



# On the Theory and Diplomacy of Environmental Treaty-Making

SCOTT BARRETT

*London Business School, Sussex Place, London NW1 4SA, UK (email: sbarrett@lbs.ac.uk)*

**Abstract.** International cooperation can be looked at in two ways: as a process and as an outcome. This paper shows how the process of treaty-making can affect treaty outcomes and how treaty design can change the rules of the game of international cooperation.

**Key words:** game theory, international environmental agreements

**JEL classification:** Q2, K33, H77

## 1. Introduction

Efforts to sustain international cooperation are almost always manifest in international agreements. Various referred to as treaties, conventions, protocols, charters, and acts, international agreements are analogous to contracts in domestic law: they stipulate what the parties to them must and must not do. But international agreements, unlike contracts, cannot be enforced by a third party; they must be self-enforcing. This is a severe constraint, and in having to obey it the parties to a treaty may not be able sustain every feasible outcome. Depending on the circumstances, they may only be able to sustain outcomes that are not much different from the noncooperative outcome.

Testing the theory is difficult because we only observe the outcome in which countries possibly sustain some cooperation. We cannot typically observe the outcomes in which there is either no cooperation or full cooperation. So we cannot easily say whether any particular agreement buys us a lot of cooperation or just a little. We can only try to infer what the counterfactual outcomes might look like.

This can be done in a number of ways: qualitatively, by the use of case studies; informally, by estimating the underlying relations in the theoretical models; and formally, by estimating econometric models of national choice. The problems that have been studied most thoroughly are stratospheric ozone depletion and acid rain, and there is a reassuring consistency in the different empirical analyses of these problems.<sup>1</sup> Moreover, this work finds some support for the most basic result of the theory: that, when the number of countries with a stake in an issue is “large”, international agreements typically improve only marginally on the anarchic out-

come (though the anarchic outcome may not differ much from the full cooperative outcome).

That the theory is difficult to test doesn't mean that it is lacking in empirical relevance. Like all applied game theory, the theory of international environmental agreements (IEAs) is especially useful in helping us to understand better the outcomes we observe.<sup>2</sup> To find that international conferences, and the treaties that they give birth to, may have little real effect is to discover something important. Of course, it might be argued that the theory is only telling us something we already knew. And sometimes it does only confirm what we already knew. But there is an advantage in formalizing our thinking on these matters. For doing so provides a check on our intuition. Sometimes it shows that our intuition is wrong.

There is another thing that theory can do for us. Treaty outcomes aren't inevitable. They depend on the acumen of diplomats, especially their skill in changing the rules of the game so that real cooperation can be sustained, despite the constraint of self-enforcement. In formalizing our thinking, the theory should not only explain the outcomes that we observe but also tell us how we could improve on these. Put differently, it should provide a framework on which we can build a strategy for treaty negotiation.

International cooperation can be looked at in two different ways: as a process and as an outcome. Ultimately, it is the outcome that we will be most interested in. But the outcome may depend on the process that gave rise to it. So we may want to model the process, too. As yet, we haven't. The theory of IEAs is largely a theory of outcomes. But even in this the theory is wanting, for there are important features of actual agreements that aren't explained by the theory.

In this paper I describe how treaties are made and what they do – not just the extent of cooperation that they are able to sustain but the specific things they require that their signatories do or not do; the things that are written down in a treaty, in fine print. I show how the theory can illuminate, but I also show areas where the theory is weak. My discussion isn't comprehensive – there isn't the space for a full treatment of these matters – but it explains the relevance of the theory to diplomacy and how the theory can be made even more relevant. I begin by discussing the treaty-making process, and the effects that this can have on treaty outcomes. I then discuss some features of treaty outcomes that serve the vital function of changing the rules of the game. I conclude with a brief discussion of the insights that the theory can offer the climate change negotiators, who are, as I write, returning home from Kyoto.

## **2. The Process of Treaty-Making**

How are treaties made? The process is complex, and not entirely linear, but for analytical purposes can be broken down into five stages.

## 2.1. PRE-NEGOTIATION

Normally, negotiations are preceded by a phase of pre-negotiation maneuvering. In the run-up to the Kyoto talks on climate change, for example, Australia claimed that it should be allowed to increase its emissions, Europe argued for deep cuts, the United States commended stabilization, and the poorer nations insisted that the rich countries had to reduce their emissions first. This is cheap talk, and in a game of cooperation it is unlikely to have much effect on the outcome of negotiations.

The making of real commitments is different. But it is very hard for countries to make real commitments. A “commitment” to stabilize carbon dioxide emissions at 1990 levels is no such thing. (Who is to punish Norway if it fails to meet its target?) The same is true of a unilateral carbon tax. Such policies can usually be changed as easily as they are first adopted. So they can be expected to have no effect on the behavior of other states at the negotiating table.

It is, in any event, hard to see how such actions could be strategic – how they could affect the behavior of other nations, to the benefit of the ones taking the actions – unless other countries were inclined to behave contrary to their self interests, whether out of a sense of moral obligation or a preference to conform.<sup>3</sup> History teaches that leadership of this kind is not often rewarded, at least not when only one or a small number of countries take the lead. It is, of course, possible that abatement by one country could lower the cost to others of abating their emissions, and so make it in the self interests of these others to increase their abatement (see Heal 1993) – that is, it is possible that abatement choices could be strategic complements. But typically this will not be so. If it were, then international cooperation would not be the problem that we understand it to be.

