International
Studies
Quarterly

International Studies Quarterly (2009) **53**, 931–953

Changing Economic Openness for Environmental Policy Convergence: When Can Bilateral Trade Agreements Induce Convergence of Environmental Regulation?

Michael M. Bechtel

ETH Zurich

and

Jale Tosun

University of Konstanz

Citizens' concerns about (international) environmental protection standards are of increasing importance to governments in industrially advanced, high-regulating countries. In almost any proposal for a trade agreement, countries with low environmental regulation standards are required to introduce higher policy standards in exchange for high-regulating countries dismantling their trade barriers and granting access to their domestic markets. Low-regulating countries often act as required and introduce legislation aimed at reducing pollution. This leads to declaratory or *de jure* policy convergence. But such legislative action is not always associated with *de facto* or actual policy convergence, since policies are not always enforced. To analyze the strategic aspect of this potential ''slippage,'' we set up a game-theoretic model with imperfect information. In the model, a high-regulating and a low-regulating country negotiate a bilateral free trade agreement with environmental provisions. We show how potential gains from trade, policy enforcement, and reputation costs, as well as domestic demands for environmental protection affect the occurrence of environmental policy convergence through conditional trade agreements. This study thereby advances our understanding of the relationship between bilateral trade and convergence of environmental policies.

Environmental pollution and its consequences seldom stop at national borders. Thus, citizens are increasingly concerned about international environmental protection and these concerns are of importance to governments in industrially advanced, high-regulating countries (Vogel 1995). As a consequence, environmental standards have become a key issue in international politics with repercussions on policy choices at both the domestic and the international level. Domestically, higher demand for environmental protection culminates in higher levels of regulation. Internationally, governments of high-regulating countries, that is countries with relatively strict environmental regulations, respond to their domestic electoral principals by putting environmental issues on the agenda

when negotiating trade agreements (Schott 2004). Two things follow from these observations. First, trade is not only about trade. Second, as in many other areas of international politics, the domestic and the international sphere are related.

Free trade agreements have proliferated in recent years (Kono 2007:178) and have considerable influence on national regulatory policies. This is because high-regulating countries use ''conditional'' trade agreements as a means for inducing their prospective low-regulating trade partners to adopt stricter environmental regulation in exchange for granting access to their domestic markets. Almost any trade agreement between a high- and a low-regulating country includes environmental provisions, which the latter is asked to accept as a prerequisite for concluding the agreement. This kind of issue-linkage contributes to create a ''level playing field,'' in which cross-country differences in regulatory policies are reduced. Such convergence of domestic policies lowers incentives to firms to relocate their pollutive production to jurisdictions with laxer standards, which abates competitive disadvantages and prevents job losses in the high-regulating country.[1] Thus, linking up trade with environmental standards allows the governments of high-regulating countries to gain domestic political support for free trade (Hufbauer and Goodrich 2004:46).

The public debate prior to the adoption of the North American Free Trade Agreement (NAFTA) illustrates the political significance of environmental issues in the context of international trade. The regulatory differences between the United States and Mexico provoked considerable public opposition to a point where it appeared that the US Congress was not willing to ratify the free trade agreement (FTA) without the accompanying supplemental environmental agreements (Rueda 2000). Also, trade agreements following NAFTA, for example the US–Jordan FTA, US–Chile FTA, US–Singapore FTA, and most recently the US–South Korea FTA, contain explicit passages that require US trading partners to significantly improve environmental standards (Rosen 2004:73). Similarly, the European Union (EU) addresses environmental policies when negotiating trade arrangements, for example with the member states of the *Mercado Comun del Sur*, that is Argentina, Brazil, Paraguay, and Uruguay (Vailland and Ons 2002:143). Furthermore, the EU has considerably influenced the regulatory policies of its 12 new member states through conditionality, that is the demand to adopt the entire *acquis communautaire* to become a member of the EU.

Both political scientists and economists have devoted great effort to explore the policy implications of trade and FTAs for environmental standards. However, they have not yet explicitly raised the question of whether such conditional FTAs are a successful instrument for inducing not only *declaratory* but also *actual* upward convergence, that is a *de facto* increase in environmental standards in low-regulating countries. Therefore, we do not know much about the interplay between domestic and international factors when governments negotiate trade agreements which contain environmental provisions. How do, for instance, domestic demands for environmental protection in high-regulating countries influence trade liberalization decisions if there is uncertainty about whether the low-regulating country will enforce compliance with stricter environmental rules domestically? Or how does the risk of detecting non-enforcement influence the high-regulating country's decision to make a trade offer in the first place? As we will show, the literature is surprisingly silent on these questions.

The empirical evidence suggests that non-enforcement is rampant in low-regulating countries. This strongly supports the idea that even when a low-regulating country promises to increase its *de jure* environmental policies, it may not implement them faithfully (Knill, Tosun, and Heichel 2008). The idea that there is a difference between declaratory and actual environmental policy (that is enforced

[1] Of course, this means that fewer jobs are created in the low-regulating country.

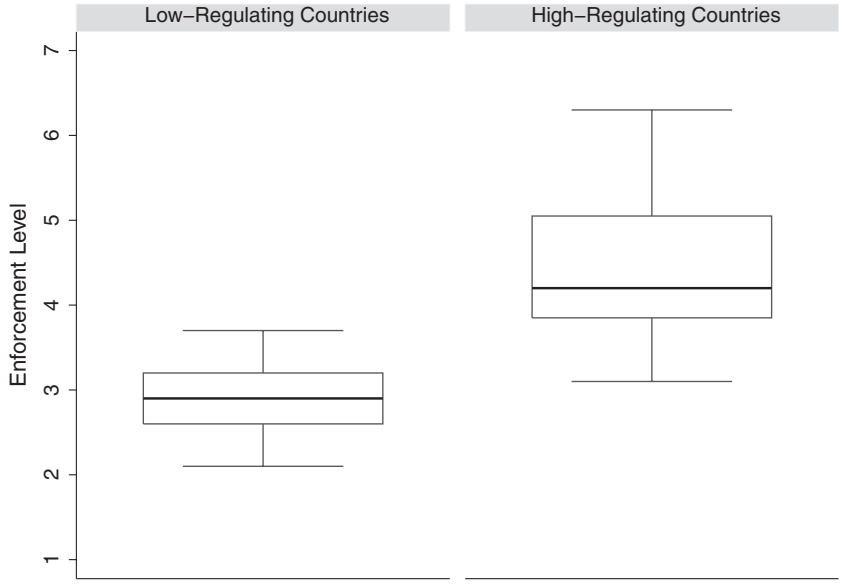

FIG 1. The Variance in Enforcement Levels
Notes. $N = 103$ (Low-Regulating Countries $N = 39$, High-Regulating Countries $N = 64$)
Data Source: 2004–2005 Global Competitiveness Report (Porter et al. 2005:611).

regulation) is also supported by the Executive Opinion Survey, a part of the 2004–2005 Global Competitiveness Report (Porter, Schwab, Sala-I-Martin, and Lopez-Claros 2005:611).[2] Figure 1 presents two box-plots illustrating the difference in regulatory enforcement. Higher levels of enforcement represent less distance between *de jure* standards and enforced regulation. While the data suggest that low-regulating countries (regulation level <3.5) implement environmental policies much less faithfully than high-regulating countries, there is considerable variance in enforcement levels. An important fact which motivates our analysis is that the difference between environmental policy output and enforced environmental regulation varies more strongly in high- than in low-regulating countries. Indeed, the robust (Levene and Brown-Forsythe) test for homogeneity of variances soundly rejects the null of homogeneous variances at the 1% level.

We argue that at least some proportion of the greater variance in high-regulating countries' enforcement levels can be explained by looking at the strategic aspects of non-enforcement. So far, the strategic dimension of enforcing environmental regulation as stipulated by bilateral trade agreements has not been explored. Most gravely, scholars have failed to analytically distinguish the different stages of trade agreement negotiations (that is offer, agreement, and enforcement) and to model their repercussions on environmental policy setting and enforcement accordingly. By highlighting this strategic dimension underlying bilateral trade relations and the (non-)enforcement of regulatory policy, we offer a more instructive theoretical model that brings together arguments which have until now been made separately in the literature.

This paper uses a game-theoretic model of imperfect information to analyze under which conditions low-regulating countries enforce stricter environmental policy standards in exchange for a bilateral free trade agreement. First, we give

[2] The enforcement variable is based on answers to the question: ''Environmental regulations in your country are (answer from 1 = confusing and enforced erratically to 7 = stable and enforced consistently and fairly).'' The regulation level variable is based on answers to the question: ''How stringent are your country's environmental regulations? (1 = lax compared with those of most countries, 7 = among the world's most stringent).''

up the (often implicit) assumption of most studies that policy enforcement is either a costless task or associated with costs which are constant across countries. We argue that low-regulating countries vary with regard to their ability and willingness to effectively enforce environmental regimes. Second, our formalization takes into account how the potential gains from trade and domestic demands for stricter environmental standards affect the occurrence of both upward policy convergence and bilateral trade. Third, and maybe most importantly, we add a portion of reality to theoretical research by introducing uncertainty on the side of the high-regulating country about the costs a low-regulating country faces when enforcing the demanded policy. As we will demonstrate, our model is capable of explaining the occurrence of *declaratory* and *actual* environmental policy convergence and how these are related to bilateral trade liberalization decisions.

