

Beyond Institutional Design: Explaining the Performance of International Organizations

Ranjit Lall*

Forthcoming, *International Organization*

Abstract

International organizations (IOs) have long been a central focus of scholarship in the field of international relations, yet we know surprisingly little about their performance. This article offers an explanation for differences in the performance of IOs and tests it using the first quantitative dataset on the topic. I argue that the primary obstacle to effective institutional performance is not deviant behavior by IO officials – as suggested by conventional “rogue-agency” analyses – but the propensity of states to use IOs to promote narrow national interests rather than broader organizational objectives. The implication is that IOs that enjoy policy autonomy vis-à-vis states will exhibit higher levels of performance. Critically, however, I argue that in the international context policy autonomy cannot be guaranteed by institutional design. Instead, it is a function of two key variables: (1) the existence of (certain types of) institutionalized alliances between IOs and actors above and below the state; and (2) the technical complexity of IO activities. I provide empirical evidence for the argument by constructing and analyzing a cross-sectional dataset on IO performance, which covers 53 IOs spanning a wide range of issue areas, and conducting a comparative case study in the realm of global food security.

* Department of Government, Harvard University. I thank Kenneth Abbott, James Alt, Anthony Atkinson, Jeffrey Frieden, Daniel Honig, Liesbet Hooghe, Torben Iversen, Robert Keohane, Gary King, Gary Marks, Walter Mattli, Anne Sartori, Beth Simmons, Duncan Snidal, Dustin Tingley, and George Yin for excellent comments and suggestions. Earlier versions of the article were presented at the 2014 American Political Science Association Annual Meeting and the Research Workshop in International Relations at Harvard University. I gratefully acknowledge financial support from the Global Economic Governance Programme at the University of Oxford and the Institute for Quantitative Social Science and Weatherhead Center for International Affairs at Harvard University.

Introduction

Do international organizations (IOs) do what they are meant to do? IOs have long been a central focus of scholarship in the field of international relations (IR), yet we know surprisingly little about their performance – that is, the extent to which they achieve their stated objectives and do so in a manner that is cost-effective and responsive to a wide range of (public and private) stakeholders. The neglect of performance issues in IR can be attributed in part to challenges of conceptualization and measurement and in part to the field’s theoretical orientations.¹ The traditional functionalist approach to understanding the existence and role of international institutions views them as “efficient” solutions to collective action problems among states stemming from asymmetric information, transaction costs, moral hazard, and other sources. The implicit assumption is that institutions will succeed in providing the benefits desired by their creators. If they failed to provide such benefits, the approach implies, there would be little reason to establish them in the first place.

While generally accepted by IR scholars, the functionalist view is belied by a growing body of empirical evidence indicating that there is considerable *variation* in the performance of IOs. Most notably, a recent wave of official government evaluations of IOs – widely viewed as the “gold standard” of institutional assessment at the international level – has found significant inter-organizational differences on every dimension of performance.² These evaluations are based on a diverse range of quantitative and qualitative data sources and employ the same methodology as assessments of local and city government performance that are widely used in the public administration literature.³ Figure 1 provides a graphical illustration of differences in the performance of 50 IOs based on numerical ratings from the two most widely cited evaluations, namely, those conducted by the British Department for

¹ Only a handful of IR studies, discussed in the next section, have focused explicitly on performance issues. The dearth of such scholarship in the pages of *International Organization* is striking, given the journal’s original aim of promoting “a comparative study of international organizations and why they have or have not worked in varying circumstances” (Katzenstein, Keohane, and Krasner 1998, 650).

² Department for International Development (United Kingdom) 2011; Ministry for Foreign Affairs (Sweden) 2011; Australian Agency for International Aid 2012; Ministry of Foreign Affairs (Denmark) 2012; and Ministry of Foreign Affairs (Netherlands) 2013. In interviews with more than 150 officials from IOs in Washington D.C., Geneva, and Rome between May 2012 and March 2015, the evaluations were often cited as the most reliable and comprehensive source of information on IO performance.

