Unilateral Influence on International Bureaucrats: An International Delegation Problem*

Johannes Urpelainen†
Department of Political Science
Columbia University
February 15, 2011

Abstract

The conventional wisdom emphasizes agency slack or bias as the central problem of international delegation. I show that the possibility of a unilateral influence contest is equally problematic. States can exert unilateral influence on autonomous international bureaucrats, either through rewards or punishments, to pursue their particular interests. A costly contest results, so some states could refuse to delegate because they expect others to be too influential. The analysis has four counterintuitive empirical implications. First, international agreements often favor institutionally weak states that are disadvantaged in the unilateral influence contest. Second, states could limit the autonomy of an international organization even if this prompts bad policies. Third, a state can sometimes profitably exchange distributional concessions for autonomy. Finally, constraints on unilateral influence are possible only if a disadvantaged state can credibly commit to compensating an advantaged state for it. A central broader contribution of the analysis is to show how power politics influences the rational design of international institutions.

*I thank Terry Chapman, Leslie Johns, Silvia Marchesi, Peter Rosendorff, Christina Schneider, the anonymous reviewers, and the editors of the Journal of Conflict Resolution for useful comments and advice. I presented this research in the poster session of the 3rd annual conference on the Political Economy of International Organizations (Georgetown University, January 2010).

†Postal: 420 W 118th St, 712 IAB, New York, NY 10027, USA. Tel: +1-734-757-0161. Email: ju2178@columbia.edu.
1 Introduction

The research program on *international delegation* is largely based on the principal-agent model.\(^1\) States are principals and international bureaucrats are agents, so the central problem of international delegation is “agency slack” (Hawkins et al. 2006b; Johns 2007; Nielson and Tierney 2003; Pollack 1997). States first bargain over the distribution of gains and then choose an international delegation contract that gives international bureaucrats the proper incentives for effective policy implementation.\(^2\)

In this article, I examine an important but poorly understood strategic problem that the canonical principal-agent model cannot capture. While the theoretical literature does address the distributional conflict over the design of international institutions, it fails to consider the possibility that states exert *unilateral influence* on international bureaucrats so as to bias policy implementation (Hawkins et al. 2006a; Martin 2006; Pollack 1997; Nielson and Tierney 2003). Reducing the value and shaping the optimal design of international delegation, unilateral influence implies that formally agreed upon institutional rules could be highly incomplete and potentially misleading proxies of the *de facto* rules of the game (Stone 2004, 2008). Consequently, standard functional theories of international institutions (Keohane 1984; Koremenos, Lipson, and Snidal 2001) may miss an integral commitment problem that states face. If states can exert unilateral influence by rewarding or punishing international bureaucrats for biased policy implementation, when and how can states capitalize on effective policy implementation through international delegation? Do the power politics of unilateral influence impede international delegation, or can states find ways to address the problem?

The question of unilateral influence is also not a theoretical curiosity. In the International Monetary Fund (IMF), the implementation of policies and enforcement of conditionality are selective. Allies of major powers can violate loan contracts without punitive consequences, despite formal rules that prescribe strict conditionality. As Stone (2004, 590) shows in his study of IMF conditionality

\(^1\)International delegation is defined here as “a grant of authority by two or more states to an international body to make decisions or take actions” (Bradley and Kelley 2008, 3).

\(^2\)Under policy implementation, I include a broad range of functions from technical assistance and coordination to dispute resolution and enforcement.
in Africa, conditionality requires credible threats to enforce it, and “[t]he obstacle to enforcing these threats is interference by the major donor countries.” In the World Trade Organization (WTO), limited legal capacity prevents developing countries from using the Dispute Settlement Understanding (DSU) to hold major powers accountable, although the very idea of legalized dispute resolution was to go beyond power politics (Busch and Reinhardt 2003). In the Global Environment Facility (GEF), wealthy donors and the World Bank dominate despite seemingly egalitarian formal voting rules because developing countries lack the resources and information to influence practical policy implementation (Streck 2001, 92).

In my formal model, two states can draft an incomplete contract to delegate policy implementation to an international organization. The contract describes the sought distribution of gains and how autonomous policy implementation by an international bureaucrat will be. Upon delegation, states engage in a zero-sum unilateral influence contest. In equilibrium, the quality of policy implementation increases with autonomy because the international bureaucrat can expend effort at a lower cost. However, an autonomous international bureaucrat cannot be easily disciplined, so the costly unilateral influence contest intensifies. By giving autonomy to an international bureaucrat, states ensure effective policy implementation. But autonomy increases the level of unilateral influence, and this is costly to both states.

The theoretical argument has several notable empirical implications. First, states sometimes deliberately formulate the delegation contract against the interests of influential states. The delegation contract forms a reference point for unilateral influence, so formal provisions that favor a disadvantaged state function as a commitment device that permits broader participation. Contravening the conventional wisdom that emphasizes the importance of aligning formal treaty provisions with the preferences of major powers (Hawkins et al. 2006a; Koremenos, Lipson, and Snidal 2001), my theoretical results reveal the possibility that written rules favor the underdog while the actual outcome favors the top dog.

Second, states must sometimes limit the autonomy of international organizations even if they ascribe a high value to competent policy implementation. If the returns to unilateral influence increase rapidly with autonomy, high levels of autonomy induce a very costly unilateral influence
contest. Consequently, states limit the autonomy of the international organization even though this results in bad policies. This finding could explain why many international organizations, such as the United Nations, are often incapable of implementing badly needed policies even though major powers seem to value such outcomes as collective security. Crucially, this result does not require the assumption that international bureaucrats hold intrinsic biases.

Third, international delegation often enables a mutually profitable exchange of distributional concessions for autonomy. If one state cares about the distribution of gains but does not value competent policy implementation, while another state is willing to concede as long as implementation is highly effective, such an exchange is possible. An example is a developing country that obtains increased foreign aid for environmental conservation and allows rigorous enforcement and intrusive monitoring by an international organization in exchange (Keohane and Levy 1996).

Finally, if states can condition international cooperation on limited unilateral influence, they achieve higher payoffs by avoiding the unilateral influence contest. However, the more influential state must be guaranteed a favorable distribution of gains in exchange for limiting its influence. Such a commitment seems to have been possible in the aftermath of the Second World War, as the United States reduced its unilateral influence in exchange for commitment to a liberal economic order by its allies (Ikenberry 2000).

These findings have notable broader implications for the voluminous literature on international cooperation and institutions (Keohane 1984; Koremenos, Lipson, and Snidal 2001). Some exceptions notwithstanding (Gruber 2000; Stone 2008; Urpelainen 2009; Voeten 2001), previous research treats international institutions as functional instruments. My findings indicate that in the case of international delegation, power politics (i) influences the rational design of delegation contracts and (ii) is necessary to explain the biased implementation of these delegation contracts. Principal-agent models of domestic and international delegation focus on the managerial aspects of the problem, and thus downplay bargaining dynamics. My results show that bargaining between states, both in formal negotiations over the delegation contract and subsequently in the form of unilateral influence contests, is an essential determinant of institutional design. Empirical analyses of the design and operation of international institutions, therefore, are highly incomplete unless they explicitly
account for relative power and unilateral influence. By characterizing the effect of power politics on the design of international institutions, my article offers a synthesis of realist and neoliberal institutionalist perspectives on international cooperation (Krasner 1991; Koremenos, Lipson, and Snidal 2001).

2 Puzzle

In a principal-agent relationship, information is asymmetric between a principal and an agent, such as a politician and a civil servant (Bendor and Meirowitz 2004; Epstein and O’Halloran 1999; Gilardi 2002; McCubbins 1985; McCubbins and Schwartz 1984; Weingast and Marshall 1988). The principal needs the agent’s expertise for competent policy implementation, but the agent could “shirk” by expending limited effort or “drift” by implementing policies that are not favorable to the principal. Consequently, the principal must design a feasible incentive structure, usually referred to as a “contract.” In the international context, the conventional wisdom is that the act of delegation is particularly problematic because states prefer not to incur “sovereignty costs” in the form of lost autonomy (Abbott and Snidal 1998, 2000; Barnett and Finnemore 2004; Hawkins et al. 2006b; Moravcsik 2000).