Previous Research

Trade and Environmental Regulation in Economics

Economists have concentrated on evaluating the ''pollution haven hypothesis'' (PHH) and the ''environmental Kuznets curve'' (EKC). The PHH states that productive activities for pollution-intensive industries are more costly in high-income than in developing countries due to more stringent environmental standards in developed countries. This motivates pollution-intensive industries to migrate to countries with weaker environmental regimes (Ferrantino 1997; Mani and Wheeler 1998). The EKC holds that there is an inverse U-shaped relationship between income and environmental degradation. From this perspective, environmental degradation—as a function of income—is initially low, reaches a maximum, and subsequently declines as an economy develops.[3]

This body of literature has generated many illuminating and important insights. The main deficiency of the economic literature with regard to our research question is that it fails to explain when regulatory policies in different countries converge. Rather, environmental policies are taken as a given. Also, economic theories give short shrift to domestic political incentives for promoting environmental standards internationally through free trade agreements. Therefore, these accounts offer only limited help for our understanding of the link between conditional trade offers and policy convergence.

Trade and Environmental Policy Convergence in Political Science

The political science literature primarily relates trade and environmental policy convergence by resorting to the concept of regulatory competition, which is based on economic theories of systems competition (Tiebout 1956). The argument behind the regulatory competition hypothesis is that jurisdictions will seek to attract mobile factors of production through reducing domestic regulatory burdens, which is expected to lead to a race-to-the-bottom scenario. In terms of standard game theory, this situation equals a prisoner's dilemma, where for all players deregulation dominates maintaining stricter environmental policies. Therefore, the Nash equilibrium is that all countries lower their environmental standards. However, the empirical evidence lends very limited support to the race-to-the-bottom hypothesis (Bhagwati 2002; Prakash and Potoski 2006). Often

[3] See Grossman and Krueger (1995), Antweiler, Copeland, and Taylor (2001), and Stern (2004) for empirical evidence.

there is in fact an upward movement in terms of environmental standards in low-regulating countries (Drezner 2001:75).[4]

Against this background, various authors argue that the absence of a race to the bottom may also be the result of countries coordinating via supranational institutions, such as the EU (Holzinger and Knill 2004), or the effectiveness of transnational communication (Holzinger, Knill, and Sommerer 2008). Another perspective, which so far has received less attention, is put forward by Hultberg and Barbiery (2004) and Drezner (2005), who emphasize the relevance of differences in negotiation power and the possibility of achieving convergence through coercion. Accordingly, great powers such as the EU or the United States can use their coercive power—including both market access and general political power—to lead other countries into accepting their preferred regulatory arrangements, which can incite a race to the top. A related argument is put forward by Vogel (1995), who claims that the impact of trade offers on regulatory standards is primarily dependent on the preferences of the more industrially advanced, wealthy states. Since industrialized countries tend to have higher regulatory standards, one might hypothesize that we should observe upward convergence of environmental regulation as a consequence of increased economic openness.[5]

Clearly, the political science literature also offers important insights. Yet, we still do not know how demands for regulatory policies influence the occurrence of FTAs and under what conditions they are a successful means to inducing environmental policy convergence in less developed, low-regulating countries. This question is all the more important, since *de jure* is not *de facto* policy convergence. Legislation on stricter environmental policies needs to be enforced to become effective. Thus, legislative action and policy enforcement are two different decisions which need to be distinguished analytically.

In what follows, we bring together arguments which up to now have been made separately in the literature and incorporate them into a unified formal model. First, our formalization takes into account how the potential gains from trade as well as domestic demands for environmental protection standards affect the occurrence of policy convergence and trade. Second, we give up the assumption implicit in most studies that policy enforcement is either a costless task or associated with costs which are constant across countries. Instead we argue that low-regulating countries vary with regard to their ability and willingness to effectively enforce environmental regimes. Third, we introduce uncertainty on the side of the high-regulating country about the costs a low-regulating country faces when enforcing a policy domestically.

The Model

There are two players: a foreign, industrially advanced, high-regulating country F, and a low-regulating country L_θ which can be of two types, so $\theta \in \Theta = \{\text{low,high}\}$. Both types face costs arising from enforcing compliance with environmental standards domestically, but these costs vary. Enforcement costs arise, for example, from administrative agencies periodically monitoring compliance and penalizing violations. Stricter environmental policy may be even more costly to enforce than legislation in other policy fields, because private actors could be more reluctant to adhere to stricter regulations. This may be especially true in the case of firms. Higher environmental standards almost

[4] The empirical literature has also assessed whether differences in environmental policy output and outcome can be explained by system type (democracies vs. autocracies). The results are ambiguous; see Bättig and Bernauer (2009) and Ward (2008) for examples.

[5] This idea has resulted in the ''trading up'' hypothesis. But see Bättig and Bernauer 2009 for contradicting evidence.

always necessitate significant long-term investment in new, ''cleaner'' production technologies. Higher production costs lower firms' profits. Thus, if firms comply with the regulation, this generates additional costs to a government as its tax revenues may decrease. But if firms do not comply, because there is a strong incentive to disregard higher environmental standards, enforcing compliance with these regulatory policies necessitates increased governmental efforts (Gordon and Hafer 2005, 2007). Moreover, regulatory legislation can lower incentives for (foreign direct) investment (Valds 1995), as it influences comparative advantages in pollutive production.

While the low-regulating country L_{high} faces high enforcement costs $\overline{c}$, these costs are low ($\underline{c}$) for L_{low}. This is because countries differ in their ability and/or willingness to effectively enforce stricter regulatory standards. Ability varies because of L_{high} having an inefficient, incompetent, or even corrupt bureaucracy (Desai 1998). Willingness varies due to differences in which groups support L's government domestically (Dai 2006). A government which represents business interests suffers from a strong loss in support if it enforces stricter environmental regulation, while a government pursuing the interests of workers and citizens concerned about environmental protection does not face such costs. A closely related aspect is that the extent to which a regulatory regime depresses tax revenues and/or deters (foreign direct) investment can vary across low-regulating countries.

Gordon and Hafer (2005) investigate the impact of corporate influence on regulatory policies *within* a country. One result is that implementation of regulatory action which imposes costs on firms varies as a function of a corporation's political expenditures. These serve as signals to the bureaucracy which convey information about the costs a firm can impose on the regulatory agency that is supposed to enforce a certain policy domestically by contesting the agency's findings and/or appeal to legislative bodies. If these costs are high enough, they can deter the bureaucracy from rigorously monitoring and enforcing regulatory policies. Since there is variation in which firms or industries can afford these kinds of actions across low-regulating countries, this underscores the need to distinguish between different types of countries with regard to enforcement costs.

The structure of the game is as follows (Figure 2): First, nature draws a type $\theta \in \Theta$ and informs the low-regulating country L about its type. While L knows its type, the foreign, high-regulating country F does not. This is to say that only the government of L knows exactly how costly domestic policy enforcement is. Let t denote the probability that F faces a type with high enforcement costs and assume that $0 < t < 1$, which is to say that there is at least some probability that either of the two types occurs. Not knowing which type nature has drawn, F can either make an offer for a bilateral conditional trade agreement (C) to the low-regulating country or not ($\neg C$). The offer is conditional on L increasing domestic environmental policy standards. If F chooses not to make an offer, both players receive 0.

If an offer is made, L can accept this offer (A), which includes enacting and enforcing the required policy domestically. If it does not agree on enacting the required legislation ($\neg A$), F will refuse to open its domestic markets. F receives some small benefit f for taking domestic demands for promoting environmental standards seriously and standing firm on these issues (Hufbauer and Goodrich 2004). This utility arises from increased government popularity among potential losers of increased economic openness, such as low-skilled workers (Scheve and Slaughter 2001; Mayda and Rodrik 2005) and voters who are interested in environmental protection (Inglehart 1997; Meyer, Bernauer, and Bechtel 2009). Consequently, F's government will be rewarded for not setting up trade relations with a country which clearly refuses to enact environmental policy standards.

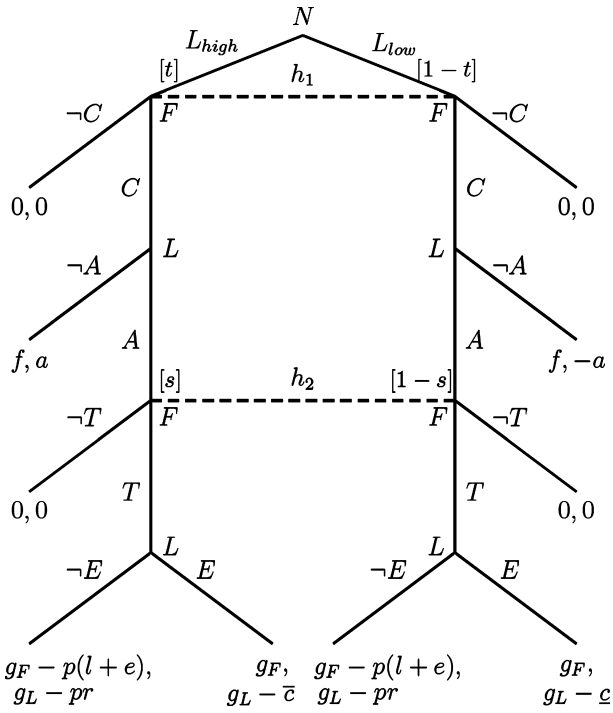

FIG 2. The Environmental Policy Convergence & Trade Game

Since L_{high} faces high enforcement costs, stating that it will implement environmental standards (A) is not in its own best interest, as this might deter (foreign direct) investment and stimulate domestic demands for such policies. Therefore, L_{high} enjoys non-announcement benefits and receives some positive utility a for refraining from announcing its willingness to introduce and enforce such policies. In contrast, L_{low} is sincerely interested in enforcing stricter policy standards and therefore receives $-a$ if it claims not to accept the offer ($\neg A$) as it includes implementation and enforcement of stricter policies domestically.[6]

If the low-regulating country promises to enforce the required policy standards, F can choose whether to trust this statement and accept the agreement or not to do so. In case F decides not to trust L, negotiations end. Thus, both countries forgo the gains from trade $g_F, g_L > 0$, where g_F denotes F's gains from trade and g_L denotes the gains from trade to L, and both receive 0. If F decides to agree on trading, L can either enforce compliance with the policy (E), which means bearing the costs of policy enforcement, or try to shirk enforcement and evade these costs. An L_{high} country faces higher costs $\overline{c}$ than L_{low} to which enforcement $\underline{c}$ is relatively less expensive. However, we assume that gains from trade are always larger than enforcement costs or $g_L > \overline{c} > \underline{c} > 0$. The assumption that g_L is greater than any other parameter reflects that it is generally beneficial for a country to engage in free trade.

