INTERNATIONAL ENVIRONMENTAL AGREEMENTS: EMISSIONS TRADE, SAFETY VALVES AND ESCAPE CLAUSES

Larry Karp et Jinhua Zhao

Presses de Sciences Po | *Revue économique*

2010/1 - Vol. 61
pages 153 à 182

ISSN 0035-2764

International Environmental Agreements: Emissions trade, safety valves and escape clauses*

Larry Karp
University of California, Berkeley

Jinhua Zhao
Michigan State University

November 6, 2009

Abstract

We explain how the structure of multi-national or multi-regional environmental agreements affect their chance of success. Trade in emissions permits has ambiguous and in some cases surprising effects on both the equilibrium level of abatement, and on the ability to persuade nations or regions to participate in environmental agreements. An escape clause policy and a safety valve policy have essentially the same properties when membership in environmental agreement is pre-determined, but they create markedly different effects on the incentives to join such an agreement. The two policies lead to a qualitative difference in the leverage that a potential member of the agreement exercises on other members.

Keywords: Kyoto Protocol, escape clause, emissions trade, cost uncertainty, participation game, International Environmental Agreement,.

JEL classification numbers C72, H4, Q54

*An earlier version of this paper benefited from comments at a December 2008 workshop, convened at UC Berkeley, on the European Commission Transatlantic Biosafety and Biodiversity collaborative research project.

Contents

1 Introduction

This paper brings together elements of the theory of international environmental agreements, trade policy, and environmental policy in an attempt to help advance the currnent international climate negotiations. International environmental agreements must be designed to encourage participation, achieve abatement efficiently, and create incentives for compliance. A design feature that addresses one of these challenges can make other challenges more difficult to meet. Although international trade of emissions permits contributes to efficient abatement, it may be inimical to participation, and do nothing to encourage compliance. Emissions trading together with a safety valve creates insurance against high costs and also helps to achieve efficient abatement, but again it may not be help in achieving either participation or compliance. We explain the role of an escape clause, which relieves a country of the responsibility to abate if it incurs some other cost that benefits signatories, e.g. pays a fine or agrees to the withdrawal of WTO-mandated trade concessions. This alternative can promote participation and provide insurance against high costs, although it does not help in achieving efficient abatement. An IEA that combines trade in permits and an escape clause may help in creating a successful IEA. It is reasonable to use carefully circumscribed trade restrictions, such as border tax adjustments, as part of an IEA. These restrictions will help to undercut a politically attractive argument for staying outside an agreement: the concern that carbon leakage will dilute most of the global gains from the agreement, while eroding participants' ability to compete in carbon-intensive sectors. The adoption of trade rules will also decrease the ability of pressure groups to demand protection once a country has joined the agreement. Finally, multilateral trade rules will mitigate the problem of carbon leakage, to the extent – currently difficult to determine – that this leakage has the potential to be a genuine problem.

We begin by discussing the broad issue of using trade policy to achieve environmental outcomes (Sections 2). Contrary to the "orthodox" position in trade economics, we think that trade policy, if used carefully, can help to achieve environmental objectives without undermining the international trade system. Section 3 discusses the relative merits of carbon taxes and a cap and trade scheme. We then explain the effects of allowing international trade in emissions permits (Section 4). Here we consider two types of situations. If the regions across which the trade occurs do not harmonize their policies, trade in permits can undercut the environmental policy in one of the regions. The more favorable case for trade in permits arises where the different regions do harmonize their policies. Even in that case, however, trade in permits might either increase or decrease incentives to impose a strict environmental policy. That is, even though trade in permits reduces abatement costs, it does not necessarily increase the equilibrium amount of abatement. Trade in permits is not necessarily "environmentally friendly". We briefly consider the case where abatement costs change endogenously with the trade regime; this possibility arises if, for example, the incentive to invest in abatement capital depends on whether permits will be tradable.

We then discuss two types of policy proposals in the case were countries' abatement costs are random: an escape clause and a safety valve (Section 5). Under the escape clause, a signatory to an IEA has the option to pay a fine rather than meet its abatement target, but there is no trade in permits. Under the safety valve, IEA members are allowed to trade permits. The IEA defends a price ceiling by distributing enough emissions permits to maintain the equilibrium price at a level no greater than the price ceiling. When the number of IEA members is pre-determined, these two policies can achieve the same distribution of environmental outcomes (abatement) and of aggregate welfare. In this respect, the two policies are (essentially) equivalent.

However, the escape clause and the safety valve have different effects on countries' incentives to join an IEA. In order to explain this difference, we take a detour, in Section 6, to review the workhorse of the economic theory of IEAs. The theory shows that self-interested countries join an IEA in order to influence other countries' decisions, i.e. in order to gain leverage over them. Therefore, a fundamental question in determining how the design of the IEA affects equilibrium membership is "How does this design affect the leverage that a country gains when it joins the IEA?" The theory identifies in a simple manner the obstacles to cooperation, thereby helping to understand how the structure of an IEA alters those obstacles.

Trade in emissions permits might serve two fundamentally different roles (Section 7). First, trade, together with the initial distribution of permits, can be a means of providing side-payments that induce countries to join the IEA. We doubt the practical importance of this possibility. Second, trade might be a means of reallocating permits in response to the realizations of random cost shocks. Although trade in permits certainly reduces expected abatement costs, conditional on the number of signatories, trade is likely to decrease the incentives for a country to join the IEA, and therefore to reduce the equilibrium size of the IEA.

The two policy proposals, the escape clause and the safety valve, have fundamentally different effects on the equilibrium size of the IEA. The safety valve does not promote membership in the IEA, and can even reduce it. A well-designed but extremely simple escape clause can increase equilibrium membership. However, trade in permits enable countries to take advantage of low cost abatement opportunities, a feature absent under the escape clause. The safety valve that we consider in detail uses a monetary fine to encourage signatories to meet their abatement target. Trade sanctions provide an alternative to the fine. The discussion of these trade sanctions brings us back to the types of issues covered in Section 2.

We emphasize the strategic elements involved in the formation of IEAs. In the final section, we recognize that there are several important benefits of IEAs unrelated to strategic considerations.

2 Trade policy and environmental policy

Environmentalists and trade economists have debated for years whether the WTO mandate should be extended in order to attempt to influence environmental policy (Guzman 2002) (Cone 2002). Until recently at least, trade economists have been broadly united in opposing such an extension. Under WTO rules this extension would require the agreement of all members, an unlikely outcome given the opposition of many countries. Even if such a consensus could be obtained, economists' general view has been that the gains from further trade liberalization are large, the current trade rules are vulnerable to erosion, and moreover the dangers of climate change are modest. As long as this perception persists, it makes sense to quarantine environmental objectives from trade policy. An alternate view is that globalization of world markets is on a sound footing and that the risks of climate change are substantial. Under this view, it makes sense to use trade policy to achieve environmental goals.[1]

Trade policy can be linked to environmental policy in order to encourage participation in an agreement; it can also be used to encourage signatories to comply and to prevent non-signatories from undermining the agreement. The most aggressive use of trade policy, requiring participation in a climate agreement in exchange for access to markets, is unlikely to be productive. Especially for developing countries, positive incentives such as the right to participate in the Clean Development Mechanism market and easier access to green technology, are more useful ways to encourage membership.

However, trade taxes do have a role in discouraging nations from undermining an international agreement. In 2007 some EU politicians proposed a "Kyoto tax", aimed primarily at the US. This tax would offset the estimated cost advantage that the US enjoys because of its decision to withdraw from the Kyoto Protocol. World Bank simulations show that the tax would have caused a substantial reduction in US exports to the EU (World Bank 2008). The unilateral imposition of the Kyoto tax would likely have contravened WTO law and would have been politically unproductive; EU Trade Commissioner Mandelson opposed the tax.

Carbon leakage is the process by which stricter emissions standards in one place encourage higher emissions elsewhere, as production of dirty goods move to places with weaker environmental standards. The proposed unilateral tax was a response to carbon leakage, a problem that many people believe currently exists and one that would likely become more severe under a stronger climate agreement. The magnitude of carbon leakage is uncertain; some estimates

[1] Karp and Zhao (2009) discusses this issue in greater detail.

put it at less than 20% (Intergovernmental Panel on Climate Change 2001) Carbon leakage is an example of the "pollution haven effect" for which the empirical evidence is mixed but generally quite weak (Copeland and Taylor 2003), (Copeland and Taylor 2004). Perhaps this evidence is weak because the magnitude of the historical change in environmental policy has not been great enough to have an appreciable effect on the location of industry; the quality of the evidence may also be a consequence of statistical and measurement problems.