Less virtuous behavior is more likely.<sup>4</sup> Countries may step up their exploitation of the shared resource to extract whatever rents may be left before the negotiations commence. Alternatively, they may increase exploitation or take other actions (including irreversible investments) to improve their post-negotiation payoff (this behavior is strategic). Whatever the motive, actions of this kind are collectively wasteful, and it will be in the interests of cooperating countries to discourage them.

Fortunately, this is not very hard to do. The incentive to behave strategically, for example, can be eliminated by mutual recognition of an historical baseline. Thus, in establishing 1990 as a baseline from which future emission reductions are to be negotiated, the climate change negotiators removed any incentive countries may have had to improve their negotiation stance by increasing their emissions. Similarly, rent-taking in the run-up to negotiations can be remedied by negotiation of a pre-agreement agreement. As an example, prior to going to arbitration, Britain agreed to reduce its pelagic seal harvest temporarily in the pre-tribunal stage, after the U.S. agreed to compensate Britain for any losses in the event that the tribunal subsequently found in favor of Britain. (As matters turned out, the tribunal did find in favor of Great Britain, and the United States paid the promised compensation.)

## 2.2. NEGOTIATION

Negotiations concerning complex and vital global issues are complicated affairs, and can involve well over 150 states. If only to reduce transactions costs, parties are likely to want to structure their negotiations, and large-scale negotiations routinely begin by specifying how the talks should proceed. This is important if the choice of process in this stage can affect the final treaty design.

Negotiation of the first climate change treaty was organized by an Intergovernmental Negotiating Committee (INC), created by a resolution of the United Nations General Assembly.<sup>5</sup> As is customary in large-scale negotiations, the climate change negotiators worked with a single negotiating text (which, for much of the time, was littered with brackets), and divided the INC into separate negotiating groups. Rather unusually, however, the INC had to meet a deadline: in a pre-negotiation maneuver, the General Assembly required that the Framework Convention on Climate Change be ready for signing at the United Nations Conference on Environment and Development in Rio de Janeiro in June 1992. In addition, though the INC rules allowed for decisions to be taken by vote, the INC chose to work by consensus. The draft text submitted on the final day of the final negotiating session, just 15 months after the first session was convened, was approved without a formal ballot.

Many features of this process could influence treaty design – including the committee arrangements, the choice of voting rules, and the imposed deadline – but here I comment only on the use of a single negotiating text. Obviously, treaty negotiations have to start somewhere. But where? There will typically be many first drafts that would be acceptable at least to a majority of countries. But only one will be selected, and though it is technically possible to backtrack or even to throw the text out and start from scratch, in practice this rarely happens. So the first draft can be expected to leave its mark on the final agreement.

This is not an original observation (see Raiffa 1982). However, I believe that history is even more important than suggested by the example of the single negotiating text. Experimental evidence confirms that the outcome of a game may depend on expectations formed in the play of previous games (Roth and Schoumaker 1983). Climate change negotiations followed quickly on the heels of the successful Montreal Protocol, and it is possible that the Framework Convention and Kyoto Protocol were partly shaped by this earlier and largely unrelated negotiation.

Negotiations sometimes proceed in stages, with a preliminary “convention” laying out general principles followed by the negotiation of “protocols” prescribing specific obligations. Typically, parties to a protocol must also be parties to the associated umbrella convention, but parties to a convention do not have to be parties also to its protocols; they can pick and choose which protocols, if any, they want to be parties to. The decision not to link different aspects of an environmental problem – like the abatement of sulfur dioxide and nitrogen oxides in an acid rain agreement – can therefore be strategic. Murdoch, Sandler and Sargent (1997) argue that, had the Convention on Long-Range Transboundary Air Pollution not provided

for protocols to be negotiated separately for the different pollutants, the negotiated reductions in sulphur emissions would have been smaller.

But sometimes it may be better to link issues, even when they are technically separate. This may be because the linkage acts like a side payment, ensuring that all parties are made better off by cooperating. Or it may be because linkage helps to sustain cooperation as an equilibrium. Carraro and Siniscalco (1994), Cesar and de Zeeuw (1994), and Folmer et al. (1993) show how prisoners' dilemma-type problems can be transformed by issue linkage. I have also shown how the decision to link trade and abatement in the Montreal Protocol was strategic (linkage is endogenous in this model; see Barrett 1997b), and transforms a prisoners' dilemma-type problem into a coordination game. These results, however, are special. There is much that we still do not understand about issue-linkage.

### 2.3. RATIFICATION

A state signals its intent to comply with an agreement by having its representative at the negotiations sign it. It is customary, however, for IEAs to become legally binding only after being ratified, and a signature does not impose a legal obligation on states to ratify (Brownlie 1990). Ratification typically requires the approval of a state's national parliament. In the United States, ratification of any treaty requires the approval of a two-thirds majority of the Senate. The theory of international cooperation routinely takes countries to be monoliths (an exception being Curran and Tulken 1997). But plainly they are not.

