

# Multilateralism, Bilateralism and Regime Design<sup>1</sup>

ALEXANDER THOMPSON AND DANIEL VERDIER  
Ohio State University

## **Abstract**

Different international regimes are built from legal instruments that vary in terms of whether they are multilateral, bilateral or a combination thereof. We investigate the reasons for such variation. The choice between multilateralism and bilateralism is a function of the trade-off between each instrument's relative flaw. Multilateralism is wasteful in incentives, as the same agreement is offered to all states regardless of their compliance costs. Bilateralism mitigates this problem by allowing for more tailored agreements but in the process multiplies transaction costs by requiring many of them. We use a formal model to generate propositions on the design of "lateralism" and the related issue of membership size, and offer illustrations in the context of four regimes: foreign direct investment, human rights, climate change, and international trade.

International regimes are built on instruments that vary in terms of whether they are multi-lateral, bilateral or a combination thereof. For example, the current trade regime has a strong multilateral component as its core, centered on the World Trade Organization (WTO), while the foreign direct investment (FDI) regime is primarily constituted of bilateral agreements. Notably, many regimes combine two or more instruments: the nuclear nonproliferation regime combines a multilateral treaty with bilateral security guarantees proffered mostly by the United States (Verdier 2008).

The question of "lateralism" relates to whether countries should be treated equally under international law or differently. Multilateralism, according to Ruggie, implements "generalized principles of conduct" whereas bilateralism "differentiates relations case-by-case based principally on a priori particularistic grounds or situational exigencies" (1992: 571). Extending this conceptual distinction, we treat as multilateral any agreement based on general obligations that apply across many states, and treat as bilateral any agreement based on obligations that apply only to particular states. How this issue is addressed in practice is usually seen as a result of prevailing norms or of bargaining power, with some states advocating more uniform rights and obligations and others calling for differential treatment.

In an attempt to understand these institutional outcomes, we point to two key factors: transaction costs and a new concept, the member surplus. The member surplus captures the idea that the multilateral strategy can be wasteful in incentives, since incentives are calculated to elicit the participation of the state that is burdened with the highest cost of compliance, thereby creating rents for the other members. The bilateral strategy, in contrast, allows the customization of rights and obligations to each individual member state. But because the bilateral strategy is more wasteful in transaction costs than the multilateral strategy, states face a trade-off.

This allows us to make three central claims: (1) multilateralism is most attractive with high transaction costs and a low member surplus, (2) bilateralism is most attractive when transaction costs are low and the member surplus is high, and (3) we encounter combinations of multilateralism and bilateralism when both transaction costs and the member surplus are high. Such combinations come in two forms, regimes that contain a mix of multilateral and bilateral agreements and regimes that rely on multilateral agreements that are customized to the needs of different members. A fourth claim is that regimes with both high transaction costs and a high member surplus exhibit

the highest rate of exclusion because of the cost of attracting members. We develop a formal model to capture the logic behind these institutional outcomes.

The next section draws insights from the existing literature. We then introduce the notions of transaction costs and member surplus and present the model and a set of predictions that flow from it. In an empirical section, we demonstrate the plausibility of our claims in the context of four prominent regimes: foreign direct investment, human rights, climate change and trade. This is followed by a consideration of alternative explanations and a concluding section that raises questions for further research.

## 1 Lateralisms in the Literature

We adopt the standard definition of international regimes, which grounds them in a set of norms or goals that guide state behavior in a given issue area (Krasner 1983; Keohane 1984). Most regimes have at their origins a market imperfection, making a mutually efficient outcome unreachable through standard competitive behavior. Our approach also complements work on regimes that views them as created by a hegemonic state (Krasner 1976), a small group of great powers (Snidal 1985), the market-oriented Anglo-Saxon subset (Cowhey and Klimenko 2000), or the North at large (Sell 2007). We assume that most regimes are organized by a small group of founders who try to enlist the cooperation of a larger group of regular members. The founders take the lead on account of their larger resources, greater interest in the regime, or principled concern over the issue.

International relations and international law scholars tend to emphasize the virtues and pre-dominance of multilateralism, based on its perceived legitimacy and its ability to treat states more uniformly (Ruggie 1992; Finnemore 2003; Blum 2008). Others, especially Realists, see a shift toward bilateralism and unilateralism in recent years, driven by the end of the Cold War and the unrivaled power of the United States.<sup>2</sup> Rather than emphasize one or the other, we look at variation in multilateralism versus bilateralism and address the question from a strategic and efficiency perspective. We build on the insight of Conybeare (1980) and Oye (1992) that decentralized bargaining offers a viable substitute for multilateral agreements in some situations. Like Kahler (2009) and Milner (2006), we also take seriously the possibility that these outcomes are not mutually exclusive but can coexist at a given point in time and even within a single regime.

Some authors do try to account for the choice of lateralism at a given time. Yarbrough and

Yarbrough (1992) argue that states prefer multilateralism in trade when there is a high risk of opportunistic behavior but must settle for bilateralism if no powerful state is willing to supply third-party enforcement. Studying the double taxation regime, Rixen (2010) argues that a multilateral organization minimizes transaction costs (see also Ruggie 1992), but argues that distributive issues are better accommodated through bilateral bargaining, provided that such bargaining produces no externalities on third parties, in which case multilateralism is more appropriate. Starting from the perspective of the federalism literature, which views decentralization as a solution to preference heterogeneity, Harstad (2007) points to a bargaining drawback inherent in this approach: countries strategically delay the revelation of their preferences, with the effect of raising transaction costs. A potential solution to this problem, he argues, is a multilateral regime that treats its members uniformly, because uniform treatment makes information revelation irrelevant.

Our paper also touches on the issue of membership—which states to include, which to exclude. Koremenos, Lipson and Snidal (2000) argue that a large membership makes cooperation easier when the issues at stake are distributive. Linked to this problem is the apparent trade-off between the depth and breadth of cooperation (Downs, Rocke, and Barsoom 1998). The central problem is that high levels of participation make enforcement difficult unless the agreement is shallow enough to appeal to the most reluctant states, who would otherwise free-ride. However, Barrett (2003) argues that full participation can be achieved if there is a collective desire to lower compliance costs by adding new members combined with the use of side payments to achieve a fairer distribution of costs. A similar logic has been applied to trade, where some argue that discriminatory bilateral agreements are a better way to promote participation when countries have asymmetric endowments (Aghion, Antràs, and Helpman 2007; Saggi and Yildiz 2010). The breadth-versus-depth dilemma also evaporates if one suspends the requirement of uniform obligations among members in a multilateral setting (Gilligan 2004).

From this vibrant debate, we retain and further develop two ideas. First, multilateralism is a solution to high transaction costs, including the costs of negotiating and enforcing agreements. Second, bilateralism may be a solution to several flaws incurred under multilateralism: free riding or exclusion in the case of a public good and inefficient uniformity in the face of preference heterogeneity, asymmetric endowments, or distributive concerns. The present theory and model tackle these ideas in a way that strives to be intuitive and mathematically tractable.

## 2 Transaction Costs and the Member Surplus

One of the key concepts that generate our results, transaction costs, is commonly used in the institutionalist literature and requires little introduction. The second concept, member surplus, is new and in need of lengthier explanation. The trade-off that these two factors produce and the various solutions that are conceivable provide the basis for our theoretical claims.

Our definition of transaction costs is borrowed from Williamson (1985: 20-22). Transaction costs are the *ex ante* costs of “drafting, negotiating and safeguarding an agreement,” with safeguarding broadly defined to include measures that render an agreement credible and self-regulating. They also include ongoing costs associated with renegotiation and with maintaining governance structures for monitoring and dispute settlement. While some of these costs are only incurred after an agreement is reached, the prospect of facing ongoing problems complicates and prolongs matters during the negotiation phase as well (Fearon 1998).

The multilateral strategy has the advantage of saving on transaction costs (Keohane 1984; Ruggie 1992; Rixen 2010). This is true because only one set of negotiations is required, and because most multilateral treaties incorporate forums that facilitate further decision-making and provide economies of scale in monitoring and dispute resolution. The multilateral strategy saves on transaction costs also because coordinated negotiations allow more bargaining possibilities to be on the table at once, promoting issue-linkage and chances that every offer finds a corresponding match. The bilateral strategy, in contrast, multiplies transaction costs because a new contract must be negotiated, safeguarded and implemented for each participant, and some participants may struggle to find an attractive counterpart. Last, by treating members identically, the multilateral strategy helps avoid the posturing and costly delays that beset bilateral negotiations in the presence of information asymmetry (Harstad 2007).

However, the multilateral strategy will be expensive for the founder in another sense—this is where the member surplus comes in. Multilateralism in its pure form offers only one deal and this deal is the same for everyone (there are exceptions to this rule, which we address later). As a result, participants are offered an incentive that is calculated to elicit the participation of the state that is burdened with the highest cost of compliance. The problem is similar to the one that occurs in competitive markets, where the law of one price for a particular good confers a surplus on all

producers who would have been willing to sell for less. This surplus is known in economics as the "producer surplus."<sup>3</sup> In direct analogy, we call it the "member surplus." Unlike the multilateral approach, the bilateral approach is immune to the member surplus because it gives to each state the incentive it needs to participate and no more. Bilateralism corresponds in market economics to an extreme case of market fragmentation, where a monopolist offers its respective reservation value to each consumer.

To see this, imagine a situation with no transaction costs. The founder would always prefer the bilateral approach to the multilateral approach. Consider Figure 1, featuring, on the horizontal axis, an ordering of all members according to their compliance cost  $z_i$  and, on the vertical axis, the cost to the founder  $t(z_i)$  of attracting participation in the regime. The slope of the curve is positive, since the higher a member's compliance cost, the higher its participation price. The slope is also convex because in most cases a majority of countries more or less share a moderate cost of compliance, except for a minority that face very high costs.

