

Self-enforcing International Environmental Agreements with Costly Monitoring for Compliance

David M. McEvoy · John K. Stranlund

Accepted: 7 June 2008 / Published online: 29 June 2008
© Springer Science+Business Media B.V. 2008

Abstract Theoretical analyses of international environmental agreements (IEAs) have often 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 compliance with these agreements once they are in place. In this paper we analyze an IEA game in which parties to an agreement finance an independent monitor who audits the compliance performance of the members of an agreement. These audits reveal instances of noncompliance so they can be sanctioned. We find that costly monitoring of compliance limits the circumstances under which international cooperation to protect the environment is worthwhile, but when IEAs do form they will often involve greater participation than IEAs that do not require costly monitoring. Consequently, costly monitoring of IEAs can produce higher international environmental quality. Moreover, under certain conditions, aggregate welfare is higher when IEAs require costly monitoring.

Keywords International environmental agreements · Self-enforcing agreements · Compliance · Monitoring

JEL Classification Q5 · H41 · C72 · F53

D. M. McEvoy (✉)
Economics Department, Appalachian State University, Raley Hall,
416 Howard Street, ASU Box 32051, Boone, NC 28608-2051, USA
e-mail: mcevoydm@appstate.edu

J. K. Stranlund
Department of Resource Economics, University of Massachusetts-Amherst,
80 Campus Center Way, 214 Stockbridge Hall, Amherst, MA 01003, USA
e-mail: stranlund@resecon.umass.edu

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 with IEAs is 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, 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 is not properly enforced, 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 focused on 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 monitor who audits, with some endogenous probability, the compliance performance of individual members.

All effective international agreements must include costly compliance monitoring.¹ For example, the Implementation Committee of the Montreal Protocol is responsible for gathering compliance information about countries that are suspected to be in violation of the Protocol (Benedick 1998).² Likewise, the Compliance Committee, one of the institutions constructed to enforce compliance with the Kyoto Protocol, is designed to monitor the compliance behavior of signatory countries and then report their findings to the Enforcement Branch, which has the power to impose sanctions (UNFCCC 2002). The Convention on International Trade in Endangered Species (CITES) makes a similar provision for monitoring for compliance through the Standing Committee, its own management body responsible for enforcement.³ The enforcement provisions included in CITES and the Montreal and Kyoto Protocols allow their respective management bodies to monitor a party's behavior only when it is suspected to be violating the treaty. In contrast, the International Whaling Commission, the management body appointed under the International Convention for the Regulation of Whaling, established the International Observers Scheme to monitor signatories' behavior. This policy required member countries to each appoint observers and finance monitoring activities on a rotating basis.⁴ Although the design of monitoring and enforcement institutions differ across international agreements, each share the common characteristic that the cost of monitoring activities must be borne by the parties themselves.

Despite the fact that costly monitoring for compliance must be part of any international agreement, theoretical analyses of IEAs usually assume that compliance can be enforced without cost. In the literature IEAs are often analyzed using the equilibrium concept of a *self-enforcing agreement*.⁵ A self-enforcing agreement made between agents is defined as a single coalition of cooperating agents from which no member of the coalition wishes to

¹ This point is argued by Heister et al. (1997) who provide a qualitative analysis of why international environmental agreements need to be enforced and the available strategies for enforcing compliance.

² Protocol on Substances that Deplete the Ozone Layer, Montreal, September 1987; Amended 1990, 1992, 1995, 1997, 1999.

³ Convention on International Trade in Endangered Species, Washington, March 1973; Amended 1979.

⁴ The International Observers Scheme was in effect from 1972 until the moratorium on whaling was imposed in 1982 (Oberthur 1998).

⁵ Other stability concepts have been employed in the analysis of IEAs, including 'core' solutions from cooperative game theory (Chandler and Tulkens 1995, 1997), renegotiation-proof equilibria (Barrett 1994; Finus and Rundshagen 1998) and farsighted stability in the context of repeated games (Osmani and Tol 2007).

withdraw (the coalition is *internally stable*) and no non-member wishes to join (the coalition is *externally stable*).⁶ This definition of coalitional stability was first proposed by D'Aspremont et al. (1983) for their analysis of cartel stability. Barrett (1994) appears to have been the first to apply it to the analysis of international environmental relations. However, the term *self-enforcing* is a bit misleading, because it does not address the problem of enforcing the compliance of the members of a cooperative agreement. Most authors who use this equilibrium concept simply assume that parties to a self-enforcing IEA comply fully with the terms of the agreement (e.g. Hoel 1992; Carraro and Siniscalco 1993, 1998; Barrett 1994, 2003; Kolstad 2007).⁷ Barrett (1994) and Finus and Rundshagen (1998) have examined enforcement mechanisms for IEAs. They focus on renegotiation-proof punishment strategies with which parties to an agreement sanction violators by jointly reducing their provision of international environmental quality in the future.

We take a different approach to the study of compliance and enforcement of IEAs that is motivated by the costly monitoring activities that are part of established international agreements. 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 body with the power to monitor their compliance behavior. Parties to an IEA that are monitored and deemed in violation of the agreement incur a costly sanction. We consider two types of sanctions—one in which an enforcement body imposes a financial sanction on detected violations, and the other in which detected violators, as well as those who choose not to participate in an IEA, face a costly deterioration of their reputation as cooperative members of the international community.⁸

Our efforts yield several new results. First, the range of international environmental problems within which a welfare-improving self-enforcing IEA can form is smaller when monitoring is necessary and costly. This follows because monitoring 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 often have more members

⁶ Models of self-enforcing IEAs have included provisions for trade sanctions (Barrett 1997a,b), side payments (Carraro and Siniscalco 1993; Carraro and Botteon 1997; Hoel and Schneider 1997), issue linkage (Folmer et al. 1993; Carraro and Siniscalco 1997; Botteon and Carraro 1998) and minimum participation requirements (Barrett 1998a, 2003; Carraro et al. 2003). For a review of many of these extensions see Wagner (2001).

⁷ There is ample evidence that many existing IEAs experience significant problems with noncompliance. For example, CITES has witnessed hundreds of infractions every year since its inception in 1979 (Carraro and Siniscalco 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). 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). 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 side 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. However, 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.

⁸ Because our model is one of enforcing compliance within stable coalitions it is related to the vast literature on the economics of law enforcement. For a review of this literature see Polinsky and Shavell (2000). There is also a large literature on the economics of enforcing environmental policies. Recent contributions on enforcing emissions standards, tradable emissions permits, and emissions taxes include Arguedas (2007), Stranlund (2007), and Macho-Stadler and Perez-Castrillo (2006), respectively. The fundamental difference between our work and the law enforcement literature is that we examine enforcement within a model of endogenous coalition formation while the typical model of enforcement involves a government enforcing a rule that it has imposed.

when the parties must pay for compliance monitoring. This 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. As a result, costly monitoring of IEAs can produce higher environmental quality. In addition, provided that the number of potential parties to an agreement is sufficiently large, social welfare can be greater under an IEA that is costly to monitor. Finally, we find that our principal results are accentuated when compliance monitoring is not perfectly accurate. 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, monitoring inaccuracy can generate higher participation and environmental quality.