Putnam (1988) reasons that the game of international negotiations is played at two levels. At the first level is the international negotiation itself. At the second is the decision by domestic parties to ratify the agreement. Where the latter is not a rubber stamp – where each country is not a monolith but represents a multiple of different interests – the outcome of this game will be nontrivial. (Note that ratification is likely to matter more for a country like the United States, because of its constitutional separation of powers, than for a parliamentary democracy.)

Though Putnam does not articulate the problem quite this way, the game can be thought of as proceeding in stages. In Stage 1, the executives of every country negotiate an agreement. In Stage 2, the agreement comes before national parliaments for ratification. It is natural to solve a problem like this using backward induction, for executives will only negotiate an agreement that they feel confident will be ratified (a point emphasized by Putnam). More sophisticated game structures may also be considered. For example, parliament may pre-empt negotiations by laying down pre-conditions for ratification or by stipulating the conditions that must be fulfilled by international treaty if certain unilateral actions are to be avoided.

During the Montreal Protocol negotiations, the U.S. Congress drafted legislation calling for trade restrictions to be imposed against nations not accepting their share of the common responsibility to reduce harmful emissions. According to the

chief U.S. negotiator at the talks, Richard Benedick (1991, p. 29), "U.S. negotiators made certain that the implications of this threat were not lost on foreign governments, pointing out that there might be a price to pay for not joining in meaningful efforts to protect the ozone layer." More recently, the U.S. Senate voted 95–0 for a resolution that the United States should not be a signatory to an agreement which would require that the industrialized countries reduce their greenhouse gas emissions, unless the agreement imposed similar obligations on developing country parties. The resolution was ostensibly intended to alter the negotiating stance of the Clinton Administration at the Kyoto talks, though it may also have altered the behavior of other countries at these negotiations, and been intended to have this effect, too.

#### 2.4. IMPLEMENTATION

After a treaty has become legally binding, it must be implemented, and this will typically require the passing of domestic legislation. Possibly, the necessity for domestic legislation may make a treaty more stable.

Many IEAs require that parties report data relating to their obligations – often, but not always, to a small administrative body set up under the agreement. Reporting is often incomplete (U.S. General Accounting Office 1992), but verification of the reported data is usually more problematic. We have only recently learned that the Soviet Union, through the offices of the KGB, falsified the data it submitted to the International Whaling Commission in the 1960s.

Where verification is difficult, monitoring will be especially important. Very often, the payoff to one country of complying with a treaty will depend on whether the other parties are also complying with it. Even the suspicion that others are not complying with a treaty may cause cooperation to unravel. Put differently, if cheating can't be detected then it can't be punished; if it can't be punished then it can't be deterred; and if it can't be deterred then cooperation can't be sustained by the treaty's bootstraps.

Of course, one implication of this is that countries will only agree to do something collectively if that something can be monitored. This is why the Limited Test Ban Treaty prohibits above-ground and not underground nuclear testing. Effective monitoring of underground testing requires on-site inspection.

Treaties sometimes specify exactly what the parties are supposed to do and when. More often they set up a regime for managing a resource, with the details of what the parties are supposed to do being decided later. The Montreal Protocol is an example of a treaty that does both. It specifies precise production limits but it does not carefully specify how parties are to be punished for failing to comply with them. Instead, suspected instances of non-compliance come before an Implementation Committee. The Committee reviews the evidence, considers the circumstances that may have resulted in non-compliance, and then makes recommendations to the parties regarding specific actions to be taken. Failure to specify in advance

the rules for taking actions may invite non-compliance, but the careful exercise of discretion may also have advantages. The trade-off between commitment to a rule and flexibility has been studied in other contexts (see, for example, Dixit 1996), but not in IEAs.

## 2.5. RENEGOTIATION

Agreements can always be renegotiated or, in the legal jargon, “adjusted.” So international agreements must be self-enforcing, even long after they have come into force (see Barrett 1994, 1997c). In equilibrium, a treaty may not be renegotiated, but our modelling should allow for this possibility.

Negotiated adjustments apply to all the original signatories, though parties may withdraw from a treaty at any time, after giving sufficient notice. As an example, the production and consumption limits in the original Montreal Protocol were adjusted by the parties at their fourth meeting, which was held in Copenhagen in 1992. These adjustments were agreed by consensus, though the Protocol allowed them to be made, “as a last resort”, by a two-thirds majority “of the Parties present and voting representing at least fifty per cent of the total consumption of the controlled substances of the Parties.”

Renegotiation may be triggered by unforeseen changes in the underlying relationships of the game, or by the resolution of uncertainties. If implementation of the original agreement were costlessly reversible, then this will only require that the self-enforcing agreement be recalculated (of course, the equilibrium number of signatories may also change). If not – if there is a problem of “lock-in” – then implementation of the original agreement may partly determine the nature of any renegotiated agreement, and of course all the parties to the agreement will take this possibility into account when they negotiate the original treaty.

Treaties can also be amended. Though all parties to the original agreement are also invited to negotiate an amendment, and though parties to the original agreement have the right to become parties to the amended agreement, they are not obligated to do so, and a state may remain a signatory to the original agreement and not become a party to the amended agreement. Any state that accedes to the treaty after the amendments have come into effect, however, must become a party to the amended agreement and not just the original agreement. The Montreal Protocol has been amended three times: in London in 1990, in Copenhagen in 1992, and in Vienna in 1995. Would this treaty look precisely the same if it were negotiated from scratch today, or does the outcome depend on the process by which the treaty came to be negotiated and renegotiated? The theory of IEAs is largely static, whereas the actual development of IEAs can be evolutionary.

