which benefits both players, and a flow or incremental variable which is costly to increase, and is nonnegative. This nonnegativity constraint implies that the value of the stock variable can never fall i.e. the stock variable is irreversible. Here, we present a model with these features, and show that it can be reformulated so that it is a special case of our base model.

Player i's payoff at time t is

$$u(c_{i,t}, c_{jt}) - \alpha(c_{i,t} - c_{i,t-1}), \tag{4.1}$$

[15] If π is discontinuous but otherwise satisfies our assumptions then asymptotic efficiency can fail. Consider an example in which player i benefits only from j's c_j, with an upwards jump in payoff at completion ($c_j = c^*$), and suffers continuous (increasing) costs from c_i. Lemma 2.1 still applies, so $c_{i,t} < c^*$, all t, and the payoff jump is never realised no matter how patient the players.

with u increasing in both arguments, and with $\alpha > 0$ being the cost of adjustment. Here, $c_{i,t}$ is to be interpreted as i's cumulative investment in, or the stock level of, the cooperative activity. We assume that the investment flow is nonnegative, which implies that the stock level of cooperation is irreversible, i.e., $c_{i,t} \geq c_{i,t-1}$, $i = 1, 2$.

We now proceed as follows. The present value payoff for i in this model is

$$\begin{aligned} \Pi_i &= u(c_{i,1}, c_{j,1}) - \alpha(c_{i,1} - c_{i,0}) + \delta[u(c_{i,2}, c_{j,2}) - \alpha(c_{i,2} - c_{i,1})] + \dots \\ &= \sum_{t=1}^{\infty} \delta^{t-1}[u(c_{i,t}, c_{jt}) - \alpha(1-\delta)c_{i,t}] + \alpha c_{i,0}. \end{aligned}$$

As initial levels of cooperation $c_{1,0}, c_{2,0}$ are fixed, we can think of this model as a special case of the model of the previous section (i.e. without adjustment costs) where per-period payoffs are

$$\pi(c, c') = u(c, c') - \alpha(1-\delta)c. \tag{4.2}$$

Of course, we require that π defined in (4.2) satisfies the conditions imposed in Section 2, and also satisfies the relevant conditions of either the differentiable or linear kinked case. If this is the case, then Propositions 3.1 and 3.2 apply directly.

We now study two important economic applications using this extension of our basic model. These are not the only topics that can be studied in this way, but they are chosen to illustrate the power and flexibility of our approach.

4.1. Dynamic Voluntary Contribution Games

There is now a small literature (Admati and Perry (1991), Fershtman and Nitzan (1991), Marx and Matthews (1998)), on dynamic games where players can simultaneously or sequentially make contributions towards the cost of a public project. The paper in this literature (Marx and Matthews (1998)) that is closest to our work is one where contributions are made simultaneously, and where the benefits from the project are proportional to the amount contributed (up to a maximum, at which point the project is completed). We will show that a special case of Marx and Matthews' model can be written as an adjustment cost game as above, and that Proposition 3.2 above can be applied to extend some of their results.

Marx and Matthews (1998) consider a model in which N individuals simultaneously make nonnegative private contributions, in each of a finite or infinite number of periods, to a public project. We assume that $N = 2$, and let $c_{i,t}$ be the cumulative contribution of a numeraire private good by i towards the public project. Individuals obtain a flow of utility $u = (1 - \delta)v(\cdot)$ from the aggregate cumulative contribution $c_{1,t} + c_{2,t}$, where $v(\cdot)$ is piecewise linear:

$$v(c_1, c_2) = \begin{cases} \lambda(c_1 + c_2) & \text{if } c_1 + c_2 < 2c^* = C^* \\ \lambda C^* + b & \text{if } c_1 + c_2 \geq C^* \end{cases}$$

where we follow as closely as possible the notation of Marx and Matthews. Thus agents get benefit λ from each unit of cumulative contribution, and an additional benefit $b \geq 0$ when the project is "completed", i.e., when the sum of cumulative contributions reaches C^*. Also, the cost to i of an increment $c_{i,t} - c_{i,t-1}$ in the cumulative contribution is simply $c_{i,t} - c_{i,t-1}$. We consider the case where $b = 0$ and the time horizon is infinite (the $b = 0$ case unravels otherwise). Also it is assumed that $0.5 < \lambda < 1$, so that it is socially efficient to complete the project (immediately, in fact), but not privately efficient to contribute anything.

Then, from (4.2), per period payoffs in the equivalent repeated game are

$$\begin{aligned} \pi(c_1, c_2) &= (1 - \delta)v(c_1, c_2) - (1 - \delta)c_1 \\ &= \begin{cases} (1 - \delta)[(\lambda - 1)c_1 + \lambda c_2] & \text{if } c_1 + c_2 < 2c^* = C^* \\ (1 - \delta)\lambda C^* - (1 - \delta)c_1 & \text{if } c_1 + c_2 \geq C^* \end{cases} \end{aligned}$$

So, $\pi_1 = (1 - \delta)(\lambda - 1) < 0$, $\pi_2 = (1 - \delta)\lambda > 0$. Thus, all the conditions of the linear kinked case are satisfied, and so Proposition 3.2 applies directly to this version of the Marx-Matthews model.

First, we can define the critical value of δ in Proposition 3.2 as

$$\hat{\delta} = \frac{-\pi_1}{\pi_2} = \frac{(1 - \lambda)}{\lambda}.$$

Two results then follow directly from our Proposition 3.2 and its proof:

1. If $\delta > \hat{\delta}$, there is a class of equilibria, indexed by the initial condition c_1; where each player's cumulative contribution c_t converges to c^*, or indeed to any value less than or equal to c^*. Along the equilibrium path, incremental contributions fall at rate $\frac{(1-\lambda)}{\delta\lambda}$.

The efficient symmetric equilibrium has initial contribution $c_1 = c^*(1 - \frac{(1-\lambda)}{\delta\lambda})$; and each player's cumulative contribution c_t converges to c^*:

2. If $\delta \leq \hat{\delta}$, then no contributions are made in any equilibrium.

Result 1 sharpens Proposition 3 and Corollary 3(ii) of Marx and Matthews, who show that for $\delta > \hat{\delta}$; there is an equilibrium with $c_t \to c^*$, and that for $\delta \approx 1$, this equilibrium is approximately efficient. In the special case of $n = 2$ and $b = 0$; we not only confirm their results, but also show that the equilibrium they construct is the efficient equilibrium for any $\delta > \hat{\delta}$: Also, Result 2 is a complete converse result to their Proposition 3.

4.2. Capacity Reduction in a Declining Industry

There is now a literature on the equilibrium evolution of capacity in an industry where demand is declining over time (See Ghemawat and Nalebuff (1990) and the references therein). For tractability, this literature assumes that product demand declines asymptotically to zero; a backward induction argument can then be used to establish the equilibrium pattern of capacity reduction by firms. Our framework allows us to deal with the more general case where demand does not decline to zero.

The model is a modification of that of Ghemawat and Nalebuff (1990). There is a duopoly where each firm $i = 1; 2$; has initial capacity at time zero of k_0. In any period, the output of firm i must be no greater than capacity, i.e., $x_{i;t} \leq k_{i;t}$. Demands and costs are as follows. At time t, each firm faces the linear inverse demand schedule $p_t = a_t - x_{1t} - x_{2t}$. There is no short-run cost of production, but there is a per-period cost of maintaining capacity $\psi > 0$, and a cost $\sigma > 0$ of scrapping capacity, with the flow cost of scrapping less than maintenance, i.e., $\sigma(1 - \delta) < \psi$. It is assumed that capacity, once withdrawn, cannot be reintroduced (for example, the capital stock may consist of specialized capital goods which are no longer manufactured).

Within a period, the production decision is delegated to myopic managers who engage in Cournot competition, so output conditional on capacity is

$$x_{i;t} = \min\{k_{i;t}; a_t/3\}; \qquad (4.3)$$

where $a_t/3$ is unconstrained Cournot output at time t: We assume that at the beginning of period 1, a_t falls permanently from a_0 to a_1, i.e., the size of the market declines once and for all.[16] We suppose that initial capital stocks have been set so as to force managers to produce at joint profit-maximizing outputs, taking into account the cost of capital, and adjustment costs, at the initial level of demand, i.e.,

$$k_0 = \frac{(a_0 - \psi + \sigma(1 - \delta))}{4}. \tag{4.4}$$

A story consistent with this is that in the (distant) past, this industry has already been hit by a negative demand shock, and has adjusted to the old long-run equilibrium.[17] Note that cutting capacity can act as a way of committing to a lower level of output than the Cournot solution. The question is, if demand falls, can the firms cut their capacities sufficiently so as to reach the joint profit maximising level?