³ The data sources and methodology are discussed in more detail in the fourth section. Notable domestic evaluations include the Government Performance Project, which produces scorecards of city government performance in the United States, and the Comprehensive Performance Assessment, which rates local government performance in the United Kingdom.

International Department (DFID) and the Australian Agency for International Development (AusAID).⁴ The axes correspond to the three dimensions of IO performance highlighted above: the achievement of stated objectives (*z*-axis), cost-effectiveness (*y*-axis), and responsiveness to a wide range of stakeholders (*x*-axis).

<< **Figure 1 around here** >>

Anecdotal evidence also attests to substantial variation in IO performance. The high-performing IOs in Figure 1, for instance, have made concrete achievements in recent years, improving the welfare of millions of people across the world. The World Food Programme (WFP) has delivered life-saving food aid to almost 100 million people in more than 70 countries every year for the past decade.⁵ The World Bank provided 117 million people with access to health services and 123 million people with access to improved water sources between 2002 and 2013.⁶ The GAVI Alliance has helped to immunize 500 million children against vaccine-preventable diseases since 2000, averting more than seven million deaths.⁷ The low-performing IOs, by contrast, can point to few such accomplishments. The International Labour Organization (ILO) has increasingly struggled to fulfill its mandate of strengthening global labor standards, with ratification rates for most of its conventions declining to less than 20 percent of member states.⁸ The United Nations Educational, Scientific, and Cultural Organization (UNESCO) has seen core functions in the areas of education and research migrate to other IOs due to its failure to achieve results on the ground.

Such variation is puzzling – and not only from the perspective of the functionalist approach. A cursory glance at Figure 1 reveals that IOs with similar memberships, mandates, organizational structures, and staff levels often perform at very different levels. The WFP and the Food and Agricultural Organization (FAO), for instance, both have a universal membership comprising almost every state; a mandate to

⁴ The figure displays an average of the two sets of ratings (which, as discussed later, are highly correlated) for overlapping IOs. A full list of the IOs is provided in Table 2.

⁵ World Food Programme annual reports 2004-13, available at <<http://www.wfp.org/policy-resources>>, accessed 15 February 2015.

⁶ See <<http://www.worldbank.org/ida/results-at-a-glance.html>>, accessed 22 December 2014.

⁷ See <<http://www.gavi.org/About/Mission/Facts-and-figures>>, accessed 22 January 2015.

⁸ International Labour Organization 2000.

promote global food security; an executive body composed of representatives of around a quarter of members elected for three-year terms; and a full-time staff of around 4,000. Yet whereas the WFP is one of the best-performing IOs in Figure 1 (with the eighth highest average rating across the three dimensions), the FAO is one of the worst (with the fifth lowest). Equally puzzling, neither high- nor low-performing IOs appear to have much in common with each other. Alongside the WFP at the top end of the spectrum, for example, is the European Bank for Reconstruction and Development (EBRD), a regional IO with 65 member states, 1,600 staff, and a nongovernmental executive body that provides development assistance in Central and Eastern Europe; and the Private Infrastructure Development Group (PIDG), an IO with just nine member states, less than 50 staff, and no separate executive body that mobilizes private investment for infrastructure projects in developing countries.

These puzzling differences in performance have assumed increased salience with the sharp expansion in the number, scope, and resources of IOs in recent decades. Between 1970 and 2005, the number of intergovernmental organizations (IGOs) increased from 199 to 354, while the number of international nongovernmental organizations (INGOs) rose from 1,993 to 7,306.⁹ During this period, IOs have branched out into diverse issue areas such as environmental protection, finance, and women's rights, supplanting traditional state-based modes of governance in unprecedented ways. At the same time, they have amassed more funds and manpower than ever before. For instance, the 50 IOs displayed in Figure 1 – a fraction of the wider universe of IOs – employ almost 110,000 full-time staff, spend more than \$50 billion per year, and possess assets worth \$1.3 trillion. As the achievements cited earlier suggest, these resources give IOs the potential to enhance the welfare of large numbers of people across the globe. What remains theoretically and empirically unclear, however, is why only *some* institutions succeed in realizing this potential.