The rest of this paper proceeds as follows. After presenting a basic model of a self-enforcing IEA in Sect. 2. In Sect. 3 we extend the model to allow for noncompliance and assume that parties to a self-enforcing IEA finance an independent monitor who checks on their compliance choices and applies a financial sanction in cases of detected noncompliance. Section 4 contains the principal results of our work. In this section we derive the consequences of costly monitoring of IEAs on the value of international cooperation, the circumstances under which a welfare-improving IEA will form, and the self-enforcing number of members of an IEA. In Sect. 5 we analyze the effects of costly monitoring on global welfare. In Sect. 6 we consider an alternative model in which IEA violators, as well as those who do not participate in the agreement, suffer a costly loss in reputation. We conclude in Sect. 7.

2 The Basic Model of a Self-enforcing IEA

In this section, we present a standard model of a self-enforcing IEA—one that does not require enforcement—to review the structure of these games and to provide a baseline for determining the effects of costly compliance monitoring on the outcomes of these games. For simplicity, throughout our analysis countries make a discrete choice whether to abate their emissions. Although a model with continuous abatement choice would allow for more general results, the discrete choice model has been adopted in a number of previous studies because of its tractability. Thus, following Barrett (2003); Ulph (2004) and Kolstad (2007), consider a situation where N identical countries each emit a uniformly mixed transboundary pollutant. Country i 's welfare is

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

where q_i is equal to one if i abates its emissions and zero if it does not, q_{-i} is the sum of the abatement decisions made by all other countries, b is the constant benefit of abatement by i or any other country, c is the constant individual cost of abatement, and A is a positive constant. Assume that $b < c$ so that no country will unilaterally abate its emissions, but that $Nb > c$ so that the countries' joint welfare will be maximized when they all abate their emissions.

Recognizing the benefits from agreeing to control their emissions, countries have the incentive to form an IEA to do so. As in Ulph (2004) and Kolstad (2007), the formation of an IEA is modeled as a two-stage game. In the first stage, each country must decide independently whether to become a party to an IEA. Assume that countries have complete information regarding each others' welfare functions and that they can form only a single cooperative coalition. In the second stage, all countries choose whether to abate their emissions. Countries that become members of an IEA in the first stage commit to decisions in the second stage that maximize their joint welfare, given the decisions of the non-member nations. In this section we assume perfect compliance with the agreement so that members

of the IEA do not have to be monitored for compliance. While the parties to an IEA carry through on their commitments to maximize their joint welfare in the second stage, non-member countries maximize their individual welfare. Since the individual cost of abatement outweighs the benefit for each country (i.e., $b < c$), all countries that are not members of an IEA will decide to not abate their emissions in the second stage.

Given the decision of all non-member countries to not abate their emissions, the members of an IEA will abate their emissions in the second stage as long as the coalition is *profitable* in the sense that each member of the coalition is at least as well off as when all countries choose not to control their emissions. Let s denote the number of parties to an IEA who all agree to abate their emissions. Moreover, let $w^p(s)$ denote the welfare of each of the parties to the IEA and let $w^{np}(s)$ denote the welfare of each of the countries that are not party to the agreement. Throughout the paper 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:

$$\begin{aligned} w^p(s) &= A + bs - c; \\ w^{np}(s) &= A + bs. \end{aligned} \tag{2}$$

Since $b < c$, $Nb > c$, and $w^p(s)$ is increasing in s , there exist coalition sizes that are strictly greater than one and weakly less than N that are profitable. The smallest of these profitable coalitions is

$$s_{\min}^{nc} = \min \{s | w^p(s) \geq w^{np}(0)\} = \min \{s | s \geq c/b\}. \tag{3}$$

Throughout, the superscript nc denotes values when the decisions of the members of an IEA do not require costly monitoring (i.e., compliance is assumed perfect). If $s \geq s_{\min}^{nc}$ from the first stage of the game, the members of the coalition will agree to abate their emissions in the second stage of the game. If $s < s_{\min}^{nc}$, the members of the IEA maximize their joint welfare by not abating their emissions in the second stage.

The subgame perfect equilibrium of the game is called a *self-enforcing agreement* in the IEA literature. Intuitively, a self-enforcing agreement is one for which no party to the agreement wants to leave the agreement and no non-party wishes to join. More formally:

Definition 1 An IEA consisting of s countries that comply with the agreement without the need for costly monitoring is self-enforcing if and only if:

$$\begin{aligned} (i) & w^p(s) \geq w^{np}(s - 1) \text{ for } s > s_{\min}^{nc}, \text{ or } w^p(s_{\min}^{nc}) \geq w^{np}(0) \text{ for } s = s_{\min}^{nc} \\ (ii) & w^{np}(s) \geq w^p(s + 1). \end{aligned} \tag{4}$$

Requirement (i) is that the agreement is *internally stable* in the sense that no party has an incentive to leave a self-enforcing agreement. Requirement (ii) is that the agreement is *externally stable* in the sense that no non-party wishes to join the IEA. The only internally stable coalition is the one with $s = s_{\min}^{nc}$ members. To see why, note that for $s > s_{\min}^{nc}$, (2) reveals that $w^p(s) \geq w^{np}(s - 1)$ implies $b \geq c$, which violates our assumption that $b < c$. On the other hand, an IEA with s_{\min}^{nc} members is internally stable, because if one member of the coalition defected then the remaining members would find it unprofitable to abate their emissions. The defector’s welfare (as well as every other country’s welfare) would then be $w^{np}(0)$, which is weakly less than $w^p(s_{\min}^{nc})$. Finally, an IEA with s_{\min}^{nc} members is also externally stable. Using requirement (ii) and the welfare functions (2), $w^{np}(s_{\min}^{nc}) \geq w^p(s_{\min}^{nc} + 1)$ implies $b \leq c$, which is satisfied because $b < c$. Since an agreement consisting of s_{\min}^{nc} members is the only internally and externally stable coalition size, it is the equilibrium number of members of a self-enforcing IEA.

Proposition 1 *Let the equilibrium number of members of a self-enforcing agreement that does not require costly monitoring be s^{nc} . Then, s^{nc} is the smallest profitable coalition; that is,*

$$s^{nc} = s_{\min}^{nc} = \min \{s | s \geq c/b\}. \quad (5)$$

The equilibrium number of members of an IEA that does not require costly monitoring is unique, strictly greater than 1, and weakly less than N . Note that [5] indicates that participation with an IEA is increasing in the cost of abatement, c , and decreasing in the individual benefit of abatement, b . 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 the coalition members' decisions to comply with the terms of the agreement once they have joined. Indeed, if countries are able to observe each others' compliance decisions perfectly and without cost, then a member country has no incentive to join an agreement and then violate its requirements. If it did so all the other participating nations would observe this violation and would realize that they would be worse off if they continued to abate their emissions. Thus, under perfect information a member of a self-enforcing IEA would not violate the terms of the agreement because this would cause the agreement to collapse.

In reality, however, nations are not able to observe the compliance behavior of others so easily. Under imperfect information about individual nations' compliance behavior, a country may be motivated to join an agreement and then decide to violate its terms. If the other participating countries do not observe this act of noncompliance, they will not automatically respond by refusing to abate their emissions. The violator may then be able to escape the cost of compliance while enjoying the benefit of cooperation of those that continue to abate their emissions.