It is convenient to assume that the decline in the market is not too large, i.e.,

$$\frac{3a_0}{4} \le a_1. \tag{4.5}$$

In this case, managers will always be constrained by capacity.[18] So, if (4.5) holds, profit in period t can be written

$$\begin{aligned} \pi_{i,t} &= a_1 k_{i,t} - k_{i,t} - k_{i,t}k_{j,t} - \psi k_{i,t} - \sigma(k_{i,t-1} - k_{i,t}) \\ &\equiv \bar{\pi}(k_{i,t}, k_{j,t}) - \sigma(k_{i,t-1} - k_{i,t}). \end{aligned}$$

So, the fully efficient capital stock at the new level of demand, k^*, maximizes $\sum_{t=1}^{\infty} \delta^{t-1}(\pi_{1,t} + \pi_{2,t})$, i.e.,

$$k^* = \frac{a_1 - \psi + \sigma(1 - \delta)}{4},$$

and adjustment should be immediate. Note that $k^* < k_0$.

Now define the level of cooperation of firm i to be the amount of capital scrapped, $c_{i,t} := k_0 - k_{i,t}$, so $c_{i,0} = 0$, $c^* = k_0 - k^*$. So, from (4.2) we can write profit as a function of cooperation levels:

$$\pi(c_{i,t}, c_{j,t}) := \bar{\pi}(k_0 - c_{i,t}, k_0 - c_{j,t}) - \sigma(1 - \delta)c_{i,t}. \tag{4.6}$$

[16] This is in contrast to Ghemawat and Nalebuff who make the assumption of a constantly declining market, an assumption which implies a backwards unravelling result and a unique equilibrium. By contrast here there will be many equilibria.

[17] Although, as we shall see, this statement is only approximately correct if δ is near 1.

[18] To see this, note that (4.5) implies $k_{it} \le k_0 = \frac{(a_0 - \psi + \sigma(1-\delta))}{4} \le \frac{a_1}{3}$ as $\psi > \sigma(1 - \delta)$ by assumption.

As $\pi(c_{i,t}, c_{j,t})$ is non-linear, the relevant case is the differentiable case. To apply Proposition 3.1, we need to verify the assumptions of the differentiable case. By direct calculation, we have:

$$
\begin{aligned}
\pi_1 &= -a_1 + \psi + 2k_{i,t} + k_{j,t} - \sigma(1-\delta); \\
\pi_2 &= k_{i,t}; \\
\pi_{11} &= -2;\ \pi_{22} = 0;\ \pi_{12} = -1.
\end{aligned}
$$

So, all the differentiable case conditions are satisfied if $\pi_1 < 0$; which in turn is satisfied if (4.5) holds and capacity (net of scrapping) costs are small[19].

Our results for the differentiable case then apply directly. In particular, on the efficient symmetric path $c_{i,t}$ rises asymptotically to **b**, where **b** is defined in (3.1) above. We can express this in terms of the capital stock: k_{it} declines asymptotically to $\hat{k}$, where $\hat{k}$ solves

$$\frac{\partial \pi_1(\hat{k}, \hat{k})}{\partial \pi_2(\hat{k}, \hat{k})} = \delta.$$

Or, using (4.6), we get:

$$\frac{a_1 - \psi - 2\hat{k} - \hat{k} + \sigma(1-\delta)}{\hat{k}} = \delta.$$

Solving, we get

$$\hat{k} = \frac{a_1 - \psi + \sigma(1-\delta)}{3+\delta} > k^*.$$

So, for $\delta < 1$, the duopolists cannot credibly reduce capacity to the new joint profit-maximizing level k^*, even asymptotically. All they can manage is to force down capital stocks to $\hat{k}$, so there will be excess capacity and output in the industry (relative to joint profit maximization), even in the long-run. As $\delta \to 1$, the amount of excess capacity goes to zero.

5. Reversible Cooperation

So far, we have assumed that cooperation is completely irreversible. This is clearly a strong assumption. In this section, we examine to what extent our results are robust to

[19] To see this note that $\pi_1 < 0$ if $k_{it} < \frac{(a_1 - \psi + \sigma(1-\delta))}{3}$. But if capacity (net of scrapping) costs are small ($\psi \simeq \sigma(1-\delta)$), $k_{it} \le k_0 = \frac{(a_0 - \psi + \sigma(1-\delta))}{4} \simeq \frac{a_0}{4} < \frac{a_1}{3} \simeq \frac{a_1 - \psi + \sigma(1-\delta)}{3}$ as required.

a relaxation of this assumption. Suppose that we modify the irreversibility constraint to

$$c_{i,t} \geq \rho c_{i,t-1}, \quad 0 \leq \rho \leq 1;$$

where the degree of irreversibility is parameterized by ρ; complete irreversibility is $\rho = 1$, and a standard repeated game is $\rho = 0$. The first—and important—point is that the effect of lowering ρ from 1 on the efficient symmetric path is not clear without further analysis, because of two effects that work in opposite directions.

The first effect of a smaller ρ is to make deviation more profitable; the deviator at t can lower his cooperation level at t to $\rho c_{t-1} < c_{t-1}$, rather than keep it at c_{t-1}. The second effect is to make punishment more severe; the worst possible perfect equilibrium punishment of the deviator is for the other player to reduce his cooperation as fast as possible over time, rather than just not increase it. A priori, it is not clear which effect will dominate. Nevertheless, we are able to show that for a small amount of reversibility the second effect dominates, and in the linear case it dominates for any degree of reversibility.

Specifically, we show that lowering ρ slightly from $\rho = 1$ relaxes the incentive constraints; that is, any path that is an equilibrium when $\rho = 1$ is also an equilibrium path when ρ is slightly lower than one, and moreover because the incentive constraints become slack, an improved path can be found, so that payoffs increase.

Consider a deviation by i from some symmetric path $\{c_t\}_{t=1}^{\infty}$ at t. The worst subgame-perfect punishment that j can impose on i is to reduce cooperation by the maximum amount in every period following t, i.e., to set $c_{j,t+1} = \rho c_t$, $c_{j,t+2} = \rho^2 c_t$, etc. Consequently, the most profitable deviation i can make is to lower his cooperation by the maximum feasible amount at t, i.e., set $c_{i,t} = \rho c_{t-1}$. So, the maximal payoff to deviation at t is

$$\Delta(\rho, c_{t-1}, c_t) := \pi(\rho c_{t-1}, c_t) + \delta\pi(\rho^2 c_{t-1}, \rho c_t) + \delta^2\pi(\rho^3 c_{t-1}, \rho^2 c_t) + \dots$$

Then, $\{c_t\}_{t=1}^{\infty}$ is an equilibrium path if and only if it satisfies for all $t \geq 1$:

$$\Delta(\rho, c_{t-1}, c_t) \leq \pi(c_t, c_t) + \delta\pi(c_{t+1}, c_{t+1}) + \delta^2\pi(c_{t+2}, \rho c_{t+2}) + \dots \qquad (5.1)$$

An efficient (symmetric) equilibrium path is defined now as the path that maximizes the utility of either agent subject to the sequence of constraints (5.1).

In order to characterize efficient payoffs, the relevant results extending Lemmas 1-4 are collected below:

Lemma 5.1. *With reversibility, there exists an efficient symmetric equilibrium sequence $\{\mathbf{b}_t\}_{t=1}^{\infty}$ such that (i) $\mathbf{b}_{t-1} \leq \mathbf{b}_t \leq c^*$ for all $t \geq 1$; (ii) if $\mathbf{b}_t < c^*$, then (5.1) holds with equality, (iii) $\{\mathbf{b}_t\}$ is the upper envelope of all equilibrium sequences which never exceed c^*.*

Proof. See Appendix. ∎

If c^* is the unique maximizer of $\pi(c, c)$, then the sequence $\{\mathbf{b}_t\}_{t=1}^{\infty}$ characterized in the lemma is the unique efficient symmetric equilibrium outcome path; otherwise there may be multiple efficient paths differing only in the interchange of efficient levels of c, but they do not differ before such levels are attained. In what follows, the 'efficient equilibrium path' is understood to refer to the one which does not exceed c^*.

Using Lemma 5.1, we now turn to discuss the impact of a small amount of irreversibility, and we begin with the differentiable case. Let $\{\mathbf{b}_t(\rho)\}_{t=1}^{\infty}$ be the efficient equilibrium path in the $\rho-$reversible game, let $\mathbf{b}_{\infty}(\rho)$ be its limit (which exists by Lemma 5.1), and let

$$\hat{\Pi}(\rho) := (1-\delta)\sum_{t=1}^{\infty} \delta^{t-1}\pi(\mathbf{b}_t(\rho), \mathbf{b}_t(\rho))$$

be the payoff from this efficient path, all for some fixed discount factor $\delta < 1$. Then we have the following:

Proposition 5.2. *In the differentiable case, provided $\mathbf{b}_{\infty}(1) > 0$, there exists $\bar{\rho}, 1 > \bar{\rho} > 0$, such that if $1 > \rho > \bar{\rho}$, then (i) if $\{\mathbf{b}_t(\rho)\}_{t=1}^{\infty}$ is the efficient equilibrium path in the irreversible case, it is also an equilibrium path in the $\rho-$reversible case; (ii) $\mathbf{b}_{\infty}(\rho) > \mathbf{b}_{\infty}(1)$ $(\equiv \mathbf{b})$; (iii) $\hat{\Pi}(\rho) > \hat{\Pi}(1)$.*

Proof. See Appendix. ∎

The reasoning behind this result is that a small amount of irreversibility relaxes the incentive constraints in every time period, allowing every components of the efficient path to be raised slightly as ρ decreases slightly from 1. This in turn implies that the limit value of the efficient path is higher, as well as the present discounted payoff from the efficient path.

