Oxford Economic Papers 46 (1994), 878–894

SELF-ENFORCING INTERNATIONAL ENVIRONMENTAL AGREEMENTS

By SCOTT BARRETT

Faculty of Economics, London Business School, Sussex Place, Regent's Park, London, NW1 4SA

Centre for Social and Economic Research on the Global Environment

1. Introduction

IT IS WELL-KNOWN that collective well-being can be increased if all countries cooperate in managing shared environmental resources like the climate and ozone layer, but that if this improved situation is attained, every country will earn even higher returns by free-riding on the virtuous behavior of the remaining cooperators. However, the fact that countries can do better if cooperation can be sustained suggests that countries have incentives to develop institutions which can punish free riding. Such institutions do in fact exist in the form of international environmental agreements (IEAs). Indeed, more than 100 IEAs are currently in force, the most recent of which include the Montreal Protocol on Substances that Deplete the Ozone Layer and the Framework Convention on Climate Change. The essential feature of IEAs is that they must be self-enforcing. No country can be forced to sign an IEA, and signatories to an IEA can always withdraw from the agreement. If IEAs can improve the management of shared environmental resources, they must make it attractive for countries to want to sign, and to want to carry out the terms of the agreement. This paper employs some concepts from game theory to explore the properties of self-enforcing IEAs.

Two different approaches are considered. The first is a model in which the number of signatories, the terms of the agreement, and the actions of non-signatories are all determined endogenously. This model assumes that signatories maximize their collective net benefits, recognizing how their choice of treaty obligations affects the actions of nonsignatories, while nonsignatories each choose their actions on the assumption that other countries will not respond. The mechanism that sustains the IEA is as follows. When a country joins the IEA, the other signatories increase their abatement levels, and hence reward the country for acceding to the agreement; when a country withdraws, the remaining signatories reduce their abatement levels, and hence punish the country for withdrawing from the agreement. These punishments and rewards are credible, because the signatories always maximize their collective net benefits. However, the magnitude of the punishments and rewards may not be sufficiently large to sustain a self-enforcing IEA consisting of many countries. Depending on the functional specification, Section 2 of this paper shows that the self-enforcing IEA may not exist, or it may consist of no more than two or three countries, or it may support up to as many countries as share the resource (depending in this last case on parameter values).

The other approach is to model the IEA as an infinitely repeated game. This approach has the attraction of increasing credible punishments compared with the other model, but it has the drawback of not solving endogenously for the number of signatories, the terms of the agreement, and the actions of non-signatories. Instead, this model can be employed to calculate the maximum number of countries that could support the full cooperative outcome. While the folk theorem guarantees that the full cooperative outcome can be sustained by any number of countries provided the rate of discount is sufficiently small, we shall see in Section 3 that this result may not hold for a self-enforcing IEA.[1] The reason is that the punishments that sustain the full cooperative outcome may be vulnerable to renegotiation, and a self-enforcing IEA must clearly be renegotiation-proof. As with the previous model, the ability of the infinitely-repeated game to sustain the full cooperative outcome depends on the functional specification and parameter values.

These different models yield one consistent and striking result. The first model shows that for global environmental problems, an IEA may achieve a high degree of cooperation, but only when the difference between global net benefits under the noncooperative and full cooperative outcomes is small. When this difference is large, a self-enforcing IEA cannot support a large number of signatory countries.[2] The second model shows that the maximum number of countries that can sustain the full cooperative outcome may be large, but only when the difference between global net benefits under the noncooperative and full cooperative outcomes is small. When this difference is large, the full cooperative outcome cannot be sustained by a self-enforcing IEA.

By necessity, the analysis imposes a number of restrictive assumptions, and the above result must be seen in the light of these. The most important assumptions are: (i) that all countries are identical; (ii) that each country's net benefit function is known by all countries, and known to be known by all countries; (iii) that the choice instrument is restricted to pollution abatement; (iv) that abatement levels are instantly and costlessly observable; (v) that the pollutant does not accumulate in the environment; and (iv) that the cost functions are independent. Further research is needed to learn whether the basic conclusion of this paper would be upset by changes in these assumptions.[3]

[1] Of course, the folk theorem only asserts that any feasible individually rational payoffs can be supported as a Nash equilibrium if the game is repeated infinitely often and players are sufficiently patient. Hence, the full cooperative outcome is only one among many supergame equilibria.

[2] Unless, of course, the agreement merely codifies the noncooperative outcome.

[3] A number of studies have considered these issues, but not within the context of the models presented in this paper. Black *et al.* (1993) consider information asymmetries in a model where the agreement specifies the minimum number of countries that must sign the agreement for it to come into force. With complete information, the solution to their model would trivially require that all countries sign the agreement. Such an agreement would not, however, be (*ex post*) self-enforcing. Barrett (1992) considers the implications for reaching a negotiated agreement when countries are different and may select different policy instruments (abatement levels, pollution taxes, and tradeable emission entitlements), but this study does not require that the agreement that is reached

2. The model of an IEA

2.1. *Linear marginal abatement benefits and costs*

Consider a world of $i = 1, \ldots, N$ identical countries, each of which emits a pollutant that damages a shared environmental resource. i's current abatement benefits are assumed to depend on current total abatement as follows

$$B_i(Q) = b(aQ - Q^2/2)N \tag{1}$$

where $B_i(Q)$ is i's abatement benefit, a and b are positive parameters, and Q is global abatement. According to (1), the marginal benefit of the first unit of abatement for i is ab/N, and that of the ath is zero.[4] The parameter b is the slope of the global marginal benefit function (found by summing (1) over all N countries, and differentiating with respect to Q).