To demonstrate the incentive parties have to violate the terms of the IEA under imperfect information regarding compliance behavior, assume that a member of an IEA with s members can fail to abate its emissions without being detected. If the party chooses to be noncompliant and the other member countries continue to abate their emissions because they can not observe this violation, the violator's payoff is $w^{nP}(s-1) = A + b(s-1)$. If it instead remained compliant its payoff is $w^P(s) = A + bs - c$. Subtracting $w^{nP}(s-1)$ from $w^P(s)$ yields $b - c$, which is negative by assumption indicating that member countries always have an incentive to violate an agreement as long as they believe that the other members are unable to observe their violation. However, with homogenous countries, each should expect that all the others will have the same motivation to violate an IEA as they do—if one member of an agreement has an incentive to violate an agreement, then all members do. Thus, under asymmetric information about countries' abatement decisions and without monitoring to check on these decisions, no self-enforcing agreement exists (except for the trivial case of $s = 0$). This motivates the implementation of a monitoring mechanism to check on the compliance decisions of the members of an IEA and a sanction to punish detected violators.

3 Self-enforcing IEAs when Compliance is Costly to Monitor

We now modify the basic model of a self-enforcing IEA to give its members the opportunity to decide whether to comply with the terms of their agreement. We assume that members cannot observe each other's compliance decisions and, therefore, each of them has an incentive to violate the requirements of the agreement in the absence of adequate monitoring. To

counteract this incentive the parties of an IEA finance an independent monitor who audits their compliance decisions. A fixed sanction is applied when violations are detected.

As before the model is analyzed as a stage game, but now the game has four stages. The first stage (*membership stage*) proceeds exactly as before; that is, each country chooses independently whether to join the IEA. In the second stage (*treaty stage*), members of the IEA jointly agree on whether to abate their emissions and, if they do agree to abate their emissions, each of them is required to contribute funds to the independent monitor. If IEA members jointly decide not to control their emissions in stage two, they do not fund the monitor, an effective IEA does not form, and the game ends. (Throughout, we refer to an *effective* IEA as one that actually leads to emissions control). If an effective IEA forms in stage two, all countries make their abatement decisions independently in stage three (*abatement stage*). Finally, in the fourth stage (*enforcement stage*) the monitor randomly audits the abatement decisions of member countries with the funding provided to it in the second stage, and a sanction is applied when the monitor finds a violation. Since the treaty, abatement, and enforcement stages are new with this study, we now describe them in detail. Because the game is solved by backward induction, we start with the last stage.

3.1 Enforcement Stage

If the game reaches this stage, an IEA with s members has formed, each member has agreed to abate their emissions, each member has contributed x dollars to fund the monitor, and all countries have made their abatement decisions. In the enforcement stage, the monitor randomly audits the abatement decisions of the IEA members with a probability that is an increasing function of the amount of funding the members contribute. Each additional dollar of funding allows the number of random audits to increase by $\alpha > 0$; 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 monitor, then the number of random audits conducted is $sx\alpha$, and the probability that any party is audited is

$$\pi = sx\alpha/s = x\alpha \quad (6)$$

We will demonstrate shortly that an effective IEA will involve a monitoring probability in the half-closed interval $(0, 1]$.

Monitoring of the countries' compliance behavior, however, may 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. Given an audit, let $\rho^1 \in [0, 1)$ denote the probability that a compliant country is judged to be noncompliant (a Type I error), and let $\rho^2 \in [0, 1)$ denote the probability that a noncompliant country is judged to be compliant (a Type II error).⁹ These probabilities are common knowledge.

In practice, there are a variety of ways noncompliant countries may be sanctioned. Much of our analysis in this regard is motivated by the enforcement mechanism of the Kyoto Protocol established under the Marrakesh Accords. These established an enforcement body (called the Compliance Committee) with the power to sanction noncompliant parties by reducing their greenhouse gas emissions quota for the next commitment period. Therefore, we assume that a party that is revealed to be noncompliant by the monitor incurs a known exogenous sanction of f . Others have criticized the enforcement strategy of the Kyoto Protocol as being non-credible (e.g. Barrett 2003), because a party to an IEA who decides not to abate its

⁹ See Kaplow and Shavell (1994) for the effects of monitoring inaccuracy in a standard model of optimal law enforcement.

emissions could just as easily choose not to pay the sanction if its violation is discovered. Moreover, a party that is found to be in violation of the requirements of an IEA could simply leave the agreement.

In Sect. 6 we consider another type of sanction that cannot be avoided so easily. There is strong agreement in the economic and political science literature that a party's willingness to jointly manage transnational resources has value beyond a single treaty (Keohane 1986; Chayes and Chayes 1991; Simmons 1998; Barrett 2003). Complying with international agreements builds a country's reputation as a cooperating member of the international community, and that reputation allows it to be a part of other beneficial agreements. Chayes and Chayes (1991, p. 320) introduce the flip-side of this argument: "A reputation for unreliability cannot be confined to the area of activity in which it is earned. It is inevitable that a state's defection from treaty rules will generate repercussions and linkages throughout the network of its relationships with others in the community. And these more diffuse responses can be calibrated more finely than formal sanctions." We would add that if failing to comply with an IEA deteriorates a country's reputation, then not participating with an agreement in first place is also likely to be damaging. In Sect. 6 we model sanctions for agreement violations and for non-participation as costly reputation effects, and show that our qualitative conclusions about the impacts of costly monitoring on self-enforcing IEAs are largely preserved with this approach.¹⁰

3.2 Abatement Stage

At this point in the game the members of an IEA have agreed to abate their emissions and have funded the monitor. In the abatement stage both members of an IEA and non-members independently choose whether to control their emissions. Since the individual cost of abatement outweighs the individual benefit, each non-member will decide not to abate their emissions in this stage. Member countries, however, make this decision by comparing the expected cost of not complying with the agreement and the benefit of noncompliance.

Assume that the parties to an IEA are risk neutral and that they comply if they are at least indifferent between compliance and noncompliance. Then, given an effective agreement consisting of s parties, in the abatement stage an individual country will comply if its expected welfare from doing so is not less than its expected welfare from noncompliance. Parties to the agreement cannot observe the others' compliance choices, so one party's violation of its commitment to abate its emissions will go unnoticed unless it is discovered by the monitor. Given the probability of an audit π , the probabilities of Type I and Type II monitoring errors ρ^1 and ρ^2 , and the sanction f , it is straightforward to show that a party's expected welfare from compliance is $w^P(s) - x - \pi\rho^1 f$, while its expected welfare from noncompliance is $w^{NP}(s-1) - x - \pi(1 - \rho^2)f$, where recall that $w^P(s)$ and $w^{NP}(s-1)$ are defined by (2). Note that $\pi\rho^1 f$ is a compliant party's expected sanction when it is falsely judged to be noncompliant and $\pi(1 - \rho^2)f$ is a noncompliant party's expected sanction from a correct determination that it is noncompliant.¹¹ A party to an IEA complies with the terms of the

¹⁰ A third possible form the sanction for noncompliance could take is a restriction on the 'club' benefits parties receive from belonging to an IEA. For example, the International Monetary Fund does not allow noncompliant members to access the pool of funds jointly provided by the signatories to the agreement (Chayes and Chayes 1991). This arrangement is similar to posting a performance bond when ratifying an agreement, which is paid back to the party if they comply with the agreement.