[Figure 1. The Member Surplus]

Assume that the founder wants to include country  $N$  and all those to its left. By construction, the  $N^{th}$  country is the most expensive member to be included in the regime. The cost of wooing countries with compliance cost less than or equal to  $N$  by means of a multilateral contract is equal to the rectangle, because the multilateral approach forces the founder to pay the transfer it pays to the  $N^{th}$  member to all other members. In contrast, by means of bilateral contracts alone, this same cost would be no more than the area situated below the curve, less than half the size of the rectangle. The area above the curve is the member surplus that results from the multilateral approach.

Now we introduce transaction costs. We posit that the transaction costs for the multilateral approach are lower than for the bilateral approach. This means that one negotiation among 100 countries is significantly less costly than the sum of the  $\binom{100}{2} = 4950$  dyadic negotiations it would take for all of the 100 countries to negotiate bilaterally with one another. This claim does not conflict with the typical assumption that absolute negotiation costs increase with the number of countries (Koremenos 2005). They do, but not as quickly as the number of dyads, with the result that in a multilateral setting the cost-per-dyad actually decreases with the number of countries.

We capture this feature formally by assuming that any single deal, bilateral or multilateral, costs constant  $T$  to process and maintain, with  $T$  positive. This simplification has no consequences for the substantive implications of our results.

There are now two possibilities. One possibility is that the  $N$  bilateral contracts could be costlier than one multilateral contract for everyone. This case is represented in Figure 2A, where the cost of the multilateral contract is the same as before (equal to the rectangle) plus the transaction cost incurred once ( $z_N t(z_N) + T$ ), while the cost of the bilateral contracts is the area situated below the lower curve augmented by the transaction cost incurred  $N$  times ( $\sum_{i=1}^N (t(z_i) + T)$ ), adding up to the upper curve. The other possibility is that the bilateral approach is the less costly of the two. As  $N$  increases (see Figure 2B), it is easy to verify that so does the relative size of the area delimited by the three points  $b$ ,  $f$ , and  $t(z_N)$  relative to the area delimited by the points  $f$ ,  $d$ , and  $e$ . In other words, the added transaction costs that arise from bilateralism are much less than the member surplus generated by multilateralism.

[Figure 2. Tradeoff between Member Surplus and Transaction Costs in Pure Regimes]

We have assumed so far that the choice between multilateralism and bilateralism was of the either-or type. We now investigate the possibility of combining instruments, that is, offering both multilateralism and bilateralism to various members. Between the high- and low-cost compliance members, it is the high-cost ones who receive the bilateral treatment. The reason is that convexity divides the domain into two sections, one where the curve is rather flat on the left and another where it is steeper on the right. The flat slope on the left means that the member surplus is less likely to be of concern than the transaction costs, and thus that multilateralism is the more appropriate instrument in that region. In contrast, the steeper slope on the right points to a serious potential cost in member surplus, a problem to which bilateralism is the more efficient response.

There is one assumption we have made but which can now be discarded, the idea that the founder seeks to include all potential members in the regime. In reality, a third possibility is open to the founder: exclusion. The founder need not provide an incentive to all countries but may exclude as many as it wants. Exclusion offers the advantage of reducing the founder's cost of building the regime and would typically affect states with the highest marginal costs of compliance—precisely those which would be candidates for bilateral deals.

The final template, of a regime combining multilateralism, bilateralism and exclusion, is shown in Figure 3. On the left-hand side of the graph, in the  $[z_1, z_x]$  segment, the incentive that has to be given to a member is sufficiently low that it makes sense to seek compliance by offering a single generic contract, minimizing transaction costs yet providing a surplus to all the members to the left of  $x$ . In the middle part of the graph, in the  $[z_{x+1}, z_y]$  interval, the incentive is too high in relation to the fixed transaction cost for overlooking the surplus. Rather than offering a more generous multilateral contract, the founder merely supplements the existing multilateral contract with bilateral ones, an approach that economizes on the member surplus (since the bilateral contracts are customized to each member of the interval) yet is wasteful in transaction costs. Finally, on the right-hand side of the graph, in the  $[z_{y+1}, z_N]$  interval, the incentive is so high that the founder is better off excluding potential members. To put it succinctly, the regime should obey the following generic condition  $0 \leq x \leq y \leq N$ , with  $x$  the member that makes the founder indifferent between offering and not offering bilateral incentives in addition to multilateral incentives, and  $y$  the member that makes the founder indifferent between including and excluding that member.

[Figure 3. A Regime Combining Instruments]

This discussion has several implications for the depth-versus-breadth trade-off. First, the pure multilateral component of a regime, by which we mean the set of uniform obligations across members, is likely to be shallow. Depth, which is only relevant in relation to members with high compliance costs, is more efficiently achieved through the adjunction of bilateral instruments than by asking for a general increase in effort level. Second, depth calls for exclusion and thus narrowness, because the logic that leads to the adjunction of bilateral instruments is the same as that behind exclusion: both features are called for to alleviate the costs of a large member surplus. Breadth is only possible if the slope of the curve in Figure 3 is close to being flat.

The notion of member surplus is a comprehensive and potentially powerful tool to explain lateralism and exclusion. Yet it is still a much-too synthetic concept, the product of a number of parameters (four in the simplified model below), which one should ascertain with greater analytical precision. To measure and operationalize member surplus as a variable, we need to determine the *functional form of the slope* of the curve drawn in Figure 3. To that effect, we build a founder-member model.

### 3 A Founder-Member Model of Instrument Choice

We posit a simple bargaining structure inspired from Aghion, Antràs, and Helpman (2007), featuring a leading country—the founder—deciding whether to negotiate with the rest of the world—the potential members. While in Aghion et al. the leader chooses between negotiating sequentially or simultaneously with other countries, here the founder chooses between negotiating bilaterally or multilaterally.<sup>4</sup> The founder has agenda-setting power, making take-it-or-leave-it offers. The founder has the ability to compensate other countries for the abandoning of suboptimal policies, while being a residual claimant on the surplus from the global provision of the public good.

The members are indexed  $i \in [1, N]$  with  $N > 1$  a positive integer. Each member  $i$  maximizes her individual utility  $u_i(s_i, s_j) = f(s_i, s_j) - c(s_i)$ , with  $s_i \geq 0$  a continuous choice variable. Function  $f$ , which is assumed to be strictly concave, makes each member  $i$ 's utility a function of both her choice  $s_i$  and the choice of all other members  $s_j$ , with  $j$  standing for all members other than  $i$ . For the sake of calculating point predictions, we opt for the following specific functional forms, combining a decreasing marginal return with a constant cost:  $f(s_i, s_j) = a_i\sqrt{s_i} - \rho \sum_{j \neq i} s_j$  and  $c(s_i) = cs_i$ , with  $c$  the marginal cost. Variable  $s_i$  may be thought of as an investment in a polluting technology or any activity producing an externality, for instance fishing in the high seas, setting protective tariffs, or curbing human rights. In each case, an investment  $s_i$  generates a negative externality  $\rho s_i$  inflicted upon every other member. Parameter  $\rho$  is the externality index; it is greater than or equal to zero, with a value of zero indicating no externalities and a value greater than zero indicating their presence. Variable  $a_i$  scales member  $i$ 's marginal gain for engaging in the activity based on the activity. (Note that the model would equally work and yield the same comparative statics if we had opted for a different mix of marginals, i.e., constant gains and marginally increasing costs, or a different sign on the externality.)

In the absence of a founder, each member maximizes  $u_i(s_i, s_j)$  with respect to her choice variable  $s_i$  such that  $s_i, s_j \geq 0$ . This version of the game yields a competitive equilibrium in which every member produces  $s_i^\# = \left(\frac{a_i}{2c}\right)^2$ . This quantity is greater than the individual production level that would maximize the social optimum,  $\sum_{i=1}^N u_i$ , which is equal to  $s_i^\circ = \left(\frac{a_i}{2(c+\rho(N-1))}\right)^2$  (see the supplemental appendix for a demonstration of both results). As one would expect, the presence of a negative externality yields a competitive equilibrium that is inefficient because members overinvest

in the activity that causes the negative externality.

Enter the founder, intent on designing a regime that would reduce excessive investment. We use the social optimum to operationalize the founder's optimum, yet it is important to note that the model and results can accommodate any notion of optimum as long as it is socially more desirable than the competitive equilibrium. The founder achieves this result by offering an incentive to each member. For the sake of simplicity, we assume the incentive to be a positive transfer—a payment— $t(s_i)$ . We posit the following functional form for transfers:  $t(s_i) = ts_i$ , with  $ts_i$  the transfer given to member  $i$  and  $t$  a positive variable standing for the subsidy rate.<sup>5</sup>

As shown in Figure 3, the transfer can be given in several ways: through a multilateral instrument in which members are treated uniformly (they are given the same transfer); through a series of bilateral agreements by which the founder is able to customize transfers to each member's need; or through a combination of multilateral and bilateral instruments, where a subset of members is treated identically and another is treated based on individual need. We model the distribution of instruments according to the template of Figure 3. As noted above, we assume that any single instrument, bilateral or multilateral, costs constant  $T$  to process.

On the founder's side, we assume that the founder values at constant  $V$  any member  $i$ 's investment that conforms with the founder's notion of what is optimal. For the sake of convenience, we also assume that  $V$  is sufficiently large for the founder not to run into a budget constraint.

The founder moves first, offering a contract to all members simultaneously. Then the members simultaneously decide to reject or accept the offer. No subset of members has the capacity to organize a coordinated response to the founder's offer. If a member rejects, there is no contract with that member. If a member accepts, the contract is executed as written; we are not giving the founder the capacity to make the signing of a contract with one (or more) member(s) contingent on the acceptance of all contracts by all other members.<sup>6</sup> There is no room for shirking once the member has accepted the founder's offer—signing an international instrument makes the commitment credible for both sides. Credibility is the result of a costly signaling game or a reputation game that is not modelled here.

