Article

cmps

Conflict Management and Peace Science
0(0) 1–25
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0738894213491182
cmp.sagepub.com
SAGE

Trade sanctions in international environmental policy: Deterring or encouraging free riding?

Alexandra e. Cirone
Columbia University, USA

Johannes Urpelainen
Columbia University, USA

Abstract
Could trade sanctions improve environmental cooperation by reducing countries' incentives to free ride? While carbon tariffs are a widely debated environmental policy, their ability to facilitate climate cooperation remains unclear. We examine game-theoretic models of environmental cooperation with and without trade sanctions. While trade sanctions prevent free riders from obtaining unfair competitive advantages, we show that they can also impede environmental cooperation. Most importantly, since trade sanctions reduce the cost of unilateral policy, they prevent environmentally inclined countries from credibly threatening to suspend cooperation if other countries defect. We use these findings to illuminate outcomes in normatively important cases such as ozone depletion and overfishing, and discuss how they cast a shadow of doubt on the use of carbon tariffs to enforce climate cooperation.

Keywords
Clean technology, international trade, leadership, nexus, trade–environment, trade sanctions, unilateral action

Introduction

Environmental cooperation is difficult. While countries would reap large collective benefits from reducing pollution and conserving resources, any individual country has a rational incentive to discount those benefits from domestic policy that accrue to foreign countries (Barrett, 2003). A cooperative solution to this problem depends on enforcing commitments, but this is difficult in an anarchic international system (Axelrod and Keohane, 1985). If two

Corresponding author:
Johannes Urpelainen, Columbia University, 420 West 118th Street, 712 International Affairs Building, New York, NY 10027, USA.
Email: ju2178@columbia.edu

or more countries agree on environmental policies that produce global benefits at high domestic costs, each country has an incentive not to comply unless noncompliance is somehow punished. For example, countries might fail to reduce trade carbon emissions or to combat overfishing because many of the benefits from climate mitigation or fisheries protection, respectively, would accrue to foreign competitor countries. The problem is made worse if policymakers emphasize the short-run electoral losses from costly pollution abatement while discounting the long-run environmental benefits.

Globally, rational free riding means that too much pollution is produced and too many resources are wasted. Climate change goes unabated, fisheries are overexploited, and rainforests are destroyed. Would trade sanctions against polluters and other free riders help? The idea behind this policy regime is that countries would punish free riders by imposing trade sanctions on their polluting products (Barrett, 1997; Biermann and Brohm, 2005; Ismer and Neuhoff, 2007; Veel, 2009). For example, countries that overfish could be prevented from selling seafood in global markets. Similarly, countries that fail to comply with climate treaties could be subjected to trade tariffs based on the carbon content of their exports. These trade sanctions would reduce the profitability of free riding, and thus improve compliance with international environmental treaties.

Unfortunately, the political feasibility and effectiveness of trade sanctions remains poorly understood. For trade sanctions to work, two conditions must be met (Schelling, 1966). First, the punishers must be able to *credibly* threaten the defector. Second, the punishment via trade sanctions must be *severe* enough to deter noncompliance. Given these requirements, what are the conditions under which trade sanctions can solve strategic compliance problems in environmental cooperation? Once achieved, are trade sanctions a panacea, or is it possible that trade sanctions could be counterproductive?

These questions are particularly topical in view of recent developments in climate policy. The European Union has been debating the use of "carbon tariffs" against other countries that fail to enact climate policies. In September 2009 Nicolas Sarkozy, the president of France, argued that

> I will not accept a system … that imports products from countries that don't respect the rules [on carbon emission reductions] … We need to impose a carbon tax at [Europe's] borders. I will lead that battle.[1]

This rhetoric notwithstanding, do we really know that trade sanctions can improve environmental cooperation? Surprisingly, previous research offers few insights into the conditions under which trade sanctions can help. Some scholars have examined the possibility that trade sanctions can improve participation in environmental agreements, but these studies mostly omit the political effects of trade sanctions (Babiker, 2005; Kemfert et al., 2004; Lessmann et al., 2009) or rely on strong assumptions, such as perfectly symmetric countries (Barrett, 1997). Others have examined the legal problems surrounding the use of trade sanctions in the World Trade Organization (Horn and Mavroidis, 2009; Veel, 2009; Zhang, 2009), but these studies do not examine the politics of trade sanctions.

We offer new insights into the utility of trade sanctions by analyzing a game-theoretic model of environmental cooperation. In the model, two countries can engage in cooperation to combat international environmental problems. To capture asymmetries in contemporary world politics, one of the countries is assumed to be relatively "enthusiastic" about environmental cooperation, while the other is more "cautious".[2] For example, in the case of ozone

depletion most industrialized countries, who could afford to develop substitutes for ozone-depleting chemicals and were vulnerable to the negative health effects of ozone depletion, would be enthusiastic, while the developing countries would be cautious (Sprinz and Vaahtoranta, 1994). Given this distinction, the model is particularly suitable for analyzing environmental cooperation between asymmetric countries, as is often the case in the North–South context.

In one variant of the model, trade sanctions are prohibited, perhaps owing to international trade rules. In another variant of the model, trade sanctions are allowed. We assume that (a) trade sanctions reduce a free rider's economic payoff when the other country implements environmental policies, and (b) the cost of unilateral environmental policy decreases. Both effects stem from reduced pollution leakage: when trade sanctions are allowed, production does not shift from high-regulating jurisdictions to low-regulating jurisdictions. In other words, trade sanctions "level the playing field" (Houser et al., 2008).

Based on the received wisdom (Babiker, 2005; Barrett, 1997), one might expect that the effects of trade sanctions are mostly beneficial. Indeed, we find that trade sanctions can improve the effectiveness of environmental cooperation by reducing the competitive advantage that free riders obtain when other countries act unilaterally. More interestingly, though, the analysis shows that trade sanctions often *impede* environmental cooperation. To begin with, trade sanctions may increase incentives to defect while foreign countries cooperate. Consider a cautious country that would be willing to cooperate without trade sanctions. When trade sanctions are allowed, this country understands that, even if it defects, pollution leakage is limited. The adverse environmental consequences of defection are limited, so defection pays off. For example, the country might understand that, without trade sanctions, its cooperation is absolutely necessary to prevent the collapse of a fishery. However, under trade sanctions, the country can continue to overfish because the foreign country's actions limit overfishing in any case.

A more serious, and theoretically novel, problem stems from the reduced credibility of punishment under trade sanctions. For example, suppose an environmentally inclined country threatens to punish other countries for free riding by suspending cooperation. If trade sanctions reduce the domestic cost of unilateral pollution abatement, these trade sanctions undermine the credibility of the enthusiast's threat. The enthusiast will continue to abate pollution, with or without reciprocal cooperation. Knowing this, the cautious country's incentive to defect looms large: why not free ride, and thus avoid domestic mitigation costs, if the enthusiast, protected by trade sanctions, continues to abate pollution? For example, if the European Union credibly commits to climate mitigation, and trade sanctions now make doing so less costly, other countries might choose to free ride because they realize that the EU will continue to abate regardless of other countries so that they can enjoy the climate benefits of European emissions reductions without doing anything. Thus, Europe's climate leadership would prove harmful to climate cooperation.

One reason why trade sanctions are currently an important topic in environmental policy is the difficulty of enforcing climate mitigation. Recently, Victor (2011) has argued that conditional policy commitments based on a credible threat of suspension upon noncompliance are essential for breaking the "global warming gridlock". Could carbon tariffs enhance the credibility of such threats? Contravening the conventional wisdom, our findings imply that trade sanctions may undermine the credibility of such threats. Suppose, for example, that the European Union could freely impose carbon tariffs against other countries, such as the China and the United States, if the latter fail to reduce their emissions. By reducing the

domestic cost of emissions reductions, this could trap the European Union into ever more ambitious unilateral policies. Such unilateral policies are not contingent on reciprocity by other countries, so they provide the rest of the world with few incentives to reduce emissions.

This article offers an important theoretical contribution by providing a simple yet nuanced account of the conditions under which trade sanctions help and hurt environmental cooperation. Empirically, the model can help understand variation in the feasibility of environmental cooperation across issue areas. This article also connects the theoretical results of the model to two real world cases of salient environmental cooperation on ozone depletion and fishing. Our model sheds light on why trade sanctions have been much more useful for enforcing the Montreal Protocol against ozone depletion than for achieving cooperation on fisheries management.

Trade sanctions in international environmental policy

By *trade sanctions*, we refer to policies that impose additional costs on foreign producers from jurisdictions with less stringent regulatory policies than those used by the sanctioning jurisdiction. For example, a European carbon tariff that forces foreign producers to pay a fee comparable to the carbon price under the emissions trading scheme of the European Union would be a trade sanction (Biermann and Brohm, 2005). Similarly, an import ban in the United States against foreign tuna produced in ways that do not equate with domestic dolphin-protection policies would constitute a trade sanction (DeSombre, 1995). Banning textiles produced using child labor would also constitute a trade sanction.