Trade policy can be used to persuade policymakers in signatory countries that they can achieve substantial reductions in GHG emissions without significant leakage and the accompanying job loss ((Stiglitz 2006)). This reorientation of trade policy carries with it the well-recognized risk of protectionist policies disguised as environmental policies ("environmental protectionism"), for which there is no simple inoculation. The extent of this risk is a matter of dispute.

The evidence of the past thirteen years shows that the WTO is capable of taking a nuanced view of the relation between trade and environmental protection (Neumeyer 2005). WTO dispute resolution panels have been willing to reject some policies ostensibly aimed at environmental health, e.g. in the disputes between the US and the EU involving hormone-fed beef and GMOs. However, the WTO does not instinctively regard environmental policies that restrict trade as disguised protectionism. The Appellate Board's 1998 decision in the "shrimp and turtle" case recognized the legality of trade restrictions used to protect the global environmental commons. Some trade economists were concerned that this decision would lead to environmental protectionism against developing countries (Bhagwati 2004), but there is little evidence of this occurring.

The shrimp-turtle decision provides modest but insufficient scope for using trade policy to achieve environmental goals within the confines of existing WTO law. Moreover, precedence has little force in WTO and other international law. Border tax adjustments (BTA) offer one means of protecting against carbon leakage and the loss of competitiveness due to environmental regulation. The efficacy of *unilateral* BTAs may be limited by the form of domestic policy. WTO law allows countries to provide export subsidies or impose import tariffs that offset certain kinds of domestic taxes. "Indirect" (or "product") taxes, which include taxes on inputs, can be adjusted at the border, whereas "direct" (or "process") taxes, such as a payroll or a profits tax cannot be adjusted. Pauwelyn (2007) explains the why the legality of using a BTA to offset a domestic carbon tax is uncertain under WTO rules. Matters become even more (legally) complicated if domestic carbon regulation uses a cap and trade or command and control policies.

One route to introducing trade policy as a lever for international climate policy is to revise the WTO. This revision would probably be opposed by most developing countries. A more practical way to introduce the trade lever is to include it in an environmental agreement rather than as part of a reform of the WTO. Trade sanctions are legal under current WTO law, provided that the sanctions are imposed against a signatory to the environmental agreement and are consistent with that agreement. The agreement should entitle signatories to impose a border tax adjustment against other signatories in order to offset a cost disadvantage, above some minimal level, caused by stricter climate policies. Climate change policies might lead to small cost increases over much of the primary and secondary sectors; the minimal level, below which costs increases are ineligible for a border tax adjustment, will prevent a general increase in tariffs. In practice, the sectors that face substantially higher costs because of environmental measures receive offsetting subsidies designed to protect the sector. These offsetting subsidies must be included in the calculation of the border tax, so that this tax provides compensation for the environmental policy's net costs to the domestic industry. This modification limits the ability to use the border tax adjustment as a means of disguised protectionism. Signatories have the right to object to border tax adjustments using the WTO dispute resolution process. Thus, the border tax adjustment protects against carbon leakage for only the sectors that face the greatest costs of complying with climate policy.

In view of the requirement to include offsetting subsidies in calculating a border tax adjustment, and the discipline imposed by the WTO dispute resolution process, the aggregate effect of the border tax adjustment would likely be small. However, the political effect could still be substantial, by making it harder to argue that leakage undercuts domestic reductions in GHG emissions and harms domestic industries.

Under our proposal, the border tax adjustment can be used only against signatories to an agreement. It would most likely be used when trading partners have very different climate-related obligations, as with trade between developed and developing countries. This fact makes it particularly important to obtain developing country participation in an environmental agreement, using the positive incentives mentioned above. Even if developing countries are not required to reduce their emissions in the short run, the border tax adjustment limits their ability to take advantage of reductions in emissions in developed countries. It therefore undercuts a major argument that the US has used for remaining outside the Kyoto Agreement.

3 The efficiency of different form of regulation

We briefly consider the merits of different forms of environmental regulation, ignoring issues raised by international trade in permits. We discuss those issues in the next section.

Environmental policy A is more efficient than a second policy B if it is cheaper to achieve a given level of environmental protection (e.g. a reduction in emissions) under policy A than

B. The Theory of the Second Best cautions that in general it might not be possible to compare the efficiency of two policies when there are externalities or other market failures, in addition to the environmental problem under consideration. However, apart from these “second best” arguments, there is broad agreement that market based policies such as taxes or cap and trade are more efficient than non-market based policies such as standards. An active field of research compares different market based policies in circumstances where these are not able to achieve the first best outcome of a fully informed regulator.

Several papers compare tax with cap and trade regulation of stock pollutants such as GHGs when regulators are imperfectly informed about firms’ costs (Hoel and Karp 2001), (Hoel and Karp 2002), (Newell and Pizer 2003), (Karp and Zhang 2006). The consensus from this literature is that taxes are more efficient than cap and trade policies for controlling GHGs. There are several reasons for this conclusion, but the most fundamental has to do with the relative magnitudes of the slopes of the marginal damage curve for GHG stocks and the marginal abatement costs for GHG emissions.

The qualitative conclusion that taxes are more efficient than cap and trade is quite robust to changes in parameter values. However, these results are based on strong assumptions about the functional forms of abatement costs and environmental damage, in particular the assumption that uncertainties do not affect the slope of the marginal cost and benefit curves. The results are not robust to generalizations with respect to functional form, and for that reason they are valuable more as a source of intuition rather than as a basis for categorical advice. A more complicated policy, e.g. an optimal two-part tax (equivalent to a cap and trade with a price ceiling), could obviously dominate either the tax or the cap and trade policy.

Even if taxes are a more efficient means of reducing GHGs (Nordhaus 2007), political considerations may make a cap and trade policy the more practical alternative (Stavins 2008). Public distaste for taxes, the lack of public understanding of the price effects of the two types of policies, and the political power of carbon-intensive industries, may favor the use of a cap and trade policy.

4 The basics of international trade in permits

Kyoto negotiators initially disagreed whether to allow international trade in permits. The argument in favor of trade mirrors the argument for market based policies in a closed economy: trade increases economic efficiency, i.e. it reduces aggregate abatement costs by shifting abatement to countries where costs are lower. The most compelling arguments against trade are based on the practical concern that it undermines emissions reductions by making it possible for some countries to sell bogus emissions credits.[2] The Kyoto Protocol resolved the dispute by allowing limited trade in permits. The same debate now arises in considering emissions trade in the next version of the Kyoto Protocol, and in considering emissions trade among countries or regions that belong to different agreements (the Kyoto Protocol, the Regional Greenhouse Gas Initiative in Eastern US, and the Western Climate Initiative involving Western US states and Canadian provinces.)

Different issues arise depending on whether trade occurs across countries or regions that set policies independently ("non-harmonized policies"), or whether it occurs within a region where countries coordinate their policies ("harmonized policies").

4.1 Non-harmonized policies

Pollution emissions are implicitly a factor of production: commodities embody pollution in the same way that they do labor or land or other factors of production. Trade in commodities is implicitly trade in factors. A famous theorem in trade theory, the Factor Price Equalization (FPE) theorem, provides conditions under which trade in commodities is a perfect substitute for trade in factors. When these conditions hold, trade in emissions permits achieves no added efficiencies, beyond those obtained from trade in commodities. That is, for an arbitrary allocation of emissions permits, under FPE the market reallocates production in order to insure that marginal abatement costs (equal to the value of marginal product of emissions) are equal across countries (Copeland and Taylor 2005).

It is extremely unlikely that the conditions of the FPE theorem exactly hold. Moreover, the cost of transporting an emissions permit from one country to another is essentially zero, whereas the cost of transporting the commodities that embody emissions is always positive. Therefore, we expect there to be some, although possibly small efficiency gain from trade in emission permits, when permits are initially allocated without regard to marginal abatement cost in different countries.