¹¹ One may wonder why a compliant country that is incorrectly determined to be noncompliant (Type I error) would passively accept the sanction. 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,

agreement if and only if

$$\begin{aligned}
 & [w^p(s) - x - \pi\rho^1 f] - [w^{np}(s - 1) - x - \pi(1 - \rho^2) f] \\
 & = \pi f(1 - \rho^1 - \rho^2) - (c - b) \geq 0
 \end{aligned}
 \tag{7}$$

Clearly, $\pi f(1 - \rho^1 - \rho^2) \geq (c - b)$ is a necessary condition for an effective IEA. If this condition did not hold, no party to an IEA would comply with its requirements and an effective agreement would not form. Note further that no amount of monitoring will be sufficient to induce compliance if $f(1 - \rho^1 - \rho^2) < (c - b)$. This indicates that low sanctions or severe monitoring errors can prevent the formation of an effective IEA. For the remainder of this analysis we assume that $f(1 - \rho^1 - \rho^2) \geq (c - b)$ so that an effective IEA can form. Since $c > b$, this condition clearly requires $\rho^1 + \rho^2 < 1$.

3.3 Treaty Stage

In this stage members of an IEA agree to abate their emissions and to fund the monitor provided that these decisions maximize their joint welfare. Because we assume that each country’s abatement decision is not directly observable by other countries, members of an IEA cannot credibly commit to abatement. This, of course, is the reason for monitoring. Each country that wishes to join an effective agreement is required to contribute funds to the monitor. These payments are perfectly observed, so a country is not able to promise to make the payment and then fail to do so.

Each party’s contribution to the monitor, x , is endogenous, so we determine this value first. If an effective agreement is to form, each party would like to contribute as little as possible while providing the monitor with sufficient resources to maintain compliance with the agreement. This requires a payment x so that $\pi f(1 - \rho^1 - \rho^2) - (c - b) \geq 0$ binds. Since $\pi = x\alpha$ from (6), the contribution to the monitor of an IEA that is required of all parties to the agreement is

$$x = (c - b) / (\alpha f(1 - \rho^1 - \rho^2)).
 \tag{8}$$

Note that this payment decreases with the size of the sanction, f , and the marginal productivity of resources devoted to monitoring, α , but is increasing in the probabilities the monitor commits a Type I or Type II error, ρ^1 and ρ^2 , and the gain from noncompliance, $c - b$. Note further that $\pi = x\alpha = (c - b) / (f(1 - \rho^1 - \rho^2))$ is the equilibrium probability the monitor audits any member of an effective IEA in the enforcement stage of the game. Our assumptions that $c > b$, $\rho^1 + \rho^2 < 1$, and $f(1 - \rho^1 - \rho^2) \geq (c - b)$ guarantee that $\pi \in (0, 1]$.

Parties to an IEA will fund the monitor and jointly agree to abate their emissions only if they will be at least as well off as without an agreement. That is, again, abatement must be profitable for the members of the agreement. Given s parties to an agreement that each expects to earn $w^p(s) - x - \pi\rho^1 f = A + bs - c - x - \pi\rho^1 f$ if they abate their emissions, and an individual country’s welfare in the absence of an IEA, A , the minimum size profitable coalition is the smallest s such that $A + bs - c - x - \pi\rho^1 f \geq A$. Substituting for x from (8) and rearranging terms yields the smallest profitable coalition when an IEA requires costly

Footnote 11 continued

for example, if the enforcement branch of the Compliance Committee finds a party to be in violation, that party is entitled to a hearing at 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 (UNFCCC 2002). This process appears to be designed to limit the probability of a Type I error.

monitoring:

$$s_{\min}^c = \min \left\{ s \mid s \geq \frac{c + \pi\rho^1 f}{b} + \frac{c - b}{b\alpha f(1 - \rho^1 - \rho^2)} \right\}. \quad (9)$$

(The superscript ‘c’ is used to denote values when compliance with an IEA requires costly monitoring). If $s \geq s_{\min}^c$ from the first stage of the game (the membership stage), the members of the coalition will agree to abate their emissions in the treaty stage of the game. Furthermore, each member pays x from (8) to the monitor in this stage and will comply with the terms of the agreement in stage three (the abatement stage). On the other hand, if $s < s_{\min}^c$ from the first stage, the members of the IEA maximize their joint welfare by not abating their emissions and not funding the monitor. Clearly, if $s < s_{\min}^c$ an effective IEA does not form and the game concludes.

4 Properties of Self-enforcing IEAs that Require Costly Monitoring

The subgame perfect equilibrium of the stage game we have just described is either a self-enforcing agreement under which all members of the agreement abate their emissions, or an effective IEA does not form. For an IEA to form, profitable coalitions must exist, which they will as long as $N \geq s_{\min}^c$; that is, using (9), $N \geq (c + \pi\rho^1 f)/b + (c - b)/(b\alpha f(1 - \rho^1 - \rho^2))$. In the model without costly monitoring of Sect. 2, profitable coalitions always exist because we assumed at the outset that $N > c/b$. However, since $c > b$ and $\rho^1 + \rho^2 < 1$, c/b is strictly less than $(c + \pi\rho^1 f)/b + (c - b)/(b\alpha f(1 - \rho^1 - \rho^2))$. This implies that while profitable coalitions always exist when monitoring is not costly, they may not exist when parties to an IEA must fund a monitor to maintain compliance with the agreement. Therefore, we have our first conclusion about the impact of costly monitoring on self-enforcing environmental agreements:

Proposition 2 *The circumstances under which an effective IEA will form are diminished by the need for costly compliance monitoring.*

That $N \geq (c + \pi\rho^1 f)/b + (c - b)/(b\alpha f(1 - \rho^1 - \rho^2))$ is required for an effective IEA to form is simply a statement about whether international cooperation to limit a global pollutant can increase aggregate welfare when we take account of the costs of compliance monitoring. Note that $(c + \pi\rho^1 f)/b + (c - b)/(b\alpha f(1 - \rho^1 - \rho^2))$ is decreasing in the marginal productivity of monitoring resources, α , and the size of the sanction, f . Decreasing either of these parameters increases the payment cooperators pay to monitor an agreement, leading to a decrease in the set of opportunities for an effective IEA. Note also that $(c + \pi\rho^1 f)/b + (c - b)/(b\alpha f(1 - \rho^1 - \rho^2))$ is increasing in the probability the monitor commits either a Type I or Type II error, indicating that the presence of monitoring errors further reduces the circumstances under which an effective IEA will form. Finally, $(c + \pi\rho^1 f)/b + (c - b)/(b\alpha f(1 - \rho^1 - \rho^2))$ is increasing in c and decreasing in b ; hence, the set of circumstances under which an effective IEA will form is smaller when abatement costs are higher and when the individual benefit from abatement is smaller.

As with the IEA model without costly monitoring in Sect. 2, a self-enforcing IEA that requires costly monitoring is the coalition size that is both internally and externally stable. And like the model in Sect. 2, it is easy to demonstrate that the only internally and externally stable coalition is the smallest profitable coalition given by (9), provided that profitable coalitions actually exist. Therefore:

Proposition 3 *Let s^c be the equilibrium number of members of a self-enforcing agreement when monitoring for compliance with the agreement is costly. Then,*

$$s^c = s_{\min}^c = \min \{s \mid s \geq (c + \pi\rho^1 f)/b + (c - b)/(b\alpha f(1 - \rho^1 - \rho^2))\}, \tag{10}$$

provided that $N \geq s_{\min}^c$. If $N < s_{\min}^c$, there are no profitable coalitions and an effective self-enforcing agreement will not form.