Our definition of trade sanctions is strictly restricted to regulatory cooperation. More generally, a large literature on "economic sanctions" examines the use of coercive measures in foreign policy (Baldwin, 1986; Drezner, 2000). The key difference between our definition and the sanctions analyzed in this literature is that our trade sanctions are specifically connected to a domestic regulatory policy. Our trade sanctions are not used to coerce foreign countries to change unrelated policies. Instead, they are used to ensure that foreign producers also comply with domestic regulatory policies. As such, they are more about "leveling the playing field" (Houser et al., 2008) than coercing foreign countries.

The literature on trade sanctions in regulatory policy has approached the issue from several perspectives. Traditionally, a key rationale for the use of trade sanctions has been the fear of a "race to the bottom" (McKenzie and Lee, 1991; Porter, 1999; Urpelainen, 2010b). Trade sanctions could help address this problem because they would prevent foreign producers from acquiring a competitive edge from lax regulatory policies (Bechtel and Tosun, 2009; DeSombre, 1995). If an industrialized country preferred a stringent policy, a trade sanction against foreign producers from low-regulating jurisdictions would ensure that both domestic and foreign producers would be subjected to identical regulatory constraints. However, some scholars have also noted that fears of a race to the bottom are overblown. In many issue areas, upward convergence, or a "race to the top", has been the more frequent outcome (Drezner, 2000; Urpelainen, 2011; Vogel, 1995).

Another reason why trade sanctions might be useful is the problem of pollution leakage. If countries enact regulatory policies, the sources of pollution may shift to nonregulating countries (Babiker, 2005; Barrett, 1997; Peters and Hertwich, 2008; Rauscher, 1997). For example, if the European Union imposes a price on carbon, heavy industries may relocate to countries that do not have carbon policies, such as Ukraine. In the absence of trade

sanctions, Ukrainian produces could export their polluting goods, such as steel, to the European Union. Trade sanctions would prevent this, and thus prevent heavy industries from relocating from the European Union to Ukraine.

The dark side of such trade sanctions is the possibility of "green unilateralism" (Bhagwati, 2000; Biermann, 2001). While trade sanctions can in principle create a level playing field between producers from different countries, they can also constitute *de facto* discrimination against foreign producers. Consider, for example, a stringent health standard for agricultural products. Such stringency does not result in trade diversion if agricultural producers in different countries rely on similar technology, but what if small farms in Africa compete with agribusiness in the United States and Europe? In this case, the stringent standards present a much bigger problem for Africans than for Americans or Europeans, as shown by Henson and Loader's (2001) survey of producers in developing countries about the importance of health and other food standards.

Another subset of the literature focuses on the possibility of using trade sanctions to increase participation in international regulatory institutions. Barrett (1997) argues that trade restrictions between members and nonmembers were an important reason for the success of the Montreal Protocol in inducing developing countries to join the fight against ozone depletion. Several other scholars have examined the use of carbon tariffs in the formation of climate mitigation agreements (Kemfert et al., 2004; Lessmann et al., 2009). Urpelainen (2010b) claims that industrialized countries are often able to induce industrializing countries to join regulatory regimes, yet their ability to induce the least developed countries to participate remains limited.

A key limitation of this research is that it does not actually consider the enforcement phase. In the models presented, it is assumed that cooperation is enforceable so long as countries are willing to join a regime. This omission is surprising given that there exists a large literature on the possibility of enforcing international cooperation through issue linkage (Lohmann, 1997; Oye, 1992; Stein, 1980) and domestic mobilization (Dai, 2010). One partial exception is Urpelainen (2010a), who examines the use of technological standards in enforcing environmental cooperation. However, he does not provide a more general theory of trade sanctions. A key broader contribution of our analysis is to shed light into the politics of enforcing regulatory cooperation through trade sanctions.

A final area of research that is directly relevant to our inquiry is the study of "unilateral leadership" in international cooperation. By unilateral leadership, we refer to the idea that a country—or a tightly integrated group of countries—implements environmental policies that mitigate international negative externalities *without* demanding reciprocity from foreign countries. One example can be found from the case of ozone depletion: the United States was one of the first countries in the world to ban the use of chlorofluorocarbons in spray cans, and the ban was implemented in 1978, years before multilateral negotiations on ozone depletion even began (Manzer, 1990). Another, more recent example is European leadership on climate policy. The European Union has committed itself to reducing carbon dioxide emissions and promoting renewable energy, regardless of what other major emitters do (Schreurs and Tiberghien, 2007).

Several scholars have proposed that regulatory pioneers can promote cooperation by offering an inspiring example (Jänicke, 2005; Skodvin and Andresen, 2006). According to this logic, unilateral leadership can have beneficial demonstration effects. By showing other countries that environmental protection is feasible at an acceptable cost, the leader can change the norms and preferences of other countries. This logic requires, of course, that the leader is actually able to implement effective policies at a relatively low cost.

Others contend that unilateral leadership is generally ineffective because foreign countries have incentives to free ride on the leader's unilateral efforts (Hoel, 1991; Victor, 2011). If the leader enacts ambitious environmental policies, other countries have even fewer incentives to act: why pay the cost of environmental protection, if the leader is already mitigating the worst consequences of the environmental problem at hand?

The strategic use of trade sanctions for unilateral leadership remains unexplored in this literature. Yet there are good reasons to believe that unilateral leadership and trade sanctions are strategically related. On the one hand, trade sanctions may influence countries' incentives to become unilateral leaders. On the other hand, trade sanctions may also reduce the need for unilateral leadership by enhancing the enforcement of international treaties. As we show below, trade sanctions may sometimes even *undermine* the enforcement of international treaties, and thus create demand for unilateral leadership.

We provide an integrated account of the benefits and disadvantages of trade sanctions as an instrument of regulatory cooperation. We present a baseline model of regulatory cooperation without trade sanctions, and then compare the equilibrium outcome to an alternative model that allows states to apply trade sanctions when they enact domestic regulatory policies. This account allows us to characterize the conditions under which international institutions that allow trade sanctions can facilitate cooperation or, less obviously, raise obstacles to it.

Model

The model examines strategic interactions between two countries that are considering environmental policies. We focus on two countries to simplify notation and reduce the number of equilibria, but the basic strategic logic also holds for multilateral cooperation. Let $i = A, B$ denote these countries. The two countries have asymmetric resources and abilities: country A could be thought of as an "enthusiastic" industrialized country, better able to shoulder the costs of policy adoption, while country B could be a "cautious" developing country that faces higher costs of policy adoption. In this setting, it is natural that country A, if any, would exercise unilateral leadership by regulating even though country B does not act. For simplicity, the game is characterized by complete information.

We examine two variants of this game. In one, the countries are able to use trade sanctions. In the other, trade sanctions are disallowed. This analysis allows us to examine whether international legal rules permitting trade sanctions undermine or improve environmental cooperation. For example, the analysis helps us examine whether international trade rules allowing the use of carbon tariffs would be useful for enforcing climate cooperation in the future. For instance, we can use the model to examine whether the World Trade Organization (WTO) could facilitate climate cooperation by explicitly allowing carbon tariffs. The game without trade sanctions would be a world with a WTO policy against carbon tariffs, while the game with trade sanctions would have the WTO allowing carbon tariffs.[3]

Throughout the analysis, we assume that the decision to impose trade sanctions is self-enforcing. In general, this assumption is not trivial because trade sanctions could carry economic costs (Martin, 1993). However, in the case of environmental policy, the assumption is plausible. If trade sanctions are perfectly legal, so that imposing them does not carry a large negative reputational cost, governments have political incentives to impose trade sanctions. This is so for the following reasons. First, the domestic industries that must pay the cost of complying with an environmental policy always have incentives to lobby for trade sanctions.

These industries, whose political clout is often considerable, reward the government for imposing trade sanctions (DeSombre, 1995; Vogel, 1995). Second, failure to impose trade sanctions would reduce the effectiveness of the environmental policy. In the absence of trade sanctions, foreign producers would continue destroying the environment and exporting the resulting goods. Thus, organized domestic environmental groups would have strong incentives to demand trade sanctions. While it seems implausible that countries would engage in full trade wars over environmental policies, requiring that foreign companies comply with domestic environmental policy is not, in and of itself, anything else than an attempt to level the playing field in an effort to remove the unfair competitive advantage of foreign producers. Thus, trade sanctions are not economically destructive to their senders.

The assumption of self-enforcing trade sanctions is also common in the literature. In the formal literature, Barrett (1997) shows that trade sanctions are a self-enforcing way to enforce treaty cooperation. Indeed, most studies of environmental treaty formation now assume that trade sanctions are self-enforcing (Kemfert, 2004; Lessmann et al., 2009). Empirically, it is also true that industrialized countries have not hesitated to impose trade sanctions on imports (Barrett, 1997; DeSombre, 1995; Vogel, 1995). In addition to being theoretically informed, our assumption thus reflects the empirical record rather well.