We now turn to the linear kinked case. We shall first characterize the sequence $\{c_t\}_{t=1}^{\infty}$ described in Lemma 5.1. From (ii) of the lemma, if $c_t < c^*$ and $c_{t+1} < c^*$ then (5.1) holds with equality at both dates, and substituting out the continuation equilibrium payoffs after $t + 1$ yields

$$\sum_{j=1}^{\infty} \delta^{j-1}(\pi_1\rho^j c_{t-1} + \pi_2\rho^{j-1}c_t) = \pi_1 c_t + \pi_2 c_t + \delta\left(\sum_{j=1}^{\infty} \delta^{j-1}(\pi_1\rho^j c_t + \pi_2\rho^{j-1}c_{t+1})\right)$$

or

$$\frac{\pi_1\rho c_{t-1} + \pi_2 c_t}{1 - \rho\delta} = \pi_1 c_t + \pi_2 c_t + \frac{\pi_1\rho c_t + \pi_2 c_{t+1}}{1 - \rho\delta};$$

which can be simplified to

$$c_{t+1} - \rho c_t = -\frac{\pi_1}{\delta\pi_2}(c_t - \rho c_{t-1}).$$

Given that $c_1 - \rho c_0 = c_1$; this can be solved for

$$c_t = (\rho^{t-1} + \rho^{t-2}a + \rho^{t-3}a^2 \ldots + \rho a^{t-2} + a^{t-1})c_1; \tag{5.2}$$

where $a = -\frac{\pi_1}{\delta\pi_2}$ as before; and note that for $\rho = 1$ (irreversibility), (5.2) reduces to (3.4). (If $\rho \neq a$ then the solution can be written $c_t = \frac{(\rho^t - a^t)}{(\rho - a)}c_1$.)

We can now prove:

Proposition 5.3. In the linear kinked case, (i) if $a(= -\frac{\pi_1}{\delta\pi_2}) < 1$ (so a non-trivial equilibrium exists with irreversibility) then payoffs in efficient symmetric equilibrium are a strictly decreasing function of ρ whenever they are below the first-best level (which they are at $\rho = 1$). Moreover if $\rho < 1$ the project is completed in finite time (i.e., $c_t = c^*$ for some $t < \infty$): (ii) If $a > 1$; then $c_t = 0$ for all t; for all $\rho \in (0, 1]$ in any symmetric equilibrium. (iii) If $a = 1$; then the project is completed asymptotically for $\rho \in (0, 1)$.

Proof. See Appendix. ∎

Recall that if $\rho = 1$; no non-trivial equilibrium exists if $a \geq 1$; while if $\rho = 0$ (repeated game) it can be checked that the first best is attainable (immediately) if $a \leq 1$; otherwise there is no non-trivial equilibrium. The path used in the proof of part (i), which satisfies (5.2) up to its maximum value, is not the efficient path unless this maximum occurs at $t = 1$; since each incentive constraint up to t^* is slack, violating Lemma 5.1(ii). So the efficient path also satisfies (5.2) so long as $c_t < c^*$; but c_1 is higher than in the construction of the proof (otherwise Lemma 5.1(iii) is violated).

6. Asymmetric Cooperation

So far, we have only considered symmetric paths, i.e., where $c_{1,t} = c_{2,t} = c_t$. A natural question is whether the agents could achieve higher (expected) equilibrium payoffs by playing asymmetrically. A further related question concerns the characteristics of efficient equilibria in a model where agents are constrained to move sequentially; as we shall see, this is a closely related issue and will be considered below.

We shall consider these questions for the linear kinked case only. Let $\{c_{1,t}, c_{2,t}\}_{t=1}^{\infty}$ be an arbitrary (possibly asymmetric) path. Then, by a similar argument to that given in Section 2, such a path is an equilibrium path if and only if for $i, j = 1, 2;\ i \neq j;$ $t = 1, 2, \ldots;$

$$\frac{\pi_1 c_{i,t-1} + \pi_2 c_{j,t}}{1 - \delta} \le \pi_1 c_{i,t} + \pi_2 c_{j,t} + \delta(\pi_1 c_{i,t+1} + \pi_2 c_{j,t+1}) + \ldots \qquad . \tag{6.1}$$

Let C_E be the set of equilibrium paths (i.e. sequences that satisfy (2.1) and (6.1)). Also, let $\Pi_i(\{c_{1,t}, c_{2,t}\}_{t=1}^{\infty})$ be the normalized (multiplied through by $(1 - \delta)$) present discounted values of payoff to i associated with a path, and let Π_E be the image of C_E in the space of normalized present discounted values of payoffs, i.e.,

$$\Pi_E = \{(\Pi_1, \Pi_2) \mid \Pi_i = \Pi_i(\{c_{1,t}, c_{2,t}\}_{t=1}^{\infty});\ \{c_{1,t}, c_{2,t}\}_{t=1}^{\infty} \in C_E,\ i = 1, 2\}$$

Our focus in on the shape of the efficient frontier of Π_E. As far as symmetric equilibria go, we know from Proposition 3.2 if $\delta \le \hat{\delta} = -\pi_1/\pi_2$, no cooperation is possible, whereas if $\delta > \hat{\delta}$, completion equilibria exist. From the symmetry assumption on payoffs, Π_E is symmetric about the 45^o line. One issue concerns the possibility that Π_E may be a non-convex set, in which case it may be optimal for the players to randomize between two pure-strategy equilibria rather than play the efficient symmetric equilibrium. The following result, which characterizes Π_E when $\delta > \hat{\delta}$; establishes that this is not the case, and moreover shows that the efficient frontier of Π_E is linear with slope -1 near the 45^{o}line, so in terms of joint payoffs, a degree of asymmetry does not matter. This part of the frontier consists of payoffs from sequences which satisfy the incentive constraints with equality (this is no longer true for efficient paths with sufficiently asymmetric payoffs).

Proposition 6.1. Assume that $\delta > \hat{\delta} = -\pi_1/\pi_2$. Then, Π_E is convex. Moreover, the efficient frontier of Π_E has the following form. There exist points $A = (\Pi'; \Pi'')$, $B = (\Pi''; \Pi')$, on the efficient frontier of Π_E with $\Pi' > \Pi'' > 0$ such that between A and B, Π_1 and Π_2 sum to a constant Σ (i.e., the frontier of Π_E is linear between A and B with slope -1). For any point on the frontier below A or above B, the sum of utilities is strictly less than Σ.

Proof. See Appendix. □

The Proposition is illustrated in Figure 1 below,

Figure 1 in here

which shows the general shape of the frontier (although we have no results about the shape of the frontier to the left of B or below A, except that it must be described by a concave function). We can also say something about how the frontier shifts as δ changes:

Proposition 6.2. The segment of the efficient frontier between A and B is increasing in δ in the sense that both Π'/Π'' and Σ are increasing in δ, and converges to the first-best frontier as $\delta \to 1$ (i.e., $\Pi''/\Pi' \to 0$ and $\Sigma \to 2(\pi_1 + \pi_2)c^*$). As $\delta \to \hat{\delta} = -\pi_1/\pi_2$ from above, $A \to B$ and $\Sigma \to 0$.

Proof. See Appendix. □

Proposition 4 is illustrated in Figure 2 below, where the solid line represents the frontier at a lower δ and the dotted line the frontier at a higher value of δ.

Figure 2 in here

Note that as $\delta \to 1$, the efficient frontier becomes linear everywhere with slope equal to minus one -1, i.e., it converges to the first-best efficient frontier. So, Proposition 6.2 generalizes Corollary 3.3 to the case of asymmetric paths, at least in the linear kinked case.

7. Sequential Moves

So far, we have assumed that players can move simultaneously. However, it may be that players can only move sequentially, e.g., Admati-Perry (1991), Gale (1997). In certain public good contribution games, the assumption made can affect the conclusions substantially. In the Admati-Perry model, where players move sequentially, a no contribution result holds when no player individually would want to complete the project, even though it might be jointly optimal to do so, but this result may disappear if the players can move simultaneously (see Marx and Matthews (1997) for a full discussion of this issue). By contrast, we shall find that in our model, equilibria in the two cases are closely related; indeed, the efficient symmetric equilibrium can "approximately" be implemented in the sequential move game.