### 3. The Outcome of Treaty-Making

Analyzing a given model is usually easy. Choosing the right model to analyze is trickier. For in doing so one must distinguish between the things one wants the theory to explain and the things one chooses instead to assume. A crucial assumption is that treaties cannot be enforced by third parties. One may ask why the international system is anarchic, but that is a higher order question. If one is interested in environmental treaties, then it is sensible to rule out third party enforcement.

Some assumptions may be unrealistic but useful as a tactic for trying to understand something else. For example, Chander and Tulkens (1995 and elsewhere) have used the concept of the core from cooperative game theory to solve for a transfer scheme that ensures that every country is at least as well off in the full agreement as it would be under any partial agreement. The agreement is sustained in their model by a strategy that will not normally be credible, but the properties of the transfer scheme are compelling nonetheless.

To take another example, a number of papers assume – usually implicitly – that signatories comply with the treaties they sign up to. It is a fact that countries do comply with the treaties they sign up to, and one can try to justify the assumption of full compliance by appealing to this observation (one can but shouldn't; see below). But the real reason for making this assumption is that it allows us to focus on a different problem, like free-rider deterrence. To show that international agreements cannot sustain much cooperation, even if countries can commit to complying with the agreements they sign up to, is to show something significant. This is one of the central messages of the theory developed to date.

However, the fact that countries comply with the agreements they sign up to *is* something that we should explain. Just because countries comply with the agreements they sign up to doesn't mean that they will comply with *any* agreement. I have recently shown that the assumption of full compliance in a model with free-rider deterrence is innocuous (Barrett 1997c). The reason, very roughly, is that the punishments needed to deter free-riding in these models are sufficient also to deter non-compliance. In a sense, once free-riding is deterred, compliance enforcement comes free-of-charge.

Tactical modelling is necessary because we can't build a model of Everything. So we subdivide the problem, analyze a bit at a time, and then try to construct from these different analyses a coherent theory of how countries cooperate. The theory of IEAs has yielded important insights, but it isn't yet complete. I discuss below how the theory can help us to understand treaty outcomes, and also how it can be made even more relevant.

#### 3.1. PARTICIPATION RESTRICTIONS

The number of countries that can potentially participate in a negotiation is taken as a given in the literature, and this assumption is often appropriate. For example,

participation in the agreement facilitating cooperative management of the Niger River basin is limited to the riparian states (of which there are 10), and it wouldn't make any sense for a model of this agreement to consider participation by any more or fewer countries.

However, in some cases, participation is not a given. The Nile is shared by 9 countries, and yet the 1959 Nile Waters Agreement was negotiated by only Egypt and Sudan; the other 7 countries were not invited to the talks. At the time, the exclusion may not have mattered much, but in the future it is likely to matter more. A more inclusive agreement on the Nile is almost certain to be needed eventually.

Rather than simply exclude countries, treaties sometimes effectively strip countries of their ability and even their desire to act. The North Pacific Fur Seal Treaty is an example. It was negotiated between the United States, Great Britain (acting for Canada), Russia, and Japan. Other countries were not invited to participate in the agreement, even though any country could catch seals outside the 3-mile territorial limit. The restriction on membership could be sustained as an equilibrium by this treaty because entry by third party states was deterred by a trade restriction. Parties were prohibited from importing seal skins from non-parties, and since all skins were processed in London, the prohibition effectively shut all non-parties out of the industry. The agreement was an equilibrium in the sense that the signatories had nothing to gain by admitting additional members, given that entry could be deterred, and third party states had no claim to a share in the gains to cooperation, since their actions did not add to these gains.

By contrast, participation in the Vienna Convention for the Protection of the Ozone Layer is open to all states and regional economic integration organizations (currently, only the European Union). Participation in the Montreal Protocol is restricted to parties to the Vienna Convention, and hence is also essentially open to all countries. Participation is open in these agreements because protection of the ozone layer is a global public good. Every signatory to these agreements would prefer that the total number of signatories be as large as possible.

Negotiations leading up to the signing of the Antarctic Treaty in 1959 were limited to the 12 nations carrying out scientific research on Antarctica during the International Geophysical Year (which ran from July 1, 1957 to December 31, 1958): Argentina, Australia, Belgium, Chile, France, Japan, New Zealand, Norway, South Africa, the Soviet Union, the United Kingdom, and the United States. Article IX of the treaty gave signatories the status of "consultative parties" only if they were one of the original 12 negotiating countries or a contracting party which "... demonstrates its interest in Antarctica by conducting substantial scientific research activity there, such as the establishment of a scientific station or the despatch of a scientific expedition." Thus, the original 12 signatories formed a club, and set a high price for membership – though not so high that entry did not occur (since the agreement was signed, more than a dozen other countries have joined as consultative parties).