This article seeks to provide an explanation for such variation. I argue that the primary obstacle to effective performance is not deviant behavior by IO officials – as suggested by “rogue-agency” analyses – but the propensity of *states* to use IOs to advance narrow national interests rather than broader organizational objectives. In

⁹ Pevehouse, Nordstrom, and Warnke 2004; and Union of International Associations 2015. The IGO figures only include formal organizations that have a permanent secretariat and were established by a formal treaty. For a partial overview of informal IGOs, see Vabulas and Snidal 2013.

other words, states are subject to a time-inconsistency problem in which incentives to pursue collective interests before the creation of IOs are replaced by incentives to pursue national interests afterward. The implication is that IOs that enjoy policy autonomy vis-à-vis states will exhibit higher levels of performance. Critically, however, I argue that in the international context policy autonomy cannot be guaranteed by institutional design. Consequently, only *de facto* – and not merely *de jure* – policy autonomy will result in effective performance. I highlight two previously overlooked sources of *de facto* policy autonomy: (1) the existence of (certain types of) institutionalized alliances between IOs and actors above and below the state; and (2) a high degree of technical complexity in IO activities.

To test the argument, I construct and analyze the first quantitative dataset on IO performance, which covers 53 IOs and is based on ratings from five of the official government assessments mentioned earlier. Data on my explanatory variables are drawn from a variety of original sources, most notably an online survey of officials from all 53 IOs. The quantitative analysis is followed by a short comparative case study of two IOs that have similar characteristics but widely varying levels of performance: the FAO and the WFP. This examination, which draws on archival data and extensive interviews with IO officials, complements the statistical tests by (among other things) providing evidence for my argument's key causal mechanisms and shedding light on patterns in performance over time as well as across IOs.

Existing Theoretical Approaches

The largest and most advanced literature on the subject of institutional performance focuses not on IOs but on national and subnational governments. In the past two decades, political scientists, economists, sociologists, and in particular public administration scholars have made considerable progress in measuring and operationalizing institutional performance at the domestic level. I later show how methodological insights from this literature can be fruitfully applied to the assessment of IO performance, enabling scholars to overcome some of the challenges that have plagued empirical research on the topic. On the explanatory front, theories of domestic institutional performance have generally focused on how characteristics of national and subnational populations (such as size, cohesion, and diversity) affect whether citizens mobilize to demand essential public services and whether authorities

are responsive to those demands.¹⁰ They thus have only limited relevance to IOs, which are insulated from direct popular control and are not formally accountable to citizens in the same way as national or subnational governments.

What other factors might influence the performance of IOs? The IR literature has devoted surprisingly little attention to this issue. The traditional theoretical paradigms in the field, while not directly addressing it, imply that IOs will be successful in fulfilling their mandates. Realism views international institutions as epiphenomenal, merely reflecting the interests of powerful nations and not independently shaping state behavior.¹¹ Insofar as they succeed in advancing these interests – which, realists suggest, is invariably the case – they can be seen to perform in an “effective” manner. Neoliberal institutionalism is more optimistic about the ability of institutions to independently influence behavior, highlighting their role in solving collective action problems among states by reducing transaction costs, providing information, and establishing a form of legal liability.¹² Yet implicit in this functionalist analysis is essentially the same assumption about performance: institutions *do* provide the benefits desired by their creators. As Stone points out, this assumption also underlies the more recent IR literatures on institutional design, legalization, and delegation, which, in the same functionalist tradition, explain institutions and their characteristics as rational responses to problems of international cooperation.¹³

Only a small number of studies have questioned this assumption. Drawing on public choice theory, Vaubel and Dreher highlight the performance problems caused by rent-seeking behavior on the part of IO officials.¹⁴ This perspective views states as a collective “principal” who strategically delegates authority to an IO “agent” to pursue a set of common objectives. Due to the costliness of monitoring and writing complete contracts, it suggests, IOs are able to acquire sufficient autonomy to advance their

¹⁰ Public administration scholars have focused on population size, density, and income. See Andrews et al. 2005; Boyne et al. 2006; and Ashworth, Boyne, and Entwistle 2010. Political scientists and sociologists have paid more attention to levels of social capital. See Putnam 1993; Cusack 1999; and Knack 2002. Economists have been primarily interested in ethnic heterogeneity. See Alesina, Baqir, and Easterly 1999; La Porta et al. 1999; and Miguel and Gugerty 2005.