Sequence of moves

Since we are interested in environmental cooperation, the game we analyze is repeated over an infinite time horizon $t = 0, \ldots, \infty$. This repetition is realistic and allows countries to condition future environmental policy on the other country's past behavior (Keohane, 1986). Let $\delta \in (0, 1)$ denote the common discount factor.[4]

Within each period t, countries $i = A, B$ simultaneously decide on an environmental policy, $P_i^t \in \{0, 1\}$.[5] For simplicity, the environmental policy is modeled as a binary choice between abating and not abating pollution. For example, the choice between abatement and nonabatement could reflect the decision to impose a carbon tax or to impose a quota on the annual fish catch.

Payoffs

Payoffs are allocated on the basis of both players' environmental policies, and countries obtain both environmental and political-economic payoffs. These are summarized in Table 1, separately for the games with and without trade sanctions. We now discuss the payoffs in different circumstances.

If both countries adopt an environmental policy at any given time, country i obtains an abatement benefit, V, normalized at value $V = 1$. However, if only one country adopts an environmental policy, both countries obtain an abatement benefit λV, where $\lambda \in (0, 1)$. Intuitively, abatement is less effective when one of the countries refuses to act. We allow the effectiveness of unilateral action, λ, to depend on the availability of trade sanctions. Without trade sanctions, $\lambda = \underline{\lambda}$, and with trade sanctions, $\lambda = \overline{\lambda}$. We assume $\underline{\lambda} < \overline{\lambda}$. This assumption states that the environmental benefit to policy adoption is higher when trade sanctions are allowed. Previous research on environmental cooperation shows that, without trade sanctions, unilateral environmental policy results in pollution leakage: production shifts from the regulating country to the nonregulating country because the production costs are lower without regulation, and the nonregulating country then exports the goods to the regulating

Table 1. Summary of payoffs.

		State B	
		$P_B = 1$	$P_B = 0$
Game without trade sanctions			
State A	$P_A = 1$	$1 - \overline{C}_A, 1 - \overline{C}_B$	$\underline{\lambda} - \overline{C}, \underline{\lambda} + \pi_B$
	$P_A = 0$	$\underline{\lambda} + \pi_A, \underline{\lambda} - \overline{C}_B$	$0, 0$
Game with trade sanctions			
State A	$P_A = 1$	$1 - \underline{C}_A, 1 - \underline{C}_B$	$\overline{\lambda} - \underline{C}_A, \overline{\lambda}$
	$P_A = 0$	$\overline{\lambda}, \overline{\lambda} - \underline{C}_B$	$0, 0$

country (Babiker, 2005; Barrett, 1997; Rauscher, 1997). This is why trade sanctions play such an important role in the politics of unilateral leadership.

Empirically, the problem of pollution leakage is a serious impediment to effective climate policy (Peters and Hertwich, 2008). For example, if Japan implements carbon tariffs, Chinese exporters to Japan must reduce their carbon emissions or pay the carbon tariff. Without carbon tariffs, Chinese exporters can continue to export to Japan without reducing their emissions. Moreover, some Japanese producers may shift production to China, so as to avoid complying with costly Japanese regulations.

The political–economic payoffs in the game are a function of the costs of adopting the environmental policy. The cost of environmental policy to country i is denoted by the parameter C_i. It represents generic costs of environmental policy, such as increased electricity prices. The cost C_i also depends on political factors. If those interest groups that would suffer from environmental policy are powerful, then the value of C_i is high. For example, in the case of climate policy, countries with a lot of politically organized heavy industry would have high costs C_i. In democratic countries, electoral considerations are also important. All else constant, governments' sensitivity to the cost of environmental policy depends on the electoral importance of those groups who expect high compliance costs. If heavy industry is located in marginal electoral districts, environmental policies that increase electricity prices are politically costly. If malapportionment in the legislature favors rural districts, then environmental policies that increase gasoline prices are politically costly.

In a world with sanctions, the baseline cost of policy is denoted $\underline{C}_i$. In the game without trade sanctions, additional cost is incurred whenever country i implements a policy but the other country $j \neq i$ fails to enact a policy. We represent this cost by $\overline{C}_i$, where $\overline{C}_i > \underline{C}_i$. This parameter represents costs to industry as a result of new domestic environmental policy that could threaten competitiveness vis-à-vis countries without similar environmental policies. This cost is not incurred if trade sanctions are allowed because they can level the playing field. Again, the difference between $\overline{C}_i$ and $\underline{C}_i$ also reflects political considerations. Suppose, for example, domestic producers face stiff competition from foreign producers. In these circumstances, it seems plausible that the cost of $\overline{C}_i$ would be particularly high. The size of $\overline{C}_i$ would also increase if the government's electoral fortunes would depend on these domestic producers, be it because the domestic producers can mobilize voters or because the domestic producers can offer campaign contributions (Grossman and Helpman, 1994).

Since country i suffers a competitive disadvantage from unilateral policy implementation without trade sanctions, it seems plausible that country j obtains a competitive edge. Thus, if

country i unilaterally imposes an environmental policy without trade sanctions, country j obtains a "free rider" bonus π_j. This bonus π_j reflects the free rider's valuation of extra competitiveness. If export industries are influential, they reward their government for failure to implement environmental policies. For example, consider a large steel producer, such as China. If the Chinese government does not impose a carbon tax while export markets (such as the European Union) do, then exporting heavy industries increase their political support to the Chinese government.

For simplicity, we assume $\pi_A < \pi_B$, so that the competitive advantage for free riding is higher for country B than country A. We also assume $C_A < C_B$ (and thus $\underline{C}_A < \underline{C}_B$ and $\overline{C}_A < \overline{C}_B$). This restriction means that country A is "enthusiastic" and country B is "cautious" in the environmental issue at hand. In enthusiastic countries, opponents of environmental regulation, such as heavy industries that do not have access to clean technology, are not as influential as in cautious countries. Moreover, governments of cautious countries emphasize economic growth over environmental protection. In the case of climate change, Victor (2011) argues that industrialized countries are generally enthusiastic while developing countries are more cautious. The former have advanced clean technologies, their citizens can afford to pay for environmental protection and they are not dependent on heavy industry for their livelihoods. In large developing countries, such as China and India, the combination of rapid industrialization and continually high poverty rates means that the government's political survival would be jeopardized by an overly aggressive climate mitigation strategy. According to Barrett (2001), this logic can also be applied to biodiversity protection and many other environmental problems.

Given these primitives, the total payoffs are easy to aggregate. When both countries adopt an environmental policy, they receive the environmental benefits of abatement and lose the costs of policy implementation. In the game with trade sanctions, each country receives $1 - \underline{C}_i$, and without trade sanctions, each country receives $1 - \overline{C}_i$. Similarly, if neither country chooses abatement, each receives a zero payoff, regardless of trade sanctions.

In the case of unilateral policy implementation, payoffs are a function of nonexcludable benefits, costs, and a bonus from free riding. For example, assume country A is the one to adopt the climate policy. Without trade sanctions, the payoff to country A is $\underline{\lambda} - \overline{C}_A$, meaning that this country assumes a reduced benefit and additional costs for abatement. The payoff to country B in this scenario, $\underline{\lambda} + \pi_j$, shows that this player receives the benefit of pollution abatement and a competition bonus as a result of free riding.

In a world with trade sanctions, we see different payoffs. Country A receives $\overline{\lambda} - \underline{C}_A$, while country B receives $\overline{\lambda}$. With trade sanctions, the incentive to enact a unilateral policy grows because the cost decreases. The incentive to free ride also decreases because the bonus π_B disappears.

Equilibrium

As a solution concept, we use subgame perfect Nash equilibrium (SPNE). We solve for a strategy profile such that country i's strategy is a best response to the strategies played by country j in every subgame. We first analyze the equilibria of a one-shot interaction (noncooperation as baseline), and then move on to discuss the implications for repeated play (possibility of cooperation). For repeated play, a formal derivation of the equilibrium conditions can be found in the Appendix.

One-shot game

While environmental cooperation is best modeled as an infinitely repeated game, it is analytically useful to first characterize the "one-shot game" without repetition. In this game, there are three possible Nash equilibria in pure strategies.[6] Of these, we begin with the possibility of mutual abatement. Suppose first trade sanctions are allowed. For both countries $i = A, B$ to enact policies in equilibrium, the following condition must hold:

$$1 - \underline{C}_i. > \overline{\lambda}.$$

Both abate as long as the environmental benefits outweigh the economic costs. To compare, in a world without sanctions we have a similar set of conditions except with the competition benefit added. Specifically, it must be that

$$1 - \overline{C}_i. > \underline{\lambda} + \pi_i$$

for each country. It is easy to see that the benefits of adopting an environmental policy, accounting for the respective costs, must be higher than the benefit received by free riding (π_i). Otherwise mutual policy formation is not possible without repeated play.