FPE is likely to be more closely approximated amongst more similar countries, e.g. those with similar relative factor endowments; it is also likely to be more closely approximated where other factors, such as capital and labor, are more mobile across countries, e.g. where there are fewer restrictions on foreign investment or labor migration. Thus, trade in permits may yield

[2] There are two less persuasive reasons for opposing trade in permits. One type of reason is based on a story of market failures. For example, domestic emissions reductions might create external benefits, e.g. from learning spillovers, that are lost if the country can buy abatement abroad. This kind of argument is not convincing in the abstract, because it can be turned on its head to argue that a different market failure results in too little trade in permits. A second type of reason for objecting to emissions trade is based on a moral absolutism that insists that emissions reductions occur domestically. Consequentialist ethics rejects that view.

smaller efficiency gains when IEA members are more similar and allow greater factor mobility.

In a general equilibrium setting, trade in permits can have more subtle effects. First, trade in emissions permits alters production costs and changes the supply of commodities, thus changing the commodity prices. If emissions trade leads to a sufficiently large fall in the price of a country's exports, or a rise in the price of its imports, the welfare loss resulting from the deterioration in the terms of trade might more than offset the direct welfare gain from trade in permits. In this case, trade in permits harms an IEA member. Of course, if the price of exports rises (or the price of imports fall) due to trade in permits, the country certainly benefits from emissions permit trade.

Finally, if there are distortions in the economy, the Theory of the Second Best alerts us to the possibility that trade in permits has ambiguous welfare effects. This observation is useful only if we have a reasonable idea of which existing distortions are likely to be altered by trade in permits.

If the degree and the type of environmental regulation differs across countries, free trade in emissions permits can undermine the stricter regulation. This effect is similar to but more pronounced than the carbon leakage described in Section 2. If two countries both restrict emissions using a cap and trade scheme, and if the equilibrium price of emissions permits in the absence of permit trade differs in the two countries (so that FPE does not hold) then trade in permits increases efficiency when the emissions caps remain fixed. However, trade can create an incentive for one or both countries to weaken their regulation in order to increase their supply of permits, thus increasing their gain from trade in permits.

If one country uses a cap and trade scheme and the other country imposes a tax, trade occurs (if permitted) whenever the autarchy equilibrium price of permits under the cap and trade scheme differs from the tax. If the autarchy permit price is lower than the tax, the cap and trade country exports permits. In this case there is a gain in efficiency and a reduction in emissions. Emissions remain constant in the tax-imposing country, because producers' marginal cost of emissions (the tax) has not changed; but emissions in the cap-and-trade country fall, since that country exports some of its emissions permits. In addition, tax revenue in the tax-imposing country falls. However, if the autarchy permit price is higher than the tax, the cap and trade country imports permits from the tax-imposing country until the permit price falls to the level of the tax. In this case, trade increases tax revenue in the tax-imposing country and reduces quota rents in the cap and trade country. More seriously, trade undermines the environmental regulation in the cap and trade country, leading to a net increase in emissions.

These examples illustrate that when policies in the different countries are not harmonized, unfettered trade in permits can easily undermine environmental policy. Since trade in permits also occurs implicitly by means of commodity trade, the latter trade can also undermine environmental policy – a point already made in Section 2. It is worth emphasizing that these remarks pertain to global (or transboundary) environmental problems, such as arise with GHGs. If the pollution externality remains within national borders, trade considerations are (relatively) unimportant.

4.2 Harmonized policies

Here we consider the opposite extreme, where a group of countries jointly set the (constrained) environmental policy to maximize collective welfare. The membership in the group is taken as given; in a later section we endogenize membership. We also take the countries' abatement costs as given; later we explain how trade might affect endogenous investment decisions that determine abatement costs. Finally, we assume that conditions for FPE do not hold, and that countries have different abatement costs.

If permits could be allocated in order to equalize marginal abatement costs, there is no advantage to trade in permits. However, in any real-world circumstance, members of an agreement are unlikely to be able to achieve such a perfect allocation. Even if there existed a single decision-maker whose objective is to maximize collective welfare, imperfect information might make the first best allocation impossible. Countries would have an incentive to overstate their abatement costs, in order to be given a larger allocation of permits. Therefore, we assume that the group allocates an equal number of permits to each member; this limitation constrains the choice of policy.

Given these assumptions, allowing trade in permits increases efficiency, for the reason given in Section 3. Because we assume that FPE does not hold, and we ignore other general equilibrium considerations (such as endogenous commodity prices) we use a partial equilibrium setting here. Permit trade makes it cheaper to achieve any level of abatement, when the countries have different marginal abatement costs. This fact might make it appear that trade in permits promotes a greater level of abatement. In this sense, trade appears to be "environmentally friendly".

This conjecture need not be true. If the IEA's objective is to maximize the amount of GHG reduction (= abatement) subject to a ceiling on total abatement costs, then trade would certainly be environmentally friendly. However, if the IEA's objective is to maximize welfare, then the optimal level of abatement depends on marginal, not on average abatement costs. Although trade certainly lowers total and average costs (when countries have different costs), it can either lower or increase marginal costs. Thus, the effect of trade on the incentive to abate is ambiguous in general (Karp 2008). When permit trade increases the equilibrium level of pollution, trade in permits is "environmentally unfriendly". The importance of this observation is that *trade in* *permits should be defended on the ground that it reduces aggregate abatement costs, not that it promotes increased abatement.*

A simple example (based on our work-in-progress) shows why trade might reduce the incentive to abate. Suppose that each country has the ability to abate at most one unit. Marginal costs in each country are constant, up to the one-unit capacity. Each country obtains 1 unit (e.g. $100 billion) of benefit for one unit of abatement that occurs anywhere in the world; this is a global pollutant. Each country's cost is a random variable with support greater than 1 and expected value equal to 2. (Costs are in the same units as benefits, so 2 represents $200 billion.) To keep the example simple, suppose that the costs are independently distributed.

After having decided whether to participate in the agreement, a country learns its own costs, but this information is not verifiable, so the group decision cannot be conditioned on individual cost realizations. Countries would have an incentive to exaggerate their costs, in order to be given a lower required level of abatement. Since the benefit to a country of abatement is 1 and the actual cost is always greater than 1, it is not in any country's self-interest to abate.

Suppose that there are 3 signatories and that an agreement requires each signatory to abate at its maximum level. We are now at a stage where the signatories decide collectively whether to modify or even eliminate the target. We assume that a supranational agent decides whether to adjust or eliminate the target in order to maximize the collective expected welfare of the signatories, ignoring the welfare of non-signatories.

First consider the case where international trade in permits is not permitted, because for example the countries did not develop the institutional structure needed to govern this trade. In this case, it is in the collective interest of the signatories to carry out the agreement: the expected total costs are $3 \times 2 = 6$ and the total benefits are $3 \times 3 = 9$.[3] Suppose instead that the countries are able to trade permits amongst themselves. In this situation, the decision problem is more complicated. The collective marginal benefit of a unit of abatement is 3 (since each of the three participants obtains the marginal benefit of 1). If the expectation of the highest cost (the third order statistic) is greater than 3, and the expectation of the second highest cost (the second order statistic) lies between 2 and 3, then it is optimal for the countries to agree on only two units of abatement. Since costs are not verifiable, the group requires each country to achieve $2/3$ of a unit of abatement, and allows them to trade in order to achieve this target. In this example, the ability to trade reduces the total level of abatement by 33%.

By choosing different numerical values, it is possible to reverse the conclusion, so that trade increases the amount of abatement. As before, suppose that there are three members, each with constant marginal benefit of abatement of 1. In this example, however, the expected

[3]The assumption of contant marginal costs and constant marginal benefits mean that in the absence of trade it is always in the interest of the countries that ratify the agreement to either abate to capacity or not to abate at all.

costs per country per unit of abatement is 4 rather than 2. The total benefit of each unit of abatement equals 3 (as before) but in the absence of trade per unit expected costs equal 4. It is in the collective interests of the countries to decide to abate nothing, when trade is prohibited. However, provided that the expectation of the lowest of three independent random cost draws (the first order statistic) is less than 3, it is optimal for the IEA to set an abatement target of (at least) 1/3 per country. In equilibrium, only the low cost country abates (when the per-country target is less than or equal to 1/3). This country has expected costs of less than 3 so expected benefits are positive. In this version of the example, there is no abatement when trade is prohibited; trade makes it optimal to achieve a positive level of abatement.

4.3 Endogenous abatement costs

When equilibrium abatement targets depend on industry abatement costs, and when those costs depend on previous lumpy investment decisions (e.g. construction of low-emission plants) the trade regime assumes a new importance. The competitive equilibrium at the stage where firms make their (lumpy) investment decisions depends on whether trade in permits will be permitted in the future (Karp 2008).