A strategy for the founder specifies the  $(t, x, y)$  regime she proposes, that is, choosing cutpoints  $x$  and  $y \in \mathbb{Z}^+$  and subsidy rate  $t \in \mathbb{R}^+$  that maximize her aggregate utility while simultaneously offering transfer levels sufficiently high to induce the  $y$  members to invest optimally. A strategy

for any member  $i$  is a mapping specifying for every combination of institution and subsidy rate an investment level in the activity causing the externality that maximizes her individual utility. The solution concept is the subgame perfect Nash equilibrium. Formally, it means for the founder and the members to simultaneously solve the program

$$P = \left\{ \begin{array}{l} 1. \max_{1 \leq x \leq y \leq N, t \geq 0} U_P = xg(x) - T + \sum_{z=x+1}^y (g(z) - T), \\ 2. \text{ with } g(i) = \delta V - t(s_i^\# - s_i), \text{ and } \begin{cases} \delta = 1 & \text{if } s_i = s_i^\circ \\ \delta = 0 & \text{if } s_i \neq s_i^\circ \end{cases}, \\ 3. \max_{s_i \geq 0} u_i = a_i \sqrt{s_i} - \rho \sum_{j \neq i} s_j - cs_i + t(s_i^\# - s_i), \text{ for all } i, j \in [1, N], \\ 4. \text{ with } s_i^\# = \left(\frac{a_i}{2c}\right)^2 \text{ for all } i \in [1, N], \text{ and } s_i^\circ = \left(\frac{a_i}{2(c+\rho(N-1))}\right)^2, \\ \text{subject to:} \\ 5. a_i \sqrt{s_i} - \rho \sum_{j \neq i} s_j - cs_i + t(s_i^\# - s_i) \geq a_i \sqrt{s_i^\#} - \rho \sum_{j \neq i} s_j - cs_i^\#, \\ \text{for all } i, j \in [1, N]. \end{array} \right.$$

The first clause formalizes the founder's maximization problem, choosing subsidy rate  $t$  and cutpoints  $x$  and  $y$  so as to offer a single multilateral treaty to members 1 to  $x$ , and bilateral contracts to members  $x + 1$  to  $y$ .<sup>7</sup> Clause 2 specifies the founder's utility function, earning positive constant  $V$  for every member who cuts activity down to the level required to implement the social optimum, at the cost of transfer  $t(s_i^\# - s_i)$  given to each member.

Clause 3 features the maximization problem for included member  $i$ , who now receives transfer  $t(s_i^\# - s_i)$ , calculated to give her an incentive to reduce activity below the competitive equilibrium,  $s_i^\#$ , whose value is reported in clause 4, along with that for the socially optimum value  $s_i^\circ$ .

Clause 5 specifies the incentive constraint for each included member, insuring that none of them has an interest in unilaterally deviating from the founder-induced optimum equilibrium.

The program is solved in the supplemental appendix. The equilibrium value of the subsidy rate,  $t^*$ , is equal to  $(N - 1)\rho$ , which can be interpreted as the externality rate, since each member causes  $N - 1$  externalities, each time with marginal impact  $\rho$ .

Assuming, to arrive at an explicit solution, the following functional form for marginal gains,  $a_i = ia$ , with  $i \in [1, N]$  and  $a > 0$ , we are ready to state the solution.

**Proposition 1** *There exists a subgame perfect equilibrium in which the founder offers*

(1) transfer  $t^* s_x^\circ$  to members  $i$  indexed 1 to  $x^*$  and in which these members invest social optimum  $s_i^\circ$ ;

(2) transfer  $t^* s_i^\circ$  to members  $i$  indexed  $x^* + 1$  to  $y^*$  and in which these members invest social optimum  $s_i^\circ$ ;

(3) no transfer to members  $i$  indexed  $y^* + 1$  to  $N$  and in which these members invest competitive equilibrium  $s_i^\#$ ;

$$\text{with } t^* = (N-1)\rho, s_i^\circ = \left(\frac{a_i}{2(c+\rho(N-1))}\right)^2, s_i^\# = \left(\frac{a_i}{2c}\right)^2, x^* \in [\underline{x}, \underline{x} + 1], y^* \in [\underline{y}, \underline{y} + 1], \underline{x} = \frac{1}{4} \sqrt{\frac{(a^2 \rho^3 (N-1)^3 + 32Tc^2(c+\rho(N-1))^2 + 2ca^2 \rho^2 (N-1)^2)}{a\rho(N-1)\sqrt{2c+\rho(N-1)}}} - \frac{1}{4}, \underline{y} = 2\frac{c}{a\rho} \frac{\sqrt{V-T}}{\sqrt{2c+\rho(N-1)}} \frac{c+\rho(N-1)}{N-1} - 1.$$

#### 4 Asymmetric Information

The model assumes complete information. What if, instead, the founder were ignorant of each member's compliance cost and dependent on their declarations? This could be the situation in environmental and trade agreements, for example, where governments are uncertain of each other's domestic constraints. Addressing the lack of progress in the Doha round of trade negotiations, one study (Narlikar and van Houten 2010: 151) notes "the real preference orderings of all the parties remain poorly understood." In the case of climate negotiations, it is sometimes difficult to accurately discern domestic political opposition to ratification or implementation (Hovi et al. forthcoming). We conjecture that in such conditions it is in the interest of each member to claim a cost greater than his actual one and thereby extract a higher payment. Asymmetric information operates like a transaction cost, disqualifying customization in favor of uniformity.

To see this, imagine that the founder has a good sense of the overall distribution of marginal-gain scalar  $a_i$ . We further assume that the founder does not know where any given member  $i$  is located on that distribution. Hence, even if the founder has a good sense of the nature and scope of the instruments that she should use to build the regime, she does not know which instrument to offer to which member. A rational strategy for any member under such circumstances is to claim to be the high marginal gain type  $a_y$ , with  $y$  being the presumed highest-marginal-gain member to be included in the regime. As a result, all agents with an actual marginal gain located on and to the left of cutpoint  $y$  would claim to be at that very cutpoint.

The founder's best response to such misrepresentation is to give up on bilateralism and offer a multilateral deal calculated to include the member with marginal gain  $a_y$ . In response to such an

offer, all agents with marginal gains inferior or equal to  $a_y$  accept the regime and all those with marginal gains above that threshold stay out. The multilateral instrument is optimal here because it functions like a partial information revelation mechanism. It leads each member to truthfully sort themselves out into a camp of members and a camp of nonmembers. The revelation is partial, however, because nothing is revealed on how agents are distributed within each camp. But note that such information is unnecessary in the context of the multilateral instrument, which treats everyone the same way and thus can be used successfully in the absence of any information on members' individual characteristics.

The situation is very different for bilateral instruments. Individualized information is needed in order to tailor the bilateral incentive to each member's type. In the case where all agents claim to be the  $a_y$  type, the bilateral strategy is suboptimal: it does not enable the founder to save on the member surplus, since all agents are given the transfer that corresponds to the  $a_y$  claim, thus multiplying transaction costs for no offsetting benefits. Bilateralism fails in the presence of asymmetric information. Our static model reproduces Harstad's (2007) dynamic result, according to which information asymmetry leads to posturing and delaying tactics during negotiations, thereby raising transaction costs.

## 5 Comparative Statics and Predictions

The model produces three sets of comparative statics that are relevant to the mix of lateralisms and exclusion:

(1) Transaction costs (information asymmetry and  $T$ ): We have modelled two sources of transaction costs, information asymmetry and the negotiating and governance costs ( $T$ ). We saw that information asymmetry raises the cost of the bilateral approach, making a multilateral treaty a more efficient method. Similarly, a rise in  $T$  increases the scope for multilateralism ( $x^*$  increases) and causes a rise in exclusion ( $y^*$  drops). In contrast, if  $T$  is low, holding everything else constant, the bilateral approach is dominant (the "Coase Theorem").

(2) Member surplus ( $a, c, \rho, N$ ): One may approximate the member surplus by the slope of the curve of Figure 3. We found, after assuming  $a_i = ia$ , with  $i \in [1, N]$ , that the curve in the model is equal to

$$t(i) = \rho(N-1) \frac{a^2}{4} \left( \frac{1}{c^2} - \frac{1}{(c + \rho(N-1))^2} \right) i^2.$$

This definition of the member surplus is rich in information. First, the slope of the curve (and thus the member surplus) is a function of the marginals: a positive function of the marginal gain scalar  $a$  and a negative function of marginal cost  $c$ . Note that  $a$  offers a measure of the distribution of actual investment gains across member states. Since a high marginal gain of investment in an externality-generating activity necessarily makes for a high compliance cost with a regime regulating that activity,  $a$  also offers a measure of the *variation in compliance costs*—whether compliance costs are uniformly or unevenly distributed across states.

Second, the slope of the curve also is a positive function of the size of the externality rate, as measured by the expression  $\rho(N-1)$ .

Third, two terms of the member surplus, the compliance-cost-variation index  $a$  and the externality rate  $\rho(N-1)$  are multiplicative. If either one of them is close to zero, it cancels the effects of the other components; the whole expression for the slope and the member surplus drops to zero.

From this definition, we can derive predictions relative to the design of the regime. Considering the marginal components, first, it is easy to check that an increase in compliance-cost-variation index  $a$  yields more exclusion (it reduces  $y^*$ ). Likewise it can be shown that an increase in  $a$  yields a decrease in multilateralism (it reduces  $x^*$ ). Parameter  $c$ , the marginal investment cost, works in the exact opposite direction. Taken together, the couple  $a$  and  $c$  determines the conditions for when a broad and shallow (and therefore multilateral) agreement is feasible: a compliance cost that is evenly distributed (low  $a$ ) or an investment cost that is large (high  $c$ ).

Considering now the second component of the member surplus, parameters  $\rho$  and  $N$ , which together determine the aggregate size of the externality, it is easy to check that both vary negatively with multilateralism ( $\partial x^*/\partial \rho < 0$  and  $\partial x^*/\partial N < 0$ ) and positively with exclusion ( $\partial y^*/\partial \rho < 0$  and  $\partial y^*/\partial N < 0$ ).

These comparative statics yield the following four predictions:

1. If transaction costs alone are high, multilateralism should be the instrument of choice.
2. If the member surplus alone is high, bilateralism should be the instrument of choice.

3. If both transaction costs and the member surplus are high, some combination of lateralisms should be the outcome.
4. If both transaction costs and the member surplus are high, exclusion is greatest. If only one of these values is high, exclusion occurs but at more moderate levels.