The second pure strategy equilibrium is when one country chooses abatement and the other does not, effectively free riding on the other player's policy choice. For this equilibrium to exist, the "enthusiastic" country A must implement a policy because it pays a lower cost. With trade sanctions, we need

$$\overline{\lambda} - \underline{C}_A > 0$$
$$\overline{\lambda} > 1 - \underline{C}_B.$$

For country A, the benefits of the policy must outweigh the individual cost, while for country B the individual cost must be greater than $1 - \overline{\lambda}$.

Without trade sanctions, for country A to cooperate even when country B does not, the benefits to adoption must be higher than both the individual costs and the cost of singularly adopting the policy. However, country B's incentive to free rider is larger because of the free rider bonus. We need

$$\underline{\lambda} - \overline{C}_A > 0$$
$$\underline{\lambda} + \pi_B > 1 - \overline{C}_B$$

The condition is more difficult to meet for country A and easier to meet for country B. Importantly, these conditions show that trade sanctions may transform the game from a PD to a unilateral action game. If trade sanctions reduce the enthusiastic country A's cost of unilateral action sufficiently, the game is no longer a PD because country A provides the environmental benefit even without cooperation. This observation plays an important role in the analysis.

If neither of the above conditions is met, mutual inaction follows. In this equilibrium, unilateral environmental policy is not profitable because the cost of enacting a policy is too high. The conditions for mutual inaction are easier to meet in the world without trade sanctions owing to competitiveness concerns.

Repeated game

For the infinitely repeated game, we are interested in the conditions necessary to sustain abatement when players use the "forgiving grim trigger" strategy. This strategy states that players adopt environmental policy until one defects, at which point both players refuse to abate for k periods before returning to mutual adoption again. During the punishment period, deviations to unilateral environmental policy from equilibrium play do not reduce the length of the punishment period. Our results do not depend on the exact punishment strategy, but for concreteness and simplicity we follow Downs and Rocke (1995) and focus on "reversionary" punishment strategies. The number k measures the most severe punishment that the two countries can credibly commit to.

To sustain mutual abatement as a SPNE, we first need to compare the stream of payoffs of abatement under carbon tariffs with the payoffs from a one-shot deviation and the resulting punishment strategy. While the below analysis applies to any payoffs of the game, the most interesting case is one where the one-stage Nash equilibrium is mutual defection. We need to verify that the threatened punishment is *effective*, or sufficiently severe to deter free riding. If trade sanctions are allowed, this requires

$$(1-\underline{C}_i)\frac{1-\delta^{k+1}}{1-\delta} \geq \overline{\lambda} + 0\frac{\delta-\delta^{k+1}}{1-\delta}.$$

On the left, we have the value of lost cooperation from the punishment. On the right, we have the gain from defection. Rearranging, we have:

$$\frac{1-\delta^{k+1}}{1-\delta} \geq \frac{\overline{\lambda}}{1-\underline{C}_i}.$$

When this condition is met, the punishment strategy would deter defection. As long as the discount factor δ is large enough to meet this condition, the deterrent threat is credible.

However, is the punishment itself *credible*? This requires that neither country has an incentive to unilaterally abate, as shown above:

$$0 > \overline{\lambda} - \underline{C}_A$$
$$0 > \overline{\lambda} - \underline{C}_B$$

In other words, this equilibrium exists as long as the corresponding equilibrium of the stage game is for both countries to not abate. If this is the case, the threat of punishment means there is no profitable one shot and both countries will abate in equilibrium.

Consider now the equilibrium without trade sanctions. In this case, an effective punishment threat requires that the value of lost cooperation exceeds the payoff from defection.

$$(1-\overline{C}_i)\frac{1-\delta^{k+1}}{1-\delta} \geq (\underline{\lambda} + \pi_i) + 0\frac{\delta-\delta^{k+1}}{1-\delta}.$$

Rearranging, we obtain

$$\frac{1-\delta^{k+1}}{1-\delta} \geq \frac{\underline{\lambda} + \pi_i}{1-\overline{C}_i}.$$

This condition is otherwise identical except that the economic incentive to defect is larger because the free rider's bonus π_i is available while the environmental consequences of defection are more severe because unilateral policy by the foreign country only produces abatement worth $\underline{\lambda}$ instead of $\overline{\lambda}$.

Again, we also need to verify that neither country has an incentive to abate unilaterally. If unilateral abatement is optimal, clearly the threat of suspending cooperation upon defection is incredible. This requires that

$$0 > \underline{\lambda} - \overline{C}_A$$
$$0 > \underline{\lambda} - \overline{C}_B$$

This condition is easier to meet than the one with trade sanctions because $\overline{C}_i > \underline{C}_i$: without trade sanctions, unilateral action is very costly. Thus, the threat to revert to noncooperation is more credible without trade sanctions.

Effects of sanctions

In this section, we summarize the empirical implications of the model. In particular, we examine when trade sanctions undermine and promote environmental cooperation. In view of this primary interest, we assume throughout that the Nash equilibrium *without* trade sanctions is mutual inaction. This requires that

$$0 > \underline{\lambda} - \overline{C}_A$$
$$0 > \underline{\lambda} - \overline{C}_B$$

This realistic condition states that neither state is willing to implement unilateral policies without trade sanctions. Given this, would allowing trade sanctions facilitate or perhaps even undermine environmental cooperation in the repeated game?

Unconditional pollution abatement

The first possibility is that trade sanctions are so powerful that they help countries dispose of any need for environmental cooperation. Each country adopts an environmental policy without an agreement based on reciprocity, because the benefits of doing so are high.

Proposition 1 (trade sanctions and unconditional pollution abatement). Consider the cautious country B. Suppose (i) the competitive gain from free riding π_B is so high while (ii) the direct cost of pollution abatement C_B is so low that

$$1 - \underline{C}_B > \overline{\lambda};$$
$$1 - \overline{C}_B < \underline{\lambda} + \pi_B.$$

Then sanctions enable otherwise impossible mutual pollution abatement without cooperation, as a Nash equilibrium of the one-stage game.

This proposition states that, if the free rider's payoff π_B is the reason why the cautious country is unwilling to enact an environmental policy, trade sanctions offer an easy way out. The

free rider's payoff π_B is no longer available, so mutual environmental policy formation is possible without a punishment scheme for defectors.

According to the proposition, trade sanctions enable mutual action without an explicit treaty prescribing reciprocal cooperation whenever the following conditions are met. First, we need to have $1 - \underline{C}_B > \overline{\lambda}$. This condition states that the direct cost of abatement C_B must be low, while the benefit of mutual action $V = 1$ relative to the payoff from free riding on the foreign country's action $\overline{\lambda}$ must be high. This condition seems plausible, for example, if the benefits from environmental policy are geographically concentrated. In this case, it is difficult to achieve a large domestic abatement payoff simply by free riding on the foreign contribution. Instead, domestic action is needed.

Second, we also need $1 - \overline{C}_B < \underline{\lambda} + \pi_B$. This condition states that unilateral action must not be optimal without trade sanctions. It holds whenever the free rider's economic payoff π_B is large. Trade sanctions remove this incentive to free ride, so when π_B is large trade sanctions can help promote mutual action without reciprocity in the repeated game.

Cooperation on pollution abatement: trade sanctions and effective punishment

The first proposition's scope is limited to environmental policies that produce large national benefits as long as trade sanctions can help countries avoid competitive disadvantage. Yet what about international negative externalities? In this case, the relevant question is whether trade sanctions can help countries enforce an agreement based on reciprocity.

Proposition 2 (trade sanctions and effectiveness of punishment). Suppose the enthusiastic country's direct abatement cost C_A obtains such an intermediate value that $0 > \overline{\lambda} - \underline{C}_A$. Consider the cautious country B:

1. If (i) the competitive gain from free riding π_B is so high while the (ii) direct cost of pollution abatement C_B obtains such a low value that

$$\frac{1-\delta^{k+1}}{1-\delta} > \frac{\overline{\lambda}}{1-\underline{C}_B}$$
$$\frac{1-\delta^{k+1}}{1-\delta} < \frac{\underline{\lambda}+\pi_B}{1-\overline{C}_B}$$

then sanctions enable otherwise impossible pollution abatement through cooperation, as an SPNE of the repeated game.

2. If (i) the competitive gain from free riding π_B is so low while the (ii) direct cost of pollution abatement C_B obtains such a high value that

$$\frac{1-\delta^{k+1}}{1-\delta} < \frac{\overline{\lambda}}{1-\underline{C}_B}$$
$$\frac{1-\delta^{k+1}}{1-\delta} > \frac{\underline{\lambda}+\pi_B}{1-\overline{C}_B}$$

then sanctions prevent otherwise feasible pollution abatement through cooperation, as an SPNE of the repeated game.

The first part of this proposition states that trade sanctions can reduce incentives to defect given the expected punishment strategy. Suppose the more enthusiastic country A can credibly commit to punishing the cautious country B, so that $0 > \overline{\lambda} - \underline{C}_A$. In this case, trade sanctions can help enforce environmental cooperation by removing the free rider's payoff π_B.