Suppose that firms anticipate that trade in permits will not be allowed in the future. Firms know that if many firms make the emissions-reducing investment, then industry-wide abatement costs will be quite low, leading to relatively strict environmental regulation in the future. The anticipated strict regulation makes the lumpy investment profitable. In this case, firms have an incentive to follow the pack: actions are "strategic complements" at the investment stage. Each firm wants to make the same decision that the bulk of firms make. In this case, there are (typically) two stable pure strategy competitive equilibria at the investment stage: all firms or no firms invest. Corresponding to these two investment equilibria, the equilibrium abatement target in the next stage is either high or it is low. Neither of these competitive equilibria are efficient.

Suppose instead that firms anticipate that trade in permits will be allowed in the future. It is still the case that the equilibrium abatement target is greater, the larger is the fraction of firms that make the emissions-reducing investment decision. With trade, however, the equilibrium price of permits falls with an increase in investment. With anticipated trade in permits, actions are strategic substitutes at the investment stage. In this case, there is a unique competitive equilibrium that consists of the socially optimal fraction of firms making the investment decision. The resulting abatement target is also socially optimal (first best).

In short, when investment decisions determine abatement costs, and when environmental policy is conditioned on those costs, trade in permits fundamentally changes the investment equilibrium. In line with the previous subsection, however, there is no presumption that trade in permits leads to a higher level of abatement, compared to no trade. Again, the environmental effect of trade in permits, taking into account the endogeneity of abatement costs, is ambiguous.

4.4 Endogenous enforcement

For cap and trade policies to effectively restrict GHG emissions, emission levels of individual firms have to be adequately monitored and enforced to match up with the permits these firms hold. Unlike the case of SO2 in the US where emissions are continuously measured at the source, carbon emissions are most likely measured through input uses and technological choices. Indirect measurement leaves more room for mis-reports, leading to imperfect monitoring and enforcement (M&E). Imperfect M&E is equivalent to a government issuing more permits than its intended level of emissions (Malik 1990) and (Stranlund and Chavez 2000). Under reasonable assumptions, the marginal abatement costs are still equalized through permit trading with lax and asymmetric M&E.

Even without international permit trading, a signatory has an incentive to loosen its M&E to suboptimal levels, as doing so allows it to free-ride on other nations' abatement. International permit trading execerbates this incentive if in equilibrium the country exports permits. If all the signatories start out with the same marginal abatement costs, and one nation reduces its M&E to a level below those of other nations, then this nation will be a net exporter of permits.

There are several approaches to overcome the incentives to loosen M&E. Third parties such as internationally licensed agents can be employed to monitor the emission levels of individual firms and "certify" the authenticity of their permits. This approach is likely to incur significant transaction costs when many firms participate in permit trading. While allowing firms to trade permits individually, a signatory nation may be required to reconcile its aggregate emissions with its aggregate permit holding, assuming that national emissions can be reliably estimated through input uses. If its aggregate emissions are found to exceed the permit holdings of all its domestic firms, the signatory has to make up the difference by purchasing permits in the international market.

Some commentators have suggested replaceing the traditional seller liability with buyer liability in order to encourage more M&E (Victor 2001). This change makes permit buyers more reluctant to purchase from signatories with lax M&E, effectively discounting the permit prices from those nations. Buyers have incentive to make sure sellers' emissions are properly monitored and enforced, and seller nations have incentive to raise M&E to minimize the discounting from buyers. The interests of buyers and seller nations are aligned towards higher M&E levels. In contrast, under seller liability, buyers do not benefit from sellers' additional M&E; in fact, buyers benefit if lax M&E reduces the permit prices.

5 Escape clauses versus safety valves

This section compares two types of policies, a safety valve and an escape clause. Here we explain the sense in which these two policies have a similar effect on outcomes. Section 8 discusses their differences. We assume that abatement costs are exogenous (e.g. we ignore investment). We also take group membership as predetermined, an assumption that we relax in Section 7. As above, the group wants to maximize collective welfare; informational or political constraints prevent it from allocating more emissions permits to high cost countries.

In order to provide a simple setting for the comparison of an "escape clause policy" and a "safety valve policy", we adopt the same model that we used for the numerical example in Section 4.2. Marginal benefits of abatement are constant. We are interested in a global pollutant, so each country's benefit depends on aggregate abatement. Each country has constant abatement costs up to an upper bound, normalized to be one unit of abatement. Countries' abatement costs are independent draws of a random variable. Countries are therefore ex ante identical (before learning their abatement costs), but they obtain different costs. The social planner knows the distribution from which the costs are drawn, but does not know the individual costs. This assumption reflects the asymmetry of information between individual countries and a global planner.

The escape clause policy prohibits trade in permits amongst countries. The policy requires that each country abate at a given level or pay a fine. When we consider the effect of the escape clause on equilibrium group membership, it becomes important to specify what is done with the revenue from the fine. Here, where we take membership as predetermined, we do not need that level of detail. We assume that the revenue from the fine is used efficiently, and individual countries ignore the end to which the revenue is put. Those assumptions, which we later drop, merely simplify comparison of the two policies when membership is predetermined.

The safety valve policy also stipulates a given level of abatement. By abating in excess of that level, a country obtains permits that it can sell to other countries that abate below the stipulated level. Moreover, the policy imposes a price ceiling. The planner defends this ceiling by distributing enough (possibly zero) additional permits, to insure that the equilibrium permit price does not exceed the ceiling. Those permits might be given freely to members, or sold at a price equal to the ceiling.

The linearity of this model means that there is no loss in generality in assuming that the (nominally) required amount of abatement per country is the maximum level, 1. With the escape clause, the value of the fine can be chosen so that only countries with sufficiently low costs actually abate. With the safety valve, the price ceiling can be chosen so that only countries with sufficiently low costs actually abate.

Both policies put a ceiling on the country's actual costs, and in that respect they provide insurance against a high cost draw. This ceiling equals the fine in the case of an escape clause, and it equals the price ceiling in the case of the safety valve. In order to make the comparison clear, suppose that the fine equals the price ceiling, so the maximum cost is the same under the two types of policies. For any set of cost realizations, the total amount of abatement is therefore the same under the two policies: under both policies, only those countries who receive cost draws lower than the ceiling abate in equilibrium. Therefore the two policies have the same environmental effect.

Provided that the revenues from the fine are redistributed efficiently, the two policies lead to the same level of efficiency. The distribution of individual country costs under the two policies differ. With the escape clause, a country's actual cost depends only on its cost realization and the level of the fine. With the safety valve, a country's cost also depends on other countries' costs, since those costs determine the equilibrium price. However, in view of the assumed risk neutrality of agents (implicit in the linearity of the model) and the assumption that the fine revenue is used efficiently, aggregate welfare under the two policies is the same.

Thus, given the assumptions described above, there is very little substantive difference between the two policies when the level of IEA membership is predetermined. However, we want to explain how different types of polices affect a country's decision whether to join an IEA. In order to achieve this objective we need a model of countries' participation decisions.

6 Review of "standard model" of IEAs

We begin by reviewing a simple version of the "standard model" of IEA formation (Barrett 2003). This review is useful for two reasons. First, the standard model makes an important point: countries decide to join an IEA to affect the decisions of other countries, i.e. *to exercise leverage*, not (in general) to alter their own decision.[4] In the standard model of IEAs, countries first decide whether to join the agreement. Conditional on the number and characteristics of members, the IEA chooses the level of abatement to maximize members' joint welfare. This model identifies obstacles to cooperation and shows how various policy changes might magnify or reduce those obstacles. We begin with the simplest version in which countries are identical, there is no uncertainty, and the countries play a one-shot game. We then relax some of these assumptions as we consider different types of agreements.

[4]Of course, in some cases joining an IEA provides a degree of committmemt, which can be useful when a country faces time-consistency problems.

Since GHGs are a global pollutant, abatement is a global public good. In the symmetric setting, all countries benefit equally from a unit of abatement. In the special case that we consider, marginal benefits of abatement are constant, here normalized to 1. Each country can produce at most one unit of abatement at constant marginal costs θ, with $\theta > 1$. This inequality means that it is a dominant strategy for a single country not to abate. There are a total of N countries, where $N > \theta$, so increased abatement always increases aggregate welfare. If M of these countries join an IEA, the IEA chooses the per country level of abatement $a \leq 1$ to maximize members' joint welfare:

$$\max_{a \leq 1} M \left(M - \theta\right) a.$$

Define $h\left(x\right)$ as the smallest integer not less than x. The linearity of the maximand means that the IEA sets $a = 0$ if $M < h(\theta)$ and it sets $a = 1$ if $M \geq h\left(\theta\right)$.