The extant literature emphasizes the implications of preference heterogeneity (Hawkins et al. 2006a; Martin 2006; Nielson and Tierney 2003). If the delegating states are in substantial disagreement, it is difficult to design a contract that induces mutually acceptable policy implementation by the international organization (Martin 2006). Consequently, preference heterogeneity should reduce the probability of successful delegation. This relationship is complicated if decisions are made by a “collective principal” through voting, as even the most powerful states must form coalitions to implement new policies (Nielson and Tierney 2003, 247). A Martin (2006, 144) warns, “[w]hen state preferences diverge ... there is more likely to be a wide range of proposals that could gain majority approval. This gives the staff room to maneuver.”

While useful, this literature is confused as to the nature of principal-agent problems in international politics. States are not a board of directors that first resolve their disputes, then write a legally binding contract to delegate, and finally discipline the bureaucrat. In addition to pursu-
ing their interests by bargaining over the delegation contract, states can exert unilateral influence on policy implementation by the bureaucrat. Consequently, the international bureaucrat could fail to implement outcomes specified in the contract also because some states reward or coerce international bureaucrats to bias policy implementation. The international delegation contract is “incomplete” in that states can undermine it to pursue their interests at the expense of other states (Gilligan, Johns, and Rosendorff 2010; Koremenos 2001; Tirole 1999).

Economic models of “common agency” capture some of these effects (Bernheim and Whinston 1986; Dixit, Grossman, and Helpman 1997; Grossman and Helpman 1994). In these models, multiple principals contribute money to an agent in exchange for favorable policies. However, existing models of common agency ignore the important possibility that a principal refuses to delegate because it anticipates unilateral influence by other other principals in the implementation phase.

International organizations are “permeable,” so states have ample opportunities to exert unilateral influence on international bureaucrats. The instruments of influence available to a state include career rewards and punishments. Voeten (2008, 428) finds that judges in the European Court of Human Rights “are subject to increased pressure on controversial cases that directly deal with the security of a country” and ascribes this finding to career incentives. States can also allocate tangible resources by conditioning funding on biased policy implementation, and they can collaborate directly with sympathetic agents (Addison, McGillivray, and Odedokun 2004; Pollack 1997; Selin 2007; Streck 2001). For instance, green Member States of the European Union can collude with the Directorate General Environment to ensure that their viewpoints are represented in policy implementation (Selin 2007). In the case of the UNEP or the GEF, wealthy donors have occasionally cut funding for these bodies out of fear that they either promote the interests of developing countries or push for overly ambitious environmental goals and policies (Ivanova 2007; Streck 2001).

States can also recruit staff that share their preferences or put ideological pressure on the international bureaucrats (Barnett and Finnemore 2004; Johns 2007; McKeown 2009). In the weapons inspections between the two Iraqi wars, Johns (2007, 267) argues that Russia, France, and China came to perceive the United Nations Special Commission created by the Security Council as “too aggressive and biased in favor of U.S. policy” and sought to exert influence “by forcing new
personnel to the agency.” Although the mandate of the Commission was previously established in international negotiations, major powers continued to bias policy implementation to advance their interests.

This effect can also be indirect if the qualifications required for a career in an international organization induce a selection effect that favors certain countries. Perhaps the most important example are the Bretton Woods institutions that hire a large number of economists. As Woods (2006, 53) shows, officials trained in American or British universities have historically had a dominant position in senior management. Such overrepresentation creates a fertile ground for the diffusion of Anglo-American ideas about economic policy, such as the neoliberal Washington Consensus (Williamson 1990).

Often states can directly influence decisions made by international bureaucrats. Consider the following description of the IMF Executive Board by Stone (2008, 595):

“The United States has a tremendous organizational advantage over other countries because it has a more extensive diplomatic corps, particularly important private financial institutions, numerous advantages in gathering information, and all of the advantages of having the IMF located in the U.S. capital, in addition to issuing the international reserve currency and commanding the resources of a superpower.”

Even flagrant forms of principal-agent collusion cannot be excluded. In the League of Nations, many supposedly neutral international bureaucrats were directly taking orders from their national government and even providing confidential information to the national capital (Mathiason 2007, 28). Especially problematic were the autocratic German and Italian regimes that were continuously undermining the operations of the League in pursuit of their national interest.

The relationship between a state’s raw bargaining power and ability to exert unilateral influence is complex. On the one hand, lucrative outside options may help exert influence on international bureaucrats. By threatening to leave an international organization, for instance, a powerful state may force the staff to adjust policies because they worry that the threatened “exit” will be detrimental for the organization. Thus, unilateral influence allows states to exercise power within international institutions (Stone 2008).
On the other hand, the covariates of unilateral influence may also differ from the covariates of raw bargaining power. While powerful states usually exert substantial influence also within international organizations, seemingly weak states can sometimes thrive in institutionalized settings. For example, many United Nations bodies, such as the Development Program (UNDP), have a bias towards developing countries simply because the staff have intrinsic developmental motivations. Similarly, an essential precondition for using the WTO dispute resolution mechanism is legal capacity, and many small industrialized countries in Europe have it in abundant supply (Busch and Reinhardt 2003). Throughout, I let “weak” and “powerful” refer to different levels of bargaining power under anarchy, while “advantaged” and “disadvantaged” refer to different opportunities for influence under the auspices of an international organization.

Indeed, one interpretation of unilateral influence is that it is a form of bargaining between states. As both Lake (2009) and Rector (2009) write, states cannot credibly commit to a fixed distribution of gains when they engage in international cooperation, such as forming a federation. Instead, they expect ex post bargaining within the international institution. Unilateral influence can be interpreted as one way for states to advance their interests after the initial cooperation decision.

Under unilateral influence, there is a tradeoff in how autonomous international organizations should be. For informed policy implementation, international bureaucrats must have autonomy (Hawkins et al. 2006a; Nielson and Tierney 2003). While such autonomy could cause agency slack, I focus here on unilateral influence. If international bureaucrats are relatively autonomous, so that their behavior is not directly constrained by such rigid governance structures as unanimity voting on minute details of various policies, states can expect great benefits if they successfully collude with them. If the United States can dictate the economic doctrine used by the Bretton Woods institutions, economic statecraft is easier (Barnett and Finnemore 2004). If the European Union dominates international regulatory institutions, European hegemony in regulatory issues is strengthened (Mattli and Büthe 2003). If the Soviet Union and its satellites can hijack the United Nations Educational, Scientific, and Cultural Organization, they score a major symbolic victory in the Cold War of ideas (Osakwe 1972).

This conceptualization of autonomy differs from the notions of “independence” or “insulation”
Autonomy aggravates the problem of unilateral influence because disadvantaged states cannot simply veto or otherwise prevent the implementation of biased policies. Independence and insulation could either mitigate the problem because international bureaucrats are not subject to threats by states, or amplify it because they can sell policy implementation for material or career rewards.

A final note on the relationship between unilateral influence and “flexibility” in international cooperation is in order (Downs and Rocke 1995; Koremenos 2001). According to the flexibility literature, states often incorporate escape clauses, renegotiation provisions, and other safety valves into international agreements. Unilateral influence may offer a new rationale for such flexibility designs: if influential states are able to bias policy implementation, escape clauses and other similar mechanisms can help weak states avoid the detrimental consequences of overly biased policies.

3 The Model

My model of unilateral influence focuses on situations in which policy implementation by international bureaucrats has distributional implications, so that a state has incentives to unilaterally bias policy implementation for private gain. The model can be applied when the international organization in focus deals with salient distributional issues, so that bargaining concerns are important. By contrast, the model is less applicable when state-principals have very homogeneous preferences regarding policy implementation. In this case, unilateral influence is hardly a problem. It may even be helpful when it allows states to informally guide policy formation in unforeseen contingencies or deal with hostile international bureaucrats, so that standard principal-agent models apply (Hawkins et al. 20066).