¹¹ Grieco 1988; and Mearsheimer 1994.

¹² Keohane 1984.

¹³ Stone 2011, 22-24. Key works in these literatures include Abbott and Snidal 2000; Koremenos, Lipson, and Snidal 2001; and Hawkins et al. 2006.

¹⁴ Vaubel 1986; and Dreher and Vaubel 2004. Along similar lines, Vreeland 2003 argues that the International Monetary Fund (IMF) colludes with governments to push through unpopular redistributive policies that enhance its resources and influence.

own goals – such as maximizing their budget and policy influence – at the expense of state interests. In other words, IOs face a moral hazard problem. The performance of IOs has also been criticized from a constructivist perspective by Barnett and Finnemore, who argue that such institutions use their rational-legal authority and control of information to carve out autonomy from their creators, resulting in “behavior that undermines [their] stated objectives.”¹⁵ Barnett and Finnemore are particularly interested in a form of dysfunctional behavior that they call “pathology,” which arises when autonomous IOs develop bureaucratic cultures that foster perverse tendencies such as the treatment of rules as ends in themselves and the adoption of incoherent worldviews that give rise to contradictory policies and mission creep. While rooted in different methodological approaches, these “rogue-agency” analyses have a key common implication, namely, that IOs that enjoy a high level of autonomy are likely to perform worse than IOs that are tightly controlled by states.¹⁶

By drawing attention to the possibility of institutional failure, these studies have significantly advanced our theoretical understanding of IO performance. In terms of explaining variation in performance, however, they suffer from two shortcomings. First, they have a notable theoretical blind spot: they fail to explain why only *some* IOs are able to acquire sufficient autonomy to engage in deviant behavior that undermines performance.¹⁷ Second, on the empirical front, they have not tested the proposition that autonomy is negatively related to performance with systematic cross-organizational data. Rather, the claim has only been examined through (qualitative and quantitative) studies of individual IOs. Even this limited evidence does not provide consistent support for it. Indeed, a recent collection of in-depth case studies conducted as a part of a symposium on IO performance found that, in the words of the latter’s editors, “[IO] bureaucracies will perform their functions...better if they are given some degree of autonomy.”¹⁸ In the next section, I develop an argument that

¹⁵ Barnett and Finnemore 1999, 716.

¹⁶ Not all applications of principal-agent theory to IOs imply that autonomy is negatively related to performance. The more typical claim is simply that autonomy may lead to outcomes that states do not anticipate or desire, which is consistent with either a negative or a positive relationship.

¹⁷ A similar criticism of Barnett and Finnemore’s analysis is made by Gutner and Thompson 2010, 229.

¹⁸ Gutner and Thompson 2010, 243. The symposium discusses several other factors that may influence IO performance – including leadership, mandate coherence, and levels of staffing and resources – in addition to helpfully highlighting some of the methodological challenges of investigating the topic (discussed later). Some of these challenges and factors are also explored in the small literature on the effectiveness of international environmental regimes. See Young 1999; and Miles et al. 2002.

both explains this apparently anomalous finding and sheds light on differences in autonomy across IOs.

The Argument

National Interests, Time Inconsistency, and the Principal's Moral Hazard

I accept a basic premise of the public choice perspective, namely, that IOs are the product of a rational and strategic effort by states to further their common interests by delegating authority to another actor – an actor with its own interests and goals. In contrast to this perspective, however, I argue that the primary obstacle to effective institutional performance is opportunistic behavior by *states themselves* rather than IO officials. In other words, it is the principal – not the agent – who experiences a moral hazard problem. This seemingly paradoxical claim follows from two insights about the structure of incentives in international delegation.