Suppose w.l.o.g. that player 1 can move at even periods and player 2 at odd periods. Then, this move structure imposes the constraint that

$$\begin{aligned} c_{1;t} &= c_{1;t-1},\ t = 1;3;5:::: \\ c_{2;t} &= c_{2;t-1},\ t = 2;4;6:::: \end{aligned} \tag{7.1}$$

Let the set of all paths that satisfy (7.1) be C^{seq}: To be an equilibrium in the sequential game, any path $\{c_{1;t}; c_{2;t}\}$ must satisfy the following incentive constraints. When player 1 moves at $t = 2;4;:::;$ he prefers to raise his level of cooperation from c_{t-2} to c_t only if

$$\frac{\pi(c_{1;t-2}; c_{2;t-1})}{1-\delta} \leq \pi(c_{1;t}; c_{2;t-1}) + \delta\pi(c_{1;t}; c_{2;t+1}) + :::;\ t = 2;4;6::: \tag{7.2}$$

Similarly, when player 2 moves at $t = 3;5:::;$ he prefers to raise his level of cooperation from $c_{2;t-2}$ to $c_{2;t}$ only if

$$\frac{\pi(c_{2;t-2}; c_{1;t-1})}{1-\delta} \leq \pi(c_{2;t}; c_{1;t-1}) + \delta\pi(c_{2;t}; c_{1;t+1}) + :::;\ t = 3;5;7::: \tag{7.3}$$

When player 2 moves at period 1, (7.3) is modified by the fact that 2 can revert to $c_0 = 0$, rather than c_{-1}, but otherwise the incentive constraint is the same, i.e.,

$$w\frac{\pi(0,0)}{1-\delta} \le \pi(c_{2,1},0) + \delta\pi(c_{2,1},c_{1,2}) + \dots . \tag{7.4}$$

Let the set of paths in C^{seq} that satisfy (7.2),(7.3) and (7.4) be $C^{seq}_E \subset C^{seq}$.

However, note that a path is in C^{seq}_E if and only if it is an (asymmetric) equilibrium path satisfying (7.1) in the simultaneous move game studied above. This is because in the simultaneous move game, the incentive constraints in the periods where agents do not have to move are automatically satisfied, as no agent likes to choose a higher $c_{i,t}$ than necessary (from π decreasing in its first argument). So, C^{seq}_E is simply that subset of C_E also in C^{seq}, i.e.,

$$C^{seq}_E = C_E \cap C^{seq}.$$

So, the set of feasible present-value payoffs Π^{seq}_E is the image of C^{seq}_E in $\Re^2$ under the payoff function , and consequently

$$\Pi^{seq}_E \subseteq \Pi_E.$$

To say more than this, we shall go to the linear kinked case, in which case we have the following. Define $A := (\Pi', \Pi'')$ as in Proposition 6.1 above, and let $\hat{\Pi}$ be the present value payoff from the efficient symmetric path in the simultaneous move game, so that $S := (\hat{\Pi}, \hat{\Pi})$ is the equal utility point on the Pareto-frontier for that game.

Proposition 7.1. Π^{seq}_E is convex. Also, A is in Π^{seq}_E; and for any fixed $\varepsilon > 0$, there is a $\delta(\varepsilon) < 1$; and a point $B = (\hat{\Pi}^{seq}_1, \hat{\Pi}^{seq}_2) \in \Pi^{seq}_E$ such that $\hat{\Pi}^{seq}_i > \hat{\Pi} - \varepsilon$; $i = 1, 2$ for $\delta \ge \delta(\varepsilon)$. Consequently, as $\delta \to 1$, the Pareto frontier of Π^{seq}_E is asymptotically linear between S and A.

Proof. See Appendix. □

This Proposition is illustrated in Figure 3 below. It shows that in the sequential move game, for low discounting, we can approximate "half" the linear part of the Pareto-frontier of the simultaneous move game, so sequential moves need not be a barrier to efficiency.

Figure 3 in here

8. Conclusions

This paper has studied a simple dynamic game where the level of cooperation chosen by each player in any period is irreversible. We have shown that irreversibility causes gradualism, i.e., any (subgame-perfect) sequence of actions involving partial cooperation cannot involve an immediate move to full cooperation, and we have refined and extended this basic insight in various ways. First, we showed that if payoffs are differentiable in actions, then (for a fixed discount factor), the level of cooperation asymptotes to a limit strictly below full cooperation, and this limit value is easily characterized. For the case where payoffs are linear up to some joint cooperation level, and constant or decreasing thereafter, the results are different — above some critical discount factor equilibrium cooperation can converge asymptotically to the fully efficient level. Below this critical discount factor, no cooperation is possible.

Later sections of the paper then extend the basic model in several directions. First, we studied an "adjustment cost" model which is applicable to a variety of economic situations, and showed that it can be reformulated so that it is a special case of our base model. We then applied the adjustment cost model to study sequential public good contribution games and capacity reduction in a declining industry.

Other extensions were to allow for irreversibility, asymmetry, and sequential moves. However, in all these variants of the base case, we have continued to assume that the underlying model is symmetric, i.e., both players have the same payoffs, given a permutation of their action variables. This is somewhat restrictive; in many situations where irreversibility arises naturally, e.g. Coasian bargaining without enforceable contracts but where actions are irreversible, payoffs will be asymmetric. Another limitation of the model is that players only have a scalar action variable; in many applications, players have several action variables, as in, for example, capacity reduction games, where firms control both capacity and output. Extending the model in these directions is a project for the future.

References

[1] Admati, A. R. and M. Perry (1991) "Joint Projects without Commitment", Review of Economic Studies, 58, 259-276.

[2] Bagwell, K. and R.W. Staiger (1997) "GATT-Think", mimeo, Columbia University.

[3] Compte, C. and P. Jehiel (1998) "When Outside Options Force Concessions to be Gradual", mimeo, C.E.R.A.S., Paris.

[4] Fershtman, C. and S, Nitzan (1991) "Dynamic Voluntary Provision of Public Goods", European Economic Review, 35, 1057-1067.

[5] Gale, D. (1997) "Monotone Games", mimeo, Department of Economics, New York University.

[6] Ghemawat, P. and B. Nalebu¤ (1990) "The Devolution of Declining Industries", Quarterly Journal of Economics, 167-186.

[7] Marx, L. and S.A. Matthews (1998) "Dynamic Voluntary Contributions to a Public Project", Discussion Paper No. 99-01, University of Pennsylvania

A. Appendix

Proof of Lemma 4. Suppose to the contrary there exists a $\{c'_t\}_{t=1}^{\infty}$ in C_{SE} with $c'_t > \hat{c}_t$ for some t. Define for all $t \geq 0$; $\tilde{c}_t = \max\{\hat{c}_t; c'_t\}$. It is clear from Assumption A1 and Lemma 2.1 (i) that

$$\pi(\tilde{c}_t; \tilde{c}_t) \geq \pi(\hat{c}_t; \hat{c}_t), \text{ all } t; \tag{A.1}$$

with at least one strict inequality, so that $\{\tilde{c}_t\}_{t=1}^{\infty}$ gives both agents a higher payoff than $\{\hat{c}_t\}_{t=1}^{\infty}$. So, if we can show that $\{\tilde{c}_t\}_{t=1}^{\infty}$ is an equilibrium sequence, this will contradict the assumed efficiency of $\{\hat{c}_t\}_{t=1}^{\infty}$ and the result is then proved.

Say the sequences $\{\hat{c}_t\}_{t=1}^{\infty}$; $\{c'_t\}_{t=1}^{\infty}$ have a crossing point at τ if $c'_{\tau-1} \leq \hat{c}_{\tau-1}$; $c'_\tau \geq \hat{c}_\tau$ with at least one strict inequality, or $c'_{\tau-1} \geq \hat{c}_{\tau-1}$; $c'_\tau \leq \hat{c}_\tau$ with at least one strict inequality. Also, define $S_t = \pi(c_t; c_t) + \delta\pi(c_{t+1}; c_{t+1}) + \ldots$; so that $\tilde{S}_t \geq \hat{S}_t; S'_\tau$ by (A.1).

There are then two possibilities at any time τ: The first is that there is no crossing point at τ. Then, either $(\tilde{c}_{\tau-1}; \tilde{c}_\tau) = (\hat{c}_{\tau-1}; \hat{c}_\tau)$ or $(\tilde{c}_{\tau-1}; \tilde{c}_\tau) = (c'_{\tau-1}; c'_\tau)$. Without loss of generality, assume the former. As $\{\hat{c}_t\}_{t=1}^{\infty}$ is an equilibrium sequence, we have $\pi(\hat{c}_{\tau-1}; \hat{c}_\tau)/(1-\delta) \leq \hat{S}_\tau$; so that $(\tilde{c}_{\tau-1}; \tilde{c}_\tau) = (\hat{c}_{\tau-1}; \hat{c}_\tau)$ and $\tilde{S}_\tau \geq \hat{S}_\tau$ together imply $\pi(\tilde{c}_{\tau-1}; \tilde{c}_\tau)/(1-\delta) \leq \tilde{S}_\tau$; i.e., the $\tau-$constraint is satisfied for $\{\tilde{c}_t\}_{t=1}^{\infty}$.