I model international delegation as follows. First, two states draft an incomplete contract that prescribes delegation of policy implementation to an international organization. Second, if they choose to delegate, each state can unilaterally influence an international bureaucrat so as to bias policy implementation toward its ideal point. Third, the international bureaucrat strategically implements a policy to achieve her preferred outcome. For ease of exposition, all model notation is summarized in Table 1.
First, each state $i = A, B$ simultaneously announces a contract $(\bar{y}, \theta)_i$. If they announce different contracts $(\bar{y}, \theta)_A \neq (\bar{y}, \theta)_B$, the game ends and both obtain an “outside option” payoff $\lambda_i > 0$ (Johns 2007; Voeten 2001). Let $\lambda_i$ be low enough for both states $i$ so that international delegation is potentially profitable in the absence of the unilateral influence contest.

If they announce an identical contract, it is denoted by $(\bar{y}, \theta)$. The first element $\bar{y} \in \mathbb{R}$ is a baseline distribution that the international bureaucrat is officially supposed to achieve through policy implementation. As $\bar{y}$ increases, state $A$ is supposed to obtain a larger share of the benefits at the expense of state $B$. Substantively, the baseline distribution comprises provisions that describe the distribution of costs and benefits, such as vote shares or general principles and norms of policy implementation.

The second element $\theta \in [0, 1]$ measures the autonomy of the international organization. As autonomy $\theta$ increases, the net cost of unilateral influence to any given state is assumed to decrease. In other words, it becomes easier for any given state to exert unilateral influence on the bureaucrat. However, the bureaucrat’s ability to competently implement policy is also enhanced. Empirically, an increase in autonomy corresponds to any changes in institutional design that allow bureaucrats to implement policies with less collective oversight by states.\footnote{Measures to “insulate” international bureaucrats, such as lifelong tenures for international judges, may have ambiguous effects. On the one hand, they may allow states to collude with sympathetic bureaucrats. On the other hand, they may reduce the effectiveness of threats.} For example, if an institutional reform deprives individual states of the ability to veto any policy change, autonomy $\theta$ increases. To illustrate the consequences, suppose the European Commission is given a lot of autonomy in policy implementation. It can capitalize on expertise to implement optimal policies, but if France is in good terms with it, other states find it difficult to prevent bargains between the Commission and France because they cannot easily interfere with the operations of the Commission.

At the second stage of the game, both states $i$ simultaneously choose how much influence $z_i \in [0, \infty)$ to exert on the international bureaucrat at cost $\alpha_i \cdot c(z_i)$, where $c$ is increasing and
strictly convex in $z_i$. Let $c(0) = c'(0) = 0$ so that neither state $i$ exerts zero influence unless influence has no effect whatsoever on the international bureaucrat. The parameter $\alpha_i > 0$ captures how costly it is for state $i$ to influence the international bureaucrat. As $\alpha_i$ increases, state $i$ becomes disadvantaged. Influence shapes the behavior of the international bureaucrat as detailed below.

At the third stage of the game, the international bureaucrat first selects an effort $e \in [0, 1]$. Following Aghion and Tirole (1997), effort constitutes the probability that policy implementation succeeds. The cost of effort is $(1 - \theta) \cdot E(e)$, where $E$ is a strictly increasing and convex function such that $E'(0) = E'(1) \to \infty$. This formulation ensures that autonomy reduces the cost of effort. For example, if the IMF staff must have the approval of all major donors for any minutiae change in a conditional loan agreement, the cost of developing new innovative policies is prohibitive.

If implementation fails, the payoff to all actors is normalized to zero. If it succeeds, the bureaucrat implicitly selects a policy to implement her ideal final distribution $y^*$. Since the model is not symmetric, it is without loss of generality to normalize the ideal point under a neutral baseline distribution, $\bar{y} = 0$, and in the absence of unilateral influence, $z_A = z_B = 0$, to zero. Given the actual baseline distribution $\bar{y}$ and unilateral influences $z_A, z_B$, the bureaucrat strategically implements

$$y^* = \bar{y} + \theta \cdot (z_A - z_B).$$  \hspace{1cm} (1)

The first term shows that the baseline distribution $\bar{y}$ influences the ideal point, as states select mutually acceptable international bureaucrats and give them proper incentives to implement the baseline distribution. The second term shows that unilateral influence by state $A$ increases the ideal point of the international bureaucrat while unilateral influence by state $B$ decreases it. This is logical because high distributions $y$ will be favorable to state $A$ at the expense of state $B$.

The payoff from successfully implementing $y^*$ to the bureaucrat is normalized to 1. Given effort $e$, the expected payoff to the bureaucrat is thus

$$e - (1 - \theta) \cdot E(e).$$  \hspace{1cm} (2)

Effort increases the probability of successful policy implementation, and this benefits the bureaucrat.
because states informally or formally reward her. This simple formulation abstracts away from exactly how the bureaucrat uses expertise and information, yet it captures the basic idea that the bureaucrat must be given incentives for policy implementation (Johns 2007).

The final distribution $y^*$ yields a policy implementation payoff $v_i(y^*)$, so the expected implementation payoff to state $i$ is

$$e \cdot v_i(y^*). \tag{3}$$

Let $v_A$ be increasing and strictly concave with $v'_i(-\infty) \to \infty$ and $v_i(-\infty) = 0$. Let $v_B$ be decreasing and strictly convex with $v'_i(\infty) \to \infty$ and $v_i(\infty) = 0$. These convenience assumptions ensure that a unique subgame equilibrium exists in the unilateral influence contest.

Combined with the cost of unilateral influence, the payoff from successful delegation for state $i$ is given by

$$e \cdot v_i(y^*) - \alpha_i \cdot c(z_i). \tag{4}$$

Intuitively, the cost of the unilateral influence contest is subtracted from the payoff to policy implementation by the international organization.

### 4 Equilibrium

My solution concept is the subgame-perfect equilibrium. An equilibrium comprises announcements $(\overline{y}, \theta)_A^*, (\overline{y}, \theta)_B^*$. For every feasible contract $(\overline{y}, \theta)^*$, it also characterizes influences $z^*_A, z^*_B$ and effort $e^*$. The final distribution $y^*$ was already characterized in the previous section.

I exclude inefficient equilibria because this is a contracting game. If possible, the equilibrium set will thus only comprise mutually profitable contracts $(\overline{y}, \theta)^*$. Otherwise all equilibria of the game are incompatible announcements $(\overline{y}, \theta)_A^* \neq (\overline{y}, \theta)_B^*$ that result in delegation failure. Intuitively, is there a delegation contract $(\overline{y}, \theta)$ that induces a delegation payoff higher than the value of the outside option $\lambda_i$ for both states $i$?

The game can be solved through backward induction. The international bureaucrat chooses
effort \( e^* \) optimally, so that the marginal cost and benefit of effort is equalized. Unless \( \theta \in \{0, 1\} \),

\[
(1 - \theta) \cdot E'(e^*) = 1.
\]

The left side is the marginal cost and the right side is the marginal benefit. If she succeeds, she implements the final distribution \( y^* \) optimally according to condition (1). Clearly, equilibrium effort \( e^* \) is increasing in autonomy \( \theta \).

Since equilibrium effort \( e^* \) is strictly increasing in autonomy \( \theta \), the payoff from policy implementation increases with it for both states \( i \).

**Claim 1.** For any given final distribution \( \tilde{y} \), the equilibrium payoff from policy implementation to state \( i \), given by \( e^* \cdot v_i(\tilde{y}) \), is strictly increasing in autonomy \( \theta \).

The cost of the unilateral influence contest notwithstanding, autonomy is unambiguously beneficial because it allows the international bureaucrat to expend effort towards successful policy implementation.