More surprisingly, it turns out that trade sanctions can also prevent environmental cooperation. To see this, suppose that π_B obtains a relatively low value, $\pi_B \to 0$. In this case, the free rider's economic payoff is not a primary impediment to environmental cooperation. Yet how can trade sanctions impede cooperation? The reason is that they *reduce* the environmental cost of defection. When trade sanctions are allowed, the defector's environmental payoff increases from $\underline{\lambda}$ to $\overline{\lambda}$ because "pollution leakage" from the other country, which continues to implement the environmental policy for one period, is reduced. If this reduced environmental cost is enough to tilt the balance in favor of defection, then trade sanctions do prevent environmental cooperation.

When should we expect trade sanctions to be harmful? Two criteria should be met. First, the free rider's economic payoff π_B should be relatively low, so that cooperation without trade sanctions is possible. Second, the incentive to defect should be relatively low, so that enforcement even without trade sanctions is possible. In this case, it is paradoxically possible that trade sanctions impede environmental cooperation.

The results are illustrated in Figure 1. This figure shows the parameter values that admit cooperation for different discount factors δ. In each figure, cooperation is possible with trade sanctions on the left side of the shaded region. With a low δ, enforcement becomes impossible in the area in which the free rider's bonus and the abatement cost reach very high levels. Cooperation is possible without trade sanctions in the region labeled "Yes", while the lower shaded region allows cooperation without, but not with, trade sanctions.

Cooperation on pollution abatement: trade sanctions and credibility of punishment

The results in the previous section were driven by the *effectiveness*: if trade sanctions increase (decrease) the effectiveness of punishment, then they promote (undermine) environmental cooperation. But environmental cooperation also requires that punishments be *credible*: the putative punisher must be able to commit to implementing them. Otherwise they are empty threats. Interestingly, the following proposition shows that trade sanctions can undermine the credibility of the punishment for environmental free riding.

Proposition 3 (trade sanctions and credibility of punishment). Suppose the enthusiastic country's direct abatement cost C_A obtains such a low value that $\overline{\lambda} - \underline{C}_A > 0$. Consider the cautious country B. If the competitive gain from free riding π_B and the direct cost of pollution abatement C_B obtain such intermediate values that

$$\frac{1-\delta^{k+1}}{1-\delta} > \frac{\underline{\lambda} + \pi_B}{1-\overline{C}_B}$$

but

$$1 - \overline{C}_B < \underline{\lambda} + \pi_B$$

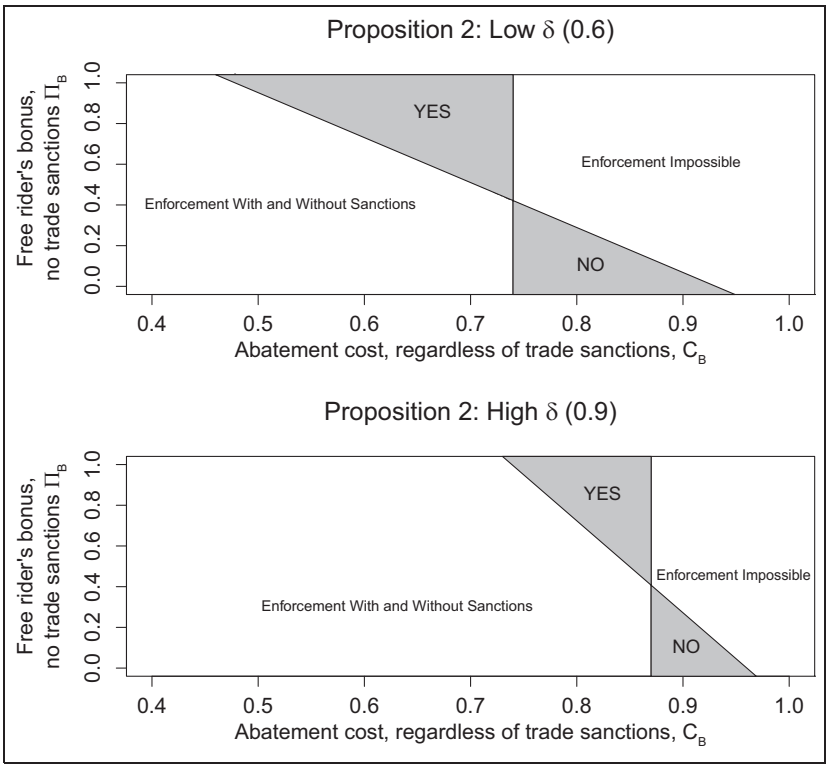

Figure 1. Proposition 2 states that trade sanctions enable and prevent environmental cooperation. In the region labeled "Enforcement With and Without Sanctions", cooperation is possible in both variants of the game because the incentive to free ride is low. In the NO region, cooperation is only possible *without* trade sanctions used because the incentive to free ride is high with trade sanctions. In the YES region, cooperation is only possible *with* trade sanctions because the incentive to free ride is high without trade sanctions. In region labeled "Enforcement Impossible", cooperation is impossible regardless of sanctions because the incentive to free ride is high. In this figure, k is fixed at 5, $\underline{\lambda}$ at 0.2, and $\overline{\lambda}$ at 0.6.

then sanctions prevent otherwise feasible pollution abatement through cooperation, as an SPNE of the repeated game. Whenever $1 - \underline{C}_B < \overline{\lambda}$, in equilibrium country A abates unilaterally.

This proposition reveals an interesting paradox of trade sanctions: even though they increase the *effectiveness* of a credible punishment threat, they may decrease the *credibility* of an effective threat. Trade sanctions allow the enthusiastic country A to reap large benefits from unilateral environmental policy. Thus, as country B considers defection, it understands that the enthusiastic country A will continue to provide environmental public goods even though it threatens to revert to noncooperation. Thus, country B can defect without worrying about the future.

Although counterintuitive, this effect turns out to be quite plausible upon closer reflection. For this credibility problem to exist, it suffices that the enthusiastic country A is willing to act unilaterally under trade sanctions. This is not implausible given that trade sanctions mitigate the adverse economic effects of unilateral action and may even produce rents for domestic

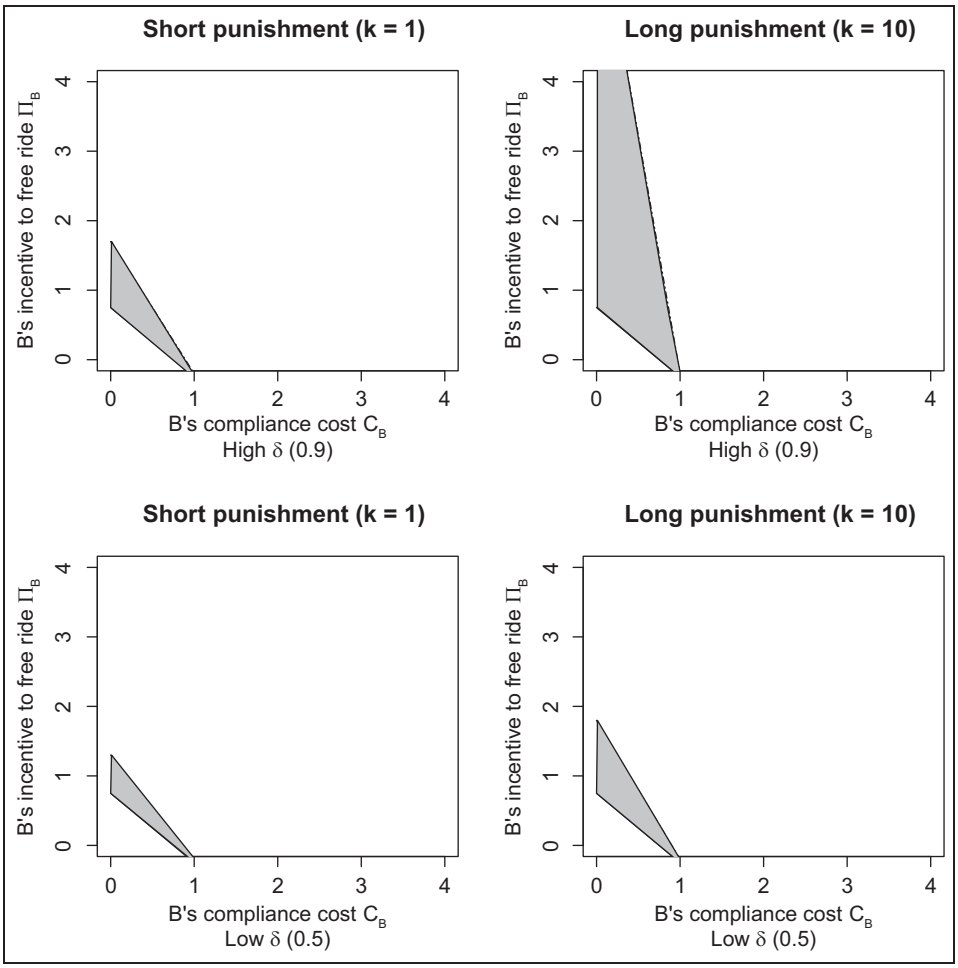

Figure 2. Proposition 3 states that trade sanctions prevent environmental cooperation when $\overline{\lambda} - \underline{C}_A > 0$ because the enthusiastic country cannot credibly commit to punishing the cautious state for free riding. Below the shaded region, mutual environmental policy is possible without the repeated game regardless of sanctions. Above the shaded region, cooperation is impossible regardless of sanctions. In the shaded region, trade sanctions *prevent* environmental cooperation. In this figure, $\underline{\lambda}$ is fixed at 0.2.

producers who already comply with regulations. Given these benefits, country *A* becomes so enthusiastic that it can no longer credibly threaten country *B* with suspension of the environmental policy.