At the participation stage of the game each country makes a binary choice: it either joins the IEA or stays out of it. At the abatement stage, nonmembers use their dominant strategy of not abating, and members follow the IEA's instructions. A non-cooperative Nash equilibrium to this game is both "internally stable" (no member wants to leave the agreement) and "externally stable" (no non-member wants to join it), taking as given other countries' participation decisions. There are two Nash equilibria: $M = 0$ and $M = h\left(\theta\right)$. Hereafter we consider only the second of these. To verify that $h\left(\theta\right)$ is a Nash equilibrium, note that each signatory's payoff is non-negative; no signatory wants to defect, because the resulting IEA would choose not to abate, leaving the defector with a zero payoff. No non-signatory wants to join, because the additional cost exceeds the additional benefit: $\theta > 1$.

This model implies that IEAs are effective only when they are unimportant; they achieve a substantial fraction of potential gains from cooperation only when potential gains are small. A reduction of abatement costs, θ need not increase welfare. More formally, we see:

1. The level of membership, $h\left(\theta\right)$, weakly increases with the membership cost, defined as the abatement cost minus the benefit of increased abatement, $\theta - 1$.

2. Global welfare is non-monotonic in abatement costs, θ. When $N \geq h\left(\theta\right)$, global welfare (the sum of members' and non-members' welfare) in the Nash equilibrium is $\left(N - \theta\right) h\left(\theta\right)$. As θ increases between integers abatement costs increase but there is no change in membership, so welfare falls; welfare has an upward jump as θ passes through an integer value, since this change leads to a discrete increase in membership and an infinitesimal increase in costs.

3. If all nations were compelled to join the IEA, global welfare (the "potential gain" from

cooperation) is $(N - \theta) N$. Relative to this grand IEA, the fraction of potential welfare achieved in equilibrium is $\frac{h(\theta)}{N}$, a non-decreasing function of costs.

4. The equilibrium membership does not depend on the number of potential members.

6.1 Dynamics

The game described above uses a static setting, in which nations decide only one time whether to participate in the IEA and then whether to abate. If the game lasts many periods and if agents care about the future, a higher level of participation can be sustained as a subgame perfect noncooperative Nash equilibrium. This more favorable outcome uses a "threat" to revert to the equilibrium in which no nation abates, following any defection. A defection occurs if any agent deviates from equilibrium play. One type of defection is for an agent not to abate when the equilibrium action is to abate.

Renegotiation proofness requires that all IEA participants would actually want to punish a defector, rather than returning immediately to cooperative behavior. Although the "good equilibria" – those with many abaters – are indeed subgame perfect, they are not, in general, "renegotiation proof" unless the number of members is quite small.

To confirm this observation, suppose that there are M IEA participants and that each has the discount factor $0 < \delta < 1$. Suppose that an equilibrium strategy requires that non-defectors (e.g. countries that have abated) punish a defector (e.g. a country that did not abate, when the equilibrium strategy called for it to abate) by not abating for a single period. during this periods the defector must abate in order for cooperation to resume in the next period. Requiring punishment to last more than a single period further reduces the equilibrium number of participants. A non-defecting IEA member prefers to carry out this punishment rather than resuming cooperation immediately if and only if

$$1 + \frac{\delta}{1-\delta} (M - \theta) \geq \frac{1}{1-\delta} (M - \theta).$$

The left side of this inequality is the payoff in equilibrium (a single period of punishment followed by cooperation) and the right side is the payoff if cooperation resumes immediately. This inequality requires $M \leq 1+\theta$ in a renegotiation-proof equilibrium. Since $1+h(\theta) \geq 1+\theta$, we see that membership in a renegotiation-proof equilibrium to the dynamic game exceeds the equilibrium in the static game by at most one country.

This result highlights the fact that the pessimistic conclusions of the standard model reflect the intrinsic difficulty of getting sovereign nations to provide public goods, together with the limitations inherent in the structure of the IEA as described by this game. The pessimistic conclusions are not due to the unrealistic assumption that the game lasts a single period. Hereafter, we use the single period setting, because its simplicity promotes clarity.

6.2 Regional agreements

It might seem that an agreement among a subset of nations would undermine the prospects for a broader agreement amongst many nations. The regional agreement imposes abatement costs on its participants, without exerting any leverage over non-participants. If the members' abatement capacity is constrained, as in our model, or if there are increasing marginal abatement costs, the regional agreement might make it more difficult for the members to entice non-members into joining a broader agreement.

The model above shows why regional agreements can actually promote greater global cooperation. This conclusion is a consequence of Result 4 above. For example, suppose that $h(\theta) < \frac{N}{2}$. In this case, if there are two participation games, each with $\frac{N}{2}$ potential IEA members, then the equilibrium size of each IEA is still $h(\theta)$. In this case, the combined number of abaters is $2h(\theta)$. Here, it is possible to achieve nearly complete cooperation simply by splitting up the pool of participants. This result shows that a regional agreement need not decrease the incentives of other countries to abate (i.e. to form other agreements).

7 The effect of emissions trade on IEA participation

Trade in emissions permits can affect equilibrium participation in an IEA. We consider two situations. "Ex ante" here refers to the status quo at the time of the participation decision. In the subsection "ex ante heterogenous countries" we assume that cost differences are common knowledge at the participation stage. In this case, trade in permits, together with the initial allocation of permits, provides an implicit side payment that can be used to induce countries to join the agreement. The more subtle and interesting case arises with "ex ante homogenous countries" where each country's abatement cost is an independent draw from a known distribution (as in Section 5); the random variables are realized after the participation decision. In this situation, trade can reduce participation in the IEA.

7.1 Ex ante heterogenous countries

Here we turn to the question of how trade affects the incentives to join an IEA when countries have known differences in abatement costs at the time they decide whether to join the IEA, and moreover the IEA can distribute permits in any manner desired We think that this amount of verifiable information and flexibility is empirically implausible, so we are skeptical of the practical importance of the results from this model. Nevertheless, the case where the potential gains from trade are known ex ante is worth considering because it provides an upper bound to the possibility of using trade in permits to encourage participation. Allowing trade, together with a judicious distribution of permits, makes it possible to increase membership. The distribution of permits provides a way to make side-payments that induce countries to join the IEA.

The intuition of using side-payments to increase membership is quite obvious, but a brief explanation is nevertheless useful. The requirement that no participant of the IEA wants to abandon the agreement – the "internal stability" requirement – states that for each participant, the payoff that the country receives in the IEA is at least as large as the payoff that it would receive if it were to abandon the IEA, holding fixed all other countries' participation decisions, but taking into account how their abatement decisions might change. (Recall that the level of abatement that the IEA chooses typically depends on the number and characteristics of its members.) When side-payments are possible, these individual participation constraints are replaced by a single constraint that requires the aggregate payoff to members to be greater than the sum of members' payoffs if they individually defect from the agreement (McGinty 2007).

Although trade in permits increases the potential size of IEA membership, it has two important limitations. First, trade typically does not lead to the grand IEA, where all countries join. This result follows from the fact that the amount of the transfers is limited by the magnitude of the gains from trade. For example, if countries are fairly similar, then the gains from trade are small, so the "budget" for the implicit side-payments arising from the allocation of permits is also small. In this case, trade in permits can create only small increases in membership. The second limitation is that to the extent that the allocation of permits, together with trade, succeeds in inducing membership, it does so by creating large implicit transfers across countries. These transfers might encounter political opposition.

In Section 6.2 we noted that using regional agreements to "split up" the grand participation game into smaller games has the potential to increase aggregate abatement. However, if these regional agreements lead to groups with relatively homogenous countries, i.e. those with similar marginal abatement costs, they decrease the scope for using permit trade to induce participation. For example, if two regions with high abatement cost form an IEA, and if this agreement causes these two regions to no longer be part of the participation game played by other countries (as assumed in Section 6.2) the regional agreement reduces the potential gains from trade among the remaining countries. In that case, the regional agreement might jeopardize the prospects for the remaining countries to form an IEA. This observation suggests that it may be valuable to have a mechanism that enables regional IEAs to interact amongst each other. The obvious mechanism is trade in permits. However, our comments in Section 4.1 on the pitfalls of trade in permits across countries or IEAs that have not harmonized their policies, applies here as well.