Comparing Proposition 3 to Proposition 1 allows us to determine how costly monitoring affects the equilibrium number of members of an IEA. When an agreement with costly monitoring forms, (10) indicates that the equilibrium size of the coalition s^c is the least s for which $s \geq (c + \pi\rho^1 f)/b + (c - b)/(b\alpha f(1 - \rho^1 - \rho^2))$. When an IEA does not require costly monitoring, (5) indicates that the size of the coalition s^{nc} is the least s for which $s \geq c/b$. Recall that $1 - \rho^1 - \rho^2 > 0$ is required for an agreement with costly monitoring to form. Therefore, $(c + \pi\rho^1 f)/b + (c - b)/(b\alpha f(1 - \rho^1 - \rho^2)) > c/b$ implies $s^c \geq s^{nc}$, and our next proposition.

Proposition 4 *If an effective IEA that is costly to monitor forms, the number of parties to the agreement will be no less, and will typically be greater, than if the IEA did not require costly monitoring.*

The intuition behind Proposition 4 is straightforward. Since funding the monitor of an IEA is an additional cost of joining the agreement, more countries are required to participate and abate their emissions to make the agreement profitable. This produces the seemingly paradoxical result that costly monitoring of IEAs can produce higher environmental quality.

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 effective IEA that is costly to monitor increases with a country’s abatement cost, c , and decreases with the individual benefit of some country’s abatement, b . More importantly for the purposes of this paper, any change in an enforcement parameter that increases the costs of monitoring will result in greater participation, as long as the IEA remains viable. Thus, participation and environmental quality increase with decreases in the marginal productivity of monitoring resources, α , and the sanction for noncompliance, f . Moreover, because monitoring inaccuracy increases the costs of maintaining compliance with an effective IEA, participation and environmental quality increase with monitoring inaccuracy as long as this does not prevent the formation of an IEA.

5 Costly Monitoring of IEAs and Aggregate Welfare

In this section we examine the differences in aggregate welfare under self-enforcing IEAs that are costly to monitor and under those for which compliance can be enforced without cost. Costly monitoring produces two countervailing effects on aggregate welfare. The first effect is the cost of monitoring itself which, holding participation with an IEA constant, diminishes aggregate welfare. However, costly monitoring can produce greater participation and higher environmental quality, which increases aggregate welfare. We demonstrate that it is possible that the latter effect dominates the former so that costly monitoring of international cooperation actually produces greater aggregate welfare.

Given s^c parties to an agreement that each earn $w^p(s^c) - x - \pi\rho^1 f = A + bs^c - c - x - \pi\rho^1 f$, and $N - s^c$ free-riding countries that each earn $w^{np}(s^c) = A + bs^c$, aggregate

welfare when cooperation requires costly monitoring is

$$\begin{aligned} W^c(s^c) &= s^c(A + bs^c - c - x - \pi\rho^1 f) + (N - s^c)(A + bs^c) \\ &= NA + s^c(Nb - c - x - \pi\rho^1 f). \end{aligned} \tag{11}$$

Recall that the superscript ‘c’ identifies variables and functions when compliance with an IEA requires costly monitoring. For convenience, let us assume that s^c is continuous. Then, from (10), $s^c = (c + \pi\rho^1 f) / (b + (c - b) / (b\alpha f(1 - \rho^1 - \rho^2)))$ as long as it is not greater than N . Substitute this and $x = (c - b) / (\alpha f(1 - \rho^1 - \rho^2))$ from (8) into $W^c(s^c)$ to obtain

$$\begin{aligned} W^c(s^c) &= NA + \left(\frac{c + \pi\rho^1 f}{b} + \frac{c - b}{b\alpha f(1 - \rho^1 - \rho^2)} \right) \\ &\quad \times \left(Nb - c - \frac{c - b}{\alpha f(1 - \rho^1 - \rho^2)} - \pi\rho^1 f \right) \end{aligned} \tag{12}$$

Aggregate welfare when a self-enforcing IEA does not require costly monitoring is $W^{nc}(s^{nc}) = NA + s^{nc}(Nb - c)$. Allowing s^{nc} to be continuous, substitute $s^{nc} = c/b$ from (5) into $W^{nc}(s^{nc})$ to obtain

$$W^{nc}(s^{nc}) = NA + (c/b)(Nb - c). \tag{13}$$

Now subtract (13) from (12):

$$\begin{aligned} W^c(s^c) - W^{nc}(s^{nc}) &= \left(\frac{c - b}{b\alpha f(1 - \rho^1 - \rho^2)} + \frac{\pi\rho^1 f}{b} \right) \\ &\quad \times \left[N - \left(\frac{2c + \pi\rho^1 f}{b} + \frac{c - b}{b\alpha f(1 - \rho^1 - \rho^2)} \right) \right]. \end{aligned}$$

Since the first term of $W^c(s^c) - W^{nc}(s^{nc})$ is positive by assumption, the sign of $W^c(s^c) - W^{nc}(s^{nc})$ is equal to the sign of the term in hard brackets, which provides our next proposition.

Proposition 5 *Costly monitoring of an IEA results in higher aggregate welfare if and only if*

$$N > \frac{2c + \pi\rho^1 f}{b} + \frac{c - b}{b\alpha f(1 - \rho^1 - \rho^2)}. \tag{14}$$

Proposition 5 indicates that costly monitoring of an effective IEA produces higher aggregate welfare when the number of potential parties to the agreement is large enough.

Figure 1 illustrates how $W^c(s^c)$ and $W^{nc}(s^{nc})$ vary with the number of potential parties to an IEA and with the accuracy of compliance monitoring. To draw this graph we have assumed, without loss of generality, that $A = 0$. Aggregate welfare when an IEA does not require costly monitoring, $W^{nc}(s^{nc})$, is zero for $N \leq c/b$. For $N > c/b$, $W^{nc}(s^{nc})$ is positive and Eq. (13) indicates that it increases linearly at rate c . The size of a self-enforcing IEA does not change as N increases; it remains constant at c/b . However, aggregate welfare increases with N because increasing N means we are increasing the number of free-riding countries, each of which benefits from the abatement efforts of the c/b parties to the agreement.