If L does not enforce compliance and gets caught, it suffers from some reputation costs $r > 0$ for having lied to its trading partner about enforcing stricter environmental policy. There are several reasons for including reputation costs in our model. First, the high-regulating country may refuse to sign future agreements which extend or deepen trade relations with L. Second, it may be argued that the low-regulating country can be punished by use of economic sanctions. A

[6] Any payoff strictly smaller than the net gain L_{low} gets from trading will lead to the same qualitative result.

third and at least equally plausible reason is that the trade agreement may itself contain a sanctioning mechanism which inflicts costs on the low-regulating country in case of violation.

We assume that these reputation costs are neither very high nor very small, that is $\overline{c}>r>\underline{c}>0$. Of course, it is unreasonable to expect that non-enforcement will always be detected. We take up this argument by discounting the reputation costs with the detection probability $p \in [0, 1]$, which is the chance of finding out that the policies stipulated in the trade agreement are not enforced by L. Although adding the parameter p means to add some complexity to the model, this is necessary because for reputation costs to become effective F needs to know whether L complies with the trade agreement and enforces policies domestically. There are several reasons for which non-compliance might remain undetected. First, it is well known that monitoring government authorities as well as non-governmental organizations often suffer from very limited resources and insufficient expertise. Second, the necessary information is often not publicly accessible and this makes it more difficult to acquire the data needed for solid monitoring. Third, even if such information is available, its reliability is questionable, since low-regulating countries with high enforcement costs might have an incentive to obfuscate non-enforcement.[7]

To maximize analytical leverage, we assume that for L_{high} it is beneficial to evade enforcement costs, while L_{low} prefers policy enforcement. In the stylized world of our model this is to say that $\underline{c}<pr<\overline{c}$. Later we will illustrate how the set of possible equilibria changes due to changes in what we call relative enforcement costs. Also, we make the assumption—as we find realistic—that for L_{high} declaring its willingness to introduce and enforce environmental policy causes minor costs, for example because this may deter some small portion of (foreign direct) investment and stimulate domestic demand for such policies. Thus, telling F that an effective regime will be introduced is not far from cheap talk. In terms of the payoff parameters this means that $0 < a < g_L - r$.

Now consider the payoffs F can receive. Trading yields gains g_F, but if L does not enforce the policies stated in the trade agreement, F suffers from $-(l + e)$. The parameter l captures the electoral punishment by potential losers of increased economic openness such as low-skilled workers. This is because trading with a developing country increases downward wage pressure and raises the probability of a worker losing his job, due to labor intensive production being relocated to countries with lower labor costs. The second parameter, environmental benefits e, reflects how much voters in F value promoting environmental protection. In sum, F's government will be punished by $-p(l + e)$ for setting up trade relations with a low-regulating country which does not comply with environmental policy standards, where $p \in [0, 1]$ is the probability of detecting non-enforcement (or L_{high}'s optimal level of non-compliance).

It is obvious that one important assumption underlying this payoff structure is that countries generally benefit from trade. Also, we make the empirically supported assumption that citizens engage in economic voting (Lewis-Beck 1988; Fair 1996). Economic voting results in F's government being rewarded by the electorate for trading with other countries, since this results in a welfare increase as a consequence of lower domestic prices of imported goods and lower unemployment due to increased foreign demand. In particular, we assume that g_F is strictly larger than any other single payoff parameter we have introduced. However, the electorate of F may punish the government if it does not promote enforcement of environmental standards.

[7] Another useful interpretation of p is that it represents the optimal level of non-compliance chosen by L_{high}. If the level of non-compliance is high, p is close to 1 and the lower the non-compliance the closer p is to 0. The case in which $p = 1$, that is non-enforcement is always detected or L chooses to perfectly comply, is then a simplified (and less realistic) version of the model we present here.

Introducing some notation, let $D_1^F = \{C, \neg C\}$ and $D_3^F = \{T, \neg T\}$ denote F's action set at the first and the third stage of the game. Likewise, $D_2^L = \{A, \neg A\}$ and $D_4^L = \{E, \neg E\}$ are the actions available to L at stage two and four, respectively. Let $\sigma_F \in \Delta D_1^F \times \Delta D_3^F$ denote a (behavioral) strategy for F which describes what F does at the first and the third stage of the game. L's strategy is given by the mapping $\sigma_L : \Theta \rightarrow \Delta D_2^L \times \Delta D_4^L$, specifying what each type θ does at the second and the fourth stage of the game. Finally, we need the belief system $\mu = \{t, s\}$, where t is the prior probability of nature drawing a high cost (L_{high}) country and s denotes F's updated belief, that is the conditional probability of facing an L_{high} country after having observed that the potential trading partner is willing to harmonize its policy.

Policy Convergence and Trade

To solve the model, we apply the concept of sequential equilibrium (Kreps and Wilson 1982). Sequential equilibrium builds on the idea that actors make small errors (trembles). Thereby, sequential equilibria are in some sense more realistic, as they build on some form of boundedly rational behavior while at the same time allowing us to discipline beliefs off the equilibrium path. A formal definition of sequential equilibrium is given in the Appendix.[8] Given our baseline payoff structure, the policy convergence and trade game has four equilibrium continua. We discuss the three major cases and present the fourth equilibrium in the Appendix.

Proposition 1 (Partially Deceptive Convergence I): *If* $t < \frac{g_F}{p(l+e)} = t^*$ *a continuum of pooling equilibria exists in which*

1. *the foreign country always makes a conditional trade offer*
2. *both low-regulating types state that they are willing to enforce stricter environmental policies*
3. *the foreign country believes this claim*
4. *the low-cost country enforces higher regulation while the high-cost country does not.*

Proof. See Appendix.

In this pooling equilibrium both types enact the policy standards F demands. However, the low-regulating country with high enforcement costs L_{high} successfully deceives F, as it does not enforce compliance with the enacted policy. This country consequently enjoys the benefits of trade without paying the costs of enforcement. Although in this case honesty is always the best policy for L_{low}, it is never for L_{high}, as this type does not enforce compliance with the stricter policy at the third stage. Therefore, while declaratory policy convergence occurs, this convergence is partially deceptive. This is because while both types agree on implementing and enforcing the policy, only one type has no incentive to renege. Thus, *de facto* policies do only partially converge upwards. The extent to which policy convergence is deceptive in letting observers believe that convergence of actual policies occurs depends on the beliefs the high-regulating country F holds about the probability of L being of a certain type. Only if $t < \frac{g_F}{p(l+e)} = t^*$ is choosing trade optimal for F, even though there are some countries which only pretend to be willing to enforce compliance with the environmental policies upon which the trade offer is conditional. This threshold t^* shifts as a function

[8] See Kreps and Wilson (1982) for a discussion of the behavioral assumptions underlying this equilibrium concept.

of the gains from trade g_F, the benefits F receives from voters for promoting international environmental standards (negatively), and the probability of catching L_{high} not enforcing compliance with the policy (negatively).

A second equilibrium exists, in which it is optimal for L_{high} to randomize between mimicking L_{low} and being honest about its lack of willingness to enforce compliance with stricter environmental standards domestically. The best F can do is to randomize between initiating trade and not not doing so, thereby sometimes preventing being tricked by L_{high}.

Proposition 2 (Partially Deceptive Convergence II): *If $t \leq t^*$, there exists a continuum of semi-separating equilibria in which*

1. *a conditional trade offer is made*
2. *the low-cost country states its willingness to introduce stricter regulation while the high-cost country randomizes over agreeing to enforce stricter regulatory policies*
3. *the foreign country sometimes believes the low-regulating country and chooses to start trading*
4. *the low-cost country enforces stricter environmental regulation while the high-cost country does not.*

Proof. See Appendix.

The 'bad' thing about this equilibrium is that F sometimes forgoes the gains from trade with a low-regulating country which would enforce stricter regulation. In sum, if a trade agreement is reached and trade is observed, this is only partially associated with actual policy convergence, as L_{high} prefers not to enforce compliance at the final stage.