The case where neither transaction costs nor member surplus are high is likely to yield indeterminacy, as it leaves the outcome open to the influence of parameters not included in our model.

[Figure 4. Theoretical Predictions]

We illustrate the logic of our theoretical argument by focusing on four prominent cases: a case of bilateralism, the FDI regime; a case of multilateralism, the human rights regime; and two cases of combined bilateralism and multilateralism, the climate change regime and the trade regime. In addition to offering variation on the dependent variables, these cases demonstrate the generalizability of the model by covering a broad range of issue areas. We include two cases of combined lateralisms because we believe these are more common in practice than the pure cases, and including two gives us more opportunity to explore the different ways in which these institutional forms can be combined. Specifically, we note that the trade regime relies mostly on *mixing* a multilateral agreement with many bilateral ones, while the climate regime relies on the *customization* of obligations within the context of multilateral instruments.

We use the predictions generated by the model and presented in the previous section to guide the empirical discussion. We begin by establishing the values of the key independent variables for each of the regimes and then assess whether our model's predictions are accurately reflected in the design of each regime. A penultimate section assesses the plausibility of alternative explanations.

## 6 Independent Variables

To operationalize our two key independent variables, transaction costs and member surplus, we look at the properties of the policy issues in question for a given regime. Transaction costs are notoriously difficult to measure and the institutionalist literature does not offer straightforward guidance. We offer an approach that builds explicitly from our theoretical model and is general enough to apply across regimes. First, transaction costs are high when negotiations must reconcile

the interests of many actors and when there is uncertainty regarding the preferences of those involved in negotiations (Weber 1997: 333; Koremenos, Lipson and Snidal 2001: 782). Because all of our cases involve a potentially large number of states (as they involve global regimes), we focus especially on the domestic level and ask whether the international negotiations in question implicate a large number of domestic interests with heterogeneous preferences. This is consistent with Moravcsik (1999: 301), who argues that bargaining and informational obstacles are most likely to come from the domestic level. Second, transaction costs are high when safeguarding the agreement is difficult. This is the case when there is an incentive to cheat after an agreement is signed and when it is difficult to observe the behavior of other actors; both factors require the establishment of ongoing monitoring and enforcement mechanisms and thus raise the costs of cooperation (Koremenos 2007; Lake 1999).

To capture the member surplus, building from our discussion of the comparative statics, we first consider whether compliance costs are uniformly or unevenly distributed across states. An uneven distribution creates a steeper cost curve for the founder and therefore a more rapidly increasing surplus as more members are added. We expect uneven compliance costs when states have varied interests or capacity levels, both of which affect the desire and ability of states to contribute to the good in question. Second, we look at the size of the international externalities generated by the activity, which magnify the potential surplus by incentivizing the founder to increase the number and size of transfers. While there is some potential for a member surplus in almost any regime, the multiplicative nature of the two components imply that the member surplus is relatively small when compliance costs are uniform or when the externality is low.

Based on these criteria, we propose that transaction costs are relatively high for three of the regimes under consideration, climate, trade, and human rights, but are lower in the case of FDI; and that the member surplus is high in climate, trade, and FDI but low in human rights.

***Transaction Costs.*** Beginning with transaction costs, we see that climate change involves high bargaining costs and high safeguarding costs. Climate policy implicates a diverse range of domestic sectors and domestic interests have consistently posed obstacles during negotiations (Sprinz and Weiss 2001). This is further complicated because governments are often uncertain of the domestic constraints in other countries (see, e.g., Hovi et al. 2012). Negotiations are also hampered by the

scientific complexity of climate change, the pervasive uncertainty behind its causes and effects, and the distributive implications of choosing different approaches to addressing the issue (Thompson 2010; Depledge 2005). States have incentives to free ride and, in the face of private information, tactical advantages to promise little (Grundig, Ward and Zorick 2001). Beyond the bargaining phase, few international regimes pose a more severe monitoring and enforcement problem than climate change, where emissions must be tracked around the world.

Trade cooperation also entails high transaction costs. Domestic interest groups are highly affected and involved and governments have incentives to adopt aggressive negotiating positions as a way to appeal to them, causing delay and deadlock (Zahrnt 2007). A further problem with trade agreements is that they are by necessity multi-faceted, involving various sectors and sub-issues (investment, intellectual property, labor, human rights, the environment, etc.). As one study of Japanese trade agreements notes, the involvement of many actors results in "very high negotiation or transaction costs" (Pekkanen, Solis and Katada 2007: 959). Safeguarding costs are also high in trade cooperation, traditionally modeled as a prisoners' dilemma, a problem compounded by the opacity of various nontariff barriers to trade that are difficult to monitor.

In the area of human rights, we argue that bargaining costs are high because the tools traditionally used to forge agreements—linkage and side-payments—are impractical. Human rights negotiations revolve around matters of principle and legitimacy and thus are "lumpy" or indivisible—an action or policy is either right or wrong, making agreement difficult to reach (Goddard 2006; Guzman and Simmons 2002). Moreover, because they involve "taboo trade-offs" (Fiske and Tetlock 1997) and depend on their normative weight to matter, they do not lend themselves to side-payments, the most common tactic for overcoming indivisibility. Transaction costs also arise from the difficulty of monitoring and enforcing human rights law. The relevant behavior is occurring at the domestic level in other states and is rarely reported. It is also difficult for individual governments to punish rights violators through the normal bilateral channel of tit-for-tat retaliation since there are no direct reciprocal benefits to be withdrawn in human rights (Hathaway 2007: 589). The result is a regime that is both difficult to monitor and lacking in mechanisms of self-regulation.

Forging agreements to govern FDI, in contrast, involves much lower transaction costs. Such agreements are focused on a narrow set of investment issues that implicate relatively few domestic interest groups. Bargaining costs are further reduced because most FDI cooperation deals with a

familiar set of substantive issues, which means that virtually all investment treaties follow the same general structure (Salacuse 2009; Haftel and Thompson forthcoming). Rather than negotiating from the ground up, most investor countries offer a "model" structure as a starting point for negotiations. This default uniformity minimizes negotiating costs and, over time, makes implementation and interpretation more straightforward (UNCTAD 1998: 24). Monitoring costs are further reduced because, in most cases, violations are noticed immediately by the affected firm, whose presence on the ground supplies a system of inexpensive monitoring.

**Member Surplus.** Turning to the member surplus variable, recall that it is a positive function of the compliance cost variation index ( $a$ ) and the size of the externality ( $\rho$ ). There is little difference among the four regimes with respect to the first component—the variation in compliance costs is high across all four—but there is a significant difference when it comes to the size of the externality, which is large in climate, trade, and FDI, but small in human rights. Given the multiplicative nature of the two subcomponents, human rights agreements, in contrast to the other issue areas, generate a low member surplus.

In climate, the cost of reducing greenhouse gas emissions varies widely across countries, making it difficult to attract participation from those who place a high value on emitting without generating a surplus for those more willing to mitigate. Developing countries, in particular, have been reluctant from the start to join the regime, fearing that efforts to tackle climate change would compromise their prospects for economic growth. Even among industrialized countries, there is substantial variation: major emitters like the United States, Canada, Japan and Australia have higher emissions per capita, rely heavily on coal for energy, and have steeper emissions projections than most European countries (Cooper et al. 1999).

We see a similar pattern in trade. Some countries benefit more from free trade than others, depending on their size, competitiveness, and on how much they value social protection in relation to liberalization (Alesina, Spolaore and Wacziarg 2005). The same is true of FDI, where some countries are much better positioned to attract and protect investment as a function of political institutions and the broader investment environment (Stasavage 2002; Asiedu 2002). As for human rights, there is no doubt that the cost of protecting these rights is much higher in repressive and nondemocratic countries (such as Burma) than in western democracies (such as Canada) that

maintain high standards under the status quo.

The four regimes differ mainly with respect to the second component of the member surplus, the size of the externality. Climate, trade, and FDI all involve direct and tangible externalities that set them apart from human rights. In climate, the key externality is greenhouse gas emissions that transcend borders and produce a market failure of global proportions. Trade policies directly impact foreign producers and may divert trade flows from one country to another (Oye 1992). Foreign investment policies and institutions shape the risk faced by all international investors and can both generate and divert FDI (Chantasawat et al. 2010), in some cases leading to competitive pressures among potential host countries (Elkins, Guzman and Simmons 2006). These are all classic cases of policy externalities that require international cooperation to accommodate.

Human rights policies do have some cross-border effects, but in comparison to economic, environmental, and security issues, human rights agreements deal mostly with internal activities that have only modest externalities. As Louis Henkin (1979: 232) notes, "Since a violation by a state of the rights of its inhabitants does not ordinarily infringe the national interests of other parties to the agreement, they have no compelling interest to scrutinize the violating behavior and call it to account." The western democracies that founded the human rights regime have a low utility for improving human rights practices in other countries and generally have been unwilling to pay a high price to induce other states to join or to comply with deeper commitments (Moravcsik 2000: 217; Donnelly 1986: 616). The low externalities and limited expectations of the regime founders produce a markedly smaller member surplus for the human rights regime.

Table 1 lists the four regimes, the values of the independent variables that we assign to them, and our corresponding theoretical predictions in terms of lateralism and exclusion.

[Table 1. Values of Independent Variables and Predictions]

## 7 Institutional Outcomes

**FDI.** Since the 19th century there has existed an international investment regime consisting of a set of "widely shared standards regarding the proper treatment of foreign capital" (Lipson 1985: 81). Nevertheless, despite efforts by the United States in particular to promote rules on seizure and compensation, there exists no multilateral treaty or comprehensive customary law on the treatment of foreign direct investment (Guzman 1998). Instead, over the last fifty years rules governing FDI

have rested overwhelmingly on more than 2,700 bilateral investment treaties (BITs). Consistent with predictions 1 and 2, relatively low transaction costs and a large member surplus combine to make bilateralism the design of choice. And because it is easy to conclude new BITs, there is only a moderate incentive to exclude states from participating in the regime (prediction 4).