In 1988, the consultative and contracting parties to the Antarctic Treaty adopted the Convention on the Regulation of Antarctic Mineral Resource Activities, which was to establish a regime for mineral exploitation, and at the same time “to keep under review the conduct of Antarctic mineral resource activities with a view to safeguarding the protection of the Antarctic environment in the interest of all mankind.” Non-parties, however, did not think that the agreement represented *their* interests, and protested. In contrast to the example of the Fur Seal Treaty, non-parties to the Antarctic Treaty arguably can claim a share to the gains from cooperation.

Probably for this reason, only six of the 38 Antarctica Treaty parties signed the 1988 Convention, which seems destined for the legal dustbin. In its place, the Antarctica Treaty members negotiated the 1991 Protocol to the Antarctic Treaty on Environmental Protection, which banned mineral exploitation for 50 years. This agreement, however, only postpones the decision of whether to allow mineral development or make Antarctica a World Park instead. Non-parties to the Antarctica Treaty are certain to demand a voice in this decision, and so the question of the membership of this treaty will inevitably resurface.

### 3.2. ENTRY INTO FORCE AND MINIMUM PARTICIPATION

International agreements usually impose additional requirements before becoming legally binding on signatories. Most multilateral agreements do not come into force until ratified by a minimum number of countries, and some do not come into force until ratified by all the parties at the bargaining table. Most multilateral treaties, and especially treaties concerned with global environmental problems, however, fix a minimum far short of the total number of negotiating countries. For example, though more than 100 countries participated in the negotiations on the Convention on Biological Diversity, only 30 countries needed to ratify the agreement before it entered into force.

The countries that must ratify an agreement before it comes into force are sometimes named. For example, the International Whaling Convention specifies that the agreement must be ratified by the Netherlands, Norway, the Soviet Union, the United Kingdom and the United States, as well as at least one other country, before entering into force. In other cases additional requirements are specified. For example, the Montreal Protocol would not enter into force until ratified by at least 11 countries “... representing at least two-thirds of 1986 estimated global consumption of the controlled substances ...” (mainly, CFCs).

At a superficial level, it is obvious why such conditions should be imposed. It would not be in the interests of a country to be bound by the obligations of a treaty until enough other states, and in some cases *particular* other states, were bound by these same obligations. Essentially, the “minimum participation clause” makes the obligations of each of its signatories a (non-linear) function of the total number of signatories.

At a deeper level, the minimum participation clause can serve as a strategic device. Suppose that the minimum participation level is given as  $k^+$ . Then we can deduce the following: if the actual number of signatories is  $k$ ; if, when countries make their choice of whether to sign a treaty, each takes the choices of all other countries as given; and if  $k < k^+ - 1$ , then accession by a nonsignatory will neither cost this country anything nor confer upon it any advantage. However, if  $k = k^+ - 1$ , then the accession would have a non-marginal effect, for it will make the agreement binding on *all* of the  $k^+$  signatories. One way to sustain full cooperation would be to set  $k^+ = N$ , while ensuring that every country is better off with the agreement than without it.<sup>6</sup> But the threat not to undertake any abatement for  $k < N$  must be credible, and in the vast majority of cases it will not be.

Typically, the actual number of signatories exceeds the minimum needed to bring an IEA into force, and this suggests that the minimum participation clause may serve as a coordinating device. In Barrett (1997b), the minimum participation level  $k^+$  is chosen endogenously. If fewer than  $k^+$  countries sign the agreement, the agreement is non-binding. So provided  $k < k^+ - 1$ , a nonsignatory has nothing to lose by acceding to the agreement. Once  $k > k^+ - 1$ , however, every country prefers to be a signatory in this model. So the minimum participation level ensures that the mutually preferred equilibrium will be selected.

### 3.3. TREATY RESERVATIONS

Some treaties allow parties to make “reservations”, or to opt out of having to be bound by certain treaty provisions. For example, the 1988 Nitrogen Oxides Protocol demands that signatories limit their emissions (or transboundary fluxes) of nitrogen oxides to a level no greater than that for 1987 “or any previous year, to be specified upon signature of, or accession to, the Protocol . . .” provided that, for any country specifying an alternative base year, the average of annual emissions or transboundary fluxes over the period 1987–1996 does not exceed the 1987 level. Of the 25 parties to this agreement, only the United States specified an alternative base year – 1978. Since U.S. emissions were greater in 1978 than in 1987, the U.S. reservation reduced the amount of abatement that it had to carry out, at least in the short-term.

Obviously, reservations can serve as a loop hole, allowing parties to agree only to do what they would have done had they not been parties to the agreement. But prohibiting reservations might not make for more effective agreements.

Indeed, the U.S. declined to sign the 1985 Sulphur Protocol because it prohibited parties from making reservations. According to the U.S. International Trade Commission (1991, pp. 5–73), “The United States played an active role in negotiating the protocol, but because the protocol did not credit the substantial progress the United States had already made in controlling sulfur dioxide emissions (reducing emissions by 24 percent since 1970), the United States elected not to participate.” By contrast, the U.S. signed the 1988 Nitrogen Oxides Protocol because it let the

U.S. opt out of a specific obligation. According to the International Trade Commission (1991, pp. 5–73), “As with the sulfur dioxide protocol, credit for prior action was again a major issue in negotiating this protocol. The United States insisted on and finally achieved some credit for its progress in controlling nitrogen oxide emissions prior to the 1987 base year.”