The dashed function in Fig. 1 denoted $W^c(s^c | \rho^1 = \rho^2 = 0)$ is aggregate welfare for a self-enforcing IEA that is costly to monitor and monitoring is perfectly accurate. The dotted function, $W^c(s^c | \rho^1 + \rho^2 > 0)$, is aggregate welfare for a self-enforcing IEA that is costly to monitor and monitoring is inaccurate. Note that $W^c(s^c)$, with and without monitoring errors, is equal to zero for a larger range of N than $W^{nc}(s^{nc})$. This follows because cooperative

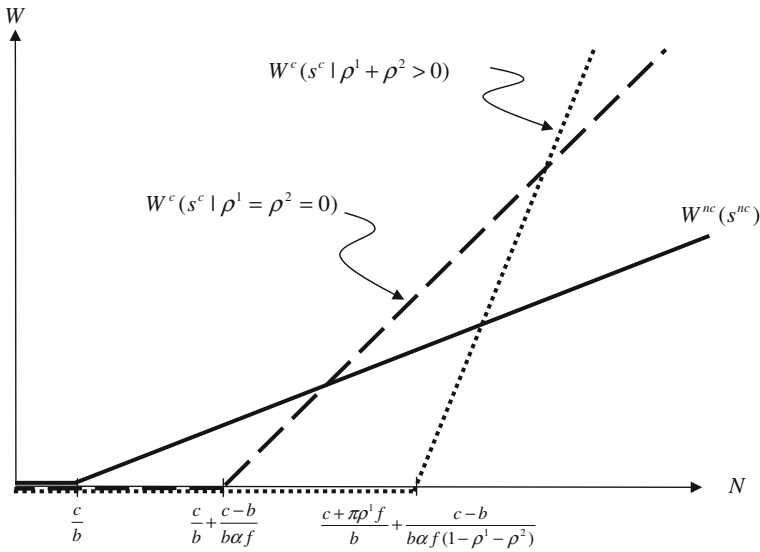


Fig. 1 A comparison of aggregate welfare under self-enforcing international environmental agreements that require costly monitoring and aggregate welfare under agreements that do not require costly monitoring

abatement efforts that require costly monitoring increase aggregate welfare under a smaller set of circumstances than when cooperation does not require monitoring (Proposition 2).

If monitoring is perfectly accurate and $N \geq c/b + (c - b)/(b\alpha f)$, the coalition $s^c = c/b + (c - b)/(b\alpha f)$ forms and aggregate welfare increases with N at rate $c + (c - b)/(\alpha f)$. For $N \geq c/b - (c - b)/(b\alpha f)$, the relationship between $W^c(s^c | \rho^1 = \rho^2 = 0)$ and $W^{nc}(s^{nc})$ depends on the positive welfare effect of higher abatement when an IEA must be monitored (because $s^c > s^{nc}$ from Proposition 4) and the negative welfare effect of the costs of monitoring. The level of N where $W^c(s^c | \rho^1 = \rho^2 = 0)$ and $W^{nc}(s^{nc})$ intersect is $2c/b + (c - b)/(b\alpha f)$. Therefore, for $N \in [c/b + (c - b)/(b\alpha f), 2c/b + (c - b)/(b\alpha f))$, the monitoring-cost effect dominates the higher-abatement effect so that aggregate welfare is lower when the equilibrium IEA requires costly monitoring. However, when N exceeds $2c/b + (c - b)/(b\alpha f)$ the higher-abatement effect dominates the monitoring-cost effect so that aggregate welfare is higher when compliance requires costly monitoring.

Monitoring errors further reduce the set of circumstance under which an effective IEA will form. For a large enough N , it is straightforward to show that $W^{nc}(s^{nc})$ and $W^c(s^c | \rho^1 + \rho^2 > 0)$ intersect at $N = (2c + \pi\rho^1 f)/b + (c - b)/(b\alpha f(1 - \rho^1 - \rho^2))$. For N greater than this value, costly and inaccurate monitoring of an IEA leads to higher aggregate welfare than if the IEA did not require monitoring. Moreover, Eq. (12) indicates that when $W^c(s^c | \rho^1 + \rho^2 > 0)$ is positive it rises more quickly than $W^c(s^c | \rho^1 = \rho^2 = 0)$. Therefore, monitoring inaccuracy results in higher aggregate welfare than under perfectly accurate monitoring if N is large enough.¹² The reason for this counterintuitive result is the now familiar monitoring cost/abatement tradeoff. Monitoring inaccuracy reduces aggregate welfare because it leads to higher monitoring costs, but inaccuracy also increases aggregate welfare because it leads to higher abatement. The latter effect dominates when the number of potential parties to an IEA is large enough.

¹² It is straightforward to show that $W^c(s^c | \rho^1 + \rho^2 > 0)$ and $W^c(s^c | \rho^1 = \rho^2 = 0)$ intersect at $N = (2c + \pi\rho^1 f)/b + (c - b)(2 - \rho^1 - \rho^2)/(b\alpha f(1 - \rho^1 - \rho^2))$.

6 Self-enforcing IEAs with Costly Monitoring and Reputation Effects

Up to this point we have assumed that an enforcement body of an IEA can impose a sanction on noncompliant members. Given the possibility that a violator may simply refuse to pay a sanction, which we discussed in Sect. 3, we now examine the consequences of an alternative model that includes reputation sanctions that cannot be avoided. Modeling costly reputation effects that result from violating the terms of an IEA forces us to also consider the possibility that countries that refuse to join an IEA in the first place may also suffer a deterioration of their reputation as cooperative members of the international community. Therefore, we now let f denote a sanction that is the value of a member country's reputation loss when it violates an IEA, and let g be the value of a country's reputation loss when it refuses to join an IEA. Since it is unreasonable to assume that all countries suffer a loss of reputation when no country joins an IEA, assume that $g = 0$ when an IEA does not form.¹³

The noncompliance reputation loss, f , enters our costly monitoring model in the same way as before. Thus, all changes in our results are due solely to the inclusion of the non-participation reputation loss, g . Moreover, as we demonstrate below, the primary effect of this non-participation sanction is that it can be large enough to induce full participation with an IEA that does not require costly monitoring. It is only in this case that there are substantive changes in the results we have obtained thus far. On the other hand, when a model of costless monitoring and a non-participation sanction predicts less than full participation, all of our qualitative results about the impacts of costly monitoring on the levels of IEA participation, environmental quality, and aggregate welfare continue to hold.

We start by revising the welfare functions in (2) to include the non-participation sanction g . For $s \geq 1$, the new welfare functions when compliance with an IEA does not require costly monitoring are now:

$$\begin{aligned} w^P(s) &= A + bs - c; \\ w^{nP}(s) &= A + bs - g. \end{aligned} \quad (15)$$

As before, $w^P(0) = w^{nP}(0) = A$. Note that

$$w^P(s) - w^{nP}(s-1) = \begin{cases} b + g - c, & \text{for } s > 1 \\ b - c < 0, & \text{for } s = 1. \end{cases}$$

Suppose for now that $b + g - c \geq 0$. Then $w^P(s) \geq w^{nP}(s-1)$ for all $s > 1$, indicating that if more than one country joins an IEA, then all other countries will join as well. By assumption, no country will unilaterally abate its emissions because $b - c < 0$; however, since $Nb > c$, each country is better off if they all form an IEA to abate their emissions than if none do. Therefore, when the reputation loss from not participating with an IEA is large enough so that $b + g - c \geq 0$, a self-enforcing IEA that does not require costly monitoring for compliance will involve all countries.