Consider next the influence contest. Autonomy \( \theta \) and the baseline distribution \( \bar{y} \) are given, but the final distribution \( y^* \) is determined in the unilateral influence contest. The following first-order condition must hold for both states \( i \):

\[
\theta \cdot e^* \cdot \frac{\partial v_i(y^*)}{\partial y} = \alpha_i \cdot \frac{\partial c(z^*_i)}{\partial z_i}.
\]

The benefits of additional influence must equal the costs, and autonomy \( \theta \) increases the benefits of additional influence. **Figure 1** illustrates the best responses. As shown in the Appendix, a unique interior equilibrium exists.

**[FIGURE 1 ABOUT HERE]**

If \( \alpha_i \) is low so that the marginal cost of influence increases only slowly for state \( i \), it can easily
influence the international bureaucrat, so it will have an upper hand against the other state \( j \). If \( \alpha_i \) and \( \alpha_j \) are increase only slowly for both states, the unilateral influence contest will be very costly.

Autonomy \( \theta \) can be harmful in the unilateral influence contest.

**Claim 2.** The equilibrium cost of unilateral influence \( \alpha_i \cdot c(z_i^*) \) is increasing in autonomy \( \theta \).

An autonomous agent can be easily influenced, so the returns to unilateral influence increase. By creating a constrained agent instead, two states can avoid the zero-sum game of unilateral influence. If a state tries to influence a constrained agent, the other state can directly intervene to undermine biased policy implementation. For instance, it could use a veto or withdraw funding.

Consider finally the decision to delegate. For a final distribution \( y^* \) and agent autonomy \( \theta \), state \( i \) is willing to delegate if and only if

\[
e^* \cdot v_i(y^*) \geq \alpha_i \cdot c(z_i^*) + \lambda_i.
\]

Unless mutually profitable policy implementation is possible, one state \( i \) obtains a strictly higher payoff \( \lambda_i \) from the *ex ante* outside option, so delegation is not feasible. Importantly, extremely asymmetric influence contests do not permit international delegation.

**Claim 3.** If a contract \( (\eta, \theta) \) induces a high or low enough final distribution \( y^* \), at least one state \( i \) cannot propose it in equilibrium.

If almost all benefits from international delegation go to state \( j \) in equilibrium, state \( i \) cannot possibly benefit from international delegation. However, it could lose because the unilateral influence contest is costly. This claim captures the commitment problem that unilateral influence prompts. If state \( j \) cannot guarantee some benefits to state \( i \), international delegation is not feasible.
5 Contracts and Outcomes

How does the design of the international delegation contract \((\bar{y}, \theta)\) influence equilibrium behavior and payoffs? The first choice that the states face is the baseline distribution \(\bar{y}\). The possibility of unilateral influence precludes credible \textit{ex ante} commitment to a final distribution \(y^*\), but the baseline distribution turns out to have an unambiguous effect on the final distribution.

**Proposition 1.** The final distribution \(y^*\) is increasing in the baseline distribution \(\bar{y}\). The influence \(z_A^*\) of the gaining state \(A\) decreases while the influence \(z_B^*\) of the losing state \(B\) increases in the baseline distribution \(\bar{y}\).

The first effect of increasing the baseline distribution \(\bar{y}\) is that the reference point for unilateral influence shifts to the advantage of state \(A\). Although changes in incentives to exert unilateral influence will partially offset this effect, the final distribution \(y^*\) must be at least as good for state \(A\). The second effect is that state \(A\) reduces its influence while state \(B\) increases its influence. The baseline distribution \(\bar{y}\) is already more favorable to state \(A\), so the value of influence decreases. State \(B\) faces an increasingly costly outcome, so it must exert additional influence.

Both effects favor state \(A\) at the expense of state \(B\). The choice of baseline distribution \(\bar{y}\) is central to the politics of international delegation because it has implications both for the distribution of gains from delegation and the cost of unilateral influence. As long as international delegation is feasible, the two states hold diametrically opposed preferences over the baseline distribution \(\bar{y}\). Unfortunately, the baseline distribution \(\bar{y}\) is an imperfect instrument. If unilateral influence cannot be prevented, both states exert some to obtain a better distribution, so the final distribution \(y^*\) drifts and both states incur a cost \(\alpha_i \cdot c(z_i^*)\). It would be better for both states if they could directly contract on the final distribution \(y^*\) and avoid the unilateral influence.

Even the direction of equilibrium influence could change with the baseline distribution \(\bar{y}\). Suppose first the baseline distribution \(\bar{y}\) is high so that state \(B\) has strong incentives to exert unilateral influence. State \(A\) obtains a favorable distribution in any case, so the most probable outcome is that the final distribution \(y^*\) will be better for state \(B\) after intense attempts to influence the in-
ternational bureaucrat: \( y^* < \overline{y} \). But if the baseline distribution \( \overline{y} \) is low so that state \( A \) has strong incentives to exert influence, the final distribution \( y^* \) drifts towards state \( A \)'s ideal point: \( y^* > \overline{y} \).

For example, if a superpower such as the United States accepts egalitarian decision rules for an international organization, it has great incentives to compel other countries to vote according to its preferences. But if the institutions rules reflect the preferences of the United States, the returns to aggressive use of power are lower.

The second choice that states face is autonomy \( \theta \). Both states gain from improved agent performance, but the effect of autonomy on the final distribution \( y^* \) or the cost of influence \( \alpha_i \cdot c(z_i^*) \) could induce distributional conflict.

**Proposition 2.** The cost of influence \( \alpha_i \cdot c(z_i^*) \) is increasing in autonomy \( \theta \). The final distribution \( y^* \) is increasing in autonomy \( \theta \) if and only if \( \alpha_B \) is low enough while \( \alpha_B \) is high enough.

As autonomy increases, the returns to successful unilateral influence increase. On the one hand, for any given influence exerted by the other state, a state has a greater incentive to exert influence because of the high benefits. On the other hand, competition will be more intense, as each state increases its influence as a response to an increase by the other state. How this shapes the final distribution \( y^* \) depends on which state is better positioned to take advantage of the increased autonomy by exerting additional influence. The relative advantage of state \( i \) depends first and foremost on the cost parameter \( \alpha_i \). For example, if the United States is in a good position to influence decisions made by the IMF Executive Board, autonomy could shift policy implementation toward the ideal point of the United States (Stone 2008).

The two states can thus reduce autonomy \( \theta \) in the delegation contract to change equilibrium behavior in two ways. First, by reducing autonomy, both states benefit from a less intense unilateral influence contest. Second, a reduced autonomy mitigates the advantage that the more influential state holds. Unfortunately, these effects come at the expense of competent policy implementation.
To analyze the design of international delegation contracts, I first find the conditions under which both states agree on the level of autonomy $\theta$ and then examine the choice of the baseline distribution $\bar{y}$.

Consider first the choice of autonomy.

**Proposition 3.** If effort $e^*$ increases rapidly (slowly) enough with autonomy $\theta$ while $\alpha_i, \alpha_j$ are high (low) enough, both states $i, j$ prefer maximal (minimal) autonomy $\theta = 1$ ($\theta = 0$).

Both states prefer autonomy if it produces a high delegation payoff by inducing competent implementation without resulting in a costly unilateral influence contest. In contrast, if autonomy does not produce a high delegation payoff, but it causes an intense unilateral influence contest, both states prefer to give the international bureaucrat limited autonomy. Even if the international bureaucrat is perfectly neutral and motivated, states may severely limit the degree of delegation to avoid a unilateral influence contest.

For instance, competent policy implementation could be particularly valuable in international environmental affairs because the consequences of pollution and resource overconsumption are highly complex (Biermann 2001; Young 2002). Additionally, many non-governmental environmental organizations constantly monitor implementation, so an international organization cannot shirk or implement biased policies in any case (Dai 2005; Raustiala 1997). Consequently, states have incentives to allow high levels of autonomy.

This result also sheds light on why states choose limited autonomy although competent policy implementation is quite valuable. Even if international organizations hold valuable expertise, the high cost of unilateral influence could dissipate too much of the surplus from international delegation. Consequently, states choose to give international bureaucrats limited freedom to implement policies, not because states do not value policy implementation, but because they understand that the international bureaucrats are sensitive to unilateral influence. For example, although the Bretton Woods institutions deal with complex issues requiring substantial expertise, the voting rules
allow major industrialized countries to easily block policies (Woods 2006).