7.2 Ex ante homogenous countries

Trade is also potentially important when countries face random abatement costs, which they learn only after the participation stage. In order to study the effect of this kind of uncertainty and country heterogeneity, consider the simplest case where, at the participation stage countries face the same distribution of abatement costs: countries are "ex ante homogeneous". Moreover, for simplicity we again assume, as in Section 5, that the country's abatement cost is constant up to a maximum level of abatement equal to 1, and the marginal benefit of abatement is 1 for each country. Country i's abatement cost, θ_i, is the realization of a random variable. Denote the expectation of this random variable as θ.

The country's abatement cost, θ_i, is private information, or in any case is not verifiable by the IEA. Therefore, the IEA cannot give different countries different abatement targets. If the IEA were to condition a country's abatement target on its announcement of its costs, the country would like to exaggerate its abatement costs. The assumption that abatement costs are non-verifiable means that the IEA discards any information the country might provide. Therefore the country has no incentive to misrepresent its costs. The IEA instructs each member to achieve the same required level of abatement, a level that depends on the number of signatories.

In the absence of trade, each country must abate at the level determined by the IEA. Since each country's expected marginal abatement cost is θ for abatement $a \leq 1$, and since the marginal benefit of abatement for the IEA as a whole is M (the number of IEA members), the optimal level of abatement in the absence of trade is the same as in Section 6. For $M < \theta$ the IEA sets $a = 0$ and for $M \geq \theta$ the IEA sets $a = 1$. The solid step function in Figure 1 graphs the equilibrium level of per country abatement, as a function of M (the number of IEA members), when trade in permits is not allowed.

With trade, the expected marginal cost of abatement varies with the level of abatement, a. For low levels of abatement, the expected marginal cost is strictly less than the expected marginal costs without trade, θ. For example, consider the case where there are 5 members, and suppose that $a = 0.2$. In this case, with trade, all of the abatement will be done by the country with the lowest abatement cost. The expected value of the lowest cost among the five draws (the first order statistic) is less than the expected cost of the average of the five draws. Therefore, with trade, the expected marginal cost of abatement at $a = 0.2$ – denote it as θ'– is strictly less than θ. A similar argument shows that at large values of a, the expected marginal costs of abatement exceeds θ. For example, when $M = 5$, if $a > 0.8$ it will be necessary for the highest cost country to produce some abatement. The expected value of the highest of five draws (the fifth order statistic) exceeds the expected value of the average of the five draws.

Now return to the example where for $a = 0.2$ the expected marginal abatement cost is

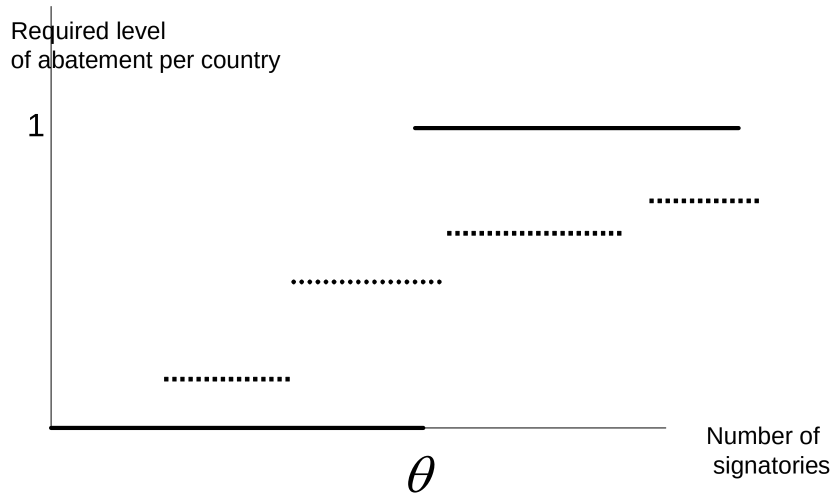

Figure 1: Equilibrium abatement level without trade (solid step function) and with trade (dotted step function) as a function of number of IEA members

$\theta' < \theta$. It is certainly possible that $\theta' < 5 < \theta$. If this inequality holds, then an IEA with 5 members would instruct no member to abate if trade is prohibited, but it would instruct each member to abate at a positive level (not less than 0.2) if trade is permitted. (We considered a variant of this example in Section 4.2.) As M increases, the expected marginal cost of a (weakly) decreases for $a < 1$; with larger M there is a higher chance of having enough low cost countries to take care of the abatement. Therefore, with trade, the equilibrium required level of abatement, per country, "resembles" the dotted step function in Figure 1. For very low levels of participation the per country required level of abatement is 0 with or without trade. For somewhat higher levels of participation, the required level of abatement is strictly positive with trade, and 0 without trade. However, for participation levels $M \geq h(\theta)$, trade reduces the equilibrium level of per country abatement. This example provides another case where trade in permits can either increase or reduce the level of abatement.

The new result here is that *trade can also decrease the equilibrium participation*. Verification of this claim is complicated, but the basic idea is straightforward (Karp 2009). Recall that a country decides to join the IEA only in order to exercise leverage, which in this circumstance requires affecting the actions of other signatories. The non-signatories always pursue their dominant strategy of not abating, so there is no possibility of affecting their actions. With trade, a possible equilibrium to the participation game has only the number of members corresponding to (for example) the lowest dotted step in Figure 1. The equilibrium to the participation game requires external and internal stability in the ex ante state (before individual firms learn their cost realizations). Signatories do not want to abandon the IEA because doing so leads to smaller – possibly zero – abatement and lower expected net benefits (ex ante internal stability). Non-signatories do not want to join because doing so would lead to a small – possibly zero – increase in average abatement by other firms, but would cause the new member to incur costs (ex ante external stability).

Without trade, equilibrium participation is $h(\theta)$, so the equilibrium amount of abatement without trade is also $h(\theta)$. With trade, even if equilibrium participation happens to be greater than $h(\theta)$, the per firm abatement is always strictly less than 1. Thus, a sufficient, but not necessary condition for trade to reduce aggregate abatement, is that equilibrium membership is less than $h(\theta)$. In this case, trade is clearly not "environmentally friendly". Even if this circumstance occurs, trade might still increase welfare, because it decreases average abatement costs.

8 Participation effects of escape clauses and safety valves

The major insight from the model described in Section 6 is that sovereign countries willingly join an IEA in order to exert leverage on other countries. In the standard model, each member of the IEA knows that if it were to defect, the IEA would reduce abatement to 0. By joining the IEA, each country recognizes that it increases the equilibrium level of abatement from 0 to $M = h(\theta)$ (the number of members). However, this leverage is quite limited, so the IEA that arises in equilibrium is quite small. There are a number of ways in which a re-designed IEA increases each country's leverage and thereby increases the size of the IEA.

Section 5 showed that (at least for our simple linear model) escape clauses and safety valves are not substantively different when we treat membership as predetermined. Both types of policies provide insurance against high cost realizations; they lead to the same level of abatement and aggregate welfare for all cost realizations. However, the two policies can have significantly different effects on the equilibrium level of participation in the IEA.

Both the standard model, and the variant with trade discussed in Section 7.2, assume that the IEA chooses the level of abatement after the participation decision. We noted that in this case the equilibrium level of participation is low. If the safety valve (the price ceiling) is chosen after the participation decision there is no significant change to the model in Section 7.2. Given the similarity of the safety valve and the escape clause, when these are chosen conditional on the level of participation, there is little point in analyzing the effect, on the participation decision, of the safety valve *chosen after participation.*

We therefore consider the two policies when these are *chosen before the participation decision*. This change in the timing of decisions is a significant departure from the standard model, and one that we regard as empirically relevant. Many aspects of IEAs are negotiated before nations decide whether to ratify the IEA. Karp and Zhao (2007) provide a more detailed defense of this revised timing.