Of course, negotiators have more instruments to hand than just the option of allowing parties to make reservations. Non-uniform obligations can be negotiated directly. Side payments can be paid. Parties to the agreement can be permitted to trade in the entitlements to emit a pollutant. The Montreal and Kyoto Protocols do all these things. And they prohibit the making of reservations.

#### 3.4. FREE-RIDER DETERRENCE

IEAs are often incomplete: typically some countries that may act to effect an outcome will not be parties to the agreement. Of course, nonsignatories need not be free-riders. The “free-rider” label should only be worn by nonsignatories that benefit directly from the collective actions of signatories.

Indeed, by this definition, most nonsignatories are not free-riders. Iceland withdrew from the Whaling Convention only after the majority of the parties to this agreement sought to ban whaling for animal welfare reasons, when a universal ban was no longer justified for commercial reasons. Rwanda and Somalia are nonsignatories to the Framework Convention on Climate Change because of civil strife and the absence of effective municipal government, not because they wanted other nations to pay the cost of reducing carbon dioxide concentrations. Opposition by the United Kingdom to the first Sulphur Protocol stemmed from a reluctance to pay for the costs of abatement, not a desire to benefit from the abatement undertaken by others.

Moreover, though the theory emphasizes the importance of free-rider deterrence, mechanisms designed to limit free-riding rarely feature in IEAs.

One possible interpretation of these facts is that free-riding isn't the problem that the theory makes it out to be. But there is an alternative interpretation: that the vast bulk of agreements demand so little of their signatories that free-riding can't severely weaken them; that, precisely because free-riding can't be easily deterred, agreements requiring that free-riding be deterred are seldom negotiated; that, if we had mechanisms that *could* deter free-riding, then IEAs would use them and protect the environment more effectively.

In their book, *The New Sovereignty*, Chayes and Chayes (1995, p. 188) claim that “free-rider problems have not seemed so urgent in practice as in the theoretical literature.” This is because “the objective of environmental treaties is not so much to achieve a particular end state as to manage the environmental resource over time, by setting in motion trends that will reverse its current decline or depletion.” They note as well the disadvantage in free-rider punishment: “Retaliation in kind, the usual prescription for free-riding, is an even less attractive option in environ-

mental contexts than elsewhere. It would itself result in further deterioration of the resource, and often the retaliator would face the same international and domestic pressures as an original defector.”

The theory of IEAs offers an alternative interpretation of these same facts. Yes, retaliation in kind *is* self-punishing, and for this reason countries may not find it in their interests to deter free-riding. But, in recognizing that the threat to deter free-riding would not be credible, they would not negotiate an agreement requiring that free-riding be deterred. They would negotiate a different and more modest agreement instead, one which did not require that specific objectives be met but which rather established a process for making step-by-step improvements. Of course, such an agreement would not be vulnerable to free-riding, and so free-riding would not *appear* to be a problem. Nonetheless, free-riding would be a problem. If a credible mechanism for deterring free-riding had been at hand, a different agreement may have been negotiated.

This alternative interpretation cannot, of course, be proved (just as the Chayes’ own interpretation cannot be proved). We can’t observe the outcome that would have been realized had a suitable mechanism for deterring free-riding existed. But this discussion shows why the theory is so useful to understanding treaty design. It tells us that negotiators ought to do what they can to deter free-riding.

But how? The theory tells us that free-riding can sometimes be deterred by sanctions. Chayes and Chayes (1995, p. 2), by contrast, argue against the use of sanctions “for the routine enforcement of treaties”, claiming that “The effort to devise and incorporate such sanctions in treaties is largely a waste of time.” At least as regards free-rider deterrence (an issue on which the Chayes do not write at length, their chief concern being with compliance; but see below), the theory of IEAs suggests otherwise. The Chayes do not examine the trade sanctions in the Montreal Protocol in any detail, but I have shown (Barrett 1997b) that the credible threat of multilateral trade sanctions has probably helped to sustain full participation in this agreement.

### 3.5. COMPLIANCE

Non-compliance can be as damaging to international cooperation as free-riding. Indeed, the example of the whaling treaty shows just how closely related the two phenomena are. Restrictions on whaling imposed by the International Whaling Commission led to an increase in the taking of whales by countries that were not parties to the treaty (the so-called “pirate” whalers). Yet at least some of the pirate whaling was noncompliance by another name, for Japanese whalers got around the IWC restrictions by re-registering their vessels under the flags of non-party states (Rose and Paleokrassis 1996). Chile, for example, became a pirate whaling nation in the 1950s, with Japanese financing, crews, and ships.<sup>7</sup>

Chayes and Chayes’ (1995) careful research shows that compliance with international agreements is the norm. This is hardly surprising, for customary law

requires that countries comply with their treaty obligations, and as the Chayes point out, unless countries could be relied upon to comply with the treaties they sign up to there would be no point in negotiating such accords in the first place. But of course this also means that parties will only agree to sign treaties that they would be sure of complying with.