A multilateral instrument to deal with FDI is potentially attractive insofar as it would set uniform rules and thereby dampen race-to-the-bottom effects that come with competition among host governments. However, the Group of 77 rejected the universal application of the "Hull Rule" (requiring "prompt, adequate and effective compensation") and subsequent efforts to negotiate a Multilateral Agreement on Investment. A multilateral approach would either generate a substantial member surplus or would have to be too watered down in an effort to include hosts with the highest compliance costs. The WTO's agreement on Trade-Related Investment Measures (TRIMs) illustrates precisely this point; it addresses only a narrow range of FDI issues and imposes only "rudimentary disciplines" (Neumayer and Spess 2005: 1571).

The alternative is bilateral treaties that can be both deeper and more customized, and therefore less wasteful in incentives. Despite their common structure, as Salacuse (2009: 126) notes, "the specific provisions of individual investment treaties are not uniform and some investment treaties restrict host country governmental action more than others." BIT provisions are tailored to the political and economic needs of signatories (in particular, of the developing-country parties) in terms of what is counted as an "investment," the standards of treatment and protection that are applied, and the nature of dispute settlement (Blake forthcoming; Allee and Peinhardt 2010). The possible downside of bilateralism, transaction costs, is mitigated by the low costs of concluding BITs.<sup>8</sup>

Our theoretical model offers no reason to expect high rates of exclusion from the regime. Indeed, BITs have proliferated rapidly along with the globalization of capital, and today 176 countries have concluded one or more BITs. Only a handful of countries, mostly least-developed and politically unstable—and therefore unattractive as destinations of FDI—lie outside the regime.

***Human Rights.*** The international human rights regime is governed by a broad and growing set of multilateral instruments, centered around the basic norm that individual human beings should be protected regardless of their nationality or location (Morsink 1999). Three core agreements, the

Universal Declaration of Human Rights (1948), the International Covenant on Civil and Political Rights (1966), and the International Covenant on Economic, Social and Cultural Rights (1966), are complemented by more specific treaties on genocide (1948), racial discrimination (1973), discrimination against women (1979), torture (1984), rights of the child (1989), rights of migrant workers (1990), rights of persons with disabilities (2006), and enforced disappearance (2006).

To accommodate the high bargaining costs that come with human rights, states have employed a multilateral approach to negotiations, often in the context of the UN. The General Assembly was used as a venue for the final negotiations of the Universal Declaration, facilitating agreement across most of the international community. The UN's Commission on Human Rights was used to generate draft texts for many human rights treaties, often with the administrative support of secretariat officials, who have proven crucial in reconciling diverse views and supplying focal points for negotiations. Given the difficulties, noted above, of using bilateral mechanisms to safeguard the regime, the primary mechanisms for promoting compliance in human rights are reputation concerns and shaming, both of which are more effective in multilateral contexts (Guzman 2008: 64; Johnston 2001). Accordingly, in addition to the Human Rights Council, the various treaties establish independent committees that provide centralized monitoring and reporting. Their findings, though not binding, do have legal and normative significance that "puts pressure on states" (Buergenthal 2006: 791). None of this would be possible with a series of bilateral treaties.

While the logic of transaction costs helps explain why the regime is primarily multilateral, there is still the possibility that multilateral instruments would be complemented by bilateralism. And, to some extent, they are, as rich countries sometimes tie the choice of foreign aid recipients and trading partners to their human rights practices (Gomez 2007; Hafner-Burton 2009). This may reflect a sense among governments that human rights practices generate greater externalities than they used to, especially to the extent that they generate more media attention and affect political stability. Still, the extent of these exceptions should not be overstated. In human rights, low externalities result in a set of broad yet shallow multilateral treaties combined with modest bilateral efforts to expand participation.

Consistent with prediction 4, the relative ease of adding members to the regime results in high participation rates: except for two recent treaties, on migrant workers and enforced disappearance, all have been ratified by more than 100 countries.

*Climate Change.* The climate case illustrates that a combination of multilateralism and bilateralism is a logical result of situations where both transaction costs and the member surplus are high (prediction 3). The main multilateral component of the regime is its two core treaties, the 1992 UN Framework Convention on Climate Change and the 1997 Kyoto Protocol, and their associated institutions. These treaties are overseen by the Climate Secretariat, whose functions are to provide information, to facilitate negotiations, and to promote implementation—all geared toward reducing transaction costs (Depledge 2005). The Secretariat also collects "national communications" from all parties, which supply information on needs and best practices and, in the case of industrialized countries, on compliance. The latter are further monitored by "expert review teams" and a Compliance Committee.

While this multilateral approach helps address transaction costs, it does not accommodate the member surplus problem. To attract participation of the states with the highest compliance costs, the Framework Convention and Kyoto embrace the principle of "common but differentiated responsibility" to exempt developing countries from assuming any costly obligations. Both treaties also require richer countries to transfer technology and financial resources to developing countries to subsidize the cost of emissions abatement. Even among industrialized countries, Kyoto customizes obligations across members to bring those with high compliance costs on board. Two mechanisms were used to achieve customization. First, targets for emissions reductions vary substantially across countries, roughly according to their ability to pay (Frankel 2010: 58-9). Second, bilateral enticements to Japan, Canada, and Russia allowed them to count additional sources of carbon sinks toward their Kyoto targets.<sup>9</sup> These allowances, which substantially reduced the compliance costs for their beneficiaries, represented a major concession by the EU—one that succeeded in prompting all three of these hold-outs to ratify the treaty.

Finally, because both transaction costs and the member surplus are high in the climate case, we expect to see relatively high levels of exclusion from the regime (prediction 4). If we conceive of participation in terms of the acceptance of emissions reduction commitments, this is clearly the case: only 38 countries have assumed binding targets. Indeed, the two largest emitters of greenhouse gases, China (because it is a developing country) and the United States (because it never ratified Kyoto), remain excluded. More generally, ratifications for Kyoto were slow in coming and the treaty did not enter into force until 2005, eight years after it was signed, reflecting the

difficulty of bringing participants on board.

*Trade.* We focus on the postwar era since this is the period during which a coherent and widespread trade regime can be said to exist. The trade regime has always involved a combination of multilateral and bilateral instruments and thus also represents a “combination” case, reflecting the importance of both transaction costs and the member surplus (prediction 3).

Historically, some of the transaction costs of bilateralism have been offset through the application of the most-favored nation (MFN) principle as a sort of multilateralizing feature of bilateral agreements. By automatically extending lower-tariff treatment to other trading partners who also have MFN status (though not to others), MFN made countries more willing to negotiate concessions bilaterally because they knew the benefits would not be impaired by subsequent, more generous concessions to other countries. However, as the volume of trade increased over time, and as the risk of opportunism increased with greater trade-related investments in the modern era (Yarbrough and Yarbrough 1992), MFN alone proved inadequate to extend and safeguard trade agreements.

The GATT enshrined the MFN principle in Article I and evolved a set of multilateral institutions that further reduced transaction costs in at least three ways. First, cooperating through one large forum reduces the number of discussions that must take place. It also provides an established and predictable venue for negotiations and offers opportunities for side-payments and issue linkage to grease the skids of agreement. Second, the GATT/WTO secretariat plays a central role in monitoring and transparency, especially through the Trade Policy Review Mechanism, which promotes compliance through reputation concerns and peer pressure (Qureshi 1995). Finally, centralized dispute settlement, which began under the GATT and has been enhanced with the WTO, has deservedly received much attention as a key feature of the multilateral trading system. All of these centralized functions provide economies of scale when it comes to reducing the costs of trade cooperation.

While multilateralism helps address some of the inefficiencies that arise from bilateralism, it comes with its own downside. Specifically, it does not effectively address the problem of variable compliance costs that lies behind the member surplus problem. As in the climate case, this is done partly through customization in the form of the GATT/WTO’s “special and differential treatment” for developing countries. It is also done through the use of separate agreements of a

bilateral and regional nature, which have proliferated in recent years. Many larger trading countries use bilateral agreements as a way to entice their less-developed counterparts to undertake free trade commitments. For instance, the United States offers Bolivia, Colombia, Ecuador and Peru duty-free access to its market on almost all goods, going beyond the U.S. Generalized System of Preferences program (USTR 2007). The EU goes even further, using bilateral trade agreements as a major policy vehicle for distributing aid and technical assistance to transition and developing economies (European Commission 2009).

Finally, prediction 4 suggests that we should see high levels of exclusion in the trade case. This is only partly true. The regime certainly began with a very limited membership—only 23 countries signed the original GATT agreement. Other than a jump in the 1960s that reflected decolonization, the growth of membership was steady but relatively slow for forty years. The rate of joining then increased dramatically during the Uruguay Round (1986-1994) to reach more than three-quarters of all states. Nevertheless, some states are still excluded from the regime for reasons consistent with our logic. These nonparticipants are mostly least-developed countries (18 WTO nonmembers fall into this category), failed states (Sudan, Iraq and Afghanistan), and statist economies (North Korea, Serbia, and Iran). The high cost of attracting these countries, even through separate bilateral agreements, leads to exclusion.

## **8 Alternative Explanations**

These empirical illustrations demonstrate that our theoretical model predicts important features of prominent international regimes across a range of issue-areas. So far, however, we have not discussed alternative arguments in a systematic way. Here we consider four alternative explanations, each associated with a distinct theoretical tradition, that could potentially explain some or all of the outcomes described above. While our variables do not trump all others in all cases, our model has at least as much explanatory power as the alternatives.

The first alternative is norm-based. Constructivists scholars argue that international affairs are increasingly multilateral in nature, a trend since World War II that flows from the perceived legitimacy of multilateralism (Ruggie 1992; Finnemore 2003). International law scholars have similarly argued that states feel obligated to pursue their international relations multilaterally (Dupuy 2000). Although undeniable, this normative bias toward multilateralism does not account for the

popularity of bilateral agreements—the vast majority of treaties are bilateral—and the variation we observe in practice. As we outlined above, the FDI regime moved away from multilateral approaches to embrace bilateralism quite explicitly. The trade regime also includes substantial bilateral components alongside its core multilateral institutions, with no signs that bilateralism is fading into history. Multilateralism is normatively appealing but far from dominant.