As we prove in the appendix, in this equilibrium the high-regulating country randomizes with probability $\sigma_F(C) = \frac{a}{g_L - pr}$ over making or not making a conditional trade offer. Thus, trade negotiations are more likely to occur, if announcement benefits (a) of the high-regulating country increase, because $\frac{\partial \sigma_F(C)}{\partial a} = \frac{1}{g_L - pr}$ and $r < g_L$. However, trade offers are less likely to be made if the low-regulating country's gains from trade increase. Interestingly, lowering reputation costs to the low-regulating country increases the probability of a trade offer.

If the proportion of low-regulating countries with high enforcement costs is too large, F prefers not to make a conditional trade offer in the first place. The following proposition characterizes this equilibrium.

Proposition 3 (Status quo): *If $t > t^*$, a continuum of semi-separating equilibria exists in which*

1. *no trade offer is made by the foreign country*
2. *if an offer were made, the low-cost country would agree to enforce higher standards, while the high-cost type would sometimes articulate its willingness to enforce stricter regulatory standards*
3. *the foreign country would then sometimes trust this statement and start trade*
4. *the low-cost country would enforce stricter environmental policies while the high-cost type would choose not to do so.*

Proof. See Appendix.

Note that this equilibrium also comprises policy divergence which occurs if either F continues to increase or L_{high} lowers its policy standards (or both). In any of these cases the variation in policies increases, that is policy divergence can be observed.

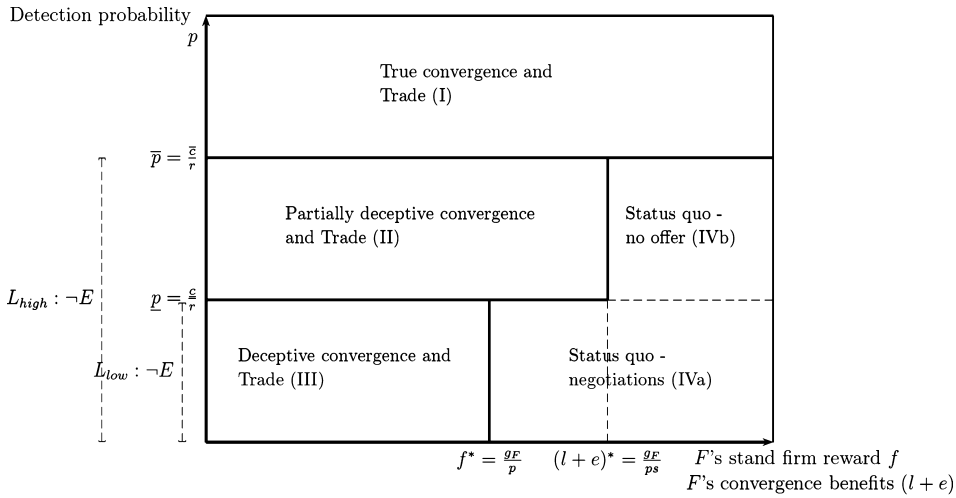

FIG 3. Equilibria of the Environmental Policy Convergence & Trade Game

As this analysis shows, there is no such thing as a ''True Convergence''-separating equilbrium in which trade is always associated with a *de facto* increase in regulatory standards assumption—at least given our baseline order imposed on the payoff parameters and that t cannot be equal to either 0 or 1. The question then is what conditions have to be met in order for such an equilibrium to exist? We address this question in the following section.

Existence of a "True Convergence" Separating Equilibrium

To explore how the equilibrium continua vary in response to changes in the probability of detecting non-enforcement, we will relax some of the assumptions regarding the order imposed on the payoff parameters. Let $\overline{p}$ denote the upper detection probability threshold such that $\overline{p} = \frac{c}{r}$ and $\underline{p}$ is the lower detection probability threshold with $\underline{p} = \frac{c}{r}$. By construction, if the detection probability is larger than the upper threshold ($p > \overline{p}$), enforcing compliance at the last stage is the optimal choice for L_{low}, and also for L_{high}, as $\overline{p} > \underline{p}$.[9] This corresponds to the upper region in Figure 3. Since F knows that a trade agreement conditional on policy harmonization will be associated with true convergence, for any value of F's convergence benefits $l + e$ and stand firm rewards f, making an offer and starting trade is the best response. Thus, in region I we can observe trade together with true policy harmonization.

If we reduce the detection probability p such that it is lower than the upper threshold $\frac{c}{r}$, but still larger than the lower critical value $\frac{c}{r}$, we are confronted with the case which has been analyzed above (II). This equilibrium breaks down if F's convergence benefits are larger than the gains from trade g_F relative to the probability p of catching L not enforcing compliance with the regime. In this case, F will not even make a conditional trade offer in the first place and the status quo is preserved (case IVb).

An interesting situation is characterized by Case III, which is an illustration of pure conflict between the low- and the high-regulating country. As $p < \underline{p}$, no type will choose to enforce compliance with the higher regulatory standards domestically. However, this setup is interesting, because bilateral trade agreements are still possible. To see this, we need to ask when it is optimal for the

[9] These unique critical values follow directly from L's best response correspondence at stage 4 (see Appendix).

high-regulating country to engage in free trade even if it knows that L will not enforce stricter regulatory standards. As can be seen from Figure 1, the stand firm benefits F receives for not continuing negotiations with a country which openly opposes higher environmental standards need to be smaller than the relative gains from trade $\frac{g_F}{p}$. This means that $pf < g_f$ and this condition actually is satisfied given our initial payoff structure, since we assumed gains from trade to be larger than any other single payoff parameter and p can only vary between 0 and 1.

Finally, Case IVa resembles a situation of unsuccessful negotiations. If the reward for standing firm on stipulating environmental policies in the trade agreement is large enough ($f > f^* = \frac{g_F}{p}$), F will make a trade offer but refuses to conclude the trade agreement at stage three. The refusal to start trade at the third stage will be anticipated by L_{high}, which then prefers to admit that it is not willing to enforce compliance with the environmental policies F demands. Since F is unwilling to trade with a country which openly dismisses environmental standards, no agreement is reached. This will give F the reward f for standing firm on environmental issues.

Clearly, equilibrium region I is particularly interesting, as this outcome represents the separating equilibrium in which a conditional trade offer is not only associated with *declaratory*, but also *de facto* policy convergence, that is true convergence. According to our analysis, the upper threshold (UT) of the detection probability determines whether a low-regulating country chooses to enforce or not to enforce stricter regulatory standards domestically. It is worth exploring the comparative statistics of this equilibrium in more detail. Figure 3 shows how the upper threshold varies as a function of both the costliness of domestic policy enforcement for a high-cost type and reputation costs.

The region above the hyperplane represents the continuum of separating equilibria. In this region, true convergence occurs between those high- and low-regulating countries which conclude a trade agreement (cases II, III, IVa, and IVb in Figure 3). The region below the hyperplane comprises pooling and partly-pooling outcomes. Since our main interest is in differentiating between separating and non-separating equilibria, only these two regions are distinguished in Figure 4.

While it would be desirable to conduct a statistical analysis to empirically evaluate the usefulness of the model, this task is beyond the scope of this paper. However, to give an impression of how the model can be applied to understand the complex dynamics of trade offers and their connectedness with environmental regulation nowadays, the following section offers a brief case study illustrating the analytical leverage of the theory.

Empirical Illustration

In the previous section, we have formally examined the conditions leading to a ''True Convergence'' equilibrium, in which trade agreements are concluded only with those low-regulating countries which have a true interest in stricter environmental standards. To illustrate our equilibrium predictions empirically, we now consider the effect of bilateral trade negotiations and agreements on environmental standards in three (formerly) low-regulating countries, namely Chile, the Czech Republic, and Mexico. These three cases are particularly instructive since they illustrate the differential impact of conditional trade agreements on environmental policy convergence. Of course, the environmental policy developments in these countries can be traced back to a plethora of factors. The purpose of these case studies is to illustrate our (reductionist) theoretical argument rather than to provide an in-depth description. We focus on the effects of key variables from our formal model on environmental policies and enforcement efforts in three countries.