Equally notable is the converse possibility that states disagree on autonomy. According to the conventional principal-agent model, autonomy is something that usually aligns state interests (Alter 2008; Hawkins et al. 2006a; Nielson and Tierney 2003). For instance, in his seminal application of the principal-agent model to the European Union, Pollack (1997, 129) claims that “supranational agents may exploit differing preferences among the member states to avoid the imposition of sanctions against shirking and to ‘push through’ legislative proposals via their formal agenda-setting powers.” If this is true, member states should have a strong collective incentive to constrain the supranational agents \textit{ex ante}, or before any particular contentious issue emerges, because they can thus at the very least avoid agency slippage.

This premise could fail under rather undemanding conditions. If a state is sensitive to poor performance but has a great advantage in the unilateral influence contest, it can but lose from reduced autonomy. If another state does not benefit from autonomy and is disadvantaged in the unilateral influence contest, it unambiguously prefers less autonomy. Distributional conflict surrounds constraints on agent autonomy, so the European principal-agent problem has dimensions that Pollack (1997) fails to acknowledge. Indeed, an important theme in the history of European integration are pervasive disagreements between major member states, such as Germany and Great Britain, on how centralized the union should be (Konstantinidis 2008; Slapin 2009).

For another example, consider a wealthy democracy and a poor autocracy that agree on democratization in exchange for foreign aid (Dunning 2004; Wright 2009). The poor autocracy is only willing to do this to receive foreign aid, so it does not really care about poor performance. If the international organization fails to monitor democratization, the ruler might even benefit. However, the wealthy democracy is interested in avoiding bad performance because the only reason it is giving foreign aid is to facilitate democratization in the poor autocracy. This distributional conflict is further aggravated if the wealthy democracy is unable to exert unilateral influence because domestic institutional constraints shrink the room for maneuver, while the poor autocracy is willing to bribe officials in the international organizations.

What about the baseline distribution \( \overline{y} \)?
Proposition 4. Fix autonomy $\theta$. If $\alpha_i$ is high enough while $\alpha_j$ is low enough, the baseline distribution $\overline{y}$ must be high (low) enough for $i = A$ ($i = B$).

Reduced autonomy mitigates the unilateral influence contest. But for any level of autonomy $\theta$, it is also true that if state $i$ is disadvantaged enough in the unilateral influence contest, the baseline distribution $\overline{y}$ must reflect its preferences. A favorable baseline distribution $\overline{y}$ reduces this disadvantage, so it could permit a credible commitment to a mutually acceptable final distribution $y^*$. This possibility is illustrated in Figure 2, which contrasts baseline and final distributions for a fixed autonomy value. In the upper panel, two equal states bargain over delegation, so the baseline distribution equals the final distribution. In the lower panel, the states are very unequal. Thus, the disadvantaged state $A$ must be given a favorable baseline distribution that is then overturned by the advantaged state through unilateral influence.

[FIGURE 2 ABOUT HERE]

This effect should be most pronounced when autonomy per se is valuable because it enables competent policy implementation. Neither state is ready to sacrifice the quality of policy implementation, so they must find a way to credibly commit to giving some gains to the disadvantaged state. They can do this by drafting a contract that specifies large gains for the disadvantaged state, understanding that this status quo will not hold after the unilateral influence contest.

As an example of high autonomy $\theta$ despite asymmetric unilateral influence, consider the WTO DSU. The complexity of international trade law gives a dispute resolution panel substantial autonomy $\theta$ in policy implementation. As Shaffer (2009, 168-169) writes, “just to read through and understand the growing WTO case law is an immense task, including for specialized academics.” Why? The increasing scope and complexity of international trade law implies that legalized dispute resolution is simply not possible without a sophisticated legal apparatus and a body of precedents. Given that states have chosen to legalize the primary mechanism for settling international trade
disputes, autonomy is plainly a necessary condition for successful implementation.

At the same time, the WTO DSU allows states to exploit legal capacity for unilateral influence. In the legalistic framework of the WTO DSU, the key to success in international trade dispute settlement is to win cases. Indeed, the complexity of the international trade law has greatly increased over time. Overall, WTO case law now comprises over 30,000 pages of decisions and other legal documents. The average length of a GATT decision was only twelve pages but the WTO panel rulings range anywhere from 100 to 500 pages. As a result, any state that does not possess substantial legal capacity or is unwilling to invest in legal expertise cannot meaningfully participate in legal dispute resolution. As Busch and Reinhardt (2003, 720) argue, the most important avenue for concessions under the WTO DSU is to settle the dispute in the shadow of international trade law, and most developing countries have consistently failed to do so: “our results indicate that the central problem for developing countries is that they are missing out on early settlement, not that they boast a worse record in winning pro-plaintiff rulings from Panels or the [Appellate Body] ... we find that this gap is due to a lack of legal capacity, not a lack of market power with which to threaten retaliation.”

If unilateral influence through exploitation of legal capacity strongly favors industrialized countries, why did developing countries accept the design of the WTO DSU? The distribution of power in the reform of the WTO DSU was, in terms of outside options, favorable to the wealthy industrialized countries. As the success of export-led growth strategies and the failure of import-substitution strategies suggests, developing countries are fundamentally dependent on open access to the international market (Frieden and Martin 2002; Milner 1999). Their ability to shape the rules of international trade is much less pronounced than in the case of international environmental affairs because they cannot credibly threaten to “exit” the world economy (Gruber 2000). Given the weak outside options that developing countries had, as well as the importance of autonomy for international trade law, it is understandable that the developing countries were willing to accept the WTO DSU despite the possibility of unilateral influence.

These findings are particularly important for two reasons. First, they contradict the conventional wisdom that delegation contracts should necessarily reflect the interests of powerful and
influential states. Hawkins et al. (2006a, 22) argue that “[w]hen institutional rules fail to reflect accurately the distribution of power, powerful states will more readily choose to act alone outside the institution” while Koremenos, Lipson, and Snidal (2001, 792) write that powerful states control international institutions through international agreements. My findings corroborate the underlying logic of power, but they also indicate that the conventional empirical predictions are only correct if powerful states cannot operate within the institution by violating the rules. Instead, delegation outcomes should reflect powerful interests. If a state has a sufficiently large disadvantage in the unilateral influence contest, no matter how weak it is outside the international organization, the delegation contract will prompt a seemingly neutral baseline policy that is then informally overturned, as was true of the WTO DSU. Conventional inferences on the advantage that major powers have in international organizations are therefore biased downward if one focuses on formal treaty provisions only.

This is also what Stone (2008) finds in his empirical analysis of IMF conditionality. Where essential American interests are threatened by stringent conditionalities, the latter are simply not being enforced, but to avoid dissipating the legitimacy of the Bretton Woods institutions, the United States exercises strategic restraint. In terms of my model, informal governance is the special case in which one of the principals enjoys a dominant position within and outside the relevant international organization.

Second, the findings provide a powerful rationalist argument for egalitarian institutional rules in international politics. Many scholars have highlighted the need to legitimize international institutions and organizations (Barnett and Finnemore 1999; Hall 1997; Hurd 1999; Najam 2005; Reus-Smit 2003; Ruggie 1982), but few have systematically examined the strategic implications of various principles and norms. My argument suggests that egalitarian principles, often seen as particularly legitimate and fair (Buchanan and Keohane 2006; Grant and Keohane 2005; Rubenstein 2007), have the added strategic benefit of committing powerful states to mutually acceptable delegation outcomes. However, the analysis also suggests that such egalitarian principles could hide an exploitative relationship or an intense unilateral influence contest.