Another plausible alternative, from within the rationalist tradition, is that institutional choices are a function of the type of good involved. The logic that widespread externalities and especially public goods problems require multilateralism is especially compelling (Hawkins et al. 2006; Koremenos, Lipson and Snidal 2001: 765). Indeed, it helps explain the outcome we see in climate change (Barrett 2003). However, one finds widespread participation in the trade and human rights regimes, which do not involve public goods. Moreover, most issues are difficult to classify in a straightforward way either because there are multiple types of goods involved or because these types are endogenous to how the issue is politically organized (Sandler 2004; Stone, Slantchev and London 2008). The climate case shows that a presumed public good at the global level can be made both divisible and excludable through the imposition of emissions caps and a system for carbon trading. For these reasons, we view our variables as more fundamental and able to explain a wider range of cases.

A third possible critique of our argument, consistent with historical institutionalism, is that it takes arbitrary "snapshots" of regimes that, in reality, are constantly evolving and partly the product of past institutional choices (Pierson 2004). We concede that the logic of path dependence helps explain some of our outcomes. For example, the prior choice to create the UN provided a ready-made forum for tackling the emerging issue of human rights in the years following World War II, setting the stage for multilateralism. In addition, the choice to go multilateral for a given regime could have a ratcheting effect, whereby it is difficult to return to a purely bilateral approach even if doing so is efficient according to our logic. However, we noted in our empirical discussion that some of our independent variables did change over time, bringing along the expected regime change. In FDI, for instance, the ascendancy of bilateralism as the preferred format coincided with a rise in the member surplus in the 1960s and 1970s, as some developing countries began demanding more control over their economies. We also noted that the standardization of BITs was one factor that reduced transaction costs over time, a good example of how prior institutional choices can

affect future outcomes. In other words, we view our approach as potentially consistent with a more historical or evolutionary approach that looks at institutions over time.<sup>10</sup>

Finally, one might expect that powerful states drive the choice of lateralism based on what maximizes their bargaining leverage in the context of a given regime. There is some evidence of this in our cases. Guzman (1998) argues that BITs have the effect of maximizing the leverage of rich senders of FDI over their lower-income counterparts, and Sell (2007) argues that rich countries have used "Trips-Plus" provisions in bilateral trade agreements to foist on developing countries intellectual property standards that go beyond WTO rules. However, the power-based approach implies that powerful states should always seek to divide and conquer their weaker cooperation partners rather than allow them to negotiate as a group. They do not and our model helps explain the choice for multilateralism by pointing to the downside of bilateralism and thus the trade-off facing even powerful states. It is certainly the case that the credible threat of bilateralism often skews multilateral outcomes in directions that advantage powerful member states (Gruber 2000; Odell 2000: 165-6), implying a more complicated and dynamic relationship between bilateralism and multilateralism.

In the end, all of these approaches have something useful to say about bilateralism and multilateralism and consideration of these additional factors gives us a richer view of institutional outcomes. Still, none of these alternative explanations calls into question the explanatory power of our fairly elegant model and none makes accurate predictions across all four regimes under consideration.

## 9 Conclusion

In this paper we shed further light on a source of variation in the design of international regimes that has attracted the attention of lawyers, economists, and political scientists. Regimes vary widely with respect to whether the legal instruments that support them are bilateral or multilateral, and to how inclusive they are. Along with a growing literature, we argue that multilateral agreements are not the only way to design regimes; bilateral agreements, as in the case of FDI, are viable alternatives. Moreover, the different instruments are not incompatible, as is too often believed, but complementary, as the trade and climate regimes illustrate.

We offer an explanation for the variation in lateralisms, arguing that transaction costs favor multilateralism, that high variation in compliance costs and high rates of externality (with the

potential of generating a "member surplus") favor bilateralism, and that combinations of lateralisms are likely when both transaction costs and the member surplus are high. With regard to membership size, we argue that the concurrence of transaction costs and member surplus is most likely to produce exclusion from a regime. Turning to empirical illustrations, we attributed the bilateral nature of the FDI regime to low transaction costs, the multilateral nature of the human rights regime to the existence of limited negative externalities (and thus a low member surplus), and the mixed nature of the trade and climate change regime to the coexistence of high transaction costs and surplus.

Our discussion of the trade and climate regimes raises new questions for future research. One concerns the status of bilateralism when combined with multilateralism. Although both the trade and climate regimes harbor a substantial bilateral component, they do so differently—outside the multilateral framework in the case of trade, embedded within the multilateral treaty in the case of climate change. The two possibilities can be thought of as *mixed* lateralisms and *customized* multilateralism, respectively. This raises the question of what approach is better for the regime founder. A multilateral treaty that treats everyone identically, supplemented with bilaterals to accommodate members with particular needs, may shelter the founder from having to justify the nature and extent of each side-payment since this can be done separately and more privately.<sup>11</sup> In contrast, a single instrument institutionalizing differential treatment has a better chance of addressing the economic inequality that exists among nations and may be more appealing to developing countries. A multilateral setting allows them to negotiate as a group, where they stand a better chance of resisting pressure from their richer counterparts, as they have done during the current, Doha Round of multilateral trade negotiations.

Another question concerns the status of non-participants. The climate regime includes members that have no ambition to contribute to the provision of the regime good. In contrast, in the trade regime, as in most other regimes, de facto non-contributors are non-participants. This raises the question of what advantage to the founders there is to including non-contributors as signatories rather than leaving them out of the regime. On the one hand, we suppose that leaving them out makes them liable to discrimination and thus offers a potential incentive to join. This is a more obvious strategy for excludable goods, as in the case of preferential trade agreements, but may also work for public goods if exclusion is endogenized. Two examples are access to emissions trading in climate and access to the institutions associated with the Law of the Sea, such as the International

Seabed Authority and the International Tribunal for the Law of the Sea. On the other hand, even if discrimination is possible, bringing non-contributors into the fold may subject them to collective coercion through majoritarian voting procedures or community pressure, steering them toward more productive participation.

## References

Alesina, Alberto, Enrico Spolaore and Romain Wacziarg. (2005) Trade, Growth and the Size of Countries. In *Handbook of Economic Growth, Vol. 1B.*, edited by Philippe Aghion and Steven N. Durlauf. Amsterdam: Elsevier.

Aghion, Philippe, Pol Andràs, and Elhanan Helpman. (2007) Negotiating Free Trade. *Journal of International Economics* 73: 1-30.

Allee, Todd and Clint Peinhardt. (2010) Delegating Differences: Bilateral Investment Treaties and Bargaining Over Dispute Settlement Provisions. *International Studies Quarterly* 54: 1-26.

Asiedu, Elizabeth. (2002) On the Determinants of Foreign Direct Investment to Developing Countries: Is Africa Different? *World Development* 30: 107-19.

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

Blum, Gabriella. (2008) Bilateralism, Multilateralism, and the Architecture of International Law. *Harvard International Law Journal* 49: 323-79.

Buergenthal, Thomas. (2006) The Evolving Human Rights System. *American Journal of International Law* 100: 783-807.

Chantasawat, Busakorn, K.C. Fung, Hitomi Iizaka and Alan Siu. (2010) FDI Flows to Latin America, East and Southeast Asia, and China: Substitutes or Complements? *Review of Development Economics* 14: 533-46.

Conybeare, John A.C. (1980) International Organization and the Theory of Property Rights." *International Organization* 34(3): 307-34.

Cooper, Adrian, Scott Livermore, Vanessa Rossi, Alan Wilson and John Walker. (1999) The Economic Impact of Reducing Carbon Emissions. *The Energy Journal* 20(Special Issue): 335-65.

Cowhey, Peter, and Mikhail M. Klimenko. (2000) Telecommunications Reform in Developing Countries after the WTO Agreement on Basic Telecommunications Services. *Journal of International Development* 12: 265-281.

Depledge, Joanna. (2005) *The Organization of Global Negotiations: Constructing the Climate Change Regime*. London: Earthscan.

Donnelly, Jack. (1986) International Human Rights: A Regime Analysis. *International Orga-*

nization 40: 599-642.

Downs, George W., David M. Rocke and Peter Barsoom. (1998) Managing the Evolution of Multilateralism. *International Organization* 52: 397-419.

Dupuy, Pierre-Marie. (2000) The Place and Role of Unilateralism in Contemporary International Law. *European Journal of International Law* 11: 19-29.

Elkins, Zachary, Andrew T. Guzman, and Beth A. Simmons. (2006) Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960-2000. *International Organization* 60: 811-846.

European Commission. (2009) EU Provides EUR 1 Billion for Trade Facilitation in Developing Countries. Directorate General for Trade, Press Release, November 12.

Fearon, James. (1998) Bargaining, Enforcement, and International Cooperation. *International Organization* 52: 269-306.

Finnemore, Martha. (2003) *The Purpose of Intervention*. Ithaca, NY: Cornell University Press.

Fiske, Alan and Philip E. Tetlock. (1997) Taboo Trade-Offs: Reactions to Transactions That Transgress the Spheres of Justice. *Political Psychology* 18: 255-97.

Frankel, Jeffrey. (2010) An Elaborated Proposal for a Global Climate Policy Architecture: Specific Formulas and Emission Targets for All Countries in All Decades. In *Post-Kyoto International Climate Policy: Implementing Architectures for Agreement*, Joseph Aldy and Robert Stavins. New York: Cambridge University Press.

Gilligan, Michael J. (2004) Is There is Broader-Deeper Trade-off in International Multilateral Agreements? *International Organization* 58: 459-484.

Goddard, Stacie E. (2006) Uncommon Ground: Indivisible Territory and the Politics of Legitimacy. *International Organization* 61: 35-68.

Gomez, Simon Peter. 2007. Human Rights and the Allocation of Foreign Aid: A Cross-National Analysis of the Last Years of the Cold War, 1980-1989. *The Social Science Journal* 44(2): 275-85.

Greif, Avner and David D. Laitin. (2004) A Theory of Endogenous Institutional Change. *American Political Science Review* 98: 633-52.