Suppose now that the non-participation loss is not so large; that is, suppose that $b + g - c < 0$. As with the model of Sect. 2, since $b < c$, $Nb > c$, and $w^P(s)$ is increasing in s , there exist coalition sizes that are strictly greater than one and weakly less than N that are profitable. Furthermore, using (15) and noting that $w^{nP}(0) = A$, the smallest of these profitable

¹³ We thank an anonymous reviewer for suggesting that we investigate the impacts on our model of non-participation and noncompliance reputation sanctions. We should note that these sanctions need not be reputation losses, but could take on other forms. For example, the Montreal Protocol deters non-participation by imposing trade restrictions on ozone-depleting chemicals between parties and non-parties. These same trade restrictions can also be used to punish noncompliant parties (see Barrett 1998b, p. 36).

coalitions is $s_{\min}^{nc} = \min \{s | w^p(s) \geq w^{np}(0)\} = \min \{s | s \geq c/b\}$, which is identical to the minimum sized profitable coalition in the model of Sect. 2 (see Eq. (3)). Moreover, this is the only internally and externally stable coalition. Hence, when $b + g - c < 0$, Proposition 1 continues to hold so that the equilibrium number of members of an IEA that does not require costly monitoring is $s^{nc} = \min \{s | s \geq c/b\}$. The only modification to Proposition 1 needed when non-participation produces a reputation loss is to state that $s^{nc} = N$ when $b + g - c \geq 0$.

Let us now determine how reputation sanctions affect the self-enforcing number of members of an IEA that requires costly compliance monitoring. As detailed in Sect. 3, the game entails four stages that proceed in exactly the same way, except that now if a member is determined to be noncompliant, the monitor makes this information public and the violator suffers a reputation loss, the value of which is f . Despite this re-interpretation of f , it enters the model exactly as in the model of Sect. 3. Therefore, a party of an IEA complies with its terms if and only if $\pi f(1 - \rho^1 - \rho^2) \geq c - b$ (Eq. (7)), and provided that $f(1 - \rho^1 - \rho^2) \geq (c - b)$, each member country must contribute $x = (c - b)/(\alpha f(1 - \rho^1 - \rho^2))$ to the monitor to maintain compliance with the agreement (Eq. (8)). Moreover, the minimum size profitable coalition is the same s_{\min}^c that is specified in Eq. (9), and profitable coalitions exist if $N \geq s_{\min}^c$.

The main effect of modeling reputation sanctions for both violations and non-participation is that the non-participation sanction may be large enough to produce full compliance with an IEA, just as in the model without costly compliance monitoring. From Sect. 3, a party's expected welfare from complying with an IEA with s members that requires costly monitoring is $A + bs - c - \pi\rho^1 f - x$. If instead the country decides not to be a member of the IEA it suffers the reputation loss of g and earns $A + b(s - 1) - g$ (provided that $s > 1$). The IEA obtains full participation if $A + bs - c - \pi\rho^1 f - x \geq A + b(s - 1) - g$, which reduces to $b + g - c - (\pi\rho^1 f + x) \geq 0$. Recall that without the need for costly monitoring, full participation is realized when $b + g - c \geq 0$. Since $\pi\rho^1 f + x > 0$, if the non-participation reputation loss is large enough to promote full participation with costly compliance monitoring, then it is large enough to promote full participation when costly monitoring is not needed. Therefore, when $b + g - c - (\pi\rho^1 f + x) \geq 0$, the self-enforcing IEA involves all countries whether compliance monitoring is costly or not.

However, in all cases in which $(b + g) - c - (\pi\rho^1 f + x) < 0$, the non-participation loss is not sufficient to induce full compliance with an IEA. Under these circumstances, it is straightforward to demonstrate that Propositions 2 and 3 of Sect. 4 are unchanged. That is, costly monitoring limits the circumstance under which an effective IEA forms, but when one does form its membership is s^c , as specified in Eq. (10). However, because the non-participation reputation loss can produce full participation with an IEA, we must modify Proposition 4 when this sanction is present. There are two circumstances under which an effective IEA that requires costly monitoring does not involve higher participation than if it did not require costly monitoring. One is the case in which the non-participation reputation loss is so large that an IEA involves full participation whether compliance monitoring is costly or not. Recall that this is the case when $b + g - c - (\pi\rho^1 f + x) \geq 0$. The other case is when $b + g - c - (\pi\rho^1 f + x) < 0$, but $b + g - c < 0$. In this case, an IEA would involve full participation if it did not require costly monitoring, but there would be less than full participation if it does require costly compliance monitoring. When both noncompliance and non-participation incur reputation losses, costly compliance monitoring of an effective IEA produces greater participation only when the non-participation reputation loss is low enough so that $b + g - c < 0$.

The fact that a loss for not participating with an IEA can produce full participation also causes us to modify our conclusions in Sect. 5 about the welfare effects of costly monitoring. There we noted that because costly compliance monitoring can increase the number of

participants and environmental quality, it can also produce higher aggregate welfare as well because the increase in environmental quality offsets the costs of monitoring under some circumstances. With both noncompliance and non-participation sanctions, this result continues to hold as long as participation is less than complete when costly monitoring is not needed. However, when a non-participation sanction is large enough to induce full compliance when monitoring is not needed, participation and environmental quality cannot be greater when costly monitoring is required. In this case costly monitoring unequivocally reduces aggregate welfare because the costs of monitoring an IEA will never be offset by an increase in environmental quality.

7 Concluding Remarks

We have analyzed games of self-enforcing IEAs when parties to such agreements finance an independent institution to monitor compliance with the IEA. We have shown that costly monitoring limits the circumstances under which international cooperation to control a trans-boundary pollutant will increase aggregate welfare. However, when an effective IEA does form, participation with the agreement and environmental quality can be greater than when an IEA does not require costly monitoring. In fact, under some circumstances, costly monitoring of an IEA is associated with higher aggregate welfare as well—not for the parties to an IEA, however, but for the countries that choose to free ride on the agreement.

Most of our results are robust to the problem of inaccurate monitoring and to different types of sanctions. We demonstrate that monitoring inaccuracy can further reduce the set of international environmental problems that can be addressed with an environmental agreement, but when an agreement is worthwhile, monitoring inaccuracy can produce higher participation and environmental quality. We also examined two types of sanctions—one that is imposed on noncompliant parties by an enforcement body, while the other involves reputation losses that are incurred by both noncompliance with an agreement and by not participating with the agreement in the first place. The qualitative results of both specifications are the same, except when the non-participation loss is so large that it induces full participation with an agreement.

Our results have important implications for multilateral management of environmental externalities. Assuming away monitoring and enforcement problems related to IEAs artificially enlarges the scope of mutually beneficial agreements between countries, and may artificially reduce the number of participating countries that are necessary to make these agreements worthwhile. In other words, for some international environmental problems, monitoring costs either require more members for an IEA to come into effect or they render mutual cooperation inefficient.

The models developed in this paper can and should be extended to include additional features. An obvious extension would be to recognize the real-world heterogeneity of countries involved in international environmental relations. Other features of international cooperation such as side payments, issue linkage, and minimum participation requirements that others have addressed, can and should be examined in the context of costly monitoring of these agreements. Certainly there are many extensions that should be addressed to gain a more complete picture of how costly monitoring affects voluntary coalition formation to protect the international environment.

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. The authors gratefully acknowledge the helpful comments of James Murphy, John Spraggon, Erin Baker, and Tim Perri. We would also like to thank Michael Rauscher and two anonymous reviewers for their helpful comments.