The GEF offers an illustration. Primarily at the urging of France and Germany, the GEF
was established to serve as the focal international organization responsible for North-South aid flows to implement environmental projects in developing countries (Streck 2001, 72). After an unsuccessful pilot phase from 1991 to 1994, the participating countries decided to restructure the GEF (Clémençon 2006, 52-53). The restructured GEF is an “institutional innovation” that strikes a delicate balance between demands made by wealthy donors and the concerns of the developing countries (Young and Boehmer-Christiansen 1997). First, the Council – the main governing body of the GEF – uses a double majority system. Any decisions must be supported both by a majority of the representatives and a group of countries responsible for the majority of contributions. Second, the composition of the Council has the developing countries in a narrow majority. Third, the GEF is a joint enterprise between the World Bank, the UNEP, and the UNDP. The World Bank was dominated by the wealthy industrialized countries, and in the international negotiations on the legal status of the restructured GEF, both the United States and Japan initially insisted that the GEF must operate within the World Bank (Sjöberg 1999, 34). As for the UNEP and the UNDP, Sjöberg (1999, 12) argues that “[t]heir support for the UN system made the developing countries natural allies with the UN agencies on a number of issues.”

The autonomy of the GEF staff is crucial for successful implementation. First, the GEF focuses on technically complex issue areas, such as global warming and biodiversity conservation (Streck 2001, 73). To implement and evaluate an environmental project, the GEF must first establish a baseline for biodiversity deterioration or the growth of greenhouse gas emissions and then estimate the effect of a proposed environmental project against this hypothetical baseline. The environmental projects must also be subjected to cost-benefit analyses and their economic and social impact must be addressed. As a 1998 evaluation of the performance of the GEF shows, both issues have posed major difficulties (Porter et al. 1998). A second complication is the requirement that the environmental projects be only funded based on their “incremental cost” beyond what the developing country should do to maximize national benefits (Clémençon 2006, 53). Wealthy donors were unwilling to help the developing countries promote their national interest, so they required that all contributions be used for globally beneficial environmental projects. Evaluating the increment of any given environmental project, however, is an enormously complicated technical task that gives
the international bureaucrats a drastic informational advantage (Young and Boehmer-Christiansen 1997, 198).

This autonomy, combined with the central position of the World Bank, allows the global North to exert unilateral influence. If the principle of incremental cost is used to determine how many funds a recipient is granted, but the recipient does not even participate in assessing the magnitude of the incremental costs, the wealthy donors have a tremendous advantage. A 2002 evaluation of GEF projects found that this is the case: the “team found that both [Implementing Agency] staff and other GEF stakeholders at the country level seemed unfamiliar with, and sometimes uncomfortable about their lack of understanding of, the economic concepts and the GEF Operational Strategy relating to the incremental costs of delivering global environmental benefits” (Christoffersen et al. 2002).

The voting rule compromise between North and South is congruent with the predictions of the model. While the wealthy donors can choose how much to contribute, the developing countries have virtually no intrinsic interest in the GEF, so they can threaten to simply not implement any environmental projects (Barrett 1994). As Streck (2001, 83) writes, “[d]eveloping countries were almost exclusively interested in the size of the fund and in extending the coverage of the GEF to areas that donor countries did not regard as clearly global problems, such as desertification and drinking water quality.” The mixed governance structure reflects neither the ideal point of North, the “one dollar one vote” system, nor the ideal point of South, the “one country one vote” system (Streck 2001, 77). Because the developing countries had a nontrivial outside option, they were able to extract concessions from the developed countries, as conventional bargaining theory predicts. However, the degree to which the governance structure of the GEF reflects principles of universality and equality, as opposed to effectiveness and competence, points to the central empirical implication of my analysis: disadvantaged states with limited influence must obtain a favorable baseline distribution $\bar{y}$ to delegate in the first place. Although the $ex \ ante$ outside option that South had was relatively strong, its ability to exert influence under the GEF was severely reduced by financial constraints and the dominance of the World Bank. As a result, the distance between the delegation contract and the practical reality, $|y^* - \bar{y}|$, is unusually high.
Consider finally the relationship between autonomy and the baseline distribution. If preferences over autonomy $\theta$ and the baseline distribution $\mathbf{y}$ vary, might two states profit from a distribution-autonomy exchange?

**Proposition 5.** Consider a delegation contract $(\theta, \mathbf{y})$ such that the autonomy parameter $\theta$ cannot be adjusted for mutual benefit. If the marginal value of effort $v_i(y^*)$ is high enough for state $i$ relative to the marginal value of distributional gains $|\partial v_j(y^*)/\partial y|$ for state $j$, both states benefit from shifting the baseline distribution $\mathbf{y}$ towards state $j$’s preference in exchange for increasing autonomy $\theta$.

Although changes in the baseline distribution $\mathbf{y}$ are always detrimental for one of the states, their willingness to pay for distributional gains could differ. State $i$ could be willing to accept a less favorable baseline distribution $\mathbf{y}$ in exchange for higher autonomy $\theta$ because the cost of poor implementation is so high. These gains from trade highlight an important relationship between the baseline distribution $\mathbf{y}$ and autonomy $\theta$. As long as the two states value different things, simultaneous adjustments are something both of them value. Interestingly, Morrow (1991) proves a related result for asymmetric alliances in which a weak state concedes “autonomy” for “security” provided by a powerful state.

Consider an example. If a wealthy donor is concerned with the allocation of foreign aid in a poor recipient country, it could propose a contract that increases the sum of foreign aid at the expense of more stringent conditions (Stone 2004; Svensson 2000; Vreeland 2003). If the recipient is in dire need of funds but the donor is not, both countries could benefit from this exchange. Increasing foreign aid is a side payment that can be loosely thought of as a shift in the policy towards the ideal point of the recipient. If the cost of the side payment is lower to the donor than the cost of changing the outcome of interest *per se*, the side payment permits international delegation that is otherwise infeasible.
7 Avoiding Unilateral Influence

Unilateral influence creates a commitment problem. States can freely choose the baseline distribution $\overline{y}$, so it would be in their common interest to replace the final distribution $y^*$ with an identical baseline distribution $\overline{y} = y^*$ and refrain from all unilateral influence. This would not change the expected policy implementation payoff, but it would save the costs of the influence contest.

Are there feasible solutions to this commitment problem? I now assume states can condition policy implementation on the distance between the baseline and final distribution $|y^* - \overline{y}|$. Intuitively, states agree on suspending international cooperation if one of them exerts undue or illegal influence. To that end, either state can prevent policy implementation if the distance $|y^* - \overline{y}|$ is strictly higher than $d$. The parameter $d$ measures the difficulty of observing and obtaining verifiable evidence on illegal influence, so it can be thought of as representing the “contractability” of unilateral influence (Bolton and Dewatripont 2005; Lipson 2003; Tirole 1999). As $d$ increases, it becomes more difficult to hold a state accountable for unilateral influence. For example, $d$ could increase if third parties, such as non-governmental organizations, lose interest in monitoring the international organization.

Does this additional commitment capacity matter? Consider delegation without conditionality. If the distance $|y^* - \overline{y}|$ is smaller than $d$ anyway, imposing conditionality has no effect on state behavior. But if $|y^* - \overline{y}|$ exceeds $d$, it is possible to reduce it to $|y^{**} - \overline{y}| = d$ through conditionality. Figure 3 shows how conditionality can prevent excessive influence.

[FIGURE 3 ABOUT HERE]

Fix some baseline distribution $\overline{y}$. On the one hand, the shift in the final distribution from $y^*$ to $y^{**}$ certainly harms one of the states. On the other hand, one might intuitively expect that the reduced cost of unilateral influence sometimes offsets this effect. This reasoning fails, however, and the cost savings from contractability cannot outweigh the loss of influence.

**Proposition 6.** Fix a contract $(\overline{y}, \theta)$. If conditionality $d$ has an effect on the final distribution, so
that $y^{**} \neq y^*$, it is unambiguously detrimental for state $A$ when $y^* > \bar{y}$ and for state $B$ when $y^* < \bar{y}$.

This proposition shows that for a given distribution of gains, as described in the delegation contract, an influential state prefers not to condition international cooperation. While it benefits somewhat from a less costly contest, the lost influence is simply too valuable.