After reviewing the record of compliance with nine different IEAs, Ausubel and Victor (1992, p. 22) conclude that, though compliance had been “fairly high . . . , much of this may be an artifact of the standards.” They illustrate their argument by noting that, in the late 1950s, Norway and the Netherlands withdrew from the IWC in a dispute over quota-setting, only to rejoin in the early 1960s when quotas were raised. Ausubel and Victor could be right that states comply with their obligations only because these obligations are weak. But the example they cite is open to a different interpretation: that the parties would have complied fully with the smaller quotas had their withdrawal been deterred (i.e., had free-riding by these countries been deterred); that the quotas were increased not to secure compliance, which could be relied upon in any event, but to secure accession by the countries that had previously withdrawn from the agreement.

The Chayes’ (1995, p. 2) central argument in *The New Sovereignty* is that “as a practical matter, coercive economic – let alone military – measures to sanction violations cannot be utilized for the routine enforcement of treaties in today’s international system, or in any that is likely to emerge in the foreseeable future.” This claim is supported by the data which the Chayes examine. But it is not consistent with all the data that might be examined. In particular, the noncompliance procedures of the Montreal Protocol do allow for the use of economic sanctions.

In 1992, the parties to the Montreal Protocol agreed to an “indicative list of measures that might be taken by a meeting of the parties in respect of non-compliance with the Protocol”, including the use of financial incentives and trade sanctions. These procedures mirror the treaty’s free-rider deterrence mechanism, in that they include both “carrots” and “sticks.” This is important in that the theory of IEAs suggests that free-riding and non-compliance are linked problems and that the sticks used to deter free-riding can also be used to deter non-compliance (Barrett 1997c). But if the theory is right about this, then the absence of sticks should not be a cause for rejoicing. The managerial model of international cooperation promoted by the Chayes is unlikely to sustain full cooperation.

#### 4. Conclusions

As I write, the climate change talks have just concluded in Kyoto. If reports in the London newspapers are to be believed, negotiations concentrated on whether the OECD countries should reduce their carbon dioxide emissions by a lot or a little. Now, it seems, the negotiators have settled this question, by a compromise as usual.

More important, perhaps, is what the negotiations didn’t settle. The parties were unable to agree on the rules that should govern emissions trading, or on whether

the poor countries must also reduce their emissions, or on how non-compliance will be punished (these matters will undoubtedly be on the agenda of future talks). The agreement doesn't mention how trade leakage will be stopped, and it relies on a participation clause to deter free-riding.<sup>8</sup> The agreement will not come into force until signed by at least 55 countries making up at least 55 percent of the total carbon dioxide emissions in 1990 by Annex I countries.

As a practical matter, it is hard to see how the agreement could enter into law without the United States being among its members. But some United States senators have asked that the treaty be sent to them immediately for ratification, so that they can reject it. The Clinton Administration has already said that it will sit on the treaty for the time being. It is hoping that, given time, public opinion will support the treaty, and force the Senate to ratify it.

So the fate of the agreement and the world climate regime more generally is uncertain. One possibility is that the treaty will not be ratified by the U.S. and so will not come into force. This may not be disastrous, for the consequence may be a renegotiation of the treaty or negotiation of a side agreement, as happened after the U.S. failed to ratify the Law of the Sea Convention. Another possibility is that the treaty will be ratified by all the key parties and so will come into force, but that little action will be taken to reduce emissions, partly perhaps as a consequence of the failure to specify a punishment mechanism. Still another possibility is that the treaty will be ratified, that effective enforcement and free-rider mechanisms will be incorporated into it over time, and that real abatement will ultimately be achieved.

One will never know, but the negotiators might have done better by working backwards; by devising carrots and sticks for deterring free-riding, leakage, and non-compliance, by determining whether these mechanisms require broad or narrow participation, and by calculating whether use of these sticks – should they be needed – would make us collectively better or worse off. Only after these matters were settled should the parties have debated seriously the merits of reducing emissions by given amounts, the appropriate allocation of emission reductions, and the use of innovative mechanisms like emissions trading for achieving these reductions. This, at least, is what I think the theory teaches us.

### **Acknowledgement**

I am grateful to two anonymous referees for helpful comments.

### **Notes**

1. Using a formal, econometric approach, Murdoch, Sandler and Sargent (1997) find that acid rain emissions in Europe are not inconsistent with voluntary emission reductions. Levy (1993), in a case study, comes to the same conclusion. Mäler (1989) estimates the underlying relations in the theoretical model, and finds that full cooperation will require greater cuts in emissions, compared with actual behavior. For related analyses of stratospheric ozone protection, see Murdoch and Sandler (1997), Parson (1993), and Barrett (1994).

2. For a survey of this literature, see Barrett (1997a).
3. Hoel and Schneider (1996) model conformism by supposing that non-signatories to a treaty bear a cost which is increasing in the number of signatories.
4. See, for example, Hoel (1991).
5. See Bodansky (1994) for a description of these negotiations.
6. This is essentially the approach taken by Chander and Tulkens (1995).
7. Chile joined the ICRW in 1979, but only after the U.S. threatened Pelly Amendment certification.
8. As signatories reduce their emissions, comparative advantage in greenhouse gas-intensive production will shift to nonsignatories, and emissions by nonsignatories can therefore be expected to rise. This is the "trade leakage" effect. It can in principle be stopped by the use of border tax adjustments or by deterring free-riding. See Hoel (1996) and Barrett (1997b).