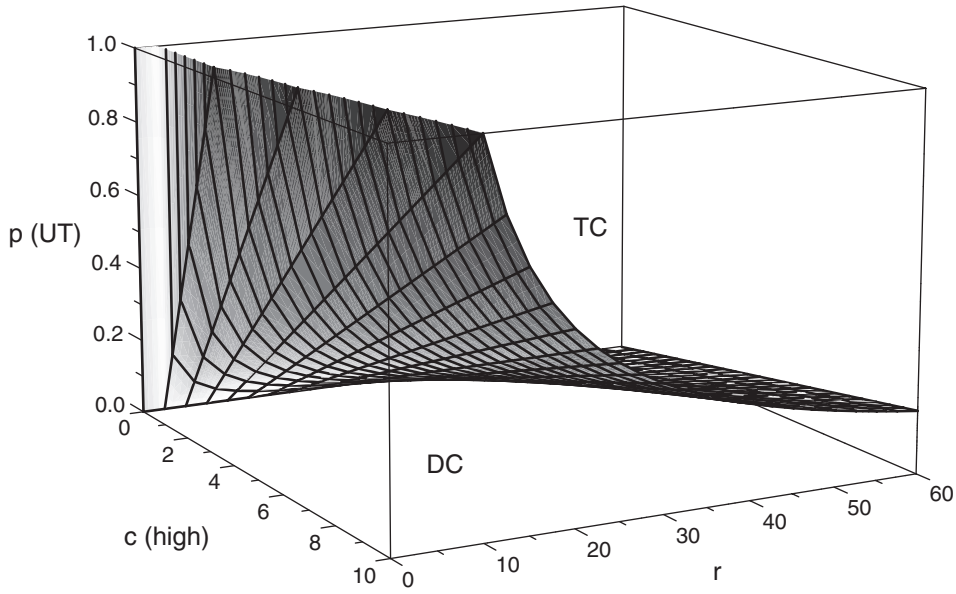

FIG 4. Simulating the Upper Threshold (UT) of the Detection Probability (p): The Effect of Enforcement (c) and Reputation Costs (r)

We first turn to the consequences of NAFTA integration during the late 1980s and early 1990s on Mexican environmental standards. Until that time, Mexico did not have a consistent environmental policy in place. This situation, however, changed, when the US Congress threatened to abort the NAFTA project, due to environmental reasons. As a response, Mexico declared its willingness to sign NAFTA's environmental side agreement and to enact United States-style environmental regulations (Knill, Tosun, and Heichel 2008:1029). At the same time, the Mexican environmental authorities also intensified monitoring and enforcement activities. But once the international scrutiny directed at Mexico in the context of the NAFTA debate ceased after 1992, the literature (Mumme 1998; Auer 2001) largely agrees that enforcement activities declined. Thus, the main deficiency of Mexican environmental policy remains the insufficient enforcement of the legal provisions (Mumme and Lybecker 2002:317). With reference to our theoretical model, the NAFTA negotiations between the United States and Mexico describe a situation in which a conditional trade offer was made and an agreement reached. However, despite considerable political pressure during the pre-accession period, the outcome can be judged as deceptive environmental policy convergence.

This particular experience with Mexico also had repercussions on the negotiations of an ''extended'' NAFTA between the United States and Chile, since various interest groups used NAFTA ''failures'' to derail other FTA negotiations during the latter half of the 1990s (Hufbauer and Goodrich 2004:46). In 1994, President Clinton formally invited Chile to join NAFTA. However, the accession negotiations that began in June 1995 came to a sudden halt by the end of the same year. This was caused by the Clinton administration's failure to achieve a renewal of its fast track authority, which had expired in 1994.[10] To a certain degree, the failure to reinstate fast track authority resulted from concerns about Chile's low environmental policy standards. In response, the Chilean government

[10] The fast track authority would have ensured the executive the power to negotiate with Chile and put the resulting agreement up to a simple yes-or-no vote in Congress, hence excluding the possibility of making amendments.

articulated its willingness to address these concerns by signing a side-agreement, but objected to the use of trade sanctions as a remedy for non-compliance (Weintraub 2004:86). This position did not convince Congress and negotiations failed.[11]

Our third example is the Czech Republic in the context of EU eastern enlargement. Since 1990, Czech environmental legislation has strengthened notably. This development predominantly resulted from attempts to bring environmental regulations in line with European standards, which represented a precondition for EU membership (Pavlinek and Pickles 2000:195). When the country received official candidate status in 1998, it had already adopted most European environmental policies, and had therefore achieved a high level of *de jure* policy convergence. As a result of the anticipatory adaptation of European standards and their satisfactory implementation and enforcement, the country became an EU member state in 2004.[12] As concerns enforcement after entering the EU, the country seems to be complying well with the European environmental standards. According to the annual report on monitoring the application of Community law, of the 685 infringement cases under examination on December 31, 2006, merely two are related to the Czech Republic (European Commission 2006). Also with regard to air pollution emissions, the Czech Republic's performance has improved drastically.[13] Thus, the evidence suggests that this represents a case of true environmental policy convergence.

How can we explain these differing results? According to our model, we must focus on the following parameters: (i) enforcement costs c, (ii) the probability of detecting non-enforcement p, and (iii) reputation costs r.

Enforcement Costs

As we have argued above, the level of enforcement costs depends on a country's (administrative) ability to effectively enforce stricter environmental policy and the extent to which the government serves interests which prefer lax environmental regimes. In the Mexican case, the government's enforcement costs seem to have been relatively high. One reason is its weak institutional infrastructure and organizational capacity (Mumme and Lybecker 2002:326). Moreover, economic austerity and the credible threat of increased unemployment due to higher production costs and, in turn, less foreign direct investment, motivated officials to undertake a nonpunitive approach to ''ensuring'' compliance (Mumme 1998:192). Lax enforcement made it easy for many industrial plants to successfully avoid compliance with higher environmental policy standards (OECD 1998:126).

In Chile, the records suggest that enforcement costs were at least as high as in the Mexican case (Mall 1998:118). The main reason was that the Chilean industry—especially the important manufacturing sector which the government presumably deemed crucial to its popularity—demonstrated its reluctance to

[11] Despite the failure of the NAFTA negotiations, Chile subsequently concluded biltateral FTAs with Canada, Mexico, and eventually with the United States in 2003. The eventual acceptance of the agreement most likely resulted from the United States perceiving the gains from trade with Chile to have increased strongly (Weintraub 2004:84). We will turn in more detail to the effects of increasing gains from trade on the likelihood of a bilateral trade offer in section 5. Formally, however, Chile never entered NAFTA and does not participate in its institutional structures, such as the Commission for Environmental Cooperation.

[12] Of course, the EU accession of the Czech Republic was not only determined by the adoption and implementation of European environmental standards, but rather of the entire *acquis communautaire* as well as the fulfillment of the so-called Copenhagen criteria. These require that a state possess the institutions to preserve democratic governance and human rights, have a functioning market economy, and accept the obligations and intent of the EU.

[13] In a recent econometric study, Earnhart and Lizal (2008:240) conclude that ''overall, based on our assessment, the implementation of tighter air protection policies appears to be the most important reason for the dramatic reduction in air pollutant emissions during the Czech economic transition period.''

adhere to stricter regulation, as this would have necessitated significant investments in ''cleaner'' production technologies (Urzua and Alvarez-Arenas 2003:37). Also, enforcement relied on a very costly coordination of various sectoral enforcement bodies, resulting in what the OECD called ''not the most effective institutional arrangement to assure compliance'' (OECD 2005:115).

In contrast to the previous two cases, enforcement costs were comparatively low in the Czech Republic. First, an effective administrative infrastructure was already in place (Moldan 1997:120). Second, the government overcame the resistance of the domestically important energy sector by offering compensation: business interests were appeased by granting large industries the maintenance of an integrated structure and a monopoly position in the Czech market. Beyond that, the compensatory bargain included an enhanced nuclear production capacity as well as governmental support for raising international capital for investments (Andonova 2004:121). In sum, enforcement costs for the Czech government from enforcing stricter environmental standards were relatively low.

Detection Probability

The detection probability reflects the likelihood of finding out that the environmental policies stipulated in the agreement are not enforced by the low-regulating country. Here international, publicly observable monitoring efforts are crucial. During the NAFTA pre-accession debate, considerable attention was directed toward the enforcement activities of the Mexican government. It notably increased the number of inspections and other ''visible'' monitoring and enforcement efforts. However, these activities were limited to the time of the United States Congressional debate and ended once it had approved the trade negotiations. There were no provisions which contained rules for setting up and maintaining monitoring mechanisms after the conclusion of the agreement. Thus, the detection probability must be considered to have been low.

In a similar vein, the US General Accounting Office reviewed the state of Chilean environmental policy, but these were also limited in time, sporadic, and issue-specific, for example focusing exclusively on Chilean pesticide rules. Things worked, however, differently for the Czech Republic. The state of environmental policy enforcement has regularly been examined by the European Commission through annual monitoring reports. Consequently, it is safe to say that the detection probability in the case of the Czech Republic is comparatively higher than had been the case for Chile. As concerns Mexico, there was indeed a short period of high attention, which the government used to send the ''right'' signals, that is enhanced enforcement activities.

Reputation Costs

Reputation costs arise from being dishonest to the potential trading partner about the domestic enforcement of stricter regulatory standards. One should expect these to be higher when the trade agreement is associated with entry into an organization which offers the possibility to link different policy fields along with redistributive measures, as is the case in the EU. While for Mexico and Chile the agreement was solely about setting up trade relations with the United States, the accession of the Czech Republic to the EU was not limited to removing trade barriers. Rather, it created the possibility to punish the Czech government for non-enforcement by linking environmental issues and decisions about the common agricultural policy. Consequently, one should expect reputation costs to have been considerably higher for the Czech Republic.

Also, reputation costs should be higher the stronger a low-regulating country is *unilaterally* economically dependent on its high-regulating trade partner, as this

increses its vulnerability to economic sanctions. Also in this respect, the Czech Republic faced relatively high reputation costs, because the EU is indeed its most important partner. In 2001, the country's exports to and imports from the EU comprised 68.9% and 61.8%, respectively, of its total exports and imports (Andonova 2004:3). As concerns Chile, its export markets were at the time of NAFTA negotiations fairly balanced among Europe, Asia, Latin America, and North America. In 2001, the United States and Canada merely absorbed 18% of Chilean exports and were the origin of 23% of imports to Chile (Weintraub 2004:82). This made Chile less vulnerable to sanctions than the Czech Republic and reduced the reputation costs. In contrast to Chile, Mexico has just recently built up an encompassing network of various FTAs. In fact, the United States has been its most important trading partner. However, this commercial relationship is reciprocal; Mexico is, besides Canada as the third NAFTA member, the primary market for United States exports (Krueger 2000; Hufbauer and Goodrich 2004). As a consequence of the economic interdependence between both countries, economic sanctions imposed by one country will hurt both, which makes them unattractive. Since other policy fields which would offer the possibility to engage in issue-linkage do not exist (as opposed to the EU), one can consider reputation costs to be low. We summarize our reasoning in Table 1.