Although the direct effect of conditionality is distributional conflict, it remains to consider changes to the delegation contract $(\bar{y}, \theta)$ that compensate the influential state that is losing:

**Proposition 7.** Fix a contract $(\bar{y}, \theta)$. If conditionality $d$ has an effect on the final distribution, so that $y^{**} \neq y^*$, the baseline policy $\bar{y}$ can always be adjusted so that the equilibrium payoff to both states strictly increases.

Changes in the baseline distribution $\bar{y}$ can be made so that the final distribution moves only slightly due to conditionality, $y^{**} \approx y^*$. The distance $|y^{**} - \bar{y}|$ is strictly smaller than $|y^* - \bar{y}|$, so both states must be exerting less influence. The disadvantaged state benefiting from conditionality gains because the final distribution $y^{**}$ is an improvement over $y^*$, while the advantaged state losing from conditionality gains because the decrease in the influence cost helps it.

Such conditionality may have been at play when nine World Bank executive directors criticized the United States and the World Bank President Robert Zoellick in January 2010.\(^4\) The Obama administration had proposed to President Zoellick that the World Bank cease funding coal plants without discussing the proposal with other shareholders. Representing 90 developing countries, the executive directors wrote that they were concerned about “unhealthy subservience of the decision making process in the Bank to the dictates of one member country brought about in an opaque and non-inclusive manner.” To the degree that such negative publicity is effective, it deters the effectiveness of American unilateral influence.

There also appears to be an intriguing commitment problem in the second order. Suppose that the disadvantaged state proposes conditionality and promises to adjust the baseline policy $\bar{y}$ in

\(^4\) *ClimateWire* January 25, 2010.
exchange. The advantaged state should certainly accept if it believes that the baseline policy \( \bar{y} \) will be changed. But if it believes the baseline policy \( \bar{y} \) cannot be changed, it expects to lose. Thus, mutually profitable conditionality requires that the influential state also be powerful in bargaining under anarchy. This seems plausible because influence and powerful generally go hand-in-hand. For instance, Ikenberry (2000) shows that the United States exercised “strategic restraint” after the Second World War in exchange for creating a legitimate liberal economic order among the Western allies according to its own preferences. By agreeing on institutional constraints that reduced the vulnerability of its Western allies, the United States was able to widen and deepen international cooperation.

8 Empirical Testing Strategies

How can the empirical implications of the model be tested? In this section, I briefly evaluate how the model predictions can be distinguished from standard principal-agent theories of international delegation (Bradley and Kelley 2008; Hawkins et al. 2006b). To begin with, the main empirical implication of my model is that states may design delegation contracts to reflect the interests of disadvantaged states, but advantaged states use unilateral influence to move the final policies towards their ideal point. Thus, the model offers opposite expectations for different cases. If states have equal influence capabilities, the delegation contract will distribute the gains evenly and approximate the final policy outcome. If states have unequal influence capabilities, the delegation contract will favor the underdog but the final policy outcome will favor the top dog. Such hypocrisy, I have found, is a consequence of the unilateral influence contest.

By contrast, standard principal-agent models offer a very different set of hypotheses. First, they only incorporate distributional conflict between states as an impediment to effective oversight. They do explain why the distance between the delegation contract and the final policy outcome may increase with distributional conflict, but they attribute the direction of policy bias to the agent’s intrinsic preferences. Thus, the unilateral influence theory can be distinguished from principal-agent models by comparing the relative performance of agency bias versus the preferences of advantaged states in a multivariate statistical model.
Second, principal-agent models imply that states design delegation contracts in view of their relative bargaining power (outside options) and agency bias. My model retains the focus on relative bargaining power, but it adds a new dimension to the problem: delegation contracts may well diverge from the expected bargaining outcome under unilateral influence, and the direction and magnitude of this divergence depends on the asymmetry of unilateral influence. My model implies that final policy outcomes are directly influenced by unilateral influence capabilities and raw bargaining power under anarchy, whereas standard principal-agent models focus on the nature of the delegation contract. In a quantitative analysis, one could compare the effect of changes in relative bargaining power or unilateral influence capabilities on delegation contracts versus actual policy implementation.

To illustrate, consider again the GEF and the WTO DSU. In the GEF, the industrialized countries had to commit to egalitarian voting rules because the developing countries had strong outside options. However, unilateral influence nonetheless allows the global North to bias policy implementation towards their own preferences. In the WTO, the developing countries did not have strong outside options, so the industrialized countries did not have to accept built-in provisions that help developing countries develop legal capacity. If the developing countries had stronger outside options, they might have demanded that the WTO DSU be accompanied by significant efforts to build legal capacity. The unilateral influence theory can explain both outcomes: although the delegation contracts may have appeared relatively equal, the outcomes were heavily biased toward the global North’s ideal point.

What about competing hypotheses? Principal-agent models predict that the distribution of power is codified in the international agreement (Hawkins et al. 2006a), so they cannot explain egalitarian voting rules that strongly favor the developing countries. Additionally, they do not shed light on how state preferences influence the direction of agency bias – they assume the main source of such bias is found in the agent’s intrinsic preferences.

Many other international relations theories are also unable to explain these outcomes. Voting rules are irrelevant for realists (Hasenclever, Mayer, and Rittberger 1997; Waltz 1979), so they can explain away egalitarian voting rules, but they cannot explain why other international organizations,
such as the IMF, have voting rules that strongly favor major powers such as the United States. Theories of gradualism that focus on the difference between “deep” and “shallow” cooperation can also explain away egalitarian voting rules by claiming that cooperation is shallow (Downs, Rocke, and Barsoom 1996), but the distributional conflict surrounding these voting rules contradicts the idea that major powers were indifferent.

9 Conclusion

I have focused on the problem of unilateral influence in international delegation. The extant literature has examined distributional conflict over the design of international agreements, but it has not investigated the time-inconsistency problems that arise if states cannot commit to restraint from unilateral influence in international organizations (Bradley and Kelley 2008; Keohane 1984; Koremenos, Lipson, and Snidal 2001). My analysis reveals an important tradeoff. By giving international organizations autonomy, states improve policy implementation. Conventional wisdom holds that the cost of this autonomy is agency slack, but the possibility of unilateral influence contests points to two additional obstacles. First, if a disadvantaged state cannot trust that a powerful state refrains from biasing policy implementation, mutually profitable delegation is not possible. Second, the unilateral influence contest is costly to both states.

These results refine the nascent social science of international delegation. While the standard principal-agent model helps us understand managerial problems in international delegation, this article focuses on genuine distributional conflicts. It combines insights from international cooperation theory, and particularly the idea that international institutions can be “rationally designed” (Koremenos, Lipson, and Snidal 2001), with the realist notion that power politics are an essential element of international relations (Drezner 2007; Gruber 2000). The theoretical and empirical implications show that the design of international institutions is a complex strategic problem. In particular, states must understand that the rules of international institutions are imperfectly enforceable. They must be chosen so that the expected effect of the unilateral influence contest is internalized. Most surprisingly, international institutions should formally benefit disadvantaged states. This and the numerous other paradoxical implications imply that previous accounts of how
international institutions are designed and operate miss an important element of the puzzle. Yet these findings also create exciting opportunities for progressive theoretical and empirical research in the future.
Appendix

Equilibria

Show that a unique subgame-perfect equilibrium exists given any mutually profitable contract \((\bar{y}, \theta)\). First, consider existence. The result is immediate for \(\theta = 0\). Fix \(\theta > 0\). A strictly increasing and continuously differentiable unique best response \(\hat{z}_j(z_i)\) exists by the properties of \(v_j, c\). By strict convexity of \(c\), the derivative \(\hat{z}'_j(z_i)\) approaches zero for high enough \(\hat{z}_j\). This guarantees that the two best responses \(\hat{z}_i, \hat{z}_j\) meet at least once. Second, consider uniqueness. Towards a contradiction, let \((z^*_A, z^*_B)\) be a subgame equilibrium and suppose there exists another subgame equilibrium \((z''_A, z''_B)\).