## References

- Ausubel, J. H. and D. G. Victor (1992), 'Verification of International Environmental Agreements', *Annual Review of Energy and the Environment* **17**, 1–43.
- Barrett, S. (1994), 'Self-Enforcing International Environmental Agreements', *Oxford Economic Papers* **46**, 878–894.
- Barrett, S. (1997a), 'Towards a Theory of International Environmental Cooperation', in C. Carraro and D. Siniscalco, eds., *New Directions in the Economic Theory of the Environment*. Cambridge: Cambridge University Press, pp. 239–280.
- Barrett, S. (1997b), 'The Strategy of Trade Sanctions in International Environmental Agreements', *Resource and Energy Economics* **19**, 345–361.
- Barrett, S. (1997c), 'Collective Rationality in the Theory of International Cooperation', mimeo. London Business School.
- Benedick, R. E. (1991), *Ozone Diplomacy: New Directions in Safeguarding the Planet*. Cambridge, MA: Harvard University Press.
- Bodansky, D. (1994), 'Prologue to the Climate Change Convention', in I. M. Mintzer and J. A. Leonard, eds., *Negotiating Climate Change: The Inside Story of the Rio Convention*. Cambridge: Cambridge University Press.
- Brownlie, I. (1990), *Principles of Public International Law*. Oxford: Oxford University Press.
- Carraro, C. and D. Siniscalco (1994), 'R&D Cooperation and the Stability of International Environmental Agreements', Working Paper 65.94, Fondazione ENI Enrico Mattei.
- Cesar, H. and A. de Zeeuw (1994), 'Issue Linkage in Global Environmental Problems', Working Paper 56.94, Fondazione ENI Enrico Mattei.
- Chander, P. and H. Tulkens (1995), 'A Core-Theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution', *International Tax and Public Finance* **2**, 279–293.
- Chayes, A. and A. H. Chayes (1995), *The New Sovereignty*. Cambridge, MA: Harvard University Press.
- Currarini, S. and H. Tulkens (1997), 'Core-Theoretic and Political Stability of International Agreements on Transfrontier Pollution', C.O.R.E. Discussion Paper No. 9793, Université Catholique de Louvain.
- Dixit, A. K. (1996), *The Making of Economic Policy: A Transaction-Cost Politics Perspective*. Cambridge, MA: MIT Press.
- Folmer, H., P. V. Mouche and S. Ragland (1993), 'Interconnected Games and International Environmental Problems', *Environmental and Resource Economics* **3**, 313–335.
- Heal, G. (1993), 'Formation of International Environmental Agreements', in C. Carraro, ed., *Trade, Innovation, Environment*. Dordrecht: Kluwer.
- Hoel, M. (1991), 'Global Environmental Problems: The Effects of Unilateral Actions Taken by One Country', *Journal of Environmental Economics and Management* **20**, 55–70.

- Hoel, M. (1995), 'Should a Carbon Tax be Differentiated Across Sectors?' *Journal of Public Economics* **59**, 17–32.
- Hoel, M. and K. Schneider (1996), 'Incentives to Participate in an International Environmental Agreement', mimeo. Department of Economics, University of Oslo.
- Levy, M. A. (1993), 'European Acid Rain: The Power of Tote-Board Diplomacy', in P. M. Haas, R. O. Keohane and M. A. Levy, eds., *Institutions for the Earth*. Cambridge, MA: MIT Press, pp. 75–132.
- Mäler, K.-G. (1989), 'The Acid Rain Game', in H. Folmer and E. van Ierland, eds., *Valuation Methods and Policy-Making in Environmental Economics*. Amsterdam: Elsevier.
- Murdoch, J. C. and T. Sandler (1997), 'The Voluntary Provision of a Pure Public Good: The Case of Reduced CFC Emissions and the Montreal Protocol', *Journal of Public Economics* **63**, 331–349.
- Murdoch, J. C., T. Sandler, and K. Sargent (1997), 'A Tale of Two Collectives: Sulphur versus Nitrogen Oxides Emission Reduction in Europe', *Economica* **64**, 281–301.
- Parson, E. A. (1993), 'Protecting the Ozone Layer', in P. M. Haas, R. O. Keohane and M. A. Levy, eds., *Institutions for the Earth*. Cambridge, MA: MIT Press, pp. 27–73.
- Putnam, R. D. (1988), 'Diplomacy and Domestic Politics: The Logic of Two-Level Games', *International Organization* **42**, 427–460.
- Raiffa, H. (1982), *The Art & Science of Negotiation*. Cambridge, MA: Harvard University Press.
- Rose, G. and G. Paleokrassis (1996), 'Compliance with International Environmental Obligations: A Case Study of the International Whaling Commission', in J. Cameron, J. Werksman and P. Roderick, eds., *Improving Compliance with International Environmental Law*. London: Earthscan.
- Roth, A. and F. Schoumaker (1983), 'Expectations and Reputations in Bargaining: An Experimental Study', *American Economic Review* **73**, 362–372.
- U.S. General Accounting Office (1992), *International Agreements are Not Well Monitored*. Washington, DC: U.S. GAO.
- U.S. International Trade Commission (1991), *International Agreements to Protect the Environment and Wildlife*. USITC Publication 2351, Washington, DC.