References

- Arguedas C (2007) To comply or not to comply? Pollution standard setting under costly monitoring and sanctioning. *Forthcoming in Environ Resour Econ*
- Barrett S (1994) Self-Enforcing International Environmental Agreements. *Oxf Econ Pap* 46(1):878–894
- Barrett S (1997a) The strategy of trade sanctions in international environmental agreements. *Resour Energy Econ* 19(1):345–361. doi:[10.1016/S0928-7655\(97\)00016-X](https://doi.org/10.1016/S0928-7655(97)00016-X)
- Barrett S (1997b) Heterogeneous international environmental agreements. In: Carraro Carlo (ed) *International Environmental Negotiations*. Edward Elgar, Cheltenham, UK and Northampton, MA
- Barrett S (1998a) On the theory and diplomacy of environmental treaty making. *Environ Resour Econ* 11(1):317–333. doi:[10.1023/A:1008243528330](https://doi.org/10.1023/A:1008243528330)
- Barrett S (1998b) Political economy of the Kyoto protocol. *Oxf Rev Econ Policy* 14(4):20–39. doi:[10.1093/oxrep/14.4.20](https://doi.org/10.1093/oxrep/14.4.20)
- Barrett S (2003) *Environment and statecraft: the strategy of environmental treaty-making*. Oxford University Press, Oxford and New York
- Benedick RE (1998) *Ozone diplomacy: new directions in safeguarding the planet*, enlarged edition. Harvard University Press, Cambridge, MA and London, England
- Botteon M, Carraro C (1998) Strategies for environmental negotiations: issue linkage with heterogeneous countries. In: Hanley N, Folmer H (eds) *Game Theory and the Environment*. Edward Elgar, Cheltenham, UK and Northampton, MA
- Carraro C, Siniscalco D (1993) Strategies for the international protection of the environment. *J Public Econ* 52(3):309–328. doi:[10.1016/0047-2727\(93\)90037-T](https://doi.org/10.1016/0047-2727(93)90037-T)
- Carraro C, Botteon M (1997) Burden sharing and coalition stability in environmental negotiations with asymmetric countries. In: Carraro C (ed) *International environmental negotiations*. Edward Elgar, Cheltenham, UK and Northampton, MA
- Carraro C, Siniscalco D (1997) R&D Cooperation and the stability of international environmental agreements. In: Carraro C (ed) *International environmental negotiations*. Edward Elgar, Cheltenham, UK and Northampton, MA, pp 1–2
- Carraro C, Siniscalco D (1998) International environmental agreements: incentives and political economy. *Eur Econ Rev* 42(3–5):561–572. doi:[10.1016/S0014-2921\(97\)00118-9](https://doi.org/10.1016/S0014-2921(97)00118-9)
- Carraro C, Marchiori C, Orefice S (2003) Endogenous minimum participation in international environmental treaties. *Nota Di Lavoro* 113.2003. Fondazione Eni Enrico Mattei Working Paper
- Chandler P, Tulkens H (1995) A core-theoretic solution for the design of cooperative agreements on trans-frontier pollution. *Int Tax Public Finance* 2(2):279–293. doi:[10.1007/BF00877502](https://doi.org/10.1007/BF00877502)
- Chandler P, Tulkens H (1997) The core of an economy with multilateral environmental externalities. *Int J Game Theory* 26(3):379–401
- Chayes A, Chayes AH (1991) Compliance without enforcement: state behavior under regulatory treaties. *Negotiation J* 7(1):311–330. doi:[10.1111/j.1571-9979.1991.tb00625.x](https://doi.org/10.1111/j.1571-9979.1991.tb00625.x)
- D’Aspremont C, Jacquemin A, Gabszewicz JJ, Weymark JA (1983) On the stability of collusive price leadership. *Can J Econ* 16(1):17–25. doi:[10.2307/134972](https://doi.org/10.2307/134972)
- Downs GW, Rocke DM, Barsoom PM (1996) Is the good news about compliance good news about cooperation? *Int Organ* 50(3):379–340
- Finus M (2004) *International cooperation to resolve international pollution problems*. Fondazione Eni Enrico Mattei Working Paper
- Finus M, Rundshagen B (1998) Toward a positive theory of coalition formation and endogenous instrument choice in global pollution control. *Public Choice* 96(1–2):145–186. doi:[10.1023/A:1005016623915](https://doi.org/10.1023/A:1005016623915)
- Folmer H, van Mouche P, Ragland S (1993) Interconnected games and international environmental problems. *Environ Resour Econ* 3(4):313–335. doi:[10.1007/BF00418815](https://doi.org/10.1007/BF00418815)
- Friends of the Earth (2006) *Climate treaty one year old, but emissions still rising*. Press Release, February 15, 2006 http://www.foe.co.uk/resource/press_releases/climate_treaty_one_year_of_14022006.html
- Heister J, Mohr E, Stahler F, Stoll P-T, Wolfum R (1997) Strategies to enforce compliance with an international CO₂ Treaty. *Int Environ Aff* 9(1):22–53
- Hoel M (1992) International environment conventions: the case of uniform reductions of emissions. *Environ Resour Econ* 2(2):141–159

- Hoel M, Schneider K (1997) Incentives to participate in an international environmental treaty. *Environ Resour Econ* 9(2):153–170
- Kaplow L, Shavell S (1994) Accuracy in the determination of liability. *J Law Econ* 37(1):1–16. doi:[10.1086/467304](https://doi.org/10.1086/467304)
- Keohane RO (1986) Reciprocity in international relations. *Int Organ* 40(1):1–27
- Kolstad C (2007) Systematic uncertainty in self-enforcing international environmental agreements. *J Environ Econ Manage* 53(1):68–79. doi:[10.1016/j.jeem.2006.08.001](https://doi.org/10.1016/j.jeem.2006.08.001)
- Macho-Stadler I, Perez-Castrillo D (2006) Optimal enforcement policy and firm's emissions and compliance with environmental taxes. *J Environ Econ Manage* 51(1):110–131. doi:[10.1016/j.jeem.2005.06.001](https://doi.org/10.1016/j.jeem.2005.06.001)
- Oberthur S (1998) *The International convention for the regulation of whaling: from over-exploitation to total prohibition*. Yearbook of international co-operation on environment and development. Earthscan Publishers, London
- Osmani D, Tol RSJ (2007) *Toward farsightedly stable international environmental agreements: part one*. Hamburg University Working Papers, FNU-140
- Polinsky AM, Shavell S (2000) The economic theory of public enforcement of law. *J Econ Lit* 38(1):45–76
- Simmons BA (1998) Compliance with international agreements. *Annu Rev Polit Sci* 1(1):75–93. doi:[10.1146/annurev.polisci.1.1.75](https://doi.org/10.1146/annurev.polisci.1.1.75)
- Stranlund JK (2007) The regulatory choice of noncompliance in emissions trading programs. *Environ Resour Econ* 38(1):99–117. doi:[10.1007/s10640-006-9058-3](https://doi.org/10.1007/s10640-006-9058-3)
- Swanson T, Johnston S (1999) *Global environmental problems and international environmental agreements: the economics of international institution building*. Edward Elgar, Cheltenham, UK and Northampton, MA
- Ulph A (2004) Stable international environmental agreements with a stock pollutant, uncertainty and learning. *J Risk Uncertain* 29(1):53–73. doi:[10.1023/B:RISK.0000031445.13939.e4](https://doi.org/10.1023/B:RISK.0000031445.13939.e4)
- 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
- Wagner UJ (2001) The design of stable international environmental agreements: economic theory and political economy. *J Econ Surv* 15(3):377–411. doi:[10.1111/1467-6419.00143](https://doi.org/10.1111/1467-6419.00143)