The best responses are strictly increasing. Set without loss of generality \((z''_A, z''_B) > (z^*_A, z^*_B)\). By the properties of \(v_A, v_B, c\), there is no final distribution \(y^*\) for which (6) is met for both states \(i = A, B\) simultaneously. Since \(c'(0) = 0\), only interior equilibria are possible, so the claim follows.

Consider now subgame-perfect equilibria at the contracting stage. Any commonly proposed contract \((\bar{y}, \theta)\) that induces a final distribution \(y^*\) such that \(e^* \cdot v_i(y^*) \geq \lambda_i\) for \(i = A, B\) is part of some subgame-perfect equilibrium of the game because no profitable deviation exists. If state \(A\) proposes \((\bar{y}_A, \theta)\) such that \(\bar{y}_A \to \infty\) while state \(B\) proposes \((\bar{y}_A, \theta)\) such that \(\bar{y}_B \to -\infty\), no profitable deviation exists either, so any such pair of proposals is a subgame-perfect equilibrium. A Pareto-frontier exists because the payoff to state \(A (B)\) given a contract \((y, \theta)\) is strictly increasing (decreasing) in \(y\) by Proposition 1 and autonomy \(\theta\) is chosen on the closed interval \([0, 1]\).

Proof of Claim 1

The payoff from policy implementation \(e^* \cdot v_i(\bar{y})\) is strictly increasing in equilibrium effort \(e^*\). Equilibrium effort \(e^*\) is strictly increasing in autonomy \(\theta\) by (5).

Proof of Claim 2

The payoff \(e \cdot v_i(y_i)\) is increasing and strictly concave in \(y_i\) for state \(i\). The effect of influence \(z_i\) on \(y_i\) is positive linear. The cost \(\alpha_i \cdot c(z_i)\) of influence \(z_i\) is increasing and strictly convex in \(z_i\). Thus, the unilateral influence contest has a unique solution. Examine (6) to see that for \(\theta' < \theta''\), this solution is found at \((z'_A, z'_B) \leq (z''_A, z''_B)\), where the inequality is strict for an interior equilibrium.
Proof of Claim 3

If $y^* \to \infty$ or $y^* \to -\infty$ given $(\overline{y}, \theta)$, at least one state $i$ obtains an equilibrium payoff that is arbitrarily close to zero, so the *ex ante* outside option $\lambda_i$ will be a profitable deviation. ■

Proof of Proposition 1

Let $\overline{y}$ increase by a small amount $\Delta \overline{y}$ and examine (6) to establish that the first-order condition cannot be simultaneously met for both states $i$ at any final policy $y^*$ lower than previously. To see that $z_A^*$ must decrease, first note that $v_A$ is strictly concave in $y$. By convexity of $c$, the unique solution $z_A^*$ to (6) given an increased $y^*$ must decrease. To see that $z_B^*$ must increase, proceed similarly to the case of state $A$. ■

Proof of Proposition 2

Examine (6) to see that with $v_i$ strictly concave and $c$ strictly convex, equilibrium influence $z_i^*$ must increase with autonomy $\theta$ for both states $i$. Now let $\alpha_A \to 0$ and $\alpha_B \to \infty$. Now the increase in $z_B^*$ is negligible while the increase in $z_A^*$ is non-negligible. As a result, $y^*$ increases. Exactly the opposite holds if $\alpha_A \to \infty$ and $\alpha_B \to 0$. ■

Proof of Proposition 3

If $e^*$ increases rapidly enough in $\theta$ while $\alpha_i, \alpha_j$ are high enough, the equilibrium influence cost $\alpha_i \cdot c(z_i^*) = \alpha_i \cdot \int z_i^* d'(z_i) dz_i$ given by (6) is negligible for both states $i$ while the policy implementation payoff (3) increases rapidly with $\theta$. Exactly the opposite holds if $e^*$ increases slowly in $\theta$ while $\alpha_i, \alpha_j$ are low enough. ■

Proof of Proposition 4

Fix $\theta$. Let $\alpha_i \to \infty$ and $\alpha_j \to 0$. Then the distance $|y^* - \overline{y}|$ will be large, so $v_i(y^*) \to 0$ unless $\overline{y}$ is high (low) for $i = A$ ($i = B$). But then (7) cannot hold for a given $\lambda_i > 0$. ■
Proof of Proposition 5

For sufficiently high $v_i(y_i)$, small increases in autonomy $\theta$ have a non-negligible effect on the payoff from delegation (4) that dominate the effect of increased cost $\alpha_i \cdot c(z^*_i)$ for any given baseline policy $\overline{y}$. By assumption, state $j$ cannot benefit from an increase in autonomy $\theta$, so a small increase in autonomy $\theta$ produces a large payoff increase for state $i$ and a small payoff decrease for state $j$.

By Claim 1, small shifts in $\overline{y}$ toward state $j$’s ideal point result in a shift of $y^*$ in that direction. For sufficiently high $| \frac{\partial v_j(y^*)}{\partial y} |$, small shifts in $\overline{y}$ toward state $j$’s ideal point have a non-negligible effect on the payoff from delegation (4) that dominate any effects on the unilateral influence contest. Thus, the shift produces a large payoff increase for state $j$ and a small payoff decrease for state $i$. ■

Proof of Proposition 6

Without loss of generality, let $y^* > \overline{y}$. Conditionality has an effect on the final policy $y^*$ if and only if $y^* > \overline{y} + d$, in which case $y^*$ decreases. Thus, $v_A(z^*_A)$ decreases while $v_B(y^*)$ increases. Additionally, $z^*_A, z^*_B$ decrease. It remains to show that the decrease in $\alpha_A \cdot c(z^*_A)$ does not dominate the decrease in $c^*(\theta) \cdot v_A(z^*_A)$. Suppose towards a contradiction this was the case. With $z^*_B$ lower than previously and $c$ strictly convex, state $A$ could have previously profitably deviated from $y^*$ to $\overline{y} + d$, so the strategy profile prior to conditionality could not have been an equilibrium, a contradiction. ■

Proof of Proposition 7

By Proposition 1, the final policy $y^*$ varies continuously and monotonically in the baseline policy $\overline{y}$. Impose conditionality $d > 0$ and set $\overline{y}$ such that the final policy $y^*$ remains unchanged. Now $| y^* - \overline{y} |$ decreases, so at least one state $i$ strictly benefits because it has to exert less influence $z^*_i$ to achieve $y^*$. For the other state $j$, the payoff remains unchanged. Now shift $\overline{y}$ slightly to the advantage of state $j$. ■
Symbol Interpretation

\( i = A, B \) States
\( \lambda_i \) Value of outside option to state \( i \)
\( \gamma \) Baseline distribution prior to unilateral influence
\( y \) Final distribution following unilateral influence
\( \theta \) Autonomy
\( z_i \) Unilateral influence by state \( i \)
\( \alpha_i \cdot c(z_i) \) Cost of unilateral influence to state \( i \)
\( v_i(y) \) Policy payoff to state \( i \)
\( e \) Effort by the international bureaucrat
\( E(e) \) Cost of effort to the international bureaucrat

Table 1. Model notation.
Figure 1. The best-response correspondences $\tilde{z}_A, \tilde{z}_B$. 
Figure 2. Baseline and final distributions in equilibrium. Under equal influence, the baseline and final distributions are identical (upper panel). Under unequal influence, the baseline distribution is biased towards the disadvantaged state A’s preferences but the final distribution will reflect more the advantaged state B’s preferences (lower panel). In the figure, $v_A = 1 + y, v_B = 1 - y$ and $\theta = \frac{1}{2}, c = 1$ are fixed. In the upper panel, $\alpha_A, \alpha_B = 1$. In the lower panel, $\alpha_A \to \infty, \alpha_B = 1$. 
Figure 3. Upon contractability, previous influences $z^*_A, z^*_B$ are replaced by lower influences $z^{**}_A, z^{**}_B$. If contractability limits the influence of state $A$, it chooses the influence $z^{**}_A$ such that $y^* = \bar{y} + d$ in equilibrium.
References