Each country's abatement costs are assumed to depend on its own abatement level and no one else's. For country i, the abatement cost function is assumed to be given by

$$C_i(q_i) = cq_i^2/2 \tag{2}$$

where $C_i(q_i)$ is i's abatement cost and q_i is i's abatement (in eq. (1), $Q = \Sigma_i q_i$). The parameter c represents the slope of each country's marginal abatement cost curve.

2.1.1. *The gains to full cooperation.* Denote i's net benefits by π_i (i.e., $\pi_i = B_i(Q) - C_i(q_i)$), and denote global net benefits by Π (i.e., $\Pi = \Sigma_i \pi_i$). The full cooperative outcome is found by choosing Q to maximize Π. The solution requires setting each country's marginal cost of abatement (MC_i) equal to the global marginal benefit of abatement (MB), and is illustrated in Fig. 1. The full cooperative abatement levels are

$$Q_c = aN/(N + \gamma), \qquad q_c = a/(N + \gamma) \tag{3}$$

where Q_c is global abatement and q_c is an individual country's abatement under the full cooperative outcome, and $\gamma \equiv c/b$.

The noncooperative or open access outcome can be computed by invoking the usual Cournot conjecture that every country chooses its abatement level

be sustainable by a self-enforcing IEA. Carraro and Siniscalco (1991) consider a model in which signatories to an IEA may choose side payments to induce nonsignatories to accede to the IEA, but for signatories to want to make such side payments, it is essential that these countries commit to certain abatement levels, and such a commitment is inconsistent with the notion of self-enforcement. Mäler (1989) and Hoel (1990) analyze cooperative solutions to the control of persistent pollutants, but which are not subject to a self-enforcement constraint. Finally, Heal (1992) considers the case where each country's abatement costs depend on the abatement actions of other countries, as well as its own, but evaluates only the Nash noncooperative equilibrium.

[4] The problem specified here does not impose an upper bound on each country's abatement effort. Clearly, emissions must be at least as great as a units, since the marginal benefit of abating a units is zero. To solve the problem we can assume that emissions exceed a by a sufficient margin so that the maximal abatement constraint is never binding.

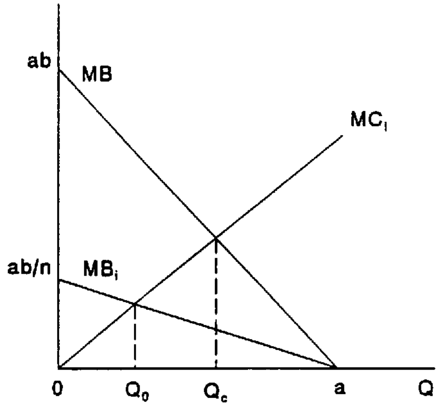

FIG. 1. The non-cooperative and full-cooperative outcomes

on the belief that the abatement levels of all other countries are given. That is, each country chooses q_i to maximize π_i, taking all $q_j, j \neq i$, as fixed. The solution requires setting each country's own marginal benefit of abatement (denoted MB_i in Fig. 1) equal to its own marginal cost of abatement. Denoting Q_o global abatement and q_o any individual country's abatement under the noncooperative or open access outcome, we find

$$Q_o = a/(1 + \gamma), \qquad q_o = a/N(1 + \gamma) \tag{4}$$

We see immediately that $Q_c > Q_o$.[5] Each country is better off at the full cooperative solution, but none has an incentive to choose q_c unilaterally.

It is clear from Fig. 1 that the difference between global net benefits under the full cooperative and non-cooperative outcomes depends on the slopes of the marginal abatement cost and benefit curves. Let Π_c and Π_o denote global net benefits under full cooperation and open access, respectively. Then we have:

Lemma 1 Let $\theta(N) = [\sqrt{(N^2 + 8N)} - N]/4$. Then (i) $\Pi_c - \Pi_o$ increases in b when $\gamma > \theta(N)$ and decreases in b when $\gamma < \theta(N)$ for c given; (ii) $\Pi_c - \Pi_o$ increases in c for γ fixed; and (iii) θ is monotonic and approaches a value of 1 as N becomes very large.

Proof See Appendix.

Lemma 1 tells us that the gains to cooperation ($\Pi_c - \Pi_o$) are larger the closer γ is to $\theta(N)$ and the larger is c. For global environmental problems, N exceeds 100 and $\theta \approx 1$. When c is small and b is large, unilateral abatement is substantial, and the gains to cooperation are relatively small. When c is large and b is small, countries will not abate their emissions by much, even if cooperation is full. When $c \approx b$ and both parameter values are small, $Q_c - Q_o$ will be large, but the gains to cooperation will be small. When $c \approx b$ and both

[5] This is a well known result. See, for example, chapter 2 of Dasgupta (1982).

parameters are large, both $Q_c - Q_o$ and $\Pi_c - \Pi_o$ will be large. Clearly, IEAs can achieve the most when these last conditions hold.

2.1.2. *The self-enforcing IEA.* Suppose that a number of countries negotiate an IEA. Clearly, the number of signatories, the terms of the IEA, and the abatement levels of nonsignatories must be determined jointly. It seems reasonable to suppose that nonsignatories will choose their abatement levels taking as given the abatement decisions of all other countries. Let α denote the proportion of countries that sign the IEA. There are then αN signatories and $(1 - \alpha)N$ nonsignatories. Let Q_n denote the abatement of all nonsignatories, and let q_n denote the abatement of any individual nonsignatory. Since all nonsignatories are identical, $Q_n = (1 - \alpha)Nq_n$. The reaction functions of nonsignatories are given by

$$Q_n(\alpha, Q_s) = (1 - \alpha)(a - Q_s)/(\gamma + 1 - \alpha) \tag{5}$$

Signatories are assumed to choose Q_s to maximize their collective net benefits, with q_s being identical for all signatory countries, subject to (5).[6] The solution is

$$Q_s^*(\alpha) = a\alpha^2 N\gamma/[(\gamma + 1 - \alpha)^2 + \alpha^2 N\gamma] \tag{6}$$

Substituting (6) into (5) yields

$$Q_n^*(\alpha) = a(1 - \alpha)(\gamma + 1 - \alpha)/[(\gamma + 1 - \alpha)^2 + \alpha^2 N\gamma] \tag{7}$$

It is easily verified that the full cooperative ($\alpha = 1$) and noncooperative ($\alpha = 0$) solutions, given by eqs (3) and (4) respectively, are special cases of (6) and (7).