First, while IO officials possess incentives to pursue their own goals, acting on such incentives need not come at the expense of realizing organizational objectives. Officials seeking to maximize their budget, for instance, are likely to enjoy *more* success if their institutions are seen to be performing effectively by states (and other potential donors). The same is true of officials seeking to maximize their policy influence: IOs typically have to rely on moral suasion to secure compliance by states, and since they are not accountable to citizens in the same way as national governments (as noted above), their surest path to legitimacy is to “effectively perform the functions invoked to justify [their] existence.”¹⁹ In addition, there is evidence from surveys and in-depth ethnographic studies that the goals of officials often *encompass* the stated aims of their IOs.²⁰ This may be because individuals who identify with these aims self-select into positions of employment or because officials tend to develop an attachment to their work through rewarding professional experience or the internalization of cultural norms within the organization. Either way, the very identity of officials – beyond any material incentives – may cause them to place a high priority on institutional success.

¹⁹ Buchanan and Keohane 2006, 422.

²⁰ See Ascher 1983; Barnett 1997; and Mathiason 2007.

The second insight is that states often possess incentives to behave in ways that compromise the achievement of organizational objectives. The public choice perspective implicitly assumes that since states establish IOs, their interests will consistently be aligned with such objectives. This assumption follows traditional principal-agent models of legislative delegation to bureaucratic agencies in the American politics literature, which imply that only agents (i.e., bureaucrats) can behave contrary to agreed-upon goals – that is, engage in moral hazard.²¹ Subsequent studies of bureaucratic delegation, however, have challenged these models by arguing that principals (i.e., politicians) often have stronger incentives to engage in moral hazard than agents. In influential work on the politics of structural choice, Moe has highlighted the tendency of politicians to take actions that impede the efficiency of bureaucratic agencies due (in part) to fears that political turnover will enable opponents to use such institutions to pursue different objectives in the future.²² More recently, building on the economics literature on time inconsistency in monetary policy, Miller and Whitford have argued that changes in political conditions can cause the *same* politician to favor different bureaucratic policies at different points in time.²³

Do analogous situations arise in the international context? As emphasized by recent work on informal governance in IOs, states are frequently prepared to interfere with IOs in ways that jeopardize institutional performance in order to safeguard core foreign policy interests.²⁴ The implication is that the delegation of authority to IOs is characterized by a similar kind of time-inconsistency problem to that highlighted by Miller and Whitford. Before the creation of IOs (period t), states have incentives to identify mutual interests and to pursue such interests by means of institutionalized cooperation. Each nation stands to gain from the centralized provision of collective goods and expects the benefits of joining an IO to exceed the costs. After IOs have been established (period $t + 1$), however, circumstances often change in unanticipated ways – due to developments both within and outside such institutions – that incentivize states to use them to advance narrow national interests rather than the broader common interests that originally motivated cooperation. In the absence of any countervailing forces, this dynamic shift in incentives leads to a suboptimal

²¹ See, for instance, Weingast and Moran 1983; and Weingast 1984.

²² Moe 1989 and 1990.

²³ Miller and Whitford 2007 and 2016.

²⁴ Stone 2011. The possibility that state interference can undermine performance is briefly mentioned by Gutner and Thompson 2010, 237.

equilibrium in which IOs – torn between national and collective interests – fail to realize organizational objectives, deploy resources inefficiently, and are unable to respond to the demands of wide range of stakeholders.

De Jure and De Facto Policy Autonomy

How can ineffective institutional performance be avoided? A central finding of the economics literature on time inconsistency in monetary policy is that actors can bind themselves to policies that they favor in period t but not period $t + 1$ by delegating policymaking authority to an *independent* agent whose preferences are aligned with the desired policy outcome in both periods.²⁵ This suggests that states can solve the time-inconsistency problem in international delegation by designing IOs to possess a high level of policy autonomy. Critically, however, states always retain the capacity to renege on previous commitments to delegate policymaking functions to IOs.²⁶

Unlike at the national level, there is, of course, no centralized mechanism for enforcing formal institutional rules protecting the agent against interference by the principal. Nor, due to collective action problems among states, is any effective system of decentralized voluntary enforcement likely to arise.