TABLE 1. Conditional Trade Agreements and Environmental Policy Convergence

	I	*II*	*III*
	Mexico-NAFTA	Chile-NAFTA 1993	Czech Republic-EU
Offer	Yes	Yes	Yes
Agreement	Yes	No	Yes
Deceptive or true convergence	Deceptive	None	True
Enforcement costs	↑	↑	↓
Detection probability	→	↓	↑
Reputation costs	↓	↓	↑

Empirically Testable Hypothesis

Despite its simplicity, the policy convergence & trade game generates a number of hypotheses which can be empirically evaluated. To facilitate subsequent research which attempts to falsify the model, in what follows we make the most important empirical implications explicit and offer possibilities to operationalize important concepts.

Conditional Trade Offers

To begin with, an intuitively plausible hypothesis to be derived from the model presented here posits that with increasing gains from trade, it becomes—ceteris paribus—more attractive for a high-regulating country to make a conditional trade offer which indeed leads to the conclusion of a FTA.

Hypothesis 1 (Successful Trade Offers): *Conditional on being in equilibrium region I, II, or III, the higher the gains from trade, the more likely a trade offer which results in a trade agreement (ceteris paribus).*

However, citizens in high-regulating countries also care about environmental issues. Consequently, *F*'s incumbent government will be punished electorally for trading with a low-regulating country which does not enforce stricter policies

domestically. Therefore, strong environmental interests make it less attractive to offer the negotiation of an FTA in the first place.

Hypothesis 2 (Environmental Issues and Trade Offers): *Conditional on being in equilibrium region I, II/IVb, or III/IVa, the more important the environmental issues in the high-regulating country, the less likely a conditional trade offer (ceteris paribus).*

As our analysis shows, the government of a high-regulating country might benefit from making a ''fake'' trade offer, that is *F* could refuse to conclude the agreement by pointing out that the low-regulating country will most likely not enforce stricter regulation. This would create unfair competition insofar as domestic goods in *F* would have to compete with those from *L* which are produced at lower costs due to lax environmental regulation. Since production in *F* becomes unattractive to firms, they will relocate their activities thereby increasing unemployment in *F*. This adds another dimension to the question when conditional trade offers are made strategically, which has up until now received no attention. The logic underlying such fake trade offers, which do not necessarily lead to the conclusion of an agreement, results in the following hypothesis.[14]

Hypothesis 3 (Fake Trade Offers): *Conditional on being in equilibrium regions III/IVa, the more citizens in the high-regulating country care about stricter international environmental regulation, the higher the probability of a fake trade offer (ceteris paribus).*

A second hypothesis about the occurrence of fake trade offers can be derived by combining our theory with standard partisan models of government. If parties deliver different policies to different electorates, those parties which serve environmental interests should be more prone to making fake trade offers.

Hypothesis 4 (Partisan Difference in Fake Trade Offer): *Conditional on being in equilibrium regions III/IVa, the more a government caters to environmental interests, the higher the probability of a fake trade offer (ceteris paribus).*

True Policy Convergence

The occurrence of *de facto* convergence of regulatory policies depends on several factors. The following hypotheses summarize their impact on the probability of true convergence.

Hypothesis 5 (Reputation Costs): *Conditional on being in equilibrium regions II/IVb, or I, a low-regulating country with high enforcement costs is more likely to enforce stricter regulation domestically if reputation costs increase (ceteris paribus).*

Several ways can be thought of to empirically evaluate this hypothesis. For example, reputation costs should be higher if the number of international organizations that both the low-regulating and the high-regulating country are members of increases, and with deeper cooperation in these organizations. This is because they provide the high-regulating country with arenas in which it can fulminate or even sanction against *L* in case non-enforcement is detected. The wider and the deeper cooperation is in the international organizations, the more

[14] We are aware of the difficulty of empirically identifying which of the unsuccessful trade offers can be considered to be fake in this sense. Nevertheless, as this is an implication of our model, we explicitly state the corresponding hypothesis here.

effectively different policy fields can be linked along with redistributive measures, for example as is the case in the EU.

Finally, the decision to enforce or not to enforce stricter regulation domestically also depends on its costliness.

Hypothesis 6 (Enforcement Costs): *Conditional on being in equilibrium regions II/III, the lower the enforcement costs, the more likely* de facto *policy convergence (ceteris paribus).*

Costs of enforcement are likely to increase, for example, if a low-regulating country has a strong polluting industry and if these interests are important to the government. This is the case if a large share of tax revenue is paid by firms belonging to that industry. Also, enforcing regulatory policies gets more costly the more they impose costs on firms. One might argue that these are a function of a corporation's political expenditures (besides other factors). Such expenditures can serve as signals to the bureaucracy which conveys information about the costs a firm can impose on the regulatory agency supposed to enforce a certain policy domestically, for example by contesting the agency's findings and/or appealing to legislative bodies (Gordon and Hafer 2007).

Hypothesis 7 (Gains from Trade): *Conditional on being in equilibrium region III, the higher the gains from trade to the low-regulating country, the less likely it is that stricter regulation will be enforced domestically (ceteris paribus).*

Conclusion

Environmental issues have become key issues in industrially advanced, high-regulating countries. As a consequence, governments in these countries have discovered conditional FTAs as a means of promoting higher policy standards internationally. Despite their growing significance in international relations, scholars have hitherto failed to more rigorously analyze under what conditions conditional FTAs are successful in achieving not only *declaratory* but also *actual* increases of environmental standards. We explore how potential gains from trade and domestic demands for environmental protection standards affect the occurrence of policy convergence and trade in the presence of uncertainty about regime enforcement. An important message of our analysis is that changing economic openness for *de facto* policy convergence is possible only under extreme, yet not completely unrealistic, conditions. In general, there is an incentive for countries to feign a true interest in policy convergence even if they will not enforce environmental policies. The situation is complicated by the fact that a high-regulating country can gain as well from the low-regulating country obfuscating domestic non-enforcement.

Several implications can be derived from the theoretical analysis presented here, which subsequent research could evaluate. Since the government of a high-regulating country can benefit from standing firm on international environmental policy standards domestically, the more citizens care about these issues, the more likely a *conditional* trade offer will be made. Also, high-regulating countries should be less likely to conclude a trade agreement the more concerned the domestic audience is about environmental issues.

More generally, our results carry implications for empirical research on cross-national policy convergence, as they highlight the risk of over-estimating convergence in empirical studies whose measures of actual regulatory standards are based on legislative output data. Furthermore, our formal analysis calls for caution against an assumption made in empirical studies, where unrevealed non-compliance is assumed to be randomly distributed across countries.

Rather, non-compliance should be viewed as a strategic choice which depends on a country's payoff structure and the behavior of its prospective trading partner. By that, the results also qualify what has been termed the ''trading-up'' hypothesis, where trade is assumed to lead to upward convergence toward more stringent environmental policies. For trade agreements between low- and high-regulating countries this relationship is very unlikely to hold in general, because for low-regulating countries it may be attractive to fake a true interest in policy convergence even if costs arising from policy enforcement are prohibitive.

Appendix

7 Proofs. [Sequential equilibrium] A sequential equilibrium in the policy convergence and trade game is an assessment (σ, μ) such that:

1. $\forall i \in I \forall h \in H \forall \sigma_i' : \pi_i((\sigma_i, \sigma_{-i}), \mu|h) \geq \pi_i((\sigma_i', \sigma_{-i}), \mu|h)$ with players i, information sets h and strategies σ_i
2. Let σ_{i_θ} denote type θ's strategy and suppose σ^n is a convergent sequence of strategy profiles with μ^n denoting the corresponding beliefs. Then, $\exists\{(\sigma^n, \mu^n)\}$:

(a) $\forall n \forall i \in I \forall d \in D_i : \sigma_i^n(d) > 0$ with action d

(b) $\forall n : s^n = \frac{t\sigma_F^n(C)\sigma_{L_{\text{low}}}^n(A)}{t\sigma_F^n(C)\sigma_{L_{\text{low}}}^n(A) + (1-t)\sigma_F^n(C)\sigma_{L_{\text{low}}}^n(A)}$

(c) $lim_{n \to \infty}(\sigma^n, \mu^n) = (\sigma, \mu)$.

To keep proofs as short as possible, we first eliminate dominated strategies as this narrows down the set of candidate equilibria.