Gruber, Llyod. (2000) *Ruling the World: Power Politics and the Rise of Supranational Institutions*. Princeton, NJ: Princeton University Press.

Grundig, Frank, Hugh Ward and Ethan B. Zorick. (2001) Modeling Global Climate Negotiations. In *International Relations and Global Climate Change*, edited by Urs Luterbacher and Detlef

Sprinz. Cambridge, MA: MIT Press.

Guzman, Andrew T. (1998) Why LDCs Sign Treaties That Hurt Them: Explaining the Popularity of Bilateral Investment Treaties. *Virginia Journal of International Law* 38: 639-88.

Guzman, Andrew T. (2008) *How International Law Works: A Rational Choice Theory*. New York: Oxford University Press.

Guzman, Andrew T. and Beth Simmons. (2002) To Settle or Empanel? An Empirical Analysis of Litigation and Settlement at the World Trade Organization. *Journal of Legal Studies* 31: S205-S235.

Hafner-Burton, Emilie N. (2009) *Forced to Be Good: Why Trade Agreements Boost Human Rights*. Ithaca, NY: Cornell University Press.

Haftel, Yoram Z. and Alexander Thompson. (Forthcoming) Delayed Ratification: The Domestic Fate of Bilateral Investment Treaties. *International Organization*.

Harstad, Bård. (2007) Harmonization and Side Payments in Political Cooperation. *American Economic Review* 97: 871-889.

Hathaway, Oona A. (2007) Why Do Countries Commit to Human Rights Treaties? *Journal of Conflict Resolution* 51: 588-621.

Hawkins, Darren G., David A. Lake, Daniel L. Nielson and Michael J. Tierney, eds. (2006) *Delegation and Agency in International Organizations*. New York: Cambridge University Press.

Henkin, Louis. (1979) *How Nations Behave*, 2nd Ed. New York: Columbia University Press.

Hovi, Jon, Detlef F. Sprinz and Guri Bang. (2012) Why the United States Did Not Become a Party to the Kyoto Protocol: German, Norwegian, and US Perspectives. *European Journal of International Relations* 18: 129-50.

Ikenberry, G. John. (2003) Is American Multilateralism in Decline? *Perspectives on Politics* 1: 533-50.

Johnston, Alastair Iain. (2001) Treating International Institutions as Social Environments. *International Studies Quarterly* 45: 487-515.

Kahler, Miles. (2009) Global Governance Redefined. In *The Challenges of Globalization*, edited by Andrew C. Sobel. New York: Routledge.

Keohane, Robert O. (1984) *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton: Princeton University Press.

Koremenos, Barbara. (2007) If Only Half of International Agreements Have Dispute Resolution Provisions, Which Half Needs Explaining? *Journal of Legal Studies* 36: 189-212.

Koremenos, Barbara, Charles Lipson and Duncan Snidal. (2001) The Rational Design of International Institutions. *International Organization* 55: 761-99.

Krasner, Stephen D. (1976) State Power and the Structure of International Trade. *World Politics* 27: 317-47.

Krasner, Stephen D., ed. (1983) *International Regimes*. Ithaca, NY: Cornell University Press.

Lake, David A. (1999) *Entangling Relations: American Foreign Policy in Its Century*. Princeton, NJ: Princeton University Press.

Lipson, Charles. (1985) *Standing Guard: Protecting Foreign Capital in the Nineteenth and Twentieth Centuries*. Berkeley: University of California Press.

Milner, Helen V. (2006) Why Multilateralism? Foreign Aid and Domestic founder-member Problems. In *Delegation and Agency in International Organizations*, edited by Darren G. Hawkins, David A Lake, Daniel L. Nielson and Michael J. Tierney. New York: Cambridge University Press.

Moravcsik, Andrew. (2000) The Origins of Human Rights Regimes: Democratic Delegation in Postwar Europe. *International Organization* 54: 217-52.

Moravcsik, Andrew. (1999) A New Statecraft? Supranational Entrepreneurs and International Cooperation. *International Organization* 53: 267-306.

Morsink, Johannes. (1999) *The Universal Declaration of Human Rights*. Philadelphia: University of Pennsylvania Press.

Narlikar, Amrita and Peiter van Houten. (2010) Know the Enemy: Uncertainty and Deadlock in the WTO. In *Deadlocks in Multilateral Negotiations: Causes and Solutions*, edited by Amrita Narlikar. New York: Cambridge University Press.

Naurin, Daniel and Rutger Lindahl. (2010) Out in the Cold? Flexible Integration and the Political Status of Euro Opt-Outs. *European Union Politics* 11: 485-509.

Neumayer, Eric. (2003) Do Human Rights Matter in Bilateral Aid Allocation? *Social Science Quarterly* 84: 650-66.

Neumayer, Eric and Laura Spess. (2005) Do Bilateral Investment Treaties Increase Foreign Direct Investment to Developing Countries? *World Development* 33: 1567-85.

Odell, John S. (2000) *Negotiating the World Economy*. Ithaca, NY: Cornell University Press.

- Oye, Kenneth A. (1992) *Economic Discrimination and Political Exchange*. Princeton, NJ: Princeton University Press.
- Pekkanen, Saadia, Mireya Solis and Saori N. Katada. (2007) Trading Gains for Control: International Trade Forums and Japanese Economic Diplomacy. *International Studies Quarterly* 51: 945-70.
- Pierson, Paul. (2004) *Politics in Time*. Princeton, NJ: Princeton University Press.
- Qureshi, Asif H. (1995) Some Lessons from 'Developing' Countries' Trade Policy Reviews in the GATT Framework: An Enforcement Perspective. *The World Economy* 18(3): 489-503.
- Rixen, Thomas. (2010) Bilateralism or Multilateralism? The Political Economy of Avoiding International Double Taxation. *European Journal of International Relations* 16: 589-614.
- Ruggie, John G. (1992) Multilateralism: The Anatomy of an Institution. *International Organization* 46: 561-98.
- Saggi, Kamal, and Hali Murat Yildiz. (2010) Bilateralism, Multilateralism, and the Quest for Global Free Trade. *Journal of International Economics* 81: 26-37.
- Salacuse, Jeswald W. (2009) *The Law of Investment Treaties*. New York: Oxford University Press.
- Sandler, Todd. (2004) *Global Collective Action*. New York: Cambridge University Press.
- Sell, Susan K. (2007) Intellectual Property and the Doha Development Agenda. In *The WTO after Hong Kong: Progress in, and Prospects for, the Doha Development Agenda*, edited by Donna Lee and Rorden Wilkinson. London: Routledge.
- Snidal, Duncan. (1985) The Limits of Hegemonic Stability Theory. *International Organization* 39(4): 579-614.
- Sprinz, Detlef F. and Martin Weiss. (2001) Domestic Politics and Global Climate Policy. In *International Relations and Global Climate Change*, edited by Urs Luterbacher and Detlef F. Sprinz. Cambridge, Mass.: MIT Press.
- Stasavage, David. (2002) Private Investment and Political Institutions. *Economics and Politics* 14(1): 41-63.
- Stone, Randall W., Branislav L. Slantchev and Tamar R. London. (2008) Choosing How to Cooperate: A Repeated Public-Goods Model of International Relations. *International Studies Quarterly* 52: 335-62.

- Thompson, Alexander. (2010) Rational Design in Motion: Uncertainty and Flexibility in the Global Climate Regime. *European Journal of International Relations* 16(2): 269-96.
- UNCTAD. (1998) *Bilateral Investment Treaties in the Mid-1990s*. New York: United Nations.
- USTR. (2007) Third Report to Congress on the Operation of the Andean Trade Preference Act as Amended. April 30. Washington: Office of the US Trade Representative.
- Vadlamannati, Krishna. (2008) Economic Interests versus Social Conscience Signing Bilateral Investment Treaties—Do Human Rights Matter? MPRA Paper No. 15431.
- Verdier, Daniel. (2008) Multilateralism, Bilateralism, and Exclusion in the Nuclear Proliferation Regime. *International Organization* 62, 3:439-76.
- Verdier, Daniel. (2009) Sanctions as Revelation Regimes. *Review of Economic Design* 13, 3:251-78.
- Weber, Katja. (1997) Hierarchy Amidst Anarchy: A Transaction Costs Approach to International Security Cooperation. *International Studies Quarterly* 41: 321-40.
- Willerton, John P., Gary Goertz and Michael O. Slobodchikoff. (2012) Treaty Networks, Nesting, and Interstate Cooperation. *International Area Studies Review* 15(1): 59-82. .
- Williamson, Oliver E. (1985) *The Economic Institutions of Capitalism*. New York: Free Press.
- Yarbrough, Beth V. and Robert M. Yarbrough. (1992) *Cooperation and Governance in International Trade*. Princeton University Press.
- Zahrnt, Valentin. (2007) Gain Claiming and Inefficiency in WTO Negotiations. *International Negotiation* 12: 363-88.

## Notes

<sup>1</sup>Author's notes: For helpful comments on earlier drafts, we thank Erin Graham, Yoram Haftel, Mihaela Papa and workshop participants at Ohio State University, Princeton University and the Harvard Law School. A supplemental appendix with formal proofs is available from the authors.

<sup>2</sup>For an overview of these arguments, see Ikenberry 2003.

<sup>3</sup>The producer surplus is the amount that producers benefit by selling at a market price that is higher than they would be willing to sell for.

<sup>4</sup>In the case of multiple founders, their interaction is left out of this type of model; they are assumed to act like a single player by means of an iterated play. We are making this assumption for reasons of tractability and so we can focus on the less-studied design issues that are our main concern.

<sup>5</sup>The model and the empirical illustrations refer to regimes that rely on positive incentives. If, instead of offering a reward, the founder merely threatened to sanction nonparticipation, the model would still work provided that one thinks of sanctions as negative prices. Generally speaking, positive incentives set the member's reservation value to zero, whereas negative incentives set the reservation value below zero. Except for this, the two forms of incentive are interchangeable (Verdier 2009).