Now assume that $\{\hat{c}_t\}_{t=1}^{\infty}$ and $\{c'_t\}_{t=1}^{\infty}$ have a crossing point at τ; and assume w.l.o.g. that

$$c'_{\tau-1} \leq \hat{c}_{\tau-1}; \; c'_\tau \geq \hat{c}_\tau: \tag{A.2}$$

Then as $\{c'_t\}_{t=1}^{\infty}$ is an equilibrium sequence, $\pi(c'_{\tau-1}; c'_\tau)/(1-\delta) \leq S'_\tau$. Also, $\tilde{S}_\tau \geq S'_\tau$ and from (A.2), $\tilde{c}_\tau = c'_\tau$. Consequently,

$$\frac{\pi(c'_{\tau-1}; \tilde{c}_\tau)}{1-\delta} \leq \tilde{S}_\tau: \tag{A.3}$$

Finally, again from (A.2), $c'_{\tau-1} \leq \hat{c}_{\tau-1} = \tilde{c}_{\tau-1}$: Using this fact, plus π decreasing in its first argument, we have $\pi(\tilde{c}_{\tau-1}; \tilde{c}_\tau) \leq \pi(c'_{\tau-1}; \tilde{c}_\tau)$; so from (A.3) the $\tau-$constraint holds for $\{\tilde{c}_t\}_{t=1}^{\infty}$. Consequently all $\tau-$constraints hold for the sequence $\{\tilde{c}_t\}_{t=1}^{\infty}$, so it is an equilibrium sequence, as required. $\square$

Proof of Lemma 5.1. (i) Take an efficient path $\{\tilde{c}_t\}_{t=1}^{\infty}$—such a sequence exists by a similar argument to that of Lemma 2—and define $\tau \geq 1$ to be the first period such that $\tilde{c}_\tau > c^*$ (if such a period does not exist, then (i) holds immediately): Define a new sequence with $\hat{c}_t := \tilde{c}_t$; for $t < \tau$; and $\hat{c}_t := c^*$ for $t \geq \tau$: $\{\hat{c}_t\}_{t=1}^{\infty}$ clearly yields as much utility as $\{\tilde{c}_t\}_{t=1}^{\infty}$ at every point, and it will be shown that it also satisfies (5.1) for all t: First, (5:1) holds at τ since $\Delta(\rho; \{\tilde{c}_{\tau-1}; \tilde{c}_\tau\}) > \Delta_\tau(\rho; \{\hat{c}_{\tau-1}; \hat{c}_\tau\})$ as $\hat{c}_\tau < \tilde{c}_\tau$ while $\hat{c}_{\tau-1} = \tilde{c}_{\tau-1}$ (and using π increasing in its second argument); moreover the RHS of (5.1) is no smaller. Likewise, for $t' > \tau$; we have $\Delta(\rho; \{c_{t'-1}; c_{t'}\}) < \Delta(\rho; \{c_{\tau-1}; c_\tau\})$ since $\hat{c}_{t'} = \hat{c}_\tau$; and $\hat{c}_{t'-1} > \hat{c}_{\tau-1}$; while continuation path payoffs (RHS of (5.1)) are the same at τ and t'. So (5.1) holds at t'; it clearly holds at $t < \tau$ as the LHS is unchanged relative to the $\{\tilde{c}_t\}_{t=1}^{\infty}$ sequence while the RHS is no smaller. The proof of $\hat{b}_{t-1} \leq \hat{b}_t$ is straightforward but tedious and is omitted. (ii) The argument is similar to the proof of Lemma 2.2. (iii) Assume the contrary, so there is an equilibrium sequence $\{c'_t\}_{t=1}^{\infty}$ yielding a higher payoff than $\{\hat{b}_t\}_{t=1}^{\infty}$, and both sequences lie below or equal to c^*. Hence the construction of Lemma 2.4 can be followed to create a new sequence $\{\tilde{c}_t\}_{t=1}^{\infty}$ which yields a higher overall payoff. That it satisfies (5.1) at each t follows from similar arguments. ∎

Proof of Proposition 5.2. (a) Let $\hat{b}_t(1) = \hat{b}_t$ to ease notation. To prove part (i), it is sufficient to show that we can find $\hat{\rho}$ such that

$$\Phi(\rho; \hat{b}_{t-1}, \hat{b}_t) < \Phi(1; \hat{b}_{t-1}, \hat{b}_t\}), \quad t = 1, 2, \ldots, \quad 1 > \rho > \hat{\rho}. \tag{A.4}$$

For then, for $1 > \rho > \hat{\rho}$, $\{\hat{b}_t\}_{t=1}^{\infty}$ satisfies the incentive constraints (5.1).

(b) Fix t; then

$$\Phi_t(\rho) - \Phi_t(1) = -\Phi'_t(1)\varepsilon + \frac{1}{2}\Phi''_t(1)\varepsilon^2 + O(\varepsilon^3), \tag{A.5}$$

where $\varepsilon := 1 - \rho$, and to ease notation, we set $\Phi_t(\rho) := \Phi(\rho; \{\hat{b}_{t-1}, \hat{b}_t\})$. Routine calculation gives:

$$\begin{aligned} \Phi'_t(1) &= A_t(1 + 2\delta + 3\delta^2 + 4\delta^3 + \ldots) &\text{(A.6)} \\ \Phi''_t(1) &= A_t(2\delta + 6\delta^2 + 12\delta^3 + \ldots) + B_t &\text{(A.7)} \end{aligned}$$

where $A_t = \pi_1 \hat{b}_{t-1} + \delta\pi_2 \hat{b}_t$, and B_t is the sum of terms involving $\pi_{11}, \pi_{22}, \pi_{12}$, and where it is understood that all derivatives of π are evaluated at $(\hat{b}_{t-1}, \hat{b}_t)$. Also the series $1 + 2\delta + 3\delta^2 + 4\delta^3 + \ldots$ and $2\delta + 6\delta^2 + 12\delta^3 + \ldots$ both converge (to $s_1, s_2 > 0$ respectively). Useful properties of A_t, B_t, proved in (c) below, are: $A_t > 0$, $B_t < 0$, $\lim_{t\to\infty} A_t = 0$, $\lim_{t\to\infty} B_t < 0$.

Consequently, we can write

$$-\Phi'_t(1)\varepsilon + \frac{1}{2}\Phi''_t(1)\varepsilon^2 = (\pi_1 \hat{b}_{t-1} + \delta\pi_2 \hat{b}_t)(-s_1\varepsilon + 0.5 s_2 \varepsilon^2) + 0.5\varepsilon^2 B_t. \tag{A.8}$$

Clearly there exists ε_t such that for ε satisfying $0 < \varepsilon < \varepsilon_t$, the RHS of (A.8) is negative. It follows from (A.5) that for $\varepsilon < \varepsilon_t$, $\Phi_t(\rho) < \Phi_t(1)$.

(c) (Properties of A_t, B_t): First we show that $A_t > 0$. We have $\hat{b}_t \geq \hat{b}_{t-1}$, so (as $\pi_2 > 0$) we only need show that

$$\pi_1(\hat{b}_{t-1}, \hat{b}_t) + \delta\pi_2(\hat{b}_{t-1}, \hat{b}_t) > 0. \tag{A.9}$$

Now, we know from Section 3 that provided the maximum attainable level of cooperation $\mathbf{b} > 0$; then $\mathbf{b}_t < \mathbf{b}$ all t; and thus $\gamma(\mathbf{b}_t) \equiv -\pi_1(\mathbf{b}_t, \mathbf{b}_t)/\pi_2(\mathbf{b}_t, \mathbf{b}_t) < \delta$, which implies

$$\pi_1(\mathbf{b}_t, \mathbf{b}_t) + \delta\pi_2(\mathbf{b}_t, \mathbf{b}_t) > 0. \tag{A.10}$$

Also, from the assumptions on π that $\pi_{11} < 0$, $\pi_{12} \le 0$, we have

$$\pi_1(\mathbf{b}_{t-1}, \mathbf{b}_t) \ge \pi_1(\mathbf{b}_t, \mathbf{b}_t); \quad \pi_2(\mathbf{b}_{t-1}, \mathbf{b}_t) \ge \pi_2(\mathbf{b}_t, \mathbf{b}_t). \tag{A.11}$$

Consequently, (A.9) follows from (A.10) and (A.11). Also note

$$\begin{aligned} \lim_{t\to\infty} A_t &= \pi_1(\mathbf{b}_{t-1}, \mathbf{b}_t)\mathbf{b}_{t-1} + \delta\pi_2(\mathbf{b}_{t-1}, \mathbf{b}_t)\mathbf{b}_t \\ &= [\pi_1(\mathbf{b}, \mathbf{b}) + \delta\pi_2(\mathbf{b}, \mathbf{b})]\mathbf{b} \\ &= 0 \end{aligned}$$

where the term in the square brackets is zero by definition of $\mathbf{b}$. The properties of B_t follow from the fact that B_t is the sum of terms involving $\pi_{11}, \pi_{22}, \pi_{12}$ with coefficients bounded (in t) above zero.

(d) We now show that the sequence $\{\rho_t\}_{t=1}^{\infty} := \{1 - \varepsilon_t\}_{t=1}^{\infty}$ can be chosen to be bounded below 1; this would imply (A.4) with $\bar{\rho} := \sup \rho_t < 1$. If such a sequence does not exist, then there must be a subsequence which w.l.o.g. we take to be $\{\rho_t\}_{t=1}^{\infty}$ itself, converging to 1; i.e., $\rho_t \to 1$ and

$$\Delta(\rho_t, \mathbf{b}_{t-1}, \mathbf{b}_t) \ge \Delta(1, \mathbf{b}_{t-1}, \mathbf{b}_t), \qquad \text{all } t. \tag{A.12}$$

But now as $t \to \infty$, $\mathbf{b}_t \to \mathbf{b}$, so from (A.5), we have

$$\begin{aligned} \Delta(\rho, \mathbf{b}, \mathbf{b}) - \Delta(1, \mathbf{b}, \mathbf{b}) &\approx \lim_{t\to\infty} \{-\Delta_t'(1)\varepsilon + \frac{1}{2}\Delta_t''(1)\varepsilon^2\} \\ &= \lim_{t\to\infty} 0.5\varepsilon^2 B_t = 0.5\varepsilon^2 \overline{B} < 0. \end{aligned}$$

So, for some fixed $\theta > 0$, there exists $\rho_\theta < 1$ such that

$$\Delta(\rho, \mathbf{b}, \mathbf{b}) < \Delta(1, \mathbf{b}, \mathbf{b}) - 3\theta, \quad 1 > \rho > \rho_\theta. \tag{A.13}$$

Also, as $t \to \infty$, $\mathbf{b}_t \to \mathbf{b}$, and $\Delta_t(\rho)$ is continuous in ρ and $\mathbf{b}_{t-1}, \mathbf{b}_t$, there exists a T_θ such that for all $t \ge T_\theta$:

$$\begin{aligned} \Delta(\rho, \mathbf{b}_{t-1}, \mathbf{b}_t) &< \Delta(\rho, \mathbf{b}, \mathbf{b}) + \theta, \quad 1 > \rho > \rho_\theta; \\ \Delta(1, \mathbf{b}, \mathbf{b}) &< \Delta(1, \mathbf{b}_{t-1}, \mathbf{b}_t) + \theta. \end{aligned} \tag{A.14}$$

Combining (A.13) and (A.14), we get

$$\Delta(\rho, \mathbf{b}_{t-1}, \mathbf{b}_t) < \Delta(1, \mathbf{b}_{t-1}, \mathbf{b}_t) - \theta, \quad 1 > \rho > \rho_\theta, \ t \ge T_\theta. \tag{A.15}$$

But (A.12) and (A.15) are in contradiction.

(e) To prove part (ii) of the Proposition, let

$$\tilde{c}_t = \begin{cases} \hat{c}_t & t < T_\theta \\ \hat{c}_t + \eta & t \geq T_\theta \end{cases}$$

Also, choose $\eta < c^* - \hat{c}$ small enough so that (by continuity)

$$\Delta(\rho; \tilde{c}_{t-1}, \tilde{c}_t) < \Delta(\rho; \hat{c}_{t-1}, \hat{c}_t) + \theta/2, \; 1 > \rho > \rho_\theta, \; t \geq T_\theta. \tag{A.16}$$

We show that $\{\tilde{c}_t\}_{t=1}^{\infty}$ is an equilibrium symmetric path in the $\rho-$reversible game, if $1 > \rho > \max\{\sup \rho_t, \rho_\theta\}$. To see this, note first that $\tilde{c}_t < c^*$, so for any t the continuation payoff from $\{\tilde{c}_t\}_{t=1}^{\infty}$ is strictly greater than that from $\{\hat{c}_t\}_{t=1}^{\infty}$. Hence, it suffices to show that the deviation payoff in the $\rho-$reversible game from $\{\tilde{c}_t\}_{t=1}^{\infty}$ is no higher than the deviation payoff from $\{\hat{c}_t\}_{t=1}^{\infty}$ in the irreversible case. But from (A.15) and (A.16), we have

$$\Delta(\rho; \tilde{c}_{t-1}, \tilde{c}_t) < \Delta(1; \hat{c}_{t-1}, \hat{c}_t) - \theta/2, \; 1 > \rho > \rho_\theta, \; t \geq T_\theta$$

as required, provided $\rho > \hat{\rho} \equiv \sup \rho_t$, (A.4) ensures (from (a)-(d) above) that (5.1) holds for $t < T_\theta$. Thus setting $\bar{\rho} = \max\{\sup \rho_t, \rho_\theta\}$ implies that (5.1) holds for all $1 > \rho > \bar{\rho}$, $t \geq 1$. Then from Lemma 5.1 (iii), $\hat{c}_\infty(\rho) \geq \hat{c}_\infty(1) + \varepsilon$.

(f) To prove part (iii), it follows immediately from the construction of $\{\tilde{c}_t\}_{t=1}^{\infty}$ that

$$\bar{\Pi} := (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} \pi(\tilde{c}_t, \tilde{c}_t) > \hat{\Pi}(1)$$

and as $\{\tilde{c}_t\}_{t=1}^{\infty}$ is an equilibrium (but not necessarily the efficient) path in the $\rho-$reversible game, $\hat{\Pi}(\rho) \geq \bar{\Pi}$ and so the result is proved. $\square$

Proof of Proposition 5.3. Let $\rho = 1$, and suppose $\{c_t\}_{t=1}^{\infty}$ is an efficient path, assuming $a < 1$, this path is increasing by earlier arguments. The derivative of $\Delta_t(\rho; \{c_t\}_{t=1}^{\infty}) \equiv (\pi_1 \rho c_{t-1} + \pi_2 c_t)/(1 - \rho\delta)$ with respect to ρ has the sign of $c_t - a c_{t-1}$, which is positive for all $t \geq 1$ as $a < 1$ and $c_t > c_{t-1} \geq 0$. Hence for any $\hat{\rho} \in [0, 1)$, $\{c_t\}_{t=1}^{\infty}$ remains an equilibrium path as the deviation payoff $\Delta_t(\hat{\rho}; \{c_t\}_{t=1}^{\infty})$ is smaller than at $\rho = 1$, while the continuation payoff is unchanged. By Lemma 5.1(i) and (iii), there exists a non-decreasing efficient path for $\hat{\rho} < 1$, say $\{\hat{c}_t\}_{t=1}^{\infty}$, which lies no lower than $\{c_t\}_{t=1}^{\infty}$ and no higher than c^* at each point. Next, the above argument can be repeated for any $\rho' < \hat{\rho} < 1$, so that at ρ', $\{\hat{c}_t\}_{t=1}^{\infty}$ is an equilibrium path. Moreover, the incentive constraint at each t is strictly looser, so that by Lemma 5.1(ii) if the first-best is not attainable at ρ, i.e., if $\hat{c}_t < c^*$ for some t, $\hat{c}_t$ is not part of an efficient equilibrium path for ρ'. The conclusion is then that at ρ', $\{\hat{c}_t\}_{t=1}^{\infty}$ is equilibrium but not efficient, i.e., there is an equilibrium path yielding a higher payoff than $\{\hat{c}_t\}_{t=1}^{\infty}$. To prove that c^* is attained in finite time, consider the path generated by

(5.2) for some choice of c_1. Note that $(\rho^{t-1} + \rho^{t-2}a + \ldots + \rho a^{t-2} + a^{t-1})$ attains a maximum at some $t^* \geq 1$, and declines to zero. Choose $c_1 = \hat{c}_1$ so that $\hat{c}_{t^*} = c^*$. If (5.2) is followed for all t, the same argument as in Lemma 2.4 establishes that the incentive constraint holds for all t as $\lim_{t\to\infty} \hat{c}_t = 0$ $(<\infty)$. (It does not matter if this path violates $\hat{c}_t \geq \rho \hat{c}_{t-1}$ beyond t^*.) Now change the path by setting $\hat{c}_t = c^*$ for $t > t^*$. Continuation payoffs are increased at each date. Deviation payoffs are the same at each date up to t^*, and since the incentive constraint is thus satisfied at t^* it must also be satisfied at all $t > t^*$. Thus this path satisfies all incentive constraints and c^* is attained in finite time. By Lemma 5.1(iii) there is an efficient path that attains c^* by t^* or earlier. (ii) If $a \geq 1$, then consider the incentive condition for a stationary path at c:

$$\frac{\pi_1 \rho c + \pi_2 c}{1 - \rho\delta} \leq \frac{\pi_1 c + \pi_2 c}{1 - \delta}. \tag{A.17}$$

Rearranging, this is equivalent to $a \leq 1$. Hence if $a > 1$, if c^* is attained, the incentive constraint is violated at c^* (likewise if a higher efficient level is attained, should one exist); if $c_t < c^*$ for all t, then the path must satisfy (5.2) for all t, implying $c_t \to \infty$ if $c_1 > 0$, , a contradiction; hence $c_1 = 0$, so $c_t = 0$ all t. If $a = 1$, (A.17) holds with equality; if c^* is attained at t, the incentive constraint at t is stricter than (A.17), and so is violated; hence $c_t < c^*$ all t, in which case (5.2) applies, and setting $c_1 = (1-\rho)c^*$ implies that $\lim_{t\to\infty} c_t = c^*$; and because the limit is finite, all incentive constraints are satisfied (as argued earlier). $\blacksquare$

Proof of Proposition 6.1. First, we show that Π_E is a convex set. First, the constraints in (6.1) are linear. Consequently, if $\{c'_{1,t}, c'_{2,t}\}_{t=1}^{\infty}$ and $\{c''_{1,t}, c''_{2,t}\}_{t=1}^{\infty}$ satisfy (6.1), a convex combination of the two must also satisfy (6.1) and so C_E is a convex set. Also, adapting Lemma 2.1, any sequence in C_E must have $c_{1,t} + c_{2,t} < 2c^*$, all i, t, so payoffs are linear in any path in C_E. It follows immediately that Π_E is a convex set also.

Let $C_{EE} \subseteq C_E$ be the set of all paths $\{c_{1,t}, c_{2,t}\}_{t=1}^{\infty}$ which satisfy the incentive constraints (6.1) with equality at each $t \geq 1$, and $\Pi_{EE} \subseteq \Pi_E$ the corresponding set of payoffs. Straightforward manipulation implies that these paths can be written as a system of two linked first-order difference equations in differences $\Delta c_{i,t} = c_{i,t} - c_{i,t-1}$,

$$\Delta c_{1,t} = a \Delta c_{2,t-1} \tag{A.18}$$
$$\Delta c_{2,t} = a \Delta c_{1,t-1} \tag{A.19}$$

where $a = \frac{-\pi_1}{\pi_2 \delta}$ as before. As $\delta > \hat{\delta}$, it follows that $a < 1$. Also, note that the initial conditions

$$\Delta c_{i,1} = c_{i,1} - c_{i,0} = c_{i,1},\ i = 1, 2$$

can be set freely. Routine manipulation of the system (A.18), (A.19) gives the solutions

$$c_{i,t} = \begin{cases} \frac{1}{1-a^2}\left[c_{i,1}(1-a^{t+1}) + a c_{j,1}(1-a^{t-1})\right], & t \text{ odd} \\ \frac{1}{1-a^2}\left[c_{i,1}(1-a^{t}) + a c_{j,1}(1-a^{t})\right], & t \text{ even} \end{cases}, \quad i, j = 1, 2,\ j \neq i. \tag{A.20}$$

Taking limits in (A.20), we get two equations that give, as $a < 1$; the limit values of $c_{1,t}; c_{2,t}$ as functions of the initial values:

$$\lim_{t\to\infty} c_{1,t} = c_{1,\infty} = \frac{1}{1-a^2}\left[c_{1,1} + ac_{2,1}\right];$$
$$\lim_{t\to\infty} c_{2,t} = c_{2,\infty} = \frac{1}{1-a^2}\left[c_{2,1} + ac_{1,1}\right]:$$

Inverting and solving, we get

$$c_{1,1} = c_{1,\infty} - ac_{2,\infty};\ c_{2,1} = c_{2,\infty} - ac_{1,\infty}: \tag{A.21}$$

Note that we can think of $c_{1,1}$ and $c_{2,1}$ as being determined by $c_{1,\infty}$ and $c_{2,\infty}$ where the latter can be freely chosen subject to the constraint that $c_{1,\infty} + c_{2,\infty} \leq 2c^*$ and that $c_{i,1} \geq 0$, $i = 1; 2$: The latter requires

$$\frac{c_{2,\infty}}{a} \geq c_{1,\infty} \geq ac_{2,\infty}: \tag{A.22}$$

C_{EE} is characterized by sequences satisfying (A.20) and (A.22) since convergent sequences satisfying (A.18) and (A.19) also satisfy (6.1) with equality as in Lemma 2.4.

Substituting (A.20) back in the payoffs gives, after some rearrangement, for $i; j = 1; 2;\ j \neq i$;

$$\begin{aligned}
\Pi_i &= (1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}\left(\pi_1 c_{i,t} + \pi_2 c_{j,t}\right)\\
&= \frac{1}{1-a^2}\left[\pi_1\left(c_{i,1} + ac_{j,1}\right) + \pi_2\left(c_{j,1} + ac_{i,1}\right)\right]\\
&\quad + \frac{(1-\delta)}{(1-a^2)(1-a^2\delta^2)}\pi_1\left[a(ac_{i,1} + c_{j,1}) + \delta a^2\left(c_{i,1} + ac_{j,1}\right)\right]\\
&\quad + \frac{(1-\delta)}{(1-a^2)(1-a^2\delta^2)}\pi_2\left[a(ac_{j,1} + c_{i,1}) + \delta a^2\left(c_{j,1} + ac_{i,1}\right)\right]:
\end{aligned}$$

Now, from (A.21), we have

$$c_{i,1} + ac_{j,1} = (1-a^2)c_{i,\infty}: \tag{A.23}$$

So, we get, after some manipulation,

$$\Pi_i = \left[1 - \frac{(1-\delta)(a + a^2\delta)}{(1-a^2\delta^2)}\right]\left(\pi_1 c_{i,\infty} + \pi_2 c_{j,\infty}\right);\ i = 1; 2$$

and so

$$\Pi_1 + \Pi_2 = \phi(\delta)(\pi_1 + \pi_2)(c_{1,\infty} + c_{2,\infty}); \tag{A.24}$$

where $\phi(\delta) := \left[1 - \frac{(1-\delta)(a+a^2\delta)}{(1-a^2\delta^2)}\right]:$

So as long as $c_{1;1} + c_{2;1} = 2c^¤$, $¦_1 + ¦_2 = Á(±)(¼_1 + ¼_2)2c^¤$, no matter how the sum $c_{1;1} + c_{2;1}$ is distributed. This says that the frontier is linear between two endpoints de…ned by the restrictions (A.22). Let A be one endpoint, de…ned by the condition that $c_{1;1} = ac_{2;1}$, and B the other endpoint, de…ned by $c_{2;1} = ac_{1;1}$ (B is symmetric to A) Combining this with $c_{1;1} + c_{2;1} = 2c^¤$ implies that A is generated by the path with endpoints

$$c_{1;1} = \frac{2ac^¤}{1+a};\ c_{2;1} = \frac{2c^¤}{1+a};$$

and therefore with payo®s $(¦^0; ¦^{00})$ where

$$\begin{aligned} ¦^0 &= \frac{2c^¤}{1+a}\left[1 ¡ \frac{(1 ¡ ±)(a + a^2±)}{(1 ¡ a^2±^2)}\right][¼_1 a + ¼_2]; \\ ¦^{00} &= \frac{2c^¤}{1+a}\left[1 ¡ \frac{(1 ¡ ±)(a + a^2±)}{(1 ¡ a^2±^2)}\right][¼_1 + a¼_2]: \end{aligned}$$

So,

$$¦^0 = ¦^{00} = \frac{¼_1 a(±) + ¼_2}{¼_1 + a(±)¼_2}: \qquad (A.25)$$

Now, it is easily checked that $¦^0; ¦^{00} > 0$ and that the RHS of (A.25) is strictly greater than 1, so $¦^0 > ¦^{00} > 0$ as claimed.

To complete the proof, we need to show that points A and B lie on the frontier of $¦_E$; the convexity of $¦_E$ then implies that the whole of line segment AB lies on this frontier. First, note that the point S where the line segment AB crosses the 45°line is generated by the symmetric path

$$c^¤_t = 0:5c_{1;t} + 0:5c_{2;t};$$

where $f c_{1;t}; c_{2;t} g^1_{t=1}$ is the path supporting A; so every incentive constraint holds with equality for $f c^¤_t g^1_{t=1}$. But then $f c^¤_t g^1_{t=1}$ is the symmetric e¢cient path characterized in Sections 2 and 3. So, S must be on the frontier since otherwise there is an asymmetric path which Pareto-dominates S;and by symmetry another path with the player indices switched which also Pareto dominates S; a convex combination of these two paths is a symmetric path which Pareto dominates S; a contradiction of the de…nition of S:

Suppose …nally that points A; B are not on the frontier of $¦_E$. Then, there must be points C; D where C (resp. D) Pareto-dominates A (resp. B) which are on the frontier of $¦_E$: But if S; C; D are all on the frontier of $¦_E$, it must be non-convex, contrary to the result already established. ¤

Proof of Proposition 6.2. From the proof of Proposition 6.1, we have

$$¦^0 = ¦^{00} = \frac{¼_1 a(±) + ¼_2}{¼_1 + a(±)¼_2}: \qquad (A.26)$$

As a is decreasing in δ, and the right-hand side of (A.26) is decreasing in a, γ'/γ'' is increasing in δ. Moreover, as $\delta \to 1$, $\gamma'/\gamma'' \to 0$; and as $\delta \to \hat{\delta}_+$, $\gamma'/\gamma'' \to 1$, as required. Likewise from (A.24) in the proof of Proposition 6.1, on the line segment AB,

$$\Gamma = \gamma_1 + \gamma_2 = \phi(\delta)(\pi_1 + \pi_2)2c^*$$

where $\phi(\delta) := \left[1 - \frac{(1-\delta)(a+a^2\delta)}{(1-a^2\delta^2)}\right]$. Rearrangement gives $\phi(\delta) = \left[1 - \frac{\hat{\delta}(1-\delta)}{\delta(1-\hat{\delta}^2)}\right]$. It is then clear that $\phi(\hat{\delta}) = 0$; $\phi(1) = 1$; and $\phi'(\delta) > 0$; $\delta \in (\hat{\delta}, 1)$; and so Γ has the desired properties on the line segment AB: $\square$

Proof of Proposition 7.1. To prove convexity of γ_E^{seq}, note that since C_E, C^{seq} are both convex, so $C_E^{seq} = C_E \cap C^{seq}$ is also convex. Consequently, γ_E^{seq} is also convex, by linearity of payoffs.

To prove A in γ_E^{seq}, we proceed as follows. Point A is generated by a path described in (A.20) with $c_{1,1} = 0$. All we have to do is show that this path is in C^{seq} as this path is already in C_E by construction. Now setting $c_{1,1} = 0$ in (A.20), we see that the path generating A satisfies:

$$c^A_{1,t} = \begin{cases} \frac{1}{1-a^2}\left[ac_{2,1}(1 - a^{t-1})\right]; & t \text{ odd} \\ \frac{1}{1-a^2}\left[ac_{2,1}(1 - a^{t})\right]; & t \text{ even} \end{cases}$$
$$c^A_{2,t} = \begin{cases} \frac{1}{1-a^2}\left[c_{2,1}(1 - a^{t+1})\right] & t \text{ odd} \\ \frac{1}{1-a^2}\left[c_{2,1}(1 - a^{t})\right]; & t \text{ even} \end{cases}$$

So, by inspection, $\{c^A_{1,t}; c^A_{2,t}\}_{t=1}^{\infty}$ has the property that player 1 only changes her level of cooperation in even periods, and player 2 in odd periods.

Next, let $\{\hat{c}_t\}_{t=1}^{\infty}$ be the (unique) symmetric efficient path in the simultaneous move game: Now define the asymmetric path $\{\hat{c}_{1,t}; \hat{c}_{2,t}\}_{t=1}^{\infty}$ in C^{seq} as follows:

$$\hat{c}_{1,t} = \hat{c}_{1,t+1} = \hat{c}_t; \ t = 0; 2; 4; 6 \ldots;$$
$$\hat{c}_{2,t} = \hat{c}_{2,t+1} = \hat{c}_t; \ t = 1; 3; 5 \ldots$$

This is simply the path where an agent whose turn it is to move at t chooses $\hat{c}_t$. Next, we show that $\{\hat{c}_{1,t}; \hat{c}_{2,t}\}_{t=1}^{\infty}$ is incentive-compatible, i.e., in C_E^{seq} in the sequential move game. Define as before $\Delta_t := \hat{c}_t - \hat{c}_{t-1}$; and recall $\Delta_t = a\Delta_{t-1}$ on the efficient path. For the player who moves at $t \geq 2$; and writing Δ for Δ_{t-1}; the constraints (7.2) and (7.3) can be written as:

$$\begin{aligned} \frac{\pi_1 c_{t-2} + \pi_2(c_{t-1} + \Delta)}{1-\delta} \leq\ & \pi_1(c_{t-2} + \Delta + a\Delta) + \pi_2(c_{t-1} + \Delta) \\ +\ & \delta(\pi_1(c_{t-2} + \Delta + a\Delta) + \pi_2(c_{t-1} + \Delta + a\Delta + a^2\Delta)) \\ +\ & \delta^2(\pi_1(c_{t-2} + \Delta + \ldots + a^3\Delta) + \pi_2(c_{t-1} + \Delta + a\Delta + a^2\Delta)) + \ldots \end{aligned} \tag{A.27}$$

or

$$\frac{\pi_2 \Delta}{1-\delta} \le \frac{(1+a)\pi_1 \Delta + (1 - \delta^2 a^2 + \delta a + \delta a^2)\pi_2 \Delta}{(1-\delta)(1-\delta^2 a^2)};$$

which holds with equality as $a = -\pi_1/(\delta\pi_2)$. Thus $\{\mathbf{b}_{1,t}, \mathbf{b}_{2,t}\}_{t=1}^{\infty}$ satisfies equilibrium conditions from $t = 2$ onwards; at $t = 1$ the constraint would hold with equality if player 2's inherited c was $-\Delta_1/a$; since it is higher, the constraint will be slack (as $\pi_1 < 0$).

The payoffs from the path $\{\mathbf{b}_{1,t}, \mathbf{b}_{2,t}\}$ are;

$$\begin{aligned}
\hat{v}_1^{seq} &= (1-\delta)\{[\pi_2 \mathbf{b}_1] + \delta[\pi_1 \mathbf{b}_2 + \pi_2 \mathbf{b}_1] + \delta^2[\pi_1 \mathbf{b}_2 + \pi_2 \mathbf{b}_3] + \ldots \\
\hat{v}_2^{seq} &= (1-\delta)\{[\pi_1 \mathbf{b}_1] + \delta[\pi_1 \mathbf{b}_1 + \pi_2 \mathbf{b}_2] + \delta^2[\pi_1 \mathbf{b}_3 + \pi_2 \mathbf{b}_2] + \ldots
\end{aligned}$$

Now since the payoffs from the efficient symmetric path in the simultaneous move game are

$$\hat{v} = (1-\delta)\{[\pi_1 \mathbf{b}_1 + \pi_2 \mathbf{b}_1] + \delta[\pi_1 \mathbf{b}_2 + \pi_2 \mathbf{b}_2] + \delta^2[\pi_1 \mathbf{b}_3 + \pi_2 \mathbf{b}_3] + \ldots;$$

we get

$$\begin{aligned}
\hat{v} - \hat{v}_1^{seq} &= (1-\delta)\{\pi_2 \mathbf{b}_1 + \delta\pi_1(\mathbf{b}_2 - \mathbf{b}_1) + \delta^2\pi_2(\mathbf{b}_3 - \mathbf{b}_2) + \delta^3\pi_1(\mathbf{b}_4 - \mathbf{b}_3) + ..\} \\
&= (1-\delta)\mathbf{b}_1\{\pi_2 \mathbf{b}_1 + \delta\pi_1 a\mathbf{b}_1 + \delta^2\pi_2 a^2 \mathbf{b}_1 + \delta^3\pi_1 a^3 \mathbf{b}_1 \ldots\} \\
&= (1-\delta)\mathbf{b}_1\left[\pi_2(1 + \delta^2 a^2 + \delta^4 a^4 + ..) + \delta a\pi_1(1 + \delta^2 a^2 + \delta^4 a^4 + ..)\right] \\
&= \frac{(1-\delta)\mathbf{b}_1}{1-\delta^2 a^2}[\pi_2 + \delta a\pi_1] \\
&< (1-\delta)\frac{\mathbf{b}_1 \pi_2}{1-(\pi_1/\pi_2)^2}
\end{aligned}$$

So, rearranging, $\hat{v} - (1-\delta)\theta < \hat{v}_1^{seq}$, $\theta > 0$. Consequently, for any $\varepsilon > 0$, $\hat{v} - \varepsilon < \hat{v}_1^{seq}$ for all $\delta \ge \delta(\varepsilon) = 1 - \varepsilon/\theta$; as required. (A similar argument applies for $i = 2$). $\square$