This logic is illustrated in Figure 2. The figure shows, for different punishment periods *k* and discount factors δ, the regions in which trade sanctions undermine the credibility of the punishment needed to sustain cooperation. In the shaded region, environmental cooperation can be enforced without trade sanctions. However, since trade sanctions prevent the enthusiastic country *A* from credibly threatening the cautious country *B* with punitive consequences for defection, environmental cooperation is no longer enforceable.

Table 2. Summary of empirical illustrations.

Case	Cautious country: compliance cost	Cautious country: incentive to free ride	Enthusiastic country: unilateral action payoff	Outcome
Ozone	Intermediate	Intermediate	High with trade sanctions, low without	Trade sanctions and successful cooperation
Fisheries	High	High	High with trade sanctions	Trade sanctions with unilateral action by the United States; many other countries free ride

Importantly, the figure shows that the adverse credibility effect of trade sanctions is maximized when the punishment period k is long and the discount factor δ is high. To see this, note that the size of the shaded region is maximized when k and δ are high. These are the conditions under which cooperation is possible without trade sanctions. When they hold, the region in which trade sanctions prevent otherwise feasible environmental cooperation is large.

Empirical illustrations

While a full test of our model is beyond the scope of this article, it is nonetheless useful to illustrate the model with a systematic comparison of two cases: ozone depletion and overfishing (Table 2). In the case of ozone depletion, trade sanctions allowed industrialized and developing countries to achieve relatively high levels of cooperation, and the addition of financial assistance sealed the deal by ensuring the participation of developing countries. We argue that the developing countries, represented by the "cautious" country *B*in our model, faced intermediate compliance costs and incentives to free ride, so trade sanctions were useful.

In the case of overfishing, the United States faces strong domestic political pressure to act unilaterally, so our model correctly predicts that multilateral measures have been relatively ineffective. The US policy of unilateral regulation is domestically popular, in large part thanks to the possibility of trade sanctions, so the government cannot afford politically to dismantle the regulations. Thus, other countries can free ride on US efforts to protect the fisheries.

Ozone depletion: trade sanctions and successful cooperation

Ozone depletion is a salient example of an environmental issue that requires international coordination to solve, and the Montreal Protocol provides a case study of the conditions under which trade sanctions can help foster cooperation. Regulation of the ozone layer is a global public good but is a very costly endeavor, especially for developing countries. Developing countries have far fewer resources, and face a tradeoff between development and ozone protection. In this case, how did sanctions play a role in eliciting successful cooperation?

The Montreal Protocol on Substances that Deplete the Ozone Layer (MP) was a treaty signed in 1987, and is notable for using trade sanctions as an enforcement mechanism to reduce the consumption of ozone-depleting substances. Signatories of the treaties are required to ban trading with nonparties in ozone-depleting substances such as CFCs, or products relating to these substances. A key element of the MP is that signatories who are found to be in compliance with phase-out schedules for ozone-depleting substances are exempt from trade sanctions, therefore basing compliance on the environmental objective at hand and not general membership (Zhang, 1998). It established the use of trade sanctions for specific environmental purposes, and thus has been found to uphold both GATT and WTO regulations (Dröge et al., 2004). Furthermore, developing countries are assisted in the objectives of the MP by the Multilateral Fund, a set of finance and technology transfer mechanisms. These mechanisms played an important role in altering the costs of compliance for cautious nations.

For developed countries, the costs of abatement varied. Sprinz and Vaahtoranta (1994) operationalize the economic costs of ozone depletion as intensity of CFC consumption in 1986, measured by net atmospheric increase in relation to GNP per capita. Abatement costs increase with higher consumption of CFCs per unit of GNP. France, Germany, Italy, Japan and the UK had relatively high costs, while emission reduction was less economically damaging in countries like Australia, Canada, Switzerland and the Nordic countries. However, regardless of economic hardship of abatement, developed countries were always better able to reduce these costs by the adoption of new technology that developed substitutes for ozone-depleting compounds. These industry-developed compounds could serve as a substitute for specific CFCs, and the investment in such technology was relatively easy, and highly beneficial, for developed nations.

The situation was quite different for developing countries, however. The ability to produce substitutes can significantly lower abatement costs, but the price of such technology for substitutes was extremely high for developing countries. Furthermore, developing countries were sparse consumers and producers of ozone-depleting substances at the time of the Montreal Protocol in the 1980s, although this was expected to change (especially for India and China). The prohibitive cost of new technology and lack of incentives for participation was reflected in the initial participation in the treaty by 1989: the only major CFC-using developing countries that had signed on were Mexico, Nigeria, and Venezuela (DeSombre, 2000–2002). The reluctance of China and India to join the Montreal Protocol illustrates that trade sanctions were not enough, and that abatement costs are an important driver of the efficacy of sanctions (as argued by our model). As of November 2011, industrialized countries had contributed more than US$2.89 billion[7] to the Multilateral Fund, significantly reducing the costs of ozone abatement for developing countries.

Considering our model, in this case there were both intermediate compliance costs and an intermediate incentive to free ride for the cautious developing countries—while countries faced potential sanctions for noncompliance that deterred free riding, aid in the form of finance and technology transfers lowered the compliance costs. As a result, developing countries were no worse off as cooperators or defectors. We can see this result in Proposition 2 and Figure 2. Without trade sanctions, there would be no incentive for developing countries to undertake costly measures of regulation, and therefore signatories could not credibly commit to ozone depletion, knowing that they faced competitive disadvantage domestically and the potential of leakage to noncompliers. Furthermore, the fact that the MP was linked to

specific environmental objectives made it easier and to target individual defectors, increasing credibility.

Without trade sanctions, it would have been hard for the industrialized countries to deter nonparticipation: developing countries could have continued to export CFCs to interested consumers, even among members of the MP. Combined with financial assistance, trade sanctions elicited successful cooperation.

Fisheries conservation: trade sanctions and unilateral regulation

International fisheries regulation is a case of when sanctions have been ineffective at inducing environmental cooperation. There are a wide range of policies relating to fishing that fall under the GATT, including conservation of tuna, salmon, dolphins and turtles, but the use of sanctions has had a less than impressive effect.

In general, the international regulation of fisheries is to preserve the sustainability of marine populations, in order to sustain fishing as a renewable resource. One of the main issues in international cooperation on fisheries relates to tuna, and the prevention of the use of driftnets. Historically, tuna have been regulated by a set of international bodies and agreements because of both the tendency of tuna to travel across ocean borders and the large number of countries across the globe that actively fish for it. Tuna is regulated by two international bodies—the Inter American Tropical Tuna Commission and the International Commission for Conservation of Atlantic Tunas. Regulations include quotas on fish stocks, as well as the prevention of driftnets for conservation purposes.

In the case of tuna regulation, the United States is by far the most enthusiastic country. In terms of seafood, the US fishing fleet participates in the large fisheries of both salmon and tuna, and the United States makes up more than half the world market for tuna (DeSombre, 2000). As one of the largest global fish producers, it has a vested interest in the fisheries industry and is incentivized to push for regulation to ensure that its fishing fleet will not be economically disadvantaged by competitor countries. As a result, it has passed a number of domestic regulations authorizing tariffs on countries that refuse to negotiate on conservation issues, as in the 1962 US Trade Expansion Act, or directly banning fish imports from countries not undertaking similar conservation efforts to the United States, as in the Tuna Conventions Act of 1950. Both fisheries depletion and noncompliance of member countries with overfishing regulations are especially costly for the United States, providing it a strong incentive to self-regulate and use sanctions to try to ensure compliance.

Yet, many domestic regulations impose costs on US fishing boats that cannot be born by foreign fishing boats (DeSombre, 2000). Compliance is costly for cautious countries, both in the loss of resources spent on new equipment and in the loss in revenues from overfishing. The incentive to free ride for cautious countries is also high. Some developing member states can bear the "costs" of sanctions—for example, states like Peru and Colombia export negligible amounts of tuna to the United States, and did not comply with US-led dolphin protection regulation. In the parallel case of salmon, quotas imposed went ignored by member states who achieved opt-out clauses, demonstrating that sole membership in an international organization is not enough to achieve compliance. Finally, the consequences of trade sanctions were made negligible by states that went around regulations by changing the way fish were caught or exporting to smaller markets.

As shown in our Proposition 3 and Figure 2, the enthusiastic country A must be able to credible threaten the cautious country B with punitive consequences for defection.

Otherwise, environmental cooperation is not enforceable because defection is not punished. The problem is that, while sanctions cannot induce cooperation, for the United States it is clearly a dominant strategy to pursue them while unilaterally regulating the fisheries. As a threat, US sanctions are too credible: the federal government has an incentive to implement them, and the accompanying fisheries regulations, *regardless* of whether other countries reciprocate or not.

In particular, US policy regarding sanctions is driven by policy coalitions of "baptists and bootleggers" (DeSombre, 1995) who push for environmental regulation. This coalition consists of domestic environmentalists who win initial adoption of environmental regulation, and industry groups who realize that they could be competitively disadvantaged without further internationalization of domestic standards. The combined support from environmentalists and industry groups in regulating both the salmon and tuna fisheries make the domestic costs of noncompliance high, so the credibility of pursuing another course of action is very weak: the government cannot afford to dismantle the unilateral regulations on fisheries.

As our model predicts, the incentives of the United States to abide by tuna regulation are so high that the use of trade sanctions undermines the credibility of punishing free riding. While we cannot empirically evaluate the case without trade sanctions, it seems plausible that the United States could more credibly threaten other countries with overfishing. After all, the trade sanctions are the reason why the industry also supports unilateral regulation. Without the trade sanctions, the industry would have much stronger incentives to demand less regulation.

Conclusion

Rapid industrialization in the global South puts unprecedented pressure on the environment. From climate change to rainforest destruction, ever increasing wealth levels pose major problems for the international community. How can countries more effectively cooperate to protect the global environment? For example, could carbon tariffs promote climate cooperation? While some scholars and practitioners have proposed that trade sanctions can help (Barrett, 1997; Lessmann et al., 2009), the conditions under which this policy instrument is beneficial remain poorly understood.

In this article, we have uncovered the *contingent* effects of trade sanctions on environmental cooperation. Unsurprisingly, trade sanctions can help countries enforce cooperation by reducing the free rider's payoff. More surprisingly, trade sanctions can also impede environmental cooperation in two ways. First, they can encourage free riding because they mitigate the environmental harm caused by free riding. A country could decide to free ride because trade sanctions reduce the adverse environmental effects of this free riding. Second, trade sanctions can undermine the credibility of a threat to suspend cooperation. If an environmentally inclined country can mitigate the economic cost of unilateral environmental policy by enacting trade sanctions, this country may not be able to credibly threaten free riders with suspension of cooperation. The domestic payoff from enacting unilateral environmental policies is simply too great when trade sanctions cushion the adverse economic impact.

This finding drives home a greater, albeit somewhat paradoxical, point: trade sanctions may enable unilateral leadership by an enthusiastic country, but this need not be good for the environment. If the trade sanctions decrease the credibility of reciprocal arrangements, then the trade sanctions may undermine cooperation based on reciprocity in the shadow of the future (Axelrod and Keohane, 1985). In such circumstances, a categorical ban on trade

sanctions may allow even the most enthusiastic countries to credibly threaten defectors with noncooperation, and this may result in improved global environmental quality, at least in the long run.

This article produces theoretical and empirical added value. Theoretically, it presents a simple model that produced a nuanced and counterintuitive summary of the variety of effects that trade sanctions can produce. Trade sanctions both deter *and* encourage free riding, depending on the setting in which they are used. They are not a panacea for environmental cooperation, and at worst they may even be harmful.

Empirically, the analysis can shed light on the dynamics of environmental cooperation in normatively important cases, such as ozone depletion and fisheries. While we have not presented a full empirical test, our model is testable and the qualitative illustrations provide ideas for measuring the relevant independent variables and outcomes. In the future, scholars of environmental politics could collect data on different countries' preferences and examine whether these preferences can explain negotiation outcomes and compliance. By exploiting variation in the legal and practical feasibility of trade sanctions, scholarship could also implicitly operationalize our distinction between games with and without the possibility of trade sanctions. Finally, the assumption of self-enforcing trade sanctions could be scrutinized by examining whether countries do impose the trade sanctions they are allowed to in the most important multilateral environmental agreements.

For readers interested in improving contemporary environmental cooperation, the largest payoff must be the policy implications. Chronic distrust prevents countries from engaging in multilateral climate cooperation. According to Victor (2011), credible contingent promises to reduce emissions and deploy clean energy could help break the negotiation gridlock. However, he never explains how such contingent promises can be made credible. Contrary to what President Sarkozy and many other policymakers in industrialized countries claim, our model suggests that the use of carbon tariffs is fraught with dangers. They could trap enthusiast countries into enacting ever more ambitious unilateral climate policies. Such unilateral climate policies can be domestically popular, but they are probably not enough to solve a genuinely global problem.

Acknowledgments

We thank Maureen Lackner, the anonymous reviewers and the editor of *Conflict Management and Piece Science* for helpful comments on a previous draft.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Notes

1. "Sarkozy Calls for Carbon Tax on Imports", *Financial Times*, 10 September 2009.
2. This terminology resembles Victor 's (2011) distinction between "enthusiastic" and "reluctant" countries in climate policy.
3. Importantly, the game we analyze is *not* a conventional Prisoner's Dilemma (PD). In a conventional PD, it is assumed that neither player has an incentive to cooperate. We have no need for such an

assumption. Indeed, our most interesting finding is the possibility that cooperation fails because one of the countries cannot, for political–economic reasons, credibly commit to punishing the other country for free riding. This result would not be possible if we analyzed a conventional PD.

4. Asymmetric time preferences would not change our main results.
5. Given that this is a repeated game, we can suppress time superscripts throughout to reduce notation.
6. Multiple equilibria may exist depending on the parameter values. In this section, we focus on equilibrium existence.
7. See the Multilateral Fund Website, http://www.multilateralfund.org/default.aspx (accessed March 2012).

References

Axelrod R and Keohane R (1985) Achieving cooperation under anarchy: Strategies and institutions. *World Politics* 38(1): 226–254.

Babiker MH (2005) Climate change policy, market structure, and carbon leakage. *Journal of International Economics* 65(2): 421–445.

Baldwin DA (1986. *Economic Statecraft*. Princeton, NJ: Princeton University Press.

Barrett S (1997) Strategy of trade sanctions in international environmental agreements. *Resource and Energy Economics* 19: 345–361.

Barrett S (2001) International cooperation for sale. *European Economic Review* 45(10): 1835–1850.

Barrett S (2003) *Environment and Statecraft: The Strategy of Environmental Treaty-making*. Oxford: Oxford University Press.

Bechtel MM and Tosun J (2009) Changing economic openness for policy convergence: When can trade agreements induce convergence of environmental regulation? *International Studies Quarterly* 53(4): 931–953.

Bhagwati JN (2000) On thinking clearly about the linkage between trade and the environment. *Environment and Development Economics* 5(4): 485–496.

Biermann F (2001) The rising tide of green unilateralism in world trade law: Options for reconciling the emerging North–South conflict. *Journal of World Trade* 35(3): 412–448.

Biermann F and Brohm R (2005) Implementing the Kyoto Protocol without the USA: The strategic role of energy tax adjustments at the border. *Climate Policy* 4: 289–302.

Dai X (2010) Global regime and national change. *Climate Policy* 10(6): 622–637.

DeSombre ER (1995) Baptists and bootleggers for the environment: The origins of United States unilateral sanctions. *Journal of Environment and Development* 4(1): 53–75.

DeSombre ER (2000. *Domestic Sources of International Environmental Policy: Industry, Environmentalists, and U.S. Power*. Cambridge, MA: MIT Press.

DeSombre ER (2000–2002) The experience of the Montreal Protocol: Particularly remarkable, and remarkably particular. *Journal of Environmental Law and Policy* 19: 49–82.

Downs GW and Rocke DM (1995) *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations*. Princeton, NJ: Princeton University Press.

Drezner DW (2000) Bargaining, enforcement, and multilateral sanctions: When is cooperation counterproductive? *International Organization* 54(1): 73–102.

Dröge S, Trabold H, Biermann F, Bröhm F and Brohm R (2004) National climate change policies and WTO law: A case study of Germany's new policies. *World Trade Review* 3(2): 161–187.

Grossman GM and Helpman E (1994) Protection for sale. *American Economic Review* 84(4): 833–850.

Henson S and Loader R (2001) Barriers to agricultural exports from developing countries: The role of sanitary and phytosanitary requirements. *World Development* 29(1): 85–102.

Hoel M (1991) Global environmental problems: The effects of unilateral actions taken by one country. *Journal of Environmental Economics and Management* 20(1): 55–70.

Horn H and Mavroidis PC (2009) Burden of proof in environmental disputes in the WTO: Legal aspects. *European Energy and Environmental Law Review* 18(2): 112–140.

Houser T, Bradley R, Staley BC, Werksman J and Heilmayr R (2008) *Leveling The Carbon Playing Field: International Competition and U.S. Climate Policy Design*. Washington, DC: Peterson Institute for International Economics and World Resources Institute.

Ismer R and Neuhoff K (2007) Border tax adjustment: A feasible way to support stringent emission trading. *European Journal of Law and Economics* 24 (2): 137–164.

Jänicke M (2005) Trend-setters in environmental policy: The character and role of pioneer countries. *European Environment* 15 (2): 129–142.

Kemfert C (2004) Climate coalitions and international trade: Assessment of cooperation incentives by issue linkage. *Energy Policy* 32 (4): 455–465.

Kemfert C, Lise W and Tol RSJ (2004) Games of climate change with international trade. *Environmental and Resource Economics* 28 (2): 209–232.

Keohane RO (1986) Reciprocity in International Relations. *International Organization* 40(1): 1–27.

Lessmann K, Marschinski R and Edenhofer O (2009) The effects of tariffs on coalition formation in a dynamic global warming game. *Economic Modelling* 26(3): 641–649.

Lohmann S (1997) Linkage politics. *Journal of Conflict Resolution* 41(1): 38–67.

Manzer LE (1990) The CFC–ozone issue: Progress on the development of alternatives to CFCs. *Science* 249(4964): 31–35.

Martin LL (1993) Credibility, costs, and institutions: cooperation on economic sanctions. *World Politics* 45 (3): 406–432.

McKenzie R and Lee D (1991) *Quicksilver Capital: How the Rapid Movement of Wealth Has Changed the World*. New York: Free Press.

Oye KA (1992) *Economic Discrimination and Political Exchange: World Political Economy in the 1930s and 1980s*. Princeton, NJ: Princeton University Press.

Peters GP and Hertwich EG (2008) CO_2 Embodied in International trade with implications for global climate policy. *Environmental Science and Technology* 42(5): 1401–1407.

Porter G (1999) Trade competition and pollution standards: "race to the bottom" or "stuck at the bottom". *Journal of Environment and Development* 8(2): 133–151.

Rauscher M (1997) *International Trade, Factor Movements, and the Environment*. New York: Oxford University Press.

Schelling TC (1966) *Arms and Influence*. New Haven, CT: Yale University Press.

Schreurs M and Tiberghien AY (2007) Multi-level reinforcement: Explaining European Union leadership in climate change mitigation. *Global Environmental Politics* 7(4): 19–46.

Skodvin T and Andresen S (2006) Leadership revisited. *Global Environmental Politics* 6(3): 13–27.

Sprinz D and Vaahtoranta T (1994) The interest-based explanation of international environmental policy. *International Organization* 48(1): 77–105.

Stein AA (1980) The politics of linkage. *World Politics* 33(1): 62–81.

Urpelainen J (2010a) Enforcing international environmental cooperation: Technological standards can help. *Review of International Organizations* 5(4): 475–496.

Urpelainen J (2010b) Regulation under economic globalization. *International Studies Quarterly* 54 (4): 1099–1121.

Urpelainen J (2011) A California effect for international externalities? *International Interactions* 37 (2): 170–189.

Veel P-E (2009) Carbon tariffs and the WTO: An evaluation of feasible policies. *Journal of International Economic Law* 12(3): 749–800.

Victor DG (2011) *Global Warming Gridlock: Creating More Effective Strategies for Protecting the Planet*. New York: Cambridge University Press.

Vogel D (1995) *Trading Up: Consumer and Environmental Regulation in a Global Economy*. Cambridge, MA: Harvard University Press.

Zhang ZX (1998) Greenhouse gas emissions trading and the world trading system. *Journal of World Trade* 32(5): 219–239.

Zhang ZX (2009) Multilateral trade measures in a post-2012 climate change regime? What can be taken from the Montreal Protocol and the WTO? *Energy Policy* 37(12): 5105–5112.

Mathematical Appendix

This Mathematical appendix provides a complete derivation of the "forgiving grim trigger" strategy for the repeated game. It also contains proofs for the three propositions.

Forgiving grim trigger in repeated play

For the repeated game, the "forgiving grim trigger" states that, if either country deviates at time t, then both countries suspend cooperation for periods $t+1, \ldots, t+k$. For this strategy to be an SPNE, we need to verify that both countries are unwilling to abate during the punishment period *and* both countries are willing to abate given the threat of suspended cooperation for k periods.

Consider the first unilateral action during the punishment period. By the definition of the forgiving grim trigger, we need to verify that neither country $i = A, B$ prefers to deviate from inaction given that the other country $j \neq i$ is also not enacting the policy *and* country i's policy has no effect on the length of the punishment period.

With trade sanctions, we need $0 > \overline{\lambda} - \underline{C}_i$ for $i = A, B$. If this condition is met, then neither country obtains an immediate benefit from acting unilaterally. Conversely, without trade sanctions, we need $0 > \underline{\lambda} - \overline{C}_i$ for $i = A, B$. As long as these conditions are met, the forgiving grim trigger is an equilibrium outcome during the punishment period.

Consider now mutual abatement during the cooperation period. With trade sanctions, the payoff from cooperation during the current and the next k periods is

$$(1 - \underline{C}_i)\frac{1 - \delta^{k+1}}{1 - \delta}.$$

The payoff stream from deviation is

$$\overline{\lambda} + 0\delta + 0\delta^2 + \ldots + \delta^k = \overline{\lambda}.$$

Therefore, we need

$$\frac{1 - \delta^{k+1}}{1 - \delta} \geq \frac{\overline{\lambda}}{1 - \underline{C}_i}.$$

Similarly, we establish the condition without trade sanctions,

$$\frac{1 - \delta^{k+1}}{1 - \delta} \geq \frac{\underline{\lambda} + \pi_i}{1 - \overline{C}_i}.$$

This completes the analysis of the equilibrium for the variants with and without trade sanctions.

Proof of Proposition 1

Proposition 1 claims that, when the gain from free riding π_B is high without trade sanctions and the cost of abatement C_B is low, sanctions enable mutual pollution abatement as the Nash equilibrium of the one-stage game, without reciprocal cooperation. For country B, for it to choose cooperation in equilibrium of the one stage, it must be that the payoff to cooperation $1 - \underline{C}_B$ is greater than the payoff from the alternate action of defection $\overline{\lambda}$. Since $\pi_A < \pi_B$ and $\underline{C}_A < \underline{C}_B$ by assumption, this suffices. ■

Proof of Proposition 2

Suppose that $0 > \overline{\lambda} - \underline{C}_A$, so that country A can credibly commit to nonabatement in the punishment stage. Proposition 2.1 claims that sanctions enable mutual pollution abatement as an SPNE of the repeated game when π_B obtains such a high value and C_B obtains such a low value that

$$\frac{1-\delta^{k+1}}{1-\delta} > \frac{\overline{\lambda}}{1-\underline{C}_B}$$

and

$$\frac{1-\delta^{k+1}}{1-\delta} < \frac{\underline{\lambda}+\pi_B}{1-\overline{C}_B}.$$

When the first condition holds, by the equilibrium analysis cooperation is possible with trade sanctions. When the second condition holds, by the equilibrium analysis cooperation is impossible with trade sanctions.

Proposition 2.2 claims that sanctions prevent mutual pollution abatement as an SPNE of the repeated game when π_B obtains such a low value and C_B obtains such a high value that

$$\frac{1-\delta^{k+1}}{1-\delta} < \frac{\overline{\lambda}}{1-\underline{C}_B}$$

and

$$\frac{1-\delta^{k+1}}{1-\delta} > \frac{\underline{\lambda}+\pi_B}{1-\overline{C}_B}.$$

The equilibrium analysis shows that in this case cooperation fails in the game with trade sanctions and succeeds in the game without trade sanctions. ■

Proof of Proposition 3

Suppose the enthusiastic country A's direct abatement cost C_A is so low that $\overline{\lambda} - \underline{C}_A > 0$. This means that the enthusiastic country A abates even during the punishment stage, in the one-stage Nash equilibrium of the game. Consequently, the forgiving grim trigger cannot operate with trade sanctions, and so the first claim of the proposition, concerning the impossibility of cooperation with trade sanctions, is proven.

To show that trade sanctions prevent cooperation, it now suffices to show that (i) an SPNE based on the forgiving grim trigger exists in the game without trade sanctions and that (ii) country B would not abate unilaterally in the one-stage Nash equilibrium of the game without trade sanctions. As shown in the equilibrium analysis of the repeated game, the first condition is met whenever

$$\frac{1-\delta^{k+1}}{1-\delta} > \frac{\underline{\lambda}+\pi_B}{1-\overline{C}_B}.$$

The second condition is met whenever

$$1-\overline{C}_B < \underline{\lambda}+\pi_B.$$

Given this, country B abates if and only if it produces immediate benefits. Therefore, country B fails to abate whenever $1-\underline{C}_B < \overline{\lambda}$. ■