8.1 Participation and the safety valve

Even the simple model described in Section 7.2 becomes complicated when we introduce a safety valve. However, the model yields a clear result under two additional assumptions. First, we assume that the cost variables are uniformly distributed. This assumption leads to a closed form expression for the distribution of order statistics, needed to analyze the safety valve. Second, we assume that the *nominal* level of abatement of members is set at the maximum level, 1 by our normalization. We explained in Section 5 why this assumption is innocuous when we consider the abatement stage of the game, taking the number of participants as given: the safety valve can be set so that abatement occurs only in countries with sufficiently low abatement costs. However, this assumption has an important implication when we consider the participation decision.

With these additional assumptions, we obtain a striking conclusion. A potential signatory has no leverage on signatories, even in the presence of the safety valve. In this context, the "lack of leverage" means that a potential signatory does not change the average abatement of existing signatories. In this case, the equilibrium level of membership in the participation game is zero. In short, the safety valve does not promote, and may even discourage participation in the IEA. The decision to set a nominal level of abatement prior to the participation decision eliminates the obvious source of leverage. A potential signatory might still exercise leverage if its participation would affect the probability distribution of prices.

8.2 Participation and the escape clause

A properly defined escape clause can increase equilibrium membership, increasing a potential signatory's leverage. We can make this point using a simple model with no cost uncertainty. In this setting, unlike the model used in Section 5, there is no need to provide insurance against unexpectedly high costs. A more complicated setting with cost uncertainty can be used to show the two roles of the escape clause, as a means of providing both insurance and leverage (Karp and Zhao 2007).

Suppose then, in the interest of a simple exposition, that abatement cost is θ, a known parameter. The terms of the IEA require each country to abate one unit or pay the fine F to exercise the escape clause. In Section 5 we did not specify what the IEA does with the revenue from the fine. Here, where our focus is on leverage (rather than insurance), the use of the fine revenue is critical. We assume that all revenue from fines is distributed equally to all IEA members, regardless of whether they chose to abate or to exercise the escape clause. Other distribution schemes are also possible.

The optimal decision for an IEA member depends on F and M, the total number of members, but not on the actions of other members; that is, the IEA member has a dominant strategy at the abatement stage of the game. Its cost of abating, less the additional benefit of abatement, is $\theta - 1$. The cost of exercising the escape clause, net of the amount returned in distribution payment, is $\frac{M-1}{M}F$. The country prefers to abate if and only if the latter is at least as great as the former, i.e. if

$$F > \frac{(\theta - 1)\,M}{M-1} \equiv f(M).$$

The function $f\,(M)$ approaches its lower bound $\theta - 1$ as M becomes large, so any fine must be greater than $\theta - 1$, i.e. it is likely to be large.

The negatively sloped function $f(M)$ divides $F - M$ plane into two regions; see Figure 2. For $F < f\,(M)$ each country prefers to exercise the escape clause. In that case, every country's total revenue from distributions of the receipts from the fine just equals the cost of the fine, so each country receives welfare 0. For $F \geq f\,(M)$ a country's welfare is higher if it abates, so all members abate. For any F, one equilibrium to the participation game is $M = 0$. The other equilibrium must satisfy $M = h\left(f^{-1}\left(F\right)\right)$, i.e. the smallest integer to the right of the graph of $f(M)$; a value of M greater than $h\left(f^{-1}\left(F\right)\right)$ is not internally stable.

At values $M = h\left(f^{-1}\left(F\right)\right) < \theta$, members would have negative payoffs. Therefore, any $F > f\,(\theta)$ leads to a unique equilibrium with 0 members. However, any $F \leq f\,(\theta)$ leads to an equilibrium with M participants, where M is the smallest integer on or to the right of the curve $f(M)$.

In short, an IEA that announces a required level of abatement of 1 and cost of exercising the escape clause $F < f\,(\theta)$ induces more members than in the game without the escape clause. By setting F to solve $h\left(f^{-1}\left(F\right)\right) = N$ the IEA can induce all countries to become signatories. To confirm this claim, suppose that the IEA chooses this value of F. If $N - 1$ countries join, the N'th country knows that if it also joins it will receive a payoff of $N - \theta > 0$, but if it stays out of the IEA all the other members will choose to exercise the escape clause, leaving the outside country with a 0 payoff. Therefore, N participants is a Nash equilibrium. Note that the equilibrium membership size increases as the fine is decreased; however, the smallest fine is strictly greater than $\theta - 1$, the difference between a country's costs of abatement and its private benefits.

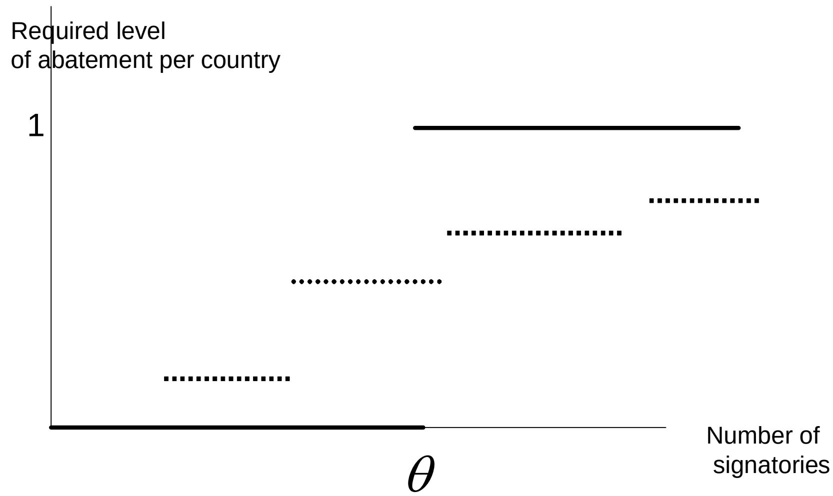

Figure 2: The graph of $f(M)$

This result may suggest that the problem of inducing membership to IEAs is trivial – which is obviously not the case. Our point is much more modest: the escape clause, together with a judicious distribution of fine revenue, provides a structure in which a potential signatory exercises leverage by joining the IEA.

9 Additional aspects of the escape clause

We explained that the escape clause offers both a means of providing insurance against high abatement costs and providing leverage. Here we discuss two additional aspects of the escape clause: its contribution to the problem of enforcement, and alternatives to the monetary fine.

9.1 The enforcement problem

The escape clause changes the nature of the enforcement problem, by requiring debt payment rather than environmental performance. Instead of having to somehow compel a signatory to achieve the level of abatement, the organization "merely" has to compel the country to pay its debt – the fine. We think that this is a simpler enforcement problem, because the process of sovereign debt collection is better developed, and certainly has a longer history, compared to the process of enforcing sovereign compliance with environmental agreements.

The enforcement aspect is particularly important when actual abatement costs are not known at the time a nation decides to participate in the IEA. In the simple model set out in Section 6, the agreement is always self-enforcing, because each participant knows that its defection causes the IEA to reduce per member emissions from 1 to 0, and the resulting welfare loss from the reduced aggregate abatement effort exceeds the country's own net cost of abatement. When costs are uncertain at the participation stage, this type of self-enforcement does not arise. In equilibrium, when θ is the ex ante expected cost of abatement, a participant has total benefits of $h(\theta)$. When costs are random, some countries face costs strictly less than θ, but others have higher costs. If a country has costs $\theta_i > h(\theta)$ its ex post welfare from being a member is negative. That country would like to withdraw from the IEA, even if its withdrawal precipitates a collapse of the IEA. When costs cannot be verified, ex post renegotiation of abatement targets is not practical, because each country has an incentive to exaggerate its owns costs.

In short, under random costs, the IEA in the standard model is not self-enforcing. Consequently, there must be some mechanism to induce countries to abide by their participation decision. Our point here is simply that it is easier to enforce payment of a monetary debt, rather than enforcement of an environmental outcome.

9.2 Alternatives to the fine

Our proposal uses a monetary fine, which leads to the problem of credible debt collection. That problem would be avoided if instead of the fine, the participants post a bond that is redeemed by achieving the target level of abatement, and forfeited otherwise. (Instead, perhaps, a portion of the bond is forfeited if the target is partially met.) This alternative is not practical, given the size of expenditures likely to be necessary to induce abatement. The fine – or the bond – is always greater than $\theta - 1$.

A second alternative replaces the fine by a "withdrawal of concessions" under WTO. This penalty is currently used when a disputant in a WTO case does not adopt the remedy determined by the WTO dispute resolution panel. This alternative would be consistent with WTO law, since it places constraints on countries that willingly enter the IEA. This use of the WTO is very different from the proposal to apply trade sanctions to countries that refuse to sign the WTO – a proposal that would run afoul of WTO law.