Lemma 4 : *Let $\sigma_i(d)$ denote the probability attached to action d by player i. In any equilibrium,*

1. *If F makes a conditional trade offer the high cost type will accept this offer: $\sigma_F(T) = 1 \Rightarrow \sigma_{L_{\text{low}}}(A) = 1$*
2. *L_{high} never enforces compliance with the policy at the final stage, i.e. $\sigma_{L_{\text{high}}}(E) = 0$*
3. *L_{low} always enforces compliance at the final stage, i.e. $\sigma_{L_{\text{low}}}(E) = 1$*
4. *L_{low} always accepts the conditional trade offer, i.e. $\sigma_{L_{\text{low}}}(A) = 1$*

Proof. 1. Since $g_l - pr > a$, it follows that $U(A) > U(\neg A)$, and therefore, $\sigma_{L_{\text{high}}}(A) = 1$. 2. Since $g_L > \overline{c} > r$, $\forall p \in [0,1] : g_L - pr > g_L - \overline{c} \Leftrightarrow U(\neg E) > U(E)$. Thus, by sequential rationality, $\sigma_{L_{\text{low}}}(I) = 0$ in any equilibrium. 3. By assumption, $\frac{\underline{c}}{r} < p$. Thus, $\underline{c} < pr \Leftrightarrow g_L - \underline{c} > g_L - pr \Leftrightarrow U(E) > U(\neg E)$. Consequently, in any equilibrium, $\sigma_{L_{\text{low}}}(E) = 1$. 4. Let τ denote the probability attached to action T by F. Given that $g_L > \underline{c} > a > 0$, clearly $\forall \tau \in [0,1] : \tau(g_L - \underline{c}) > -a \Leftrightarrow U(A) > U(\neg A)$. Therefore, $\sigma_{L_{\text{low}}}(A) = 1$ in any equilibrium.

In the following lemma we derive the belief consistency condition from sequential equilibrium for our specific application. Thus, only Perfect Bayesian Equilibrium (PBE) which satisfy this consistency condition are sequential.

Lemma 5 (Consistent beliefs) : *Suppose (σ^*, μ) is a PBE. In any sequential equilibrium, (a) if $\sigma_{L_{\text{high}}}(A) = \sigma_{L_{\text{low}}}(A) \Rightarrow s = t$ and (b) if $\sigma_{L_{\text{high}}}(A) = \gamma \wedge \sigma_{L_{\text{low}}}(A) = 1 \Rightarrow s = \frac{t\gamma}{1+t(\gamma-1)}$.*

Proof. Let σ^n be a sequence of strategy profiles such that σ^n is completely mixed and $lim_{n\to\infty}\sigma^n = \sigma^*$. Along the sequence Bayes' rule yields $s^n = \frac{t\sigma^n_{L_{\text{high}}}(A)}{t\sigma^n_{L_{\text{high}}}(A)+(1-t)\sigma^n_{L_{\text{low}}}(A)}$. (a) Suppose $\sigma_{L_{\text{high}}}(A) = \sigma_{L_{\text{low}}}(A) \in \sigma^*$. Then, $lim_{n\to\infty} s^n = t$. (b) Suppose $\sigma_{L_{high}}(A) = \gamma \wedge \sigma_{L_{\text{low}}}(A) = 1 \in \sigma^*$. It follows that $lim_{n\to\infty} s^n = \frac{t\gamma}{1+t(\gamma-1)}$.

We now show that the strategies stated in the equilibrium propositions are optimal choices and that the beliefs satisfy the consistency conditions. This demonstrates that these equilibria are sequential.

Proof of proposition 1. Strategies: $\sigma_{L_{\text{high}}} = A, \neg E$ and $\sigma_{L_{\text{low}}} = A, E$ follow from lemma 4. For F it is sequentially rational to play C at the first stage if $U(C) \geq U(\neg C) \Leftrightarrow s(g_F - p(l+e) + (1-s)g_F \geq 0 \Leftrightarrow s \leq \frac{g_F}{p(l+e)} = s^*$. Thus, at the first stage, $\sigma_F = \begin{cases} C & \text{if } s \leq \frac{g_F}{p(l+e)} \\ \neg C & \text{otherwise} \end{cases}$. Analogously, at the third stage $U(T) \geq U(\neg T) \Leftrightarrow t(g_F - p(l+e) + (1-t)g_F \geq 0$, which yields $t \leq \frac{g_F}{p(l+e)} = t^*$. Therefore, at the third stage, $\sigma_F = \begin{cases} T & \text{if } t \leq \frac{g_F}{p(l+e)} \\ \neg T & \text{otherwise} \end{cases}$. Since $t \leq \frac{g_F}{p(l+e)}$, C is optimal. Beliefs: T is optimal since $s \leq \frac{g_F}{p(l+e)}$. Information set h_2 is on the equilibrium path and Bayes' rule directly applies. Since both types play A we get $s = t$, which satisfies the consistency condition (a).

Proof of proposition 2. Strategies: $\sigma_{L_{\text{low}}} = A, E$ and $\sigma_{L_{\text{high}}} = \neg E$ follow from lemma 4. For showing that L_{high} playing A with probability $\gamma = \frac{g_F(1-t)}{t(p(l+e)-g_F)}$ is optimal we use the belief consistency condition $s = \frac{t\gamma}{1+t(\gamma-t)}$. Substituting the expression for γ yields $s = \frac{t\frac{g_F(1-t)}{t(p(l+e)-g_F)}}{1+t(\frac{g_F(1-t)}{t(p(l+e)-g_F)}-1)} = \frac{\frac{g_F(1-t)}{p(l+e)-g_F}}{\frac{p(l+e)-g_F+g_F(1-t)-t(p(l+e)-g_F)}{p(l+e)-g_F}} = \frac{g_F}{p(l+e)}$. For F playing C at the first stage, we need $U(C) = t[\gamma(\tau(g_F - p(l+e)) + (1-\tau)0) + (1-\gamma)(l+e)] + (1-t)\tau g_F > U(C) = 0$, which reduces to $t < \frac{g_F}{p(l+e)} = t^*$. Thus, playing C is sequentially rational. F behaves optimally by playing T with probability τ such that this makes L_{high} indifferent. Since $U(A) = \tau(g_L - pv) + (1-\tau)0 = \frac{a}{g_L - pr}(g_L - pr) = a = U(\neg A)$, randomizing with $\tau = \frac{a}{g_L - pr}$ is indeed sequentially rational. Beliefs: $t \leq t^*$ and by construction $s = \frac{g_F}{p(l+e)}$ satisfies the consistency condition from lemma 5 part (b).

Proof of proposition 3. Strategies: $\sigma_{L_{low}} = (A, E)$ and L_{high} playing $\neg E$ again follows from lemma 4. The rest of the strategies are sequentially rational by proof of proposition 2. Beliefs: Again, the initial belief must be $t > t^*$ and by the consistency condition part (b) $s = \frac{g_F}{p(l+e)}$.

Varying the Probability of Detecting Non-Enforcement

For completeness we prove (i) that the detection probability threshold values $\overline{p}$ and $\underline{p}$ illustrate how the set of equilibria changes are unique and that (ii) $\overline{p} > \underline{p}$. Also, (iii) we prove that the critical values for $l + e$ in cases I, III, and IVa are unique and that F's optimal strategy only depends on these values. Note that in what follows only existence is explicitly shown, because uniqueness directly follows from the monotonicity of the utility functions. Since equilibria II and IVb correspond to the initial setup of the policy convergence and trade game, the proofs can be found above.

Proof of Threshold Values and Rationality of Strategies. (i) Let $c \in \{\overline{c}, \underline{c}\}$. Consider the decision of L whether to enforce the policy at the final stage for an arbitrary type $\theta \in \Theta$. Let $p > \frac{c}{r}$. Thus, $pr > c \Leftrightarrow g_L - pr < g_L - c \Leftrightarrow U(\neg E) < U(E)$. The corresponding best reply correspondences are $\sigma_{L_{low}} =$ if $p > \frac{\underline{c}}{r} = \underline{p} \neg E$ otherwise and $\sigma_{L_{\text{high}}} =$ if $p > \frac{\overline{c}}{r} = \overline{p} \neg E$ otherwise. (ii) By assumption $\overline{c} > \underline{c}$. Using the result from (a) we can write $\frac{\overline{c}}{r} > \frac{\underline{c}}{r} \Leftrightarrow \overline{p} > \underline{p}$. (iii) Consider equilibrium I. Since both types enforce compliance at the final stage, for F playing T is optimal, because $g_F > 0 \Leftrightarrow U(T) > U(\neg T)$. As $\forall t \in (0,1) : t(l+e) + (1-t)0 > 0 \Leftrightarrow U(C) > U(\neg C)$. Thus, F will play (C, T). Consider equilibrium III and IVa, in which no type implements the policy. Suppose $l + e < \frac{g_F}{p} \Leftrightarrow 0 < g_F - p(l+e) \Leftrightarrow U(\neg T) < U(T)$. Thus, F will play T if $l + e < \frac{g_F}{p}$. Now consider F's decision at the first stage. (Equilibrium III) Suppose F plays T at stage 3, then $\forall t(0,1) : t(l+e) + (1-t)0 > 0 \Leftrightarrow U(C) > U(\neg C)$. (Equilibrium IVa) Suppose F plays $\neg T$. This will induce L_{high} to play $\neg A$ at the second stage, because $U(\neg A) = 0 < a = U(A)$. Suppose $f < \frac{g_F}{p}$. Thus, $0 < g_F - pf \Leftrightarrow U(\neg C) < U(C)$, which shows that F will play C if $f < \frac{g_F}{p} = f^*$.

Remaining Equilibrium

There is one remaining continuum of sequential equilibria based on $t = t^*$ being satisfied, in which any randomization φ over making and not making a conditional trade offer is sequentially rational.

Proposition 6 (Status quo) : *If* $t = \frac{g_F}{p(l+e)}$, *a continuum of status quo preserving equilibria exists in which* $\sigma_F = (\varphi, \tau)$, $\sigma_{L_{\text{high}}} = (A, \neg E)$, $\sigma_{L_{\text{low}}} = (A, E)$, *and* $\mu = \{\frac{g_F}{p(l+e)}, t\}$ *with* $\phi \in (0, 1) \wedge \tau \in (\frac{a}{g_L - pr}, 1)$.

Proof. For optimality of σ_F see proof of proposition 1. Consistency of beliefs follows from lemma 5. Optimality of strategies for both types follows from lemma 4.

References

ANDONOVA, LILIANA B. (2004) *Transnational Politics of the Environment: The European Union and Environmental Policy in Central and Eastern Europe: EU Integration and Environmental Europe.* Cambridge: MIT Press.

ANTWEILER, WERNER, BRIAN COPELAND, AND SCOTT M. TAYLOR. (2001) Is Free Trade Good for the Environment? *American Economic Review* 91 (4): 877–908.

AUER, MATTHEW. (2001) Energy and Environmental Politics in Post-Corporatist Mexico. *Policy Studies Journal* 29 (3): 437–455.

BÄTTIG, MICHÉLE B., AND THOMAS BERNAUER. (2009) National Institutions and Global Public Goods: Are Democracies More Cooperative in Climate Change Policy? *International Organization* 63 (2): 281–308.

BHAGWATI, JADISH. (2002) *Free Trade Today.* Princeton: Princeton University Press.

DAI, XINYUAN. (2006) The Conditional Nature of Democratic Compliance. *Journal of Conflict Resolution* 50 (5): 690–713.

DESAI, UDAY. (1998) Environment, Economic Growth, and Government. In *Ecological Policy and Politics in Developing Countries*, edited by Uday Desai. Albany, NY: State University of New York Press, pp. 1–45.

DREZNER, DANIEL W. (2001) Globalization and Policy Convergence. *International Studies Review* 3 (1): 53–78.

DREZNER, DANIEL W. (2005) Globalization, Harmonization, and Competition: The Different Pathways to Policy Convergence. *Journal of European Public Policy* 12 (5): 841–859.

EARNHART, DIETRICH, AND LUBOMIR LIZAL. (2008) Pollution Reductions in the Czech Republic. *Post-Communist Economies* 20 (2): 231–252.

EUROPEAN COMMISSION. (2006) *XXIVth Report on Monitoring the Application of Community Law [COM(2007) 398].* Brussels: European Commission.

Fair, Ray C. (1996) Econometrics and Presidential Elections. *Journal of Economic Perspectives* 10 (3): 89–102.

Ferrantino, Michael. (1997). International Trade, Environmental Quality and Public Policy. *World Economy* 20 (1): 43–72.

Gordon, Sanford C., and Catherine Hafer. (2005) Flexing Muscle: Corporate Political Expenditures as Signals to the Bureaucracy. *American Political Science Review* 99 (2): 245–261.

Gordon, Sanford C., and Catherine Hafer. (2007) Corporate Influence and the Regulatory Mandate. *Journal of Politics* 69 (2): 300–319.

Grossman, Gene M., and Alan B. Krueger. (1995) Economic Growth and the Environment. *Quarterly Journal of Economics* 110 (2): 353–377.

Holzinger, Katharina, and Christoph Knill. (2004) Regulatory Competition and Regulatory Cooperation in Environmental Policy: Individual and Interaction Effects. *Journal of Public Policy* 24 (1): 25–47.

Holzinger, Katharina, Christoph Knill, and Thomas Sommerer. (2008) Environmental Policy Convergence: The Impact of International Harmonization, Transnational Communication, and Regulatory Competition. *International Organization* 62 (4): 553–587.

Hufbauer, Gary, and Ben Goodrich. (2004) Lessons from NAFTA. In *Free Trade Agreements: US Strategies and Priorities*, edited by Jeffrey J. Schott. Washington, DC: Institute for International Economics, pp. 37–50.

Hultberg, Patrik T., and Edward B. Barbiery. (2004) Cross-Country Policy Harmonization with Rent-Seeking. *Contributions to Economic Analysis and Policy* 3 (2): 1–20.

Inglehart, Ronald. (1997) *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton: Princeton University Press.

Knill, Christoph, Jale Tosun, and Stephan Heichel. (2008) Balancing Competitiveness and Conditionality: Environmental Policy-Making in Low-Regulating Countries. *Journal of European Public Policy* 15 (7): 1019–1040.

Kono, Daniel Y. (2007) When Do Trade Blocs Block Trade? *International Studies Quarterly* 51 (1): 165–181.

Kreps, David M., and Robert Wilson. (1982) Sequential Equilibrium. *Econometrica* 50 (4): 863–894.

Krueger, Anne O. (2000) NAFTA's Effects: A Preliminary Assessment. *World Economy* 23 (6): 761–775.

Lewis-Beck, Michael S. (1988) Comparative Economic Voting: Britain, France, Germany, Italy. *American Journal of Political Science* 30 (2): 315–346.

Mall, Bradley. (1998) The Effect of NAFTA's Environmental Provisions on Mexican and Chilean Policy. *International Lawyer* 32 (1): 153–187.

Mani, Muthukumara, and David Wheeler. (1998) In Search of Pollution Havens? Dirty Industry in the World Economy: 1960–1995. *Journal of Environment and Development* 7 (3): 215–247.

Mayda, Anna Maria, and Dani Rodrik. (2005) Why Are Some People (and Countries) More Protectionist than Others? *European Economic Review* 49 (6): 1393–1430.

Meyer, Reto, Thomas Bernauer, and Michael M. Bechtel. (2009) Green Protectionism? Assessing the Micro-level Implications of the Trade-Environment Nexus. Working paper. Available at https://www.ib.ethz.ch/people/mbechtel/papers/GreenProtect.pdf. (Accessed March 8, 2009.)

Moldan, Bedrich. (1997) Czech Republic. In *The Environmental Challenge for Central European Economies in Transition*, edited by Jürg Klarer and Bedrich Moldan. Chichester: John Wiley & Sons, pp. 107–130.

Mumme, Stephen P. (1998) Environmental Policy and Politics in Mexico. In *Ecological Policy and Politics in Developing Countries*, edited by Uday Desai. Albany, NY: State University of New York Press, pp. 183–203.

Mumme, Stephen P., and Donna Lybecker. (2002) Environmental Capacity in Mexico: An Assessment. In *Capacity Building in National Environmental Policy*, edited by Hartmut Weidner and Martin Jänicke. Berlin: Springer, pp. 311–328.

OECD. (1998) *Environmental Performance Review of Mexico*. Paris: OECD.

OECD. (2005) *Environmental Performance Review of Chile*. Paris: OECD.

Pavlinek, Petr, and John Pickles. (2000) *Environmental Transitions: Transformation and Ecological Defence in Central and Eastern Europe*. London: Routledge.

Porter, Michael E., Klaus Schwab, Xavier Sala-I-Martin, and Augusto Lopez-Claros, Eds. (2005) *The Global Competitiveness Report 2004-2005*. New York: Palgrave Macmillan.

Prakash, Aseem, and Matthew Potoski. (2006) Racing to the Bottom? Trade, Environmental Governance, and ISO 14001. *American Journal of Political Science* 50 (2): 350–364.

Rosen, Howard. (2004) Free Trade Agreements as Foreign Policy Tools: The US-Israel and US-Jordan FTAs. In *Free Trade Agreements: US Strategies and Priorities*, edited by Jeffrey J. Schott. Washington, DC: Institute for International Economics, pp. 51–78.

Rueda, Andres. (2000) Tuna, Dolphins, Shrimp & Turtle: What About Environmental Embargoes under NAFTA? *Georgetown International Environmental Law Review* 12 (3): 647–692.

Scheve, Kenneth F., and Matthew J. Slaughter. (2001) What Determines Individual Trade-Policy Preferences? *Journal of International Economics* 54 (2): 267–292.

Schott, Jeffrey J. (2004) Assessing US FTA Policy. In *Free Trade Agreements: US Strategies and Priorities*, edited by Jeffrey J. Schrott. Washington, DC: Institute for International Economics, pp. 359–381.

Stern, David I. (2004) The Rise and Fall of the Environmental Kuznets Curve. *World Development* 32 (8): 1419–1439.

Tiebout, Charles. (1956) A Pure Theory of Local Expenditure. *Journal of Political Economy* 64 (5): 416–424.

Urzua, Osvaldo, and Juana Alvarez-Arenas. (2003) Chile. In *Towards Sustainable Development in Industry? Reports from Seven Developing and Transition Economies*, edited by Ralph A. Luken, and Paul Hesp. Cheltenham: Edward Elgar, pp. 19–42.

Vailland, Marcel, and Alvaro Ons. (2002) Preferential Trading Agreements Between the European Union and South America: The Political Economy of Free Trade Zones in Practice. *World Economy* 25 (10): 1433–1468.

Valds, Alberto. (1995) Joining an Existing Regional Trade Agreement from the Perspective of a Small Open Economy: Chile's Accession to NAFTA and MERCOSUR. *American Journal of Agricultural Economics* 77 (5): 1292–1297.

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

Ward, Hugh. (2008) Liberal Democracy and Sustainability. *Environmental Politics* 17 (3): 386–409.

Weintraub, Sidney. (2004) Lessons From the Chile and Singapore Free Trade Agreements. In *Free Trade Agreements: US Strategies and Priorities*, edited by Jeffrey J. Schott. Washington, DC: Institute for International Economics, pp. 79–92.