Circumventing formal safeguards on policy autonomy in order to protect national interests, however, is not costless for states. In addition to undermining the effectiveness and legitimacy of IOs and thus reducing the benefits they derive from the advancement of collective interests, it can cause them significant reputational damage (which may spread to other issue areas).²⁷ At the design stage, therefore, states will seek to strike a balance between giving IOs sufficient policy autonomy to effectively pursue organizational objectives and retaining sufficient control to protect national interests without having to contravene formal rules. As suggested above, however, while states may be able to foresee the negative consequences of rule violations, they cannot fully anticipate the risks or potential costs of future threats to national interests in any given issue area. That is, they operate in what contract theorists describe as an environment of “procedural incompleteness.”²⁸ As a result,

²⁵ The delegation solution was first proposed by Rogoff 1985.

²⁶ Stone 2011, 29-30.

²⁷ Ibid, 16-19.

²⁸ See Grossman and Hart 1986; Hart and Moore 1990; and Hart 1995. As noted earlier, this assumption also underpins principal-agent analyses of IOs, which emphasize the incomplete nature of the contract between such institutions and states.

they will often misjudge the optimal balance between policy autonomy and control and subsequently be forced to breach formal rules on a regular basis. Incentives to defect are strengthened by the general difficulty of altering formal rules in IOs (or “recontracting”) due to the typical supermajority vote requirement for adopting rule amendments as well as the high transaction costs and collective action problems associated with reaching agreement among large and diverse set of actors.²⁹

The emergence of informal policymaking practices that are at odds with formal rules entails that there will only be a weak correlation between an IO’s *de jure* policy autonomy, or the policy autonomy that it is designed to possess by its creators, and its *de facto* policy autonomy, or the policy autonomy that it possesses in practice.

Nevertheless, this correlation is still likely to be positive both because the costs of breaching formal rules increase with the size of the gap between *de jure* and *de facto* policy autonomy and because states will sometimes (largely by chance) succeed in designing IOs with the ideal balance between autonomy and control and thus avoid having to contravene formal rules in period $t + 1$. By providing a solution to the time-inconsistency problem in international delegation, *de facto* policy autonomy can naturally be expected to have a strong positive association with institutional performance. Since *de jure* policy autonomy has a weak positive association with *de facto* policy autonomy, however, we should expect it to have a similar relationship with performance.³⁰ In other words, in the international context institutional design will not be sufficient to guarantee policy autonomy and thus effective performance.

What are the specific characteristics of *de jure* and *de facto* policy autonomy? Policy autonomy can be defined more precisely as the ability of IOs to decide which mandate-related problems to focus on and what substantive measures they take to address such problems in the absence of interference from states.³¹ This definition suggests that policy autonomy has three characteristics. An IO can be classified as

²⁹ Koremenos, Lipson, and Snidal 2001, 794.

³⁰ In this respect, my argument differs from Stone’s theory, which implies that both relationships will be relatively strong because powerful states allow formal rules to prevail in most circumstances in exchange for the right to override such rules when their interests are intense. This logic makes sense in the context of IOs such as the IMF – Stone’s principal focus – where policy decisions typically concern the allocation of institutional resources to individual members and thus rarely have a direct impact on powerful states. In most of the IOs in my dataset, however, policies usually affect the entire membership, incentivizing powerful states to breach formal rules with a *high* level of frequency.

³¹ For similar definitions, see Abbott and Snidal 1998; Barnett and Finnemore 1999; and Haftel and Thompson 2006. Stone has a more expansive conception of IO autonomy that encompasses all executive, legislative, and judicial functions that can be delegated to such institutions.

possessing a high level of *de jure* policy autonomy if it is designed to possess these three characteristics – a function of its formal rules – and a high level of *de facto* policy autonomy if it actually exhibits them in practice.