The replacement of the fine with a withdrawal of WTO "concessions" gives rise to the same kind of leverage as does the fine. By joining the IEA a participant increases the actual cost to any country of exercising the escape clause; the additional participant increases the amount of concessions withdrawn from the country exercising the escape clause. This alternative also has the advantage that it does not require that IEA members collect the fine from a member that exercises the escape clause; the IEA members who achieve the target can impose the penalty without the cooperation of the delinquent country.

A possible disadvantage of this alternative is that the country that imposes the trade sanctions also suffers diminished gains from trade; however, countries appear willing to suffer those losses: witness the fact that liberalized trade is viewed as a "concession" in WTO-speak, even though it is one that typically benefits the country making the concession. Of course, some countries are reluctant to exercise their rights to withdraw concessions, either because they recognize the resulting costs to themselves, or because they fear some kind of retribution from a more powerful trading partner.

A third alternative is to give NGOs standing in bring a complaint against a signatory that violates its IEA agreement. This type of extension of rights already has precedence in international law. For example, under NAFTA law, groups (other than signatories) are allowed to bring a complaint before a NAFTA tribunal in the event that a country violates its own national environmental law in the pursuit of foreign investment. Also, many investment treaties, including Chapter 11 of NAFTA, give investors the right to bring suit against signatories in the event that the government violates the treaty, e.g. by expropriating investments or failing to honor rights of establishment. These suits can be brought even if the offending government action occurs at a sub-national (e.g. province or state) level. In other words, the investment treaties open the door to investor-to-state disputes, unlike the WTO which permits only state-to-state disputes. This extension is important because a government has to balance many interests, and may not be willing to represent the interests of a particular investor that feels it has been treated illegally. An IEA that gives NGOs legal standing – analogous to that of investors in the investment context – together with a tribunal structure (similar to ICSID) would ease the problem of enforcement.

9.3 Partial escapes

We could extend the proposal to allow "partial escapes", under which the country can achieve a portion of its abatement commitment, and then pay a fine that depends on the gap between its target and its achievement. Depending on how this partial escape is constructed, it puts a ceiling on the marginal cost of abatement. Our simple proposal that allows only a "total escape" puts a ceiling on aggregate abatement costs.

10 Non-strategic effects of multinational cooperation

Our paper emphasizes strategic issues involving cooperation among nations. We close by recognizing several non-strategic reasons that this cooperation might promote improved climate change policies.

1. Carbon leakage. Leakage results from cost differences; it is less likely to occur when other countries/regions impose similar restrictions. Therefore, by instituting climate change policies, the US (for example) decreases the cost of similar policies in the EU and increases the environmental benefits of those policies, making it more likely that the European policies will be sustained.

2. Learning from Policy experience. A country/region is able to learn from policy experiments elsewhere. International cooperation makes it easier for countries to transfer this knowledge. For example, the EU experience with carbon trading is likely to be useful in a broader context.

3. Economies of scale. Reducing the costs of GHG abatement is likely to require large fixed costs in R&D and investment. An increase in the scope of regulation, such as occurs when other countries/regions require GHG abatement, increases the demand for the new technology and spreads the fixed costs of its development over more units of abatement, thereby decreasing the average costs of abatement.

11 Policy implications

Environmental agreements involving different countries or regions confront three types of challenges: encouraging participation, achieving abatement efficiently, and creating incentives for compliance. Our single most important point is that a design feature that addresses one of these challenges can make other challenges more difficult to meet. For example international trade of emissions permits contributes to efficient abatement, but it may be inimical to participation, and do nothing to encourage compliance. Emissions trading together with a safety valve creates insurance against high costs and also helps to achieve efficient abatement, but again it may not be help in achieving either participation or compliance. We discussed the idea of using an escape clause, which relieves a country of the responsibility to abate if it incurs some other cost that benefits signatories, e.g. pays a fine or agrees to the withdrawal of WTO-mandated trade concessions. This alternative can promote participation and provide insurance against high costs, although it does not help in achieving efficient abatement. An IEA that combines trade in permits and an escape clause may help in creating a successful IEA.

Our second important recommendation is that it is reasonable to use carefully circumscribed trade restrictions, such as border tax adjustments, as part of an IEA. These restrictions will help to undercut a politically attractive argument for staying outside an agreement: the concern that carbon leakage will dilute most of the global gains from the agreement, while eroding participants' ability to compete in carbon intensive sectors. The adoption of rules will also decrease the ability of pressure groups to demand protection once a country has joined the agreement. Finally, multilateral rules will mitigate the problem of carbon leakage, to the extent – currently difficult to determine – that this leakage has the potential to be a genuine problem.

References

BARRETT, S. (2003): *Environment and Statecraft*. Oxford University Press.

BHAGWATI, J. (2004): *In Defense of Globalization*. Oxford University Press, Oxford.

CONE, S. M. (2002): "The Environment and the World Trade Organization," *New York Law School working paper*.

COPELAND, B., AND M. TAYLOR (2003): *Trade and the Environment: Theory and Evidence*. Princeton University Press.

——— (2004): "Trade, Growth and the Environment," *Journal of Economic Literature*, XLI(1), 7–71.

——— (2005): "Free trade and global warming: a trade theory view of the Kyoto Protocol," *Journal of Environmental Economics and Management*, 49(2), 205–234.

GUZMAN, A. (2002): "Global Governance and the WTO," *UC Berkeley School of Law Research Paper No. 89*.

HOEL, M., AND L. KARP (2001): "Taxes and Quotas for a Stock Pollutant with Multiplicative Uncertainty," *Journal of Public Economics*, 82, 91–114.

HOEL, M., AND L. KARP (2002): "Taxes Versus Quotas for a Stock Pollutant," *Resource and Energy Economics*, 24, 367 – 384.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (2001): "Climate Change 2001: Mitigation Working Group III contribution to the IPPC 3rd assessment report," http://www.ipcc.ch/SPM040507.pdf.

KARP, L. (2008): "Correct (and misleading) arguments for using market-based pollution control technologies," http://are.berkeley.edu/ karp/marketbasedJuly08.pdf.

——— (2009): "Incentives to join international environmental agreements with permit trading and safety valves," Giannini Foundation working paper.

KARP, L., AND J. ZHANG (2006): "Regulation with Anticipated Learning about Environmental Damage," *Journal of Environmental Economics and Management*, 51, 259–280.

KARP, L., AND J. ZHAO (2007): "A proposal to reform the Kyoto Protocol: escape clauses and foresight," http://are.berkeley.edu/ karp/KyotoReformFeb07.pdf.

——— (2009): “Suggestions for the road to Copenhagen,” Giannini Foundation working paper.

MALIK, A. (1990): “Markets for Pollution Control when Firms are Noncompliant,” *Journal of Environmental Economics and Management*, 18, 97–106.

MCGINTY, M. (2007): “International environmental agreements among asymmetric nations,” *Oxford Economics Papers*, 59, 45 – 62.

NEUMEYER, E. (2005): “The WTO and the environment: its past record is better than critics believe, but the future outlook is bleak,” *Global Environmental Politics*, 4, 1–8.

NEWELL, R., AND W. PIZER (2003): “Regulating Stock Externalitiues Under Uncertainty,” *Journal of Environmental Economics and Management*, 45, 416–32.

NORDHAUS, W. D. (2007): “To tax or not to tax: alternative approaches to slowing global warming,” *Review of Environmental Economics and Policy*, 1(1), 26–44.

PAUWELYN, J. (2007): “U.S. Federal Climate Policy and Competitiveness Concerns: The Limits and Options of International Trade Law,” Nicholas Institutute for Environmental Policy Solultions.

STAVINS, R. (2008): “Addressing climate change with a comprehensive US cap-and-trade system,” Harvard University, John F Kennedy School of Government.

STIGLITZ, J. E. (2006): “A New Agenda for Global Warming,” *Economists’ Voice*, pp. 1–4.

STRANLUND, J. K., AND C. A. CHAVEZ (2000): “Effective Enforcement of a Transferable Emissions Permit System with a Self-Reporting Requirement,” *Journal of Regulatory Economics*, 18, 113–131.

VICTOR, D. G. (2001): *The Collapse of the Kyoto Protocol and the Struggle to Slow Global Warming*. Princeton University Press.

WORLD BANK (2008): *International Trade and Climate Change*. World Bank.