The remaining problem is to determine $\alpha^* N$, or the number of signatories to the self-enforcing IEA. In determining α^*, it will prove helpful to invoke a concept of stability developed for the analysis of cartels (d'Aspremont *et al.* 1983). Let π_s be the net benefit of every signatory and π_n that of every nonsignatory.

Definition An IEA consisting of αN signatories is self-enforcing if

$$\pi_n(\alpha - 1/N) \leqslant \pi_s(\alpha) \quad \text{and} \quad \pi_n(\alpha) \geqslant \pi_s(\alpha + 1/N) \tag{8}$$

If (8) holds, no signatory will want to withdraw unilaterally from the agreement; such a withdrawal would reduce the signatory's abatement level, and hence its costs, but its defection from the IEA would weaken the agreement, and the resulting loss in benefits would more than offset the reduction in costs gained by the withdrawal. Similarly, no nonsignatory, acting alone, would want

[6] The assumption that the q_s are identical is easily justified. If the problem of signatories agreeing on individual country obligations is modelled as a cooperative game, then it can be shown that if all signatories are identical, the Nash bargaining solution will require that each country undertake the same level of abatement.

TABLE 1
Stability analysis for hypothetical example

α	q_s	q_n	π_s	π_n	Q	Π
0.0	—	8.0	—	472.0	80.0	4,720.0
0.1	1.9	8.5	476.8	468.1	78.7	4,690.0
0.2	4.2	8.7	474.0	466.6	78.2	4,681.2
0.3	6.7	8.4	472.3	468.9	78.9	4,699.4
0.4*	8.9*	7.6*	472.2*	474.9*	81.1*	4,738.1*
0.5	10.5	6.3	473.7	482.5	84.2	4,781.2
0.6	11.3	4.9	476.4	489.4	87.7	4,816.0
0.7	11.5	3.6	479.5	494.3	91.0	4,839.8
0.8	11.1	2.5	482.7	497.3	93.8	4,855.9
0.9	10.5	1.6	485.4	498.8	95.9	4,867.8
1.0	9.8	—	487.8	—	97.6	4,878.0

to accede to the IEA; although the recruit's benefits would rise, its abatement costs would rise even more.[7]

The concept is best understood by considering an example. Let $N = 10$, $a = 100$, $b = 1$, and $c = 0.25$; and denote global net benefits by $\Pi(\alpha)$; i.e. $\Pi(\alpha) = \alpha N \pi_s + (1 - \alpha) N \pi_n$. Table 1 shows the net benefit and abatement levels corresponding to each possible α, and Fig. 2 illustrates the example. To solve for α^*, start at $\alpha = 0$ and compare $\pi_n(0)$ with $\pi_s(0.1)$. Clearly, $476.8 > 472.0$, and hence it will play a nonsignatory to accede to the agreement. Now compare $\pi_n(0.1)$ with $\pi_s(0.2)$. Again, $474.0 > 468.1$, and it will pay a nonsignatory to cooperate. Continuing in this way one finds that nonsignatories always do better by acceding when $\alpha < 0.4$. Likewise, starting at $\alpha = 1.0$, one finds that signatories always do better by withdrawing from the agreement whenever $\alpha > 0.4$. Hence, an IEA consisting of four signatories is the only self-enforcing IEA for this problem.[8]

[7] The concept of self-enforcement defined by (8) was first employed by Barrett (1989), but has also been used by Carraro and Siniscalco (1991) and Hoel (1992). The essential feature of (8) is that it does not consider whether coalitions of two or more countries would want to accede to or withdraw from the IEA. Heal (1992) defines a related concept which he calls the minimum critical coalition. This is a group of countries S, which (i) adopts an abatement policy yielding each member of S positive net benefits, and (ii) does not contain a proper subset of countries which yields its members positive net benefits.

[8] Notice that $\pi_s(\alpha^* + 1/N) - \pi_s(\alpha^*) > \pi_n(\alpha^*) - \pi_s(\alpha^* + 1/N)$ in this example. This implies that signatories to a self-enforcing IEA have an incentive to offer a side payment to nonsignatories to join the agreement. The difficulty is that the agreement to offer a side payment is not self-enforcing. While the signatories to the self-enforcing IEA have a collective incentive to offer a side payment to a nonsignatory, each of the original signatories would do better still by withdrawing from the agreement since $\pi_n(\alpha^*) > \pi_s(\alpha^* + 1/N)$ (by definition).

Carraro and Siniscalco (1991) propose a way out of this dilemma. They show that if the signatories to the self-enforcing IEA can commit themselves to cooperate whatever the magnitude of α, then the total number of signatories can be increased (although the full cooperative outcome may still be unattainable). The snag in this proposal, as noted in footnote 2, is that the commitment is not credible; the expanded agreement is not self-enforcing. If commitment could be made credible, then the full cooperative outcome would be easy to sustain: every country would simply commit to choose the full cooperative abatement level q_c.

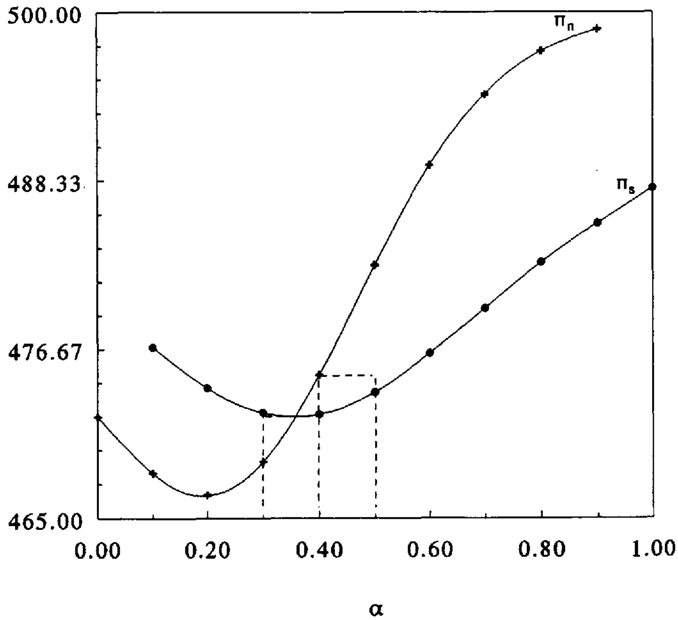

FIG. 2. Graphical analysis of self-enforcing IEA

Unfortunately, a full characterization of the solution cannot be obtained analytically for this functional specification. However, simulations reveal a very simple and compelling relationship between γ and α^*, and between γ and the ability of an IEA to close the gap between the full cooperative and non-cooperative outcomes.

2.1.3. *Simulations.* Table 2 presents solutions to $\alpha^* N$ for various values of b and c when $N = 100$.[9] It is clear from the simulations that the number of signatories to the self-enforcing IEA increases as γ decreases. It is encouraging to find that a self-enforcing IEA can consist of many signatories, if only when γ is small. However, Lemma 1 demonstrated that if γ is small, then the difference between global net benefits under the full cooperative and noncooperative outcomes will also be small. Hence, IEAs signed by many countries do not increase global net benefits by very much compared with the non-cooperative outcome. Indeed, Tables 3 and 4, which give the total abatement and net benefit figures corresponding to a selection of nine cells from Table 2, confirm that the self-enforcing IEA increases global abatement and global net benefits by very little when γ is small. These results suggest that IEAs signed by a large number of countries have very little effect.

Table 2 shows that when γ is large, few countries will sign the self-enforcing

[9] The value taken by the parameter a does not affect the results in Table 2.

TABLE 2
Number of signatories out of 100 for various values of c and b

	c										
	0.01	*0.05*	*0.10*	*0.25*	*0.50*	*1.00*	*2.00*	*5.00*	*10.00*	*20.00*	*100.00*
0.01	3	3	3	2	2	2	2	2	2	2	2
0.05	6	3	3	3	3	2	2	2	2	2	2
0.10	11	4	3	3	3	3	2	2	2	2	2
0.25	21	6	4	3	3	3	3	2	2	2	2
0.50	35	11	6	4	3	3	3	3	2	2	2
1.00	51	18	11	5	4	3	3	3	3	2	2
2.00	68	30	18	9	5	4	3	3	3	3	2
5.00	84	51	35	18	11	6	4	3	3	3	2
10.00	92	68	51	30	18	11	6	4	3	3	3
20.00	96	81	68	46	30	18	11	5	4	3	3
100.00	100	96	92	81	68	51	35	18	11	6	3

TABLE 3
*Global abatement under the noncooperative, IEA, and full co-operative outcomes**

		c		
		0.01	*1*	*100*
	0.01	500.0	9.9	0.1
		503.9	10.1	0.1
		990.1	500.0	9.9
b	1	990.1	500.0	90.9
		990.2	503.9	95.2
		999.9	990.1	909.1
	100	999.9	990.1	500.0
		1,000.0	990.2	503.9
		1,000.0	999.9	990.1

* Assumes $N = 100$ and $a = 1{,}000$. The top number in each cell is Q_o, the middle $Q(\alpha^*)$, and the bottom Q_c.

IEA. Since IEAs typically specify a minimum number of signatories,[10] this means that an agreement is unlikely to be reached when γ is large. But Tables 3 and 4 indicate that the difference in global abatement and global net benefits under the full cooperative and noncooperative outcomes is trivial in absolute terms when γ is large. Hence, failure to reach agreement is of no great consequence when γ is large.

[10] For example, the Framework Convention on Climate Change will not come into force until 50 countries have ratified the agreement. See Black *et al.* (1993) for an analysis showing why countries might want to specify a minimum number of signatories.

TABLE 4
*Global net benefits under the noncooperative, IEA, and full cooperative outcomes**

		c		
		0.01	*1*	*100*
	0.01	3,738	98	1
		3,757	100	1
		4,950	2,500	50
b	1	499,902	373,750	9,803
		499,903	375,659	9,990
		499,950	495,050	250,000
	100	49,999,949	49,990,197	37,375,000
		49,999,950	49,990,293	37,565,851
		49,999,950	49,995,000	49,504,950

* Assumes $N = 100$ and $a = 1{,}000$. The top number in each cell is Π_0, the middle $\Pi(\alpha^*)$, and the bottom Π_c.

The situation is very different where c and b are both large. As Lemma 1 showed, it is here that the difference between global net benefits under the noncooperative and full cooperative outcomes is large. Table 4 confirms this. But from Table 2 we see that few countries are likely to sign an IEA for such a pollutant.

Proposition 1 For global environmental problems characterized by eqs (1) and (2), the self-enforcing IEA will be signed by a lot of countries, each undertaking substantial abatement, when γ is small, but under these circumstances the IEA increases global net benefits by very little compared with the noncooperative outcome. Cooperation would increase net benefits substantially when c and b are both large, but under these circumstances the self-enforcing IEA cannot sustain a large number of signatories.

Table 5 shows the magnitude of free riding for these simulations. Signatories abate more than they would in the absence of an IEA, while nonsignatories abate less. When $\gamma = 1$, only three countries cooperate. Every signatory abates significantly more than under the noncooperative outcome, and every nonsignatory just a little less. But because there are so many more nonsignatories in this case, the free-rider effect is huge in total and the IEA has little overall impact. When $\gamma = 0.01$, there are 51 signatories. In this case the free-rider effect is small—but so too is the additional abatement of signatories. Again, the self-enforcing IEA has little overall impact. Finally, Table 6 shows the net benefit levels corresponding to the abatement levels in Table 5. All countries receive higher net benefits when there exists a self-enforcing IEA compared with the noncooperative outcome, but nonsignatories take a free ride on signatories.

It must be emphasized that Proposition 1 applies to environmental problems

TABLE 5

*Country abatement under the noncooperative, IEA, and full cooperative outcomes**

		c: 0.01	1	100
	0.01	5.00	0.10	—
		7.55	0.20	—
		4.96	0.10	—
		9.90	5.00	0.10
b	1	9.90	5.00	0.10
		10.00	7.55	0.20
		9.80	4.96	0.10
		10.00	9.90	5.00
	100	10.00	9.90	5.00
		10.00	10.00	7.55
		—	9.80	4.96
		10.00	10.00	9.90

* Assumes $N = 100$ and $a = 1{,}000$. The top number in each cell is q_o, the second $q_s(\alpha^*)$, the third $q_n(\alpha^*)$, and the bottom q_c.

TABLE 6

*Country net benefits under the noncooperative, IEA, and full cooperative outcomes**

		c: 0.01	1	100
	0.01	37.38	0.98	0.01
		37.41	0.99	0.01
		37.57	1·00	0.01
		49.50	25.00	0.50
b	1	4,999.02	3,737.50	98.03
		4,999.02	3,740.84	98.50
		4,999.04	3,757.07	99.93
		4,999.50	4,950.50	2,500.00
	100	499,999.50	499,901.97	373,750.00
		499,999.50	499,901.98	374,083.96
		—	499,903.92	375,707.21
		499,999.50	499,950.00	495,049.50

* Assumes $N = 100$ and $a = 1{,}000$. The top number of each cell is π_o, the second $\pi_s(\alpha)$, the third $\pi_n(\alpha^*)$, and the bottom π_c.

for which N is large. As N becomes small, the effect of withdrawal or accession by any country on the abatement level of the other cooperators increases, and hence free riding can be punished more effectively. To take an example, suppose $N = 5$, $a = 100$ and $c = b = 1$. Then the self-enforcing IEA consists of three countries with $Q_o = 50$, $Q(\alpha^*) = 63$ and $Q_c = 83$; and with $\Pi_0 = 3{,}500$, $\Pi(\alpha^*) = 3{,}869$, and $\Pi_c = 4{,}162$. In this case, the self-enforcing IEA is able to close the gap between the noncooperative outcomes quite substantially.

2.2. *Constant marginal benefits and linear marginal costs*

Suppose that maginal benefits are constant; i.e.

$$B_i(Q) = \omega Q \tag{9}$$

Then, if the abatement cost function is given by e.g. (2), the self-enforcing IEA can be solved analytically.

Proposition 2 For the model specification given by eqs (2) and (9), the self-enforcing IEA consists of 2 countries when $N = 2$ and 3 countries when $N \geqslant 3$.

Proof See Appendix.

2.3. *Constant marginal benefits and logarithmic marginal costs*

Suppose now that abatement benefits continue to be given by eq. (9) but that total abatement costs for country i are now given by

$$C_i(q_i) = x\sigma[(1 - q_i/x)\ln(1 - q_i/x) + q_i/x] \tag{10}$$

The marginal abatement cost function is then

$$MC_i(q_i) = -\sigma \ln(1 - q_i/x) \tag{11}$$

If the parameter x is taken to equal emissions in the absence of any abatement effort, q_i/x is percentage abatement. In that case, the marginal abatement cost function (11) is the same form as was estimated by Nordhaus (1990) for US carbon dioxide emissions.[11] Since Nordhaus's model also assumes constant marginal benefits, eqs (9) and (10) describe Nordhaus's model of the greenhouse effect. However, Nordhaus does not disaggregate to the individual country level. Here it is assumed that all countries are identical and have the same cost and benefit functions.[12] The solution to this problem, like the previous one, can be derived analytically.

[11] The cost function specification is described in footnote 10 of Nordhaus (1990), p. 19.

[12] Since Nordhaus (1990) calculates a global marginal benefit, the parameter ω in eq. (9) can be taken to equal each country's (identical) share of the global marginal benefit.

Lemma 2 For the model specification given by eqs (9) and (10),

(i) $\pi_n(0) = \pi_s(1/N) = \pi_n(1/N)$
(ii) $\pi_s(2/\mathrm{N}) > \pi_n(1/N)$
(iii) $\pi_n(\alpha) > \pi_s(\alpha + 1/N) \forall \alpha \geqslant 2/N$

Proof See Appendix.

Lemma 2 implies:

Proposition 3 For the model specification given by eqs (9) and (10), the self-enforcing IEA consists of 2 countries for $N \geqslant 2$.

2.4. *Linear marginal benefits and constant marginal costs*

Suppose now that country i's benefit function is given by (1) and its cost function by

$$C_i(q_i) = dq_i \tag{12}$$

Then the self-enforcing IEA may be solved easily.

Proposition 4 For the model specification given by eqs (1) and (12), a self-enforcing IEA consisting of 2 or more countries does not exist.

Proof See Appendix.

With this specification, any increase in Q_s is entirely offset by a reduction in Q_n. There is therefore nothing to be gained by cooperating.

Clearly, the results obtained using the specification given by eqs (1) and (2) are not generally robust. According to the specifications examined in this section, the number of signatories may be limited to just a few whatever parameter values characterise the problem being investigated.[13] However, the results of all specifications considered here are consistent in one respect: when N is large, an agreement achieving substantial emission reductions over and above the non-cooperative levels is not self-enforcing.

3. IEAs as repeated games

The preceding model is unable to sustain the full cooperative outcome in general because its punishments, though credible, are small. It is well known that the full cooperative outcome of the one-shot game can be sustained as a sub-game perfect equilibrium of the infinitely repeated game if the rate of discount is sufficiently small. However, this section will show that repetition may not allow IEAs to sustain the full cooperative outcome, even for arbitrarily small discount rates, because the punishments that sustain cooperation in infinitely repeated games may be vulnerable to renegotiation, and such punishments could not be supported by a self-enforcing IEA.

[13] d'Aspremont *et al.* (1983), Hoel (1992), and Carraro and Siniscalco (1991) also solve for models yielding two or three cooperators.

Let π_i now denote country i's average payoff in the infinitely repeated game, and consider the following functional specification

$$\pi_i = b(aQ - Q^2/2)/N - dq_i - (ab - dN)^2/2bN \tag{13}$$

where the constant term ensures that each country is guaranteed a payoff of at least zero. Proposition 4 tells us that specification (13) will not allow a stable IEA to improve on the noncooperative outcome. We shall see that this is not necessarily so when the game is infinitely repeated.

The full cooperative outcome yields a global payoff of $d(N-1) \times (2ab - d(N+1))/2b$. Hence, the set of feasible, individually rational payoffs is

$$V^* = \{v \mid v_i \geqslant 0, \sum v_i \leqslant d(N-1)(2ab - d(N+1))/2b\}$$

Following Farrell and Maskin (1989), a payoff vector v may be said to be (weakly) renegotiation proof (for discount rates near zero) if abatement levels q_j^j, q_k^j can be chosen to punish j for cheating, such that

$$\max_{q_j} \{(b/N)[a(q_j + (N-1)q_k^j) - (q_j + (N-1)q_k^j)^2/2] - dq_j - (ab - dN)^2/2bN\} \leqslant v_j \tag{14}$$

$$(b/N)[a(q_j^j + (N-1)q_k^j) - (q_j^j + (N-1)q_k^j)^2/2] - dq_k^j - (ab - dN)^2/2bN \geqslant v_k \tag{15}$$

Condition (14) ensures that any one country cannot gain by cheating on the agreement.[14] Condition (15) ensures that countries have no incentive to renegotiate. For this problem, it also makes sense to require that the payoff vector v be Pareto efficient in the punishment phase.[15] If we further assume that j chooses for q_j^j its minimax abatement level, $q_j^j = (ab - dN)/b$, then $q_k^j = \mathrm{d}/b$. Substituting, we require for every country i

$$\mathrm{d}[2abN - \mathrm{d}(N+1)^2]/2bN \geqslant v_i \geqslant \mathrm{d}^2(N-1)/b \tag{16}$$

Under the full cooperative outcome, every country receives a payoff of $\mathrm{d}(N-1)(2ab - \mathrm{d}(N+1))/2bN$. Substituting this value into (16), we find:

Proposition 5 For the specification given by eq. (13), the full cooperative outcome can be sustained as a self-enforcing IEA for a common discount rate close to zero if the number of countries N does not exceed

$$\bar{N} = \min(ab/d - 1, 2ab/3d - \tfrac{1}{3})$$

It is easy to see that $\bar{N}$ is increasing in a and b, and decreasing in d. $\bar{N}$ is calculated for various values of b and d in Table 7.[16] The number of countries that can sustain the full cooperative outcome varies substantially, depending

[14] As elsewhere in the paper, I am assuming that countries do not form coalitions.
[15] Farrell and Maskin (1989) refer to such an equilibrium as being strongly perfect.
[16] The table considers parameter values which yield only interior solutions.

TABLE 7
*Illustration of maximum number of countries that can sustain the full cooperative outcome**

a	b	d	$\bar{N}$	$(\Pi_c - \Pi_o)/(N-1)^2$
100	1	0.99	67	0.49
100	1	0.75	88	0.28
100	1	0.65	102	0.21
100	2	1.98	67	0.98
100	2	1.50	88	0.56
100	2	1.30	102	0.42
1,000	1	9.9	67	49
1,000	1	7.5	88	28
1,000	1	6.5	102	21
1,000	2	19.8	67	98
1,000	2	15.0	88	56
1,000	2	13.0	102	42

* Assumes $N = 100$, to ensure interior solutions.

on the parameter values, and is higher the larger is b and the smaller is d. It can be shown that $\Pi_c - \Pi_o = d^2(N-1)^2/2b$ for specification (13). The last column in Table 7 indicates that when $\bar{N}$ is large, the difference $\Pi_c - \Pi_o$ is small, and that when $\bar{N}$ is small, the difference $\Pi_c - \Pi_o$ may be very large. These results are reminiscent of those obtained in Section 2.1. Clearly, even infinite repetition may not allow the IEA to sustain the full cooperative outcome when IEAs must be self-enforcing.[17]

4. Conclusions

This paper has shown that self-enforcing international environmental agreements (IEAs), which establish rules for managing shared environmental resources, may not be able to improve substantially upon the noncooperative outcome. Two different modelling approaches support this conclusion. The model of a self-enforcing IEA, which solves jointly for the number of signatories, the terms of the IEA, and the actions of nonsignatories, shows that, depending on the functional specification, a self-enforcing IEA may not exist, or it may not be able to sustain more than two or three signatory countries, in which case the IEA cannot increase global net benefits substantially when the number of countries that share the resource is very large. For one functional specification, the self-enforcing IEA can sustain a large number of signatories—up to as many as share the resource—but only when the difference in net benefits between the noncooperative and full cooperative outcomes is very small. When this difference is very large, the self-enforcing IEA can sustain only a small number of signatories, and hence can only marginally improve upon the noncooperative

[17] Barrett (1994) carries out a similar analysis for the conservation of biological diversity, a global public good, but using a different functional form. Though this model can only be evaluted using simulation, the analysis arrives at a similar conclusion.

outcome when the environmental resource is shared globally. The other model, which takes the IEA to be an equilibrium to an infinitely repeated game, but one which is renegotiation-proof, shows that the full cooperative outcome can be sustained by a large number of countries, but only when the difference in global net benefits between the noncooperative and full cooperative outcomes is small. When this difference is large, the full cooperative outcome can be sustained by only a few countries, or possibly by none at all.

Despite the arguably harsh assumptions that underpin these models, it is tempting to consider an actual IEA in the light of this analysis. The Montreal Protocol on Substances that Deplete the Ozone Layer, originally negotiated in 1987, has now been signed by about 80 countries, and in its original form committed each signatory to cut its production and consumption of the so-called hard CFCs by 50%.[18] This agreement is understandably considered to be a substantial achievement of international cooperation.[19] However, when seen in the context of this paper, an analysis of ozone depletion by the US Environmental Protection Agency (1988) suggests that the Montreal Protocol may not have increased global net benefits substantially compared with the noncooperative outcome. Assuming 'substantial levels of participation by other nations', the EPA estimates that the US would receive a benefit, primarily in the form of a reduction in the incidence of skin cancers, of $3,575bn, and incur a cost of $21bn, if it acceded to the treaty. If the US cut its production and consumption of CFCs by 50% unilaterally, the cost to the US would remain unchanged, but the benefit to the US would fall to only $1,373bn. While one would ideally like to know the marginal costs and benefits, these total benefit and cost figures strongly suggest that the US had an incentive to carry out the terms of the Montreal Protocol unilaterally. There is no compelling reason to believe that these benefit and cost estimates would be wildly different for most other countries in relative terms, and so, if the EPA's figures are broadly correct and consistent with the beliefs of the Montreal Protocol negotiators, it seems that this agreement may not have achieved as much as is commonly claimed. That conclusion is consistent with the analysis presented in this paper.

[18] The agreement has since been amended, and strengthened in a number of respects.
[19] But see the analysis by Bohm (1990).

REFERENCES

D'ASPREMONT, C. A., JACQUEMIN,, J., GABSZEWEIZ, J., and WEYMARK, J. A. (1983). 'On the Stability of Collusive Price Leadership', *Canadian Journal of Economics*, **16**, 17–25.

BARRETT, S. (1989). 'On the Nature and Significance of International Environmental Agreements', mimeo, London Business School.

BARRETT, S. (1992). 'International Environmental Agreements as Games', in R. Pethig (ed.), *Conflict and Cooperation in Managing Environmental Resources*, Springer-Verlag, Berlin.

BARRETT, S. (1994). 'The Biodiversity Supergame', *Environmental and Resource Economics*, **4**, 111–22.

Black, J., Levi, M. D., and de Meza, D. (1992). 'Creating a Good Atmosphere: Minimum Participation for Tackling the "Greenhouse Effect"', *Economica*, **60**, 281–93.

Bohm, P. (1990). 'Efficiency Aspects of Imperfect Treaties on Global Public Bads: Lessons from the Montreal Protocol', mimeo, Department of Economics, University of Stockholm.

Carraro, C. and Siniscalco, D. (1991). 'Strategies for the International Protection of the Environment', mimeo, Fondazione E.N.I. 'Enrico Mattei'.

Dasgupta, P. (1982). *Control of Resources*,, Harvard University Press, Cambridge, MA.

Farrell, J. and Maskin, E. (1989). 'Renegotiation in Repeated Games', *Games and Economic Behaviour*, **1**, 327–60.

Heal, G. (1992). 'International Negotiations on Emission Control', *Structural Change and Economic Dynamics*, forthcoming.

Hoel, M. (1990). 'CO_2 and the Greenhouse Effect: A Game Theoretic Exploration', mimeo, Department of Economics, University of Oslo.

Hoel, M. (1992). 'International Environment Conventions: The Case of Uniform Reductions of Emissions', *Environmental and Resource Economics*, **2**, 141–59.

Mäler, K.-G. (1989). 'The Acid Rain Game, 2', mimeo, Department of Economics, Stockholm School of Economics.

Nordhaus, W. D. (1990). 'To Slow or Not to Slow: The Economics of the Greenhouse Effect', mimeo, Department of Economics, Yale University.

US Environmental Protection Agency (1988). *Regulatory Impact Analysis: Protection of Stratospheric Ozone*, Washington, D.C.

APPENDIX

Proof of Lemma 1 (i) The potential gains to cooperation are

$$\Pi_c - \Pi_o = a^2 c\gamma(N-1)^2/2N(N+\gamma)(\gamma+1)^2 \tag{A.1}$$

Differentiating with respect to γ while holding c constant gives

$$\partial(\Pi_c - \Pi_o)/\partial\gamma \gtreqless 0 \quad \text{if} \quad (2\gamma^2 + \gamma N - N) \lesseqgtr 0 \tag{A.2}$$

This quadratic equation has one positive solution, $\gamma = \theta(N)$.

(ii) From (A.1) we see immediately that $\partial(\Pi_c - \Pi_o)/\partial c\,|_{\gamma=\bar{\gamma}} > 0$.

(iii) Notice that $\theta(N) = [\sqrt{(N^2 + 8N)} - N]/4$ can be rewritten

$$(1-\theta)/(2\theta^2) = 1/N \tag{A.3}$$

As $N \to \infty$, $1/N \to 0$, and hence $\theta \to 1$. □

Proof of Proposition 2 The solutions to the appropriate optimization problems are

$$q_n = \omega/c, \qquad q_s = \omega\alpha N/c$$

Substituting, we obtain

$$\pi_n(\alpha) = \omega^2 N(1 - \alpha + \alpha^2 N - 1/2N)/c \tag{A.4a}$$

$$\pi_s(\alpha) = \omega^2 N(1 - \alpha + \alpha^2 N/2)/c \tag{A.4b}$$

$$\pi_n(\alpha - 1/N) = \omega^2 N(1 - 3\alpha + 3/2N + \alpha^2 N)/c \tag{A.46c}$$

$$\pi_s(\alpha + 1/N) = \omega^2 N(1 - 1/2N + \alpha^2 N/2)/c \tag{A.4d}$$

Substituting (A.4a) and (A.4d) into (8) yields $\alpha \geqslant 2/N$. Substituting (A.4b) and (A.4c) into (8) and solving for the resulting quadratic equation yields $3/N \geqslant \alpha \geqslant 1/N$. Since we require that both of these conditions hold, the self-enforcing IEA consists of $3 \geqslant \alpha N \geqslant 2$ countries. Clearly, if $N = 2$, then both countries will sign the IEA. Upon substitution, it is easy to show that $\pi_s(3/N) = \pi_n(2/N)$. If it is assumed that an additional country will joint the IEA when $\pi_s(\alpha + 1/N) = \pi_n(\alpha)$, then the self-enforcing IEA will consist of three countries whenever $N \geqslant 3$. □

Proof of Lemma 2 (i) The solutions to the appropriate optimization problems are

$$q_n = a(1 - e^{-\omega/\sigma}), \qquad q_s = a(1 - e^{-\omega\alpha N/\sigma})$$

Substituting , we obtain

$$\pi_n(\alpha) = [\omega N - \sigma - \alpha\omega N\, e^{-\omega\alpha N/\sigma} + e^{-\omega/\sigma}(\alpha\omega N - \omega N + \sigma + \omega)] \tag{A.5a}$$

$$\pi_s(\alpha) = a[\omega N - \sigma - \omega N(1 - \alpha)\, e^{-\omega/\sigma} + \sigma\, e^{-\omega\alpha N/\sigma}] \tag{A.5b}$$

Part (i) of Lemma 2 can be proved by substituting the appropriate values for α into eqs (A.5a) and (A.5b).

(ii) Setting $\alpha = 2/N$ in (A.5b) and $\alpha = 1/N$ in (A.5a), and substituting, the inequality in (ii) implies

$$\omega/\sigma > 1 - e^{-\omega/\sigma} \tag{A.6}$$

Let $z = \omega/\sigma$. When $z = 0$, both sides of inequality (A.6) equal zero (and hence the inequality does not hold at $z = 0$). The LHS of inequality (A.6) rises with z at rate 1, while the RHS rises at rate e^{-z}. The LHS of (A.6) therefore rises more rapidly with z than the RHS. Since both sides of (A.6) equal zero when $z = 0$, the inquality holds $\forall z > 0$.

(iii) Upon substitution, the inequality in (ii) implies

$$-\alpha\omega N\, e^{-\omega\alpha N/\sigma} + \sigma\, e^{-\omega/\sigma} > \sigma\, e^{-\omega/\sigma}\, e^{-\omega\alpha N/\sigma} \tag{A.7}$$

$\forall \alpha \geqslant 2/N$.

First determine that (A.7) holds for $\alpha = 2/N$. Upon substituting, one obtains

$$1 - e^{-2\omega/\sigma} > 2w\, e^{-\omega/\sigma}/\sigma \tag{A.8}$$

Again, let $z = \omega/\sigma$. When $z = 0$, both sides of (A.8) equal zero. The derivative of the LHS is $2\,e^{-2z}$, while that of the RHS is $2\,e^{-z}(1 - x)$. When $\alpha = 2/N$, the derivative of the LHS of (A.8) exceeds that of the RHS provided $z > 1 - e^{-z}$. But this condition is the same as (A.6), which has already been shown to hold. Hence (iii) holds for $\alpha = 2/N$. Now take the derivative of both sides of (A.7) with respect to α. The derivative of the LHS of (A.7) exceeds the derivative of the RHS if

$$\alpha\omega N/\sigma > 1 - e^{-\omega/\sigma} \tag{A.9}$$

We know (A.9) holds for $\alpha = 1/N$. Hence, it must hold for $\alpha > 1/N$. Since (A.7) holds for $\alpha = 2/N$, it must therefore also hold for all $\alpha \geqslant 2N$. □

Proof of Proposition 4 Each non-signatory will equate its own marginal benefit with its own constant marginal cost, yielding the non-signatory reaction function

$$Q_n = a - dN/b - Q_s \tag{A.10}$$

Signatories will equate their combined marginal benefit with their individual marginal costs, yielding the reaction function

$$Q_s = a - d/\alpha b - Q_n \tag{A.11}$$

Solving (A.10) and (A.11) yields $\alpha N = 1$. From (A.10) we see that $Q_n + Q_s = a - dN/b$, which is the same abatement level achieved when $Q_s = 0$. □