<sup>6</sup>In most multilateral treaties a clause stipulates a minimum number of signatories for the treaty to enter into force. This design feature, however, lacks generalizability, for it is not feasible when the regime relies in part or in toto on bilateral treaties.

<sup>7</sup>To build the summation term, we took advantage of the mathematical identity between offering each of the  $y - x - 1$  members (1) a multilateral treaty and a supplementary bilateral treaty or (2) a bilateral treaty with incentives that subsume those of the multilateral treaty.

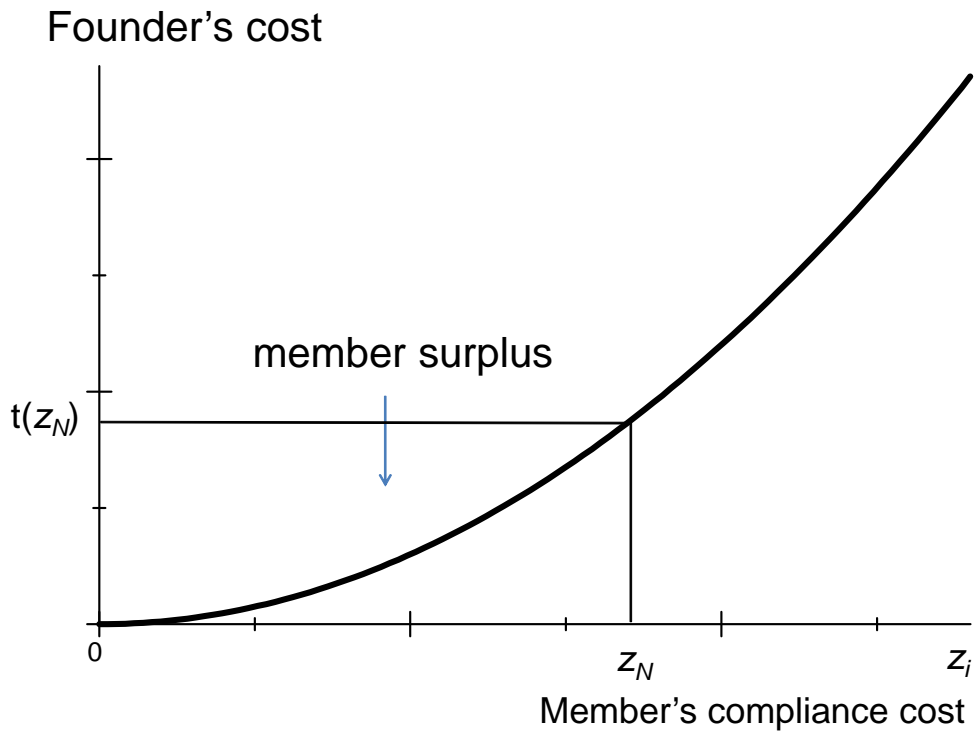
<sup>8</sup>It should be noted that most BIT disputes go through the International Centre for Settlement of Investment Disputes. The FDI regime is therefore not a case of pure bilateralism.

<sup>9</sup>For a discussion, see Thompson 2010: 285-286.

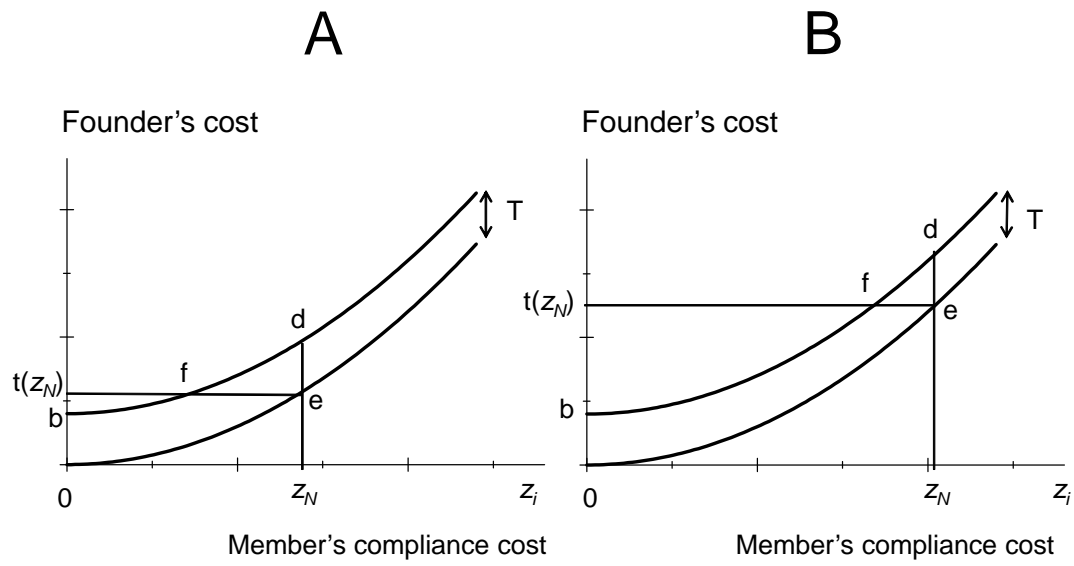
<sup>10</sup>On the compatibility of rational-choice institutionalism with a more historical approach, see Greif and Laitin (2004) and Thompson (2010).

<sup>11</sup>For example, see Willerton, Goertz, and Slobodchikoff 2012.

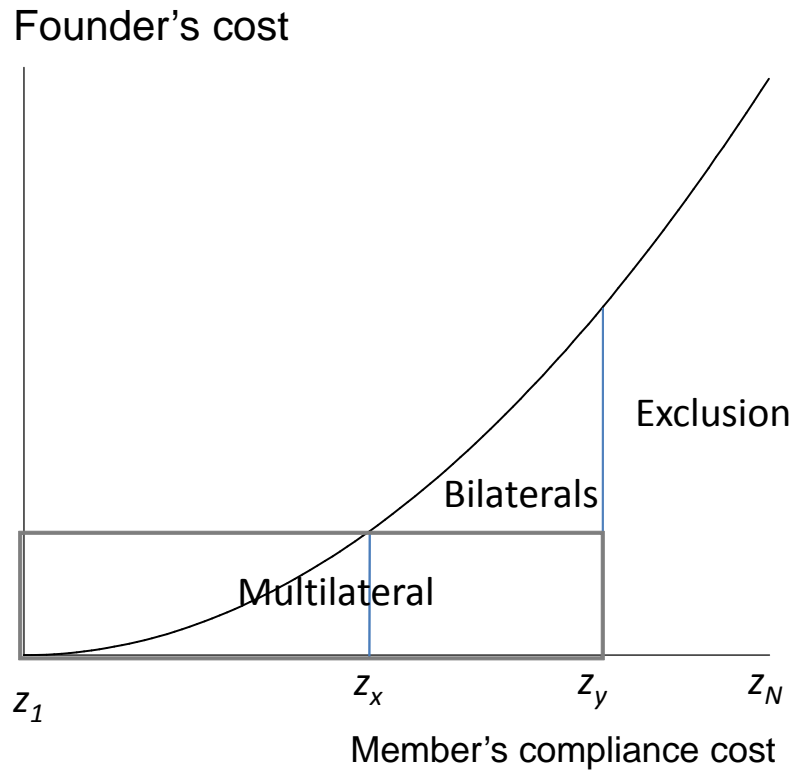
[Figure 1. The Member Surplus]



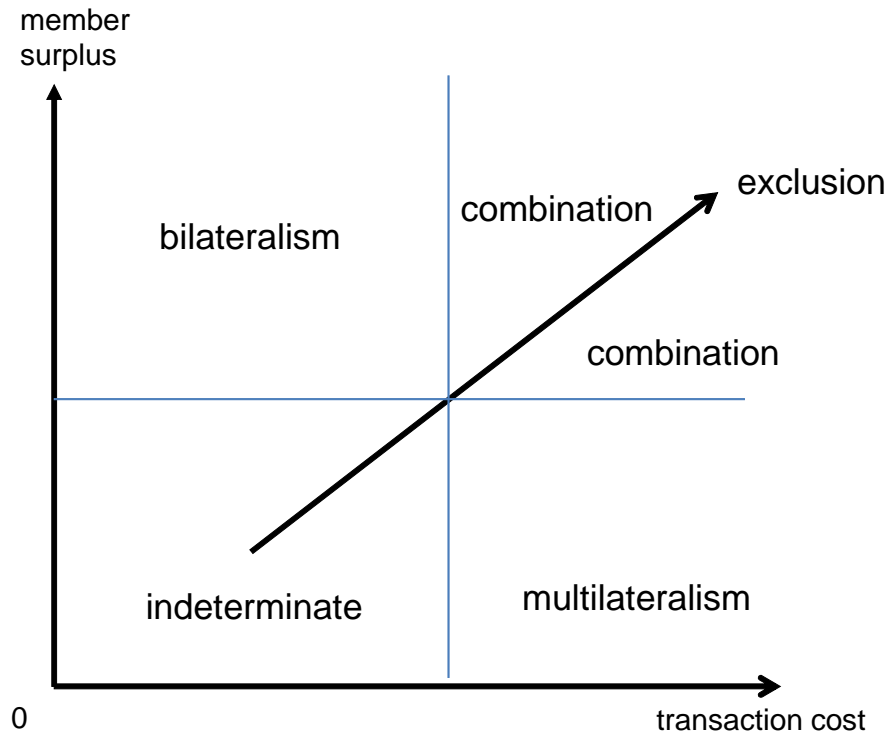
[Figure 2. Tradeoff between Member Surplus and Transaction Costs in Pure Regimes]



[Figure 3. A Regime Combining Instruments]



[Figure 4. Theoretical Predictions]



[Table 1. Values of Independent Variables and Predictions]

Table 1. Values of Independent Variables and Predictions

Regime	Independent Variables			Predictions	
	Transaction Costs	Member Surplus $a/c$ $\rho N$		Lateralism	Exclusion
FDI	Low	High	High	Bilateral	Moderate
Human Rights	High	High	Low	Multilateral	Moderate
Climate	High	High	High	Combination	Higher
Trade	High	High	High	Combination	Higher