The first characteristic is the ability of IO officials to *set the policy agenda*, which has three specific components: (1) the ability to propose new programs, projects, regulations, and other policies; (2) the ability to draft the annual budget; and (3) the ability to prepare the work program for IO governing bodies.³² By exercising these abilities, IO officials can circumscribe the range of choices available to states and thus shift policy outcomes away from national interests (i.e., the status quo) and toward organizational objectives (i.e., their ideal point).³³ The second characteristic is the ability of individual states to *veto policy decisions*, which is a function of decision-making procedures in IO governing bodies. In governing bodies that employ majority voting – whether absolute or qualified – it is generally not possible for any one state to block decisions. In governing bodies that take decisions by consensus, by contrast, every state effectively possesses a veto.³⁴ The third characteristic is the access of IOs to *nonstate sources of financing*. When states enjoy a monopoly on IO funding, they have the potential to prevent the adoption and implementation of policies that clash with national interests. By contrast, when IOs can draw on alternative sources of income, such as contributions from nonstate actors and independently earned revenue, they will often be able to pursue such policies even if states restrict their funding.

The argument thus far can be summarized in the following three hypotheses:

H1: An IO's de facto policy autonomy – which is higher when officials set the policy agenda, governing body decisions are made by majority voting, and states do not monopolize funding – is positively and strongly related to its performance.

H2: An IO's de jure policy autonomy – which is higher when its formal rules stipulate that officials set the policy agenda, governing body decisions are made by majority

³² See Hooghe et al. forthcoming.

³³ See Johnson 2014.

³⁴ Koremenos, Lipson, and Snidal 2001; Haftel and Thompson 2006; and Hooghe et al. forthcoming. Note that policy autonomy also increases with the degree of preference divergence among governing body members, though only when majority voting is employed (since consensus-based decision making allows “outlier” states to veto decisions even when there is a low degree of preference divergence overall). See Copelovitch 2010.

voting, and states do not monopolize funding – is positively but weakly related to its de facto policy autonomy.

H3: An IO's de jure policy autonomy is positively but weakly related to its performance.

Sources of De Facto Policy Autonomy

These hypotheses naturally raise the question: Where does *de facto* policy autonomy come from, if not institutional design? While the factors affecting *de jure* policy autonomy have received considerable attention from IR scholars, the sources of *de facto* policy autonomy have yet to be systematically investigated.³⁵ I highlight two hitherto neglected determinants of *de facto* policy autonomy: (1) the existence of (certain types of) institutionalized alliances between IOs and actors above and below the state; and (2) the technical complexity of IO activities.

Institutionalized alliances. Institutionalized alliances arise when IOs enlist the voluntary assistance of subnational and supranational actors – including NGOs, businesses, public-private partnerships (PPPs), and IGOs – who share their goals and possess complementary material, informational, organizational, and other resources. Such collaboration is mutually beneficial: in return for their assistance in developing and implementing policies, IOs provide partners with access to contacts and networks, normative guidance, and the legitimacy that comes with endorsement by a multilateral body.³⁶ While recent scholarship has highlighted the importance of institutionalized alliances in filling “gaps” in the operational capacities of IOs, it has paid less attention to their crucial role in protecting the latter against state interference.³⁷ This oversight is perhaps unsurprising, as the notion that an institution’s autonomy can be *enhanced* by links with external stakeholders runs counter to both the pluralist literature on the state, which views societal influence as a hindrance to governmental independence, and the American politics literature on bureaucratic delegation, which treats interest

³⁵ See Haftel and Thompson 2006; Johnson 2014; and Hooghe et al. forthcoming. As discussed earlier, Stone’s work on informal governance is a notable exception.

³⁶ Abbott and Snidal 1998.

³⁷ Abbott et al. 2015.

group involvement in agency rulemaking as a design choice intended to facilitate the achievement of delegated goals – a logic that has recently been applied to IOs.³⁸

Why do institutionalized alliances reinforce rather than undermine IO policy autonomy? Since partners derive benefits from collaboration and have similar goals to IOs – a key reason they are enlisted in the first place – they possess incentives to assist IOs in addressing not only the *practical* challenge of policy formulation and implementation but also the *political* challenge of state interference. Specifically, partners can protect IO policy autonomy in two ways. First, and most importantly, they can alter the payoff to states of pursuing national interests. This can be achieved via several means, including lobbying policymakers at the domestic level; drawing public attention to the adverse consequences of institutional failure and thus raising the “audience costs” of interference; and neutralizing state pressure at the international level by mobilizing counter-coalitions of other subnational and supranational actors and supporting them with information, