University of Massachusetts Amherst
Department of Resource Economics
Working Paper No. 2006-6
http://www.umass.edu/resec/workingpapers

Enforcing 'Self-Enforcing' International Environmental Agreements

David M. McEvoy [1] and John K. Stranlund [2]

Abstract:

Theoretical analyses of international environmental agreements (IEAs) have typically employed the concept of self-enforcing agreements to predict the number of parties to such an agreement. The term self-enforcing, however, is a bit misleading. The concept refers to the stability of cooperative agreements, not to enforcing these agreements once they are in place. Most analyses of IEAs simply ignore the issue of enforcing compliance by parties to the terms of an agreement. In this paper we analyze an IEA game in which parties to an agreement finance an independent enforcement body with the power to monitor the parties' compliance to the terms of the IEA and impose penalties in cases of noncompliance. This approach is broadly consistent with the enforcement mechanism of the Kyoto Protocol under the Marrakesh Accords. We find that costly enforcement limits the circumstances under which international cooperation to protect the environment is worthwhile, but when IEAs do form they will involve greater participation than IEAs that do not require costly enforcement. Consequently, costly enforcement of IEAs is associated with higher international environmental quality. Moreover, under certain conditions, aggregate welfare is higher when IEAs require costly enforcement. These conclusions are accentuated when monitoring for compliance to IEAs is inaccurate.

Keywords: International environmental agreements, self-enforcing agreements, compliance, enforcement.
JEL Classification: Q5, H41, C72, F53

---

[1]David M. McEvoy, Department of Resource Economics
University of Massachusetts, Stockbridge Hall, 80 Campus Center Way, Amherst, MA 01003
E: dmcevoy@resecon.umass.edu P: 617-671-8147 F: 413-545-5853

[2]John K Stranlund, Department of Resource Economics
University of Massachusetts, Stockbridge Hall, 80 Campus Center Way, Amherst, MA 01003
E: stranlund@resecon.umass.edu P: 413-545-6328 F: 413-545-5853

August 2006

Enforcing 'Self-Enforcing' International Environmental Agreements

DAVID M. McEVOY*
Department of Resource Economics
University of Massachusetts-Amherst

JOHN K. STRANLUND
Department of Resource Economics
University of Massachusetts-Amherst

Acknowledgements: The authors acknowledge funding for this project from the Center for Public Policy and Administration, University of Massachusetts, Amherst, and the Cooperative State Research Extension, Education Service, U. S. Department of Agriculture, Massachusetts Agricultural Experiment Station, and the Department of Resource Economics under Project No. MAS00871. James Murphy provided valuable comments in the early stages of this work.

* Correspondence to: David M. McEvoy, 80 Campus Center Way, 404 Stockbridge Hall, Amherst, MA 01003, USA, Phone: 617-671-8147, Fax: 413-545-5853, E-mail: dmcevoy@resecon.umass.edu.

Enforcing 'Self-Enforcing' International Environmental Agreements

Abstract: Theoretical analyses of international environmental agreements (IEAs) have typically employed the concept of *self-enforcing agreements* to predict the number of parties to such an agreement. The term *self-enforcing*, however, is a bit misleading. The concept refers to the stability of cooperative agreements, not to enforcing these agreements once they are in place. Most analyses of IEAs simply ignore the issue of enforcing compliance by parties to the terms of an agreement. In this paper we analyze an IEA game in which parties to an agreement finance an independent enforcement body with the power to monitor the parties' compliance to the terms of the IEA and impose penalties in cases of noncompliance. This approach is broadly consistent with the enforcement mechanism of the Kyoto Protocol under the Marrakesh Accords. We find that costly enforcement limits the circumstances under which international cooperation to protect the environment is worthwhile, but when IEAs do form they will involve greater participation than IEAs that do not require costly enforcement. Consequently, costly enforcement of IEAs is associated with higher international environmental quality. Moreover, under certain conditions, aggregate welfare is higher when IEAs require costly enforcement. These conclusions are accentuated when monitoring for compliance to IEAs is inaccurate.

Keywords: International environmental agreements, self-enforcing agreements, compliance, enforcement.

JEL Codes: Q5, H41, C72, F53

1. Introduction

International environmental agreements (IEAs) made between sovereign nations seeking to manage shared environmental and natural resources are susceptible to two sources of free-riding. First, because participation to IEAs are voluntary, countries can decide at will whether to become a party to an agreement. If an IEA only requires a subset of countries to join before entering into force, as is typically the case, then incentives exist for some countries to stay out of the agreement and free-ride off the provision of the cooperating others. Second, if compliance with the terms of the agreement are not enforced well, parties to an IEA will have an incentive to violate the terms of the agreement and free-ride off those countries that do comply. Previous theoretical analyses of IEAs have usually investigated the impact the first form of free-riding has

on the effectiveness of IEAs, while avoiding altogether the possibility that member countries may not fully comply with their commitments. In this paper, we deal directly with the issue of noncompliance within IEAs by constructing and analyzing a game in which parties to an IEA finance an independent enforcement body with the power to monitor parties' compliance behavior and impose penalties in cases of noncompliance.

There is ample evidence that many existing IEAs experience significant problems with noncompliance. For example, the Convention on International Trade in Endangered Species (CITES) has witnessed hundreds of infractions every year since its inception in 1979 (Swanson and Johnston 1999; Finus 2004). Similarly, customs officers throughout the world frequently intercept flows of CFC products even though these substances have been banned under the Montreal Protocol since 1991 (Finus 2004). As a more recent example, although the performance of the Kyoto Protocol cannot formally be evaluated until the end of the first commitment period (2008-2012), many countries are currently on pace to significantly exceed their quotas (Friends of the Earth 2006).[1]

Not surprisingly, many IEAs include provisions for enforcement. The critical element of these enforcement mechanisms is that they must be negotiated and included as part of the agreement. In addition, as Hovi and Areklett (2004, pg 3) point out, "… enforcement must either be carried out by the parties themselves, or by some institution erected, accepted and empowered by the parties." An example of the first approach is contained in the Montreal Protocol, which enforces compliance by threatening the use of trade sanctions on defecting parties (Benedick

[1] Although it is universally accepted that some level of noncompliance with IEAs exists, there remains debate as to the magnitude and consequences of noncompliance. On one end of the debate, Chayes and Chayes (1991) argue that high levels of compliance are often observed in IEAs without formal enforcement mechanisms and therefore enforcement is not a real concern. On the other end, Downs et al. (1996) argue that high levels of compliance may result because the lack of enforcement motivates parties to form shallow agreements to begin with.

1998; Heister 1997). It is one of an estimated 19 IEAs that uses trade sanctions as a mechanism for motivating compliance (United States International Trade Commission 1991).[2]

An example of an "institution erected, accepted and empowered by the parties" to enforce compliance with an IEA has been implemented within the Kyoto Protocol under the Marrakesh Accords. An independent Compliance Committee, consisting of members elected by the parties, has been formed to oversee monitoring for compliance and applying sanctions in cases of noncompliance. If a party to the Kyoto Protocol is found in violation, the Compliance Committee has the power to impose a penalty in the form of a reduction in next period's greenhouse gas emissions quota (UNFCC 2002). Although it is too early to evaluate the effectiveness of this type of enforcement mechanism, its structure is unique in the sense that parties have agreed to invest in an institution with the power to monitor and sanction noncompliance.

Conceptually, IEAs have often been modeled with the equilibrium concept of a *self-enforcing agreement*. A self-enforcing agreement made between agents, as first proposed by D'Aspremont et al. (1983) and later coined by Barrett (1994) for use in his analysis of international environmental relations, is defined as a single coalition from which no member wishes to withdraw (the coalition is *internally stable*) and no nonmember wishes to join (the coalition is *externally stable*).[3] The term *self-enforcing* is a bit misleading, however, because it does not address the problem of enforcing compliance by parties to a cooperative agreement. In fact, most authors who use this equilibrium concept simply assume that parties to a self-enforcing agreement comply fully with the terms of the agreement (e.g. Barrett 1994, 2003;

[2] Barrett (2003) estimates that a much larger number of IEAs make provisions for trade sanctions, however, many of those agreements use bans on traded items as part of the agreement's goal, not its method of enforcement.
[3] Models of self-enforcing IEAs have included provisions for trade sanctions (Barrett 1997a; 1997b), side payments (Carraro and Siniscalco 1993; Carraro and Botteon 1997 and Hoel and Schneider 1997), issue linkage (Folmer 1993; Carraro and Siniscalco 1997; Botteon and Carraro 1998) and minimum participation requirements (Barrett 1998a; 2003 and Carraro et al. 2003). For a review of many of these extensions see Wagner (2001).

Carraro and Siniscalco 1993, 1998; Hoel 1992). A few studies do include costly enforcement mechanisms within the framework of a repeated IEA (Barrett 1994, 2003; Finus and Rundshagen 1998). These models use reciprocal punishment strategies in which parties to an agreement may punish violators in future rounds by jointly reducing the level of the international public good.

We take a different approach to the study of compliance and enforcement of IEAs that is motivated by the enforcement mechanism of the Kyoto Protocol under the Marrakesh Accords. We maintain the concept of a self-enforcing agreement to determine the equilibrium number of parties to an IEA. However, to counteract the incentive that parties have to violate the terms of an IEA, they finance and empower an independent enforcement agency with monitoring and sanctioning capabilities.

Our efforts yield a number of new results. First, the range of international environmental problems within which a self-enforcing IEA can form is smaller when enforcing cooperation is necessary and costly. This follows because enforcement increases the cost of cooperation, which in turn limits the set of situations under which a treaty can actually increase international welfare. Second, when an IEA is expected to form, it will have more members when enforcement is costly, hence, costly enforcement of IEAs is associated with higher environmental quality. Greater participation in IEAs that are costly to enforce occurs because the additional cost of being a party to an IEA must be offset with an increase in the level of environmental quality, which is realized through an increase in the number of parties to the agreement. In addition, provided that the number of countries involved is sufficiently large, social welfare will be greater under an IEA that is costly to enforce. If the number of countries involved is large enough, the aggregate benefit from increased participation levels the parties' costs of enforcing their agreement. Finally, we find that when monitoring for compliance is not perfectly accurate, the

principal results of our model are accentuated. That is, monitoring errors further reduce the set of situations in which multilateral cooperation is welfare enhancing and thus limit the opportunities for an IEA to form. However, when an agreement does form it will involve a higher level of participation.

The paper proceeds as follows: Section 2 presents the basic model of a self-enforcing IEA without enforcement costs. In section 3 we introduce the endogenous enforcement mechanism, and in section 4 draw conclusions about the effects of costly enforcement on the equilibrium number of parties to an agreement and on aggregate welfare. In section 5 we extend the model to allow for errors in monitoring. In section 6 we provides some concluding remarks.

2. The basic model of a self-enforcing IEA

In this section we present a standard model of a self-enforcing IEA in order to review the structure of these games and to provide a baseline for determining the effects of costly enforcement on the outcomes of these games in later sections. Following Barrett (2003), consider a situation where *N* identical countries each emit a uniformly mixed transboundary pollutant. Country *i*'s welfare is

$$w_i = A + b(q_i + q_{-i}) - cq_i \,, \qquad [1]$$

where q_i is equal to one of *i* abates its emissions and is zero if it does not, q_{-i} is the sum of the abatement decision by all other countries, *b* is the constant marginal benefit of abatement, *c* is the cost of abatement, and *A* is a positive constant. Assume that the underlying structure of the countries' interactions with each other is an *N*-player prisoners' dilemma. That is, all countries have a dominant strategy to not abate in a noncooperative Nash equilibrium. This will be true

when $b < c$. However, the countries' joint welfare will be maximized when they all abate their emissions. This requires $Nb > c$.

Recognizing the benefits from jointly agreeing to abate their emissions, the countries have the incentive to form an international environmental agreement to do so. Because participation in an IEA is voluntary, countries can join or not join the agreement at will. In this section, however, parties to any IEA that forms will comply with the terms of the agreement, that is they will abate their emissions, without the need for an enforcement mechanism.

Let *s* denote the number of parties to an IEA. Moreover, let $w^p(s)$ denote the common welfare of each of the parties to the IEA, and let $w^{np}(s)$ denote the common welfare of each of the countries that are not party to the agreement. Throughout, the superscript *p* signals that the country in question is a party to an IEA, while the superscript *np* signals that the country is not a party to the agreement. From [1] we have:

$$w^p(s) = A + bs - c;$$

$$w^{np}(s) = A + bs. \qquad [2]$$

Let s^{nc} denote the equilibrium number of parties to an IEA. (The superscript *nc* identifies s^{nc} as the equilibrium number of parties when the terms of an IEA are not costly to enforce). The definition of a *self-enforcing voluntary agreement* in this setting is:

Definition: An IEA consisting of s^{nc} countries that can enforce the IEA without cost is self-enforcing if and only if:

$$(i)\ \ w^p(s^{nc}) \geq w^{np}(s^{nc} - 1)$$
$$(ii)\ \ w^{np}(s^{nc}) \geq w^p(s^{nc} + 1).^4 \qquad [3]$$

[4] Although the concept of a self-enforcing agreement is most often adopted for analyses of international environmental agreements, it was first developed to study the stability of cartels by D'Apremont et al. (1983). See

Requirement (*i*) of a self-enforcing IEA is that no party has an incentive to leave the agreement; that is, the agreement is *internally stable*. Requirement (*ii*) is that no non-party wishes to join the IEA; that is, the agreement is *externally stable*. The equilibrium value of s^{nc} follows easily from these two conditions.

Using the welfare functions [2] and the external stability condition (*ii*), $w^{np}(s) - w^{p}(s+1) = c - b \geq 0.$ Since we have assumed that $b < c$, the external stability condition is always satisfied. Determining the values of *s* that satisfy the internal stability condition (*i*) involves examining the consequences of two possible outcomes of a single country that leaves an IEA. The first is when this single defection does not cause all the other parties to the IEA to also defect. In this case, using [2] and the internal stability condition (*i*), $w^{p}(s) - w^{np}(s-1) = b - c \geq 0,$ which violates our assumption that $b < c$. This implies that if the number of parties to an IEA is such that it would remain intact if one party defected, then the IEA is not internally stable. Thus, the self-enforcing number of parties to an agreement must be such that one defection would make the agreement collapse. In this case, the internal stability condition (*i*) is written as $w^{p}(s) - w^{np}(0) = bs - c \geq 0.$ The self-enforcing number of parties to an agreement that is enforced without cost is the minimum size coalition that satisfies this inequality. Formally:

$$s^{nc} = \min s \mid w^{p}(s) - w^{np}(0) \geq 0 = \min s \mid s \geq c/b. \qquad [4]$$

Our assumption that $b < c$ implies that $s^{nc} > 1$, while the assumption that $Nb - c > 0$ implies that [4] will be satisfied for some $s^{nc} \leq N.$ Condition [4] indicates that participation with an IEA is increasing in the cost of abatement, *c*, and decreasing in the individual benefit of abatement, *b*.

Diamontoudi (2005) for a recent contribution to this literature. Dawson and Segerson (2003) have used the concept of a self-enforcing agreement to model voluntary domestic environmental policies.

Thus, Barrett's (1994, 2003) claim that international cooperation to protect the environment will be greatest when it is needed least is easily verified.

It is clear that the concept of a self-enforcing agreement applies to the stability of a cooperating coalition, not to parties' decisions to comply with the terms of the agreement once they've joined. However, Barrett (1998b, pg 36) claims: "The binding constraint on international cooperation is free-rider deterrence, not compliance enforcement. Once free-riding [nonparticipation] can be deterred, compliance can be enforced free of charge." This is a very strong statement that follows if countries are able to observe each other's compliance decisions perfectly and without cost. In this setting a country has no incentive to join an agreement and then not comply with its requirements. If it did so all the other participating nations would automatically observe this violation, would realize that they would then be worse off if they stayed with the agreement, and therefore would leave the agreement. Under perfect information a party to an IEA would not violate the terms of the agreement, because this violation would cause the agreement to collapse.

Realistically, however, nations cannot observe each other's abatement perfectly and without cost, nor can a country's abatement decision be inferred from an aggregate measure (e.g., the global concentration of a pollutant). In these cases, a country may be motivated to join an agreement and then decide to violate its terms. If the other participating countries are not able to observe this act of noncompliance, they will not automatically leave the agreement. The violator, then, is able to escape the cost of compliance while enjoying the benefit of cooperation of those that remain with the agreement. Asymmetric information among parties to an IEA about their compliance decisions motivates the implementation of some enforcement mechanism to counteract the incentive to violate the terms of the agreement.

3. Endogenous enforcement of compliance to an IEA

We now give parties to an IEA the opportunity to violate the terms of the agreement, as well as the opportunity to invest in a third-party enforcement body that is charged with maintaining compliance to the agreement. The enforcement mechanism works as follows; each party to the agreement pays x dollars to a third-party enforcement body that is capable of monitoring the parties with probability π and penalizing noncompliant parties with a fine *f*.

The monitoring capability of the enforcer is a montonically increasing function of the amount of funding provided by the parties. Suppose that monitoring consists of random audits of the parties. Audits are perfectly accurate in the sense that an audit always uncovers a violation if one has occurred and does not 'discover' a violation when one hasn't occurred. (We examine the consequences of inaccurate monitoring in section 5). Each dollar of additional enforcement funding allows the number of random audits to increase by α; that is, α is the constant marginal productivity of resources devoted to monitoring. If *s* parties to an agreement each provide *x* to fund the enforcer, then the number of random audits the enforcer conducts is $sx\alpha$, and the probability that any party is audited is

$$\pi = sx\alpha / s = x\alpha . \qquad [5]$$

Clearly, constraining π to be between zero and one requires $\alpha \in [0,\ 1/x]$. We will maintain this assumption throughout. The expected penalty for noncompliance is $\pi f = x\alpha f$. The fine, *f*, is constrained (by convention, norm, or law) to be no more than $\overline{f}$.

Assume that the parties to an IEA are risk neutral, and that they comply with the terms of the agreement if they are at least indifferent between compliance and noncompliance. Then, given an agreement consisting of *s* parties, an individual party will comply if its payoff from doing so is not less than its expected payoff from noncompliance. A party's payoff from

compliance is $w^p(s) - x$, and its expected payoff from noncompliance is $w^{np}(s-1) - x - x\alpha f$, where recall that $w^p(s)$ and $w^{np}(s-1)$ are defined by [2]. Therefore, a party to an IEA complies with the terms of the agreement if and only if $[w^p(s) - x] - [w^{np}(s-1) - x - x\alpha f] =$ $x\alpha f - (c-b) \geq 0$. Intuitively, if the expected penalty, $x\alpha f$, is less than the gain from noncompliance, $c-b$, then all parties to an IEA will violate its terms and no self-enforcing agreement is possible. On the other hand, if $x\alpha f$ is no less than $c-b$, each party to an IEA will comply with its requirements. Clearly, $x\alpha f - (c-b) \geq 0$ is a necessary condition for a viable IEA. Indeed, a self-enforcing IEA that is costly to enforce is not internally stable unless this condition holds.

Now let us determine each party's contribution to the enforcer of an IEA. Clearly, each would like to contribute as little as possible while providing the enforcer with sufficient resources to maintain compliance with the agreement. This requires a payment *x* so that $x\alpha f - (c-b) \geq 0$ binds, yielding $x = (c-b)/\alpha f$. Moreover, since *x* is monotonically decreasing in the fine for noncompliance, the parties to the agreement will choose the fine to be as high as possible; that is, $f = \overline{f}$.[5] Thus, the contribution to the enforcer of an IEA that is required of all parties to the agreement is

$$x = (c-b)/\alpha \overline{f}. \qquad [6]$$

Note that this payment decreases with the maximal fine, $\overline{f}$, and the marginal productivity of resources devoted to monitoring, α, but is increasing in the gain from noncompliance, $c-b$.

[5] The idea that the penalty for noncompliance should be set as high as possible to conserve on monitoring costs when agents are risk neutral is common in the literature on the economics of law enforcement. See Polinsky and Shavell (2000) for a review of this literature.

4. Self-enforcing IEAs with costly enforcement

We are now ready to analyze the consequences of costly enforcement of international environmental agreements. The first effect of costly enforcement is that it changes the set of circumstances under which cooperation increases aggregate welfare. Given *s* parties to an agreement that each earn $w^p(s) - x = A + bs - c - x$, and $N - s$ free-riding countries that each earn $w^{np}(s) = A + bs$, aggregate welfare when cooperation is costly to enforce is

$$\begin{aligned} W^c(s) &= s(A + bs - c - x) + (N - s)(A + bs) \\ &= NA + s(Nb - c - x). \end{aligned} \qquad [7]$$

(The superscript '*c*' identifies variables and functions when cooperation is costly to enforce). Note that $W^c(s)$ is linearly decreasing in *s* if $Nb - c - x < 0$, and linearly increasing if $Nb - c - x > 0$. Use [6] to substitute for *x* in the latter inequality and rearrange the result to obtain

$$N > \frac{c}{b} + \frac{c - b}{b\alpha \bar{f}}. \qquad [8]$$

If [8] holds, then any coalition of countries that cooperate to abate their emissions will increase aggregate welfare. Moreover, aggregate welfare is maximized if all countries cooperate to abate their emissions. If the inequality in [8] is reversed, then the costs of enforcing a cooperative agreement are high enough to make cooperation by any subset of countries inefficient. Recall that we assumed $N > c/b$ under costless enforcement of an agreement so that any coalition of cooperating countries would increase aggregate welfare, and the grand coalition of cooperating countries maximized their joint welfare. Since $c > b$, the second term on the right side of [8] is strictly positive, which yields the following proposition:

Proposition 1: The set of values of *N*, *b*, and *c* for which international environmental agreements increase aggregate welfare is smaller when enforcement of these agreements is costly.

Since enforcing cooperation entails an additional cost of forming cooperative agreements, the set of circumstances under which cooperation will increase aggregate welfare is smaller than when cooperation can be enforced without cost. It is straightforward to show that $c/b + (c-b)/b\alpha\overline{f}$ is increasing in *c* and decreasing in *b*. Thus, the set of circumstances under which cooperation is worthwhile is larger when abatement costs lower and the benefits of abatement are higher. Moreover, $c/b + (c-b)/b\alpha\overline{f}$ is decreasing in α and $\overline{f}$. Increasing either of these parameters decreases the payment cooperators pay to enforce an agreement, leading to an enlargement of the set of opportunities for welfare-enhancing cooperation.

Now let us determine the equilibrium coalition size when IEAs are costly to enforce. As in the case of costless enforcement, a self-enforcing equilibrium is the minimum size coalition of cooperating countries for which the welfare of each of these countries is not less than when no country abates their pollution. That is, letting s^c denote the size of a self-enforcing IEA with costly enforcement, $s^c = \min s \mid w^p(s) - x \geq w^{np}(0)$. Using the welfare functions [2] and substituting for *x* from [6], the equilibrium size coalition under costly enforcement is:

$$s^c = \min s \mid s \geq \frac{c}{b} + \frac{c-b}{b\alpha\overline{f}}. \qquad [9]$$

An IEA with costly enforcement will form if and only if $N \geq c/b + (c-b)/b\alpha\overline{f}$, because then there exists a coalition $s^c \leq N$ that satisfies [9]. On the other hand, when enforcement is costless the term $(c-b)/b\alpha\overline{f}$ disappears, so that an IEA without costly enforcement will form as long as $N \geq c/b$. Since $(c-b)/b\alpha\overline{f} > 0$, we have the following proposition:

Proposition 2: If compliance to IEAs is costly to enforce, then the set of values of *N*, *b*, and *c* for which an IEA will form is smaller than when enforcement is costless.

The condition under which an IEA with costly enforcement would be expected to form is identical to the condition that determines whether cooperative management of the international environmental resource will increase aggregate welfare (equation [8]). This is a simple statement that implies that a cooperative agreement among some subset of countries can be expected to form when cooperation among nations is worthwhile. Propositions 1 and 2 reveal that the circumstances under which cooperative agreements are expected emerge are limited by the need for costly enforcement.

However, when an IEA is expected to form, costly enforcement implies that the number of parties to the agreement will typically be higher. When an IEA with costly enforcement forms, [9] indicates that the equilibrium size of the coalition s^c is the least *s* for which $s \geq c/b + (c-b)/b\alpha\overline{f}$. When an IEA is not costly to enforce, [4] indicates that the equilibrium coalition s^{nc} is the least *s* for which $s \geq c/b$. Again, since $(c-b)/b\alpha\overline{f} > 0$, $s^c \geq s^{nc}$. Therefore:

Proposition 3: If an IEA that is costly to enforce forms, membership in the IEA will be no less, and will typically be greater, than if the IEA could be enforced without cost.

The intuition behind this result is straightforward. Since contributing to enforcement is an additional cost of joining an IEA, more countries are required to participate in an IEA to make

the agreement worthwhile. Note that if an IEA forms, then costly enforcement is associated with increased environmental quality because more countries agree to abate their emissions.

Generally speaking, participation in an IEA increases with the costs of participation and decreases with the benefit that participation provides to all the countries. Thus, participation with an IEA that is costly to enforce increases with a country's abatement cost, *c*, and decreases with individual benefit of some country's abatement, *b*. Moreover, participation decreases as the cost of enforcement is reduced because either the marginal productivity of monitoring resources, α, increases or the maximal penalty for noncompliance, $\overline{f}$, increases.

Finally, let us examine possible difference in aggregate welfare under self-enforcing IEAs that are costly to enforce and under those that are not costly to enforce. From [7] and [9], social welfare when a self-enforcing IEA requires costly enforcement is $W^c(s^c) =$ $NA + s^c(Nb - c - x)$. For convenience, let us assume that *s* is continuous. Then, from [9], $s^c = c/b + (c-b)/b\alpha\overline{f}$. Substitute this and $x = (c-b)/\alpha\overline{f}$ from [6] into $W^c(s^c)$ to obtain

$$W^c(s^c) = NA + \left(\frac{c}{b} + \frac{c-b}{b\alpha\overline{f}}\right)\left(Nb - c - \frac{c-b}{\alpha\overline{f}}\right). \qquad [10]$$

Welfare when a self-enforcing IEA does not require costly enforcement is $W^{nc}(s^{nc}) =$ $NA + s^{nc}(Nb - c)$, which upon substitution of $s^{nc} = c/b$ from [4] becomes

$$W^{nc}(s^{nc}) = NA + (c/b)(Nb - c). \qquad [11]$$

Subtract [11] from [10] to obtain

$$W^c(s^c) - W^{nc}(s^{nc}) = \left(\frac{c-b}{\alpha\overline{f}}\right)\left[N - \left(\frac{2c}{b} + \frac{c-b}{b\alpha\overline{f}}\right)\right].$$

Since the first term of $W^c(s^c) - W^{nc}(s^{nc})$ is positive, its sign is equal to the sign of the term in hard brackets. Therefore, we have our final proposition.

Proposition 4: Aggregate welfare is higher when IEAs are costly to enforce if and only if

$$N > \frac{2c}{b} + \frac{c-b}{b\alpha \overline{f}}. \qquad [12]$$

Figure 1 illustrates how $W^c(s^c)$ and $W^{nc}(s^{nc})$ vary with the number of potential parties to an IEA. To draw this graph we have assumed, without loss of generality, that $A = 0$. Aggregate welfare when an IEA does not require costly enforcement, $W^{nc}(s^{nc})$, is zero for $N \le c/b$. For $N > c/b$, $W^{nc}(s^{nc})$ increases linearly at rate c (from equation [11]). The size of a self-enforcing IEA does not change as N increases—it remains constant at c/b—but aggregate welfare increases with N because increasing N means we are increasing only the number of free-riding countries, each of which benefits from the abatement efforts of the c/b parties to the agreement.

The bold dashed function in Figure 1 is aggregate welfare for a self-enforcing IEA that is costly to enforce. Note that $W^c(s^c)$ is equal to zero for a larger range of N than $W^{nc}(s^{nc})$; that is, up to $c/b + (c-b)/b\alpha \overline{f}$. This follows because cooperative abatement efforts that are costly to enforce increase aggregate welfare under a smaller set of circumstances than when cooperation can be enforced without cost (Proposition 1), and consequently, IEAs will only form under this smaller set of circumstances (Proposition 2). When $N \ge c/b + (c-b)/b\alpha \overline{f}$, the coalition $s^c = c/b + (c-b)/b\alpha \overline{f}$ forms and aggregate welfare increases with N at rate s^c (from equation [10]). It is important to reiterate, however, that the increase in aggregate welfare that comes from increasing N goes entirely to the free-riding countries, and not to the members of s^c.

For $N > c/b + (c-b)/b\alpha\overline{f}$, the relationship between $W^c(s^c)$ and $W^{nc}(s^{nc})$ depends on two countervailing factors, the positive welfare effect of higher abatement when an IEA must be enforced (because $s^c > s^{nc}$ from Proposition 3) and the negative welfare effect of the costs of this enforcement. For $N \in \left[c/b + (c-b)/b\alpha\overline{f}, 2c/b + (c-b)/b\alpha\overline{f}\right)$, the enforcement-cost effect dominates the higher-abatement effect so that aggregate welfare is lower when the equilibrium IEA requires costly enforcement. However, when N exceeds $2c/b + (c-b)/b\alpha\overline{f}$ the higher-abatement effect dominates the enforcement-cost effect so that aggregate welfare is higher when compliance to an agreement requires costly enforcement than when compliance can be enforced without cost.

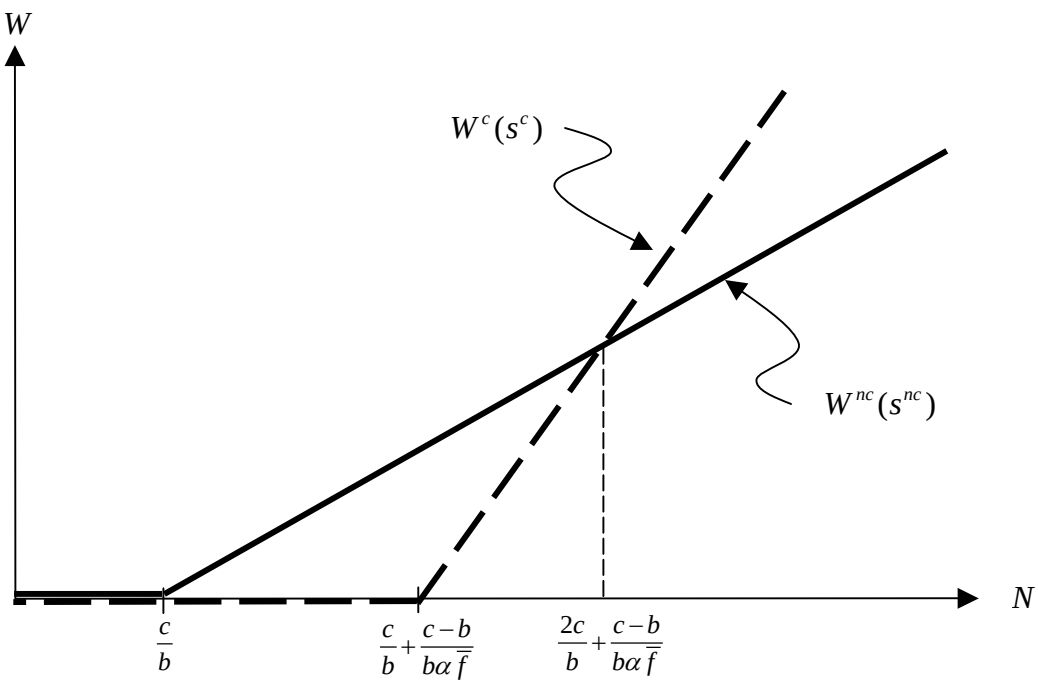

Figure 1: Aggregate welfare from self-enforcing international environmental agreements when they are costly to enforce and when they are not costly to enforce.

5. Inaccurate monitoring

We have assumed to this point that the third-party enforcer of an IEA is capable of monitoring member countries' actions with complete accuracy. That is, an audit always reveals a country's true compliance status. However, monitoring of countries' compliance behavior is likely to be subject to a host of possible errors, including errors due to erroneous or missing data, reporting errors, and errors in evaluating available performance data. In this section we relax the assumption of perfectly accurate monitoring to examine how inaccurate monitoring might affect the structure of self-enforcing IEAs.[6] Two types of monitoring errors are possible: given an audit, a compliant country may be judged to be noncompliant (a Type I error) and a noncompliant country may be judged to be compliant (a Type II error). Let $\rho^1 \in (0,1)$ and $\rho^2 \in (0,1)$ denote the probabilities that the third-party enforcer commits a Type I error and Type II error, respectively. These probabilities are common knowledge among the countries and the enforcer.

Maintaining the assumption that countries are risk neutral and simply imposing the result that parties to an agreement will instruct the third-party enforcer to impose the maximal fine in cases on noncompliance, it is straightforward to show that the expected welfare of a compliant country when monitoring is not perfectly accurate is

$$w^p(s) - x - \pi\rho^1 \overline{f} \ . \qquad [13]$$

Note that $\pi\rho^1$ is the probability that a compliant firm is audited and judged to be noncompliant, and $\pi\rho^1 \overline{f}$ is the country's expected sanction from the possibility of a false judgment of

[6] See Kaplow and Shavell (1994) for the effects of monitoring inaccuracy in a standard model of optimal law enforcement.

noncompliance.[7] Similarly, the expected welfare of a noncompliant country can be shown to be

$$w^p(s-1) - x - \pi(1-\rho^2)\overline{f}\,. \qquad [14]$$

In [14], $\pi(1-\rho^2)$ in [14] is the probability that a noncompliant country is correctly determined to be noncompliant, and $\pi(1-\rho^2)\overline{f}$ is the expected penalty it faces.

Again, a country that is party to an IEA consisting of *s* members will comply with the terms of the agreement if and only if it expected payoff from compliance is not less that its expected payoff from noncompliance. Using [2] to substitute for $w^p(s)$ and $w^p(s-1)$ in [13] and [14], respectively, a country is compliant if and only if $\pi\overline{f}(1-\rho^1-\rho^2) \geq c-b$. Clearly, given a monitoring intensity and the maximal available fine, the presence of monitoring errors of both types weakens deterrence. In fact, if $\rho^1+\rho^2 \geq 1$, then countries will never comply with the terms of an IEA, indicating that IEAs cannot form when monitoring errors are severe enough. From here on let us assume that $\rho^1+\rho^2 < 1$ to allow an IEA to form.

Because monitoring inaccuracy weakens deterrence, the payment by each member of an IEA to the enforcer will be higher. To see this, recall that the enforcer's monitoring probability is determined by $\pi = \alpha x$. The minimum payment required of each member of an IEA is then the solution to $x\alpha\overline{f}(1-\rho^1-\rho^2) = c-b$; that is,

$$x = \frac{(c-b)}{\alpha\overline{f}(1-\rho^1-\rho^2)}\,. \qquad [15]$$

[7] One may wonder why a compliant country that is incorrectly determined to be noncompliant (Type I error) will simply accept the penalty imposed on it. At the very least, a compliant country that faces a sanction for noncompliance will attempt to prove that it has, in fact, satisfied the terms of the agreement. Under the Kyoto Protocol, for example, if the enforcement branch of the Compliance Committee finds a party to be in violation, that party is entitled to a hearing in which it can present additional evidence and expert testimony. Further, if a party is still deemed noncompliant after the appeal process, it is given the chance to submit a final written testimony to the enforcement branch (UNFCC 2002). This process appears to be designed to limit the probability of a Type I error. Even if this probability is zero ($\rho^1 = 0$), the results of this section hold as long as the probability of a Type II error is positive ($\rho^2 > 0$), at least qualitatively.

Clearly, inaccurate monitoring increases the payment required of each party to an IEA to maintain compliance with the agreement.

Since inaccuracy is associated with increased enforcement costs, the circumstances under which cooperative agreements can increase international welfare, and hence, are expected to form, are limited by monitoring inaccuracy. Recall from [7] that aggregate welfare can be expressed as $NA + s(Nb - c - x)$, which reveals that cooperation increases aggregate welfare as long as $Nb - c - x > 0$. Substituting [15] for x and solving $Nb - c - x > 0$ for N yields

$$N > \frac{c}{b} + \frac{(c-b)}{b\alpha \overline{f}(1-\rho^1-\rho^2)}. \qquad [16]$$

Since the right hand side of [16] is increasing in ρ^1 and ρ^2, the range of international environmental problems that can be addressed with cooperative management is decreasing in these probabilities.

Provided that [16] holds, a self-enforcing agreement with a positive number of participating countries will form. Once again, the self-enforcing equilibrium is the minimum size coalition for which the welfare of each of the participating countries is not less than their welfare when no coalition forms; that is, $s^c = \min s \mid w^p(s) - x \geq w^{np}(0)$. Substituting [15] for x yields

$$s^c = \min s \mid s \geq \frac{c}{b} + \frac{c-b}{b\alpha \overline{f}(1-\rho^1-\rho^2)}. \qquad [17]$$

Since the last term of [17] is increasing in ρ^1 and ρ^2, monitoring inaccuracy will tend to produce higher levels of participation in IEAs that form. The intuition behind this result parallels the explanation of Proposition 3 in the previous section. Monitoring errors increase the cost of multilateral cooperation. If an IEA is to form, therefore, greater participation is required so that the increase in international environmental quality offsets the additional enforcement costs.

6. Concluding remarks

We have analyzed a game of self-enforcing international environmental agreements when parties to such an agreement finance an independent enforcement body to maintain compliance with the agreement. We have shown that costly enforcement limits the circumstances under which international cooperation to control a transboundary pollutant will increase aggregate welfare. Consequently, the circumstances under which an IEA can be expected to form are limited by costly enforcement. However, when an IEA is expected to form, participation with the agreement will typically be greater than when an IEA does not require costly enforcement. As a result, costly enforcement is associated with higher international environmental quality. In fact, under some circumstances, costly enforcement of an IEA is associated with higher aggregate welfare—not for the parties to an IEA, however, but for the countries that choose to free-ride on the agreement. We have also shown that these results are accentuated when monitoring for compliance to an agreement is subject to error. That is, monitoring inaccuracy further reduces the circumstances under which self-enforcing IEAs can form, but when they do form, participation will be higher.

Our results have important implications for multilateral management of environmental externalities. By assuming away enforcement problems related to IEAs, the scope of mutually beneficial agreements between countries will be artificially enlarged and the minimum required coalition of countries will be artificially reduced. In other words, for some international environmental problems, once enforcement costs are included, IEAs will either require more members before coming into effect or may not be worthwhile at all.

While costly enforcement of IEAs is expected to result in improved environmental quality, the increased participation requirement may pose an additional problem for multilateral

cooperation. The formation of an IEA is largely a problem of coordination among the countries involved. Although the simple model proposed in this paper assumes countries can perfectly coordinate their actions, in reality coordination problems typically increase when more countries need to be involved. Therefore, the increased membership requirement resulting from the enforcement problem may exacerbate the fundamental coordination, which, in turn, could jeopardize the formation of welfare-enhancing IEA.

Additionally, when IEAs require costly enforcement, the benefits to nonparticipation free-riding are higher. Nonmembers escape both the cost of providing the public good and the additional cost of financing enforcement of cooperation, while benefiting from higher environmental quality. If countries involved in forming an IEA have a strong aversion to inequity, the large benefits captured by the free-riders may prevent the agreement from materializing. Greater welfare inequity may motivate countries to sacrifice individual and collective gains to prevent a self-enforcing agreement from forming in order to deny the gains to free-riders.

A logical next step in analyzing how costly enforcement and free-rider incentives affect the stability of international environmental agreements is to observe how people actually play these games. It is well known that insights derived from theory are limited in that they abstract from a number of other potential influences on behavior. The coordination and inequity problems previously described are just two examples. Shogren (2006) argues for using economics experiments to analyze strategic interactions among players involved in forming self-enforcing agreements. Conducting controlled experiments may provide useful empirical evidence about the performance of these agreements.

Finally, although the model developed in this paper offers new and interesting insights by introducing costly enforcement into the standard IEA game, it could easily be extended to include additional features. The welfare functions of the parties, for example, could be constructed to better reflect the real heterogeneity of countries involved in a particular environmental problem. Additionally, the analysis could investigate the effects of allowing for the possibility of multiple self-enforcing agreements forming concurrently. Obviously, there are a host of other extensions that can and should be addressed to gain a more complete picture of how costly enforcement affects voluntary coalition formation to protect the international environment.

References

Barrett, Scott, 2003. *Environment and Statecraft: The Strategy of Environmental Treaty-Making*. Oxford and New York: Oxford University Press.

Barrett, Scott, 1998a. "On the Theory and Diplomacy of Environmental Treaty Making." *Environmental and Resource Economics*. 11(1), 317-333.

Barrett, Scott, 1998b. "Political Economy of the Kyoto Protocol." *Oxford Review of Economic Policy*. 14(4), 20-39.

Barrett, Scott, 1997a. "The Strategy of Trade Sanctions in International Environmental Agreements." *Resource and Energy Economics*. 19(1), 345-361.

Barrett, Scott, 1997b. "Heterogeneous international environmental agreements." *International Environmental Negotiations*. Carlo Carraro (ed.). Edward Elgar: Cheltenham, U.K. and Northampton MA.

Barrett, Scott, 1994. "Self-Enforcing International Environmental Agreements." *Oxford Economic Papers* 46(1), 878-94.

Benedick, Richard E., 1998. *Ozone Diplomacy: New Directions in Safeguarding the Planet*, Enlarged Edition. Cambridge, MA and London, England: Harvard University Press.

Botteon, Michele and Carlo Carraro. 1998. "Strategies for Environmental Negotiations: Issue Linkage with Heterogeneous Countries." *Game Theory and the Environment.* Hanley, Nick and Henk Folmer (eds.) Edward Elgar: Cheltenham, U.K. and Northampton MA.

Carraro, Carlo and Michele Botteon, 1997. "Burden Sharing and Coalition Stability in Environmental Negotiations with Asymmetric Countries." *International Environmental Negotiations*. Carlo Carraro (ed.). Edward Elgar: Cheltenham, U.K. and Northampton MA.

Carraro, Carlo; Carmen Marchiori and Sonia Oreffice, 2003. "Endogenous Minimum Participation in International Environmental Treaties." *Nota Di Lavaro 113.2003. Fondazione Eni Enrico Mattei Working Paper*.

Carraro, Carlo and Domenico Siniscalco, 1998. "International Environmental Agreements: Incentives and Political Economy." *European Economic Review* 42(3-5), 561-572.

Carraro, Carlo and Domenico Siniscalco, 1997. "R&D Cooperation and the Stability of International Environmental Agreements." *International Environmental Negotiations*. Carlo Carraro (ed.). Edward Elgar: Cheltenham, U.K. and Northampton MA.

Carraro, Carlo and Domenico Siniscalco, 1993. "Strategies for the International Protection of the Environment." *Journal of Public Economics*. 52(3), 309-328.

Chayes, Abram and Antonia Handler Chayes, 1991. "Compliance without Enforcement: State Behavior under Regulatory Treaties." *Negotiation Journal.* 7(1), 311-330.

D'Aspremont, Claude; Alexis Jacquemin; Jean Jaskold Gabszewicz and John A. Weymark, 1983. "On the Stability of Collusive Price Leadership." *The Canadian Journal of Economics*. 16(1), 17-25.

Dawson, Na Li and Kathleen Segerson, 2003. "Voluntary Agreements with Industries: Participation Incentives with Industry-wide Targets." *University of Connecticut, Department of Economics, Working papers*: 2004-06.

Diamantoudi, Effrosyni, 2005. “Stable Cartels Revisted.” *Economic Theory*. 26(4), 907-921.

Downs, George W.; Rocke, David M. and Peter M. Barsoom, 1996. “Is the Good News about Compliance Good News about Cooperation?” *International Organization.* 50(3), 379-40.

Finus, Michael, 2004. “International Cooperation to Resolve International Pollution Problems.” *Fondazione Eni Enrico Mattei Working Paper*.

Finus, Michael and Bianca Rundshagen, 1998. “Toward a Positive Theory of Coalition Formation and Endogenous Instrument Choice in Global Pollution Control.” *Public Choice*. 96(1-2), 145-186.

Folmer, Henk.;van Mouche, Pierre and Shannon Ragland, 1993. “Interconnected Games and International Environmental Problems.” *Environmental and Resource Economics*. 3(4), 313-335.

Friends of the Earth, 2006. “Climate Treaty One Year Old, but Emissions Still Rising.” Press Release, February 15, 2006. ttp://www.foe.co.uk/resource/press_releases/climate_treaty_one_year_ol_14022006.html

Heister, Johannes; Ernst Mohr; Frank Stahler; Peter-Tobias Stoll and Rudiger Wolfum, 1997. “Strategies to Enforce Compliance with an International CO2 Treaty.” *International Environmental Affairs*. 9(1), 22 – 53.

Hoel, Michael and Kerstin Schneider, 1997. “Incentives to Participate in an International Environmental Treaty.” *Environmental and Resource Economics*. 9(2), 153-170.

Hoel, Michael, 1992. “International Environment Conventions: The Case of Uniform Reductions of Emissions.” *Environmental and Resource Economics*. 2(2), 141-159.

Hovi, Jon and Ivar Areklett, 2004. “Enforcing the Climate Regime: Game Theory and the Marrakesh Accords.” *International Environmental Agreements: Politics, Law and Economics*. 4(1), 1-26.

Kaplow, Louis and Steven Shavell. 1994. “Accuracy in the Determination of Liability.” *Journal of Law and Economics* 37(1): 1-16.

Polinsky, A. Mitchell and Steven Shavell. 2000. ”The Economic Theory of Public Enforcement of Law.” *Journal of Economic Literature* 38(1): 45-76.

Shogren, Jason F., 2006. “Book Review: Environment and Statecraft.” *American Journal of Agricultural Economics*. 88(2), 519-520.

Swanson, Timothy and Sam Johnston, 1999. *Global Environmental Problems and International Environmental Agreements: The Economics of International Institution Building*. Cheltenham, U.K and Northampton, MA: Edward Elgar.

United Nations Framework Convention on Climate Change, 2002. “Report on the Conference of the Parties: VII Session. Part Two: Actions Taken by the Conference of the Parties.” Marrakesh, January 2002.

United States International Trade Commission, 1991. *International Agreements to Protect the Environment and Wildlife*. USITC Publication 2351. Washington, D.C.

Wagner, Ulrich J., 2001. “The Design of Stable International Environmental Agreements: Economic Theory and Political Economy.” *Journal of Economic Surveys*. 15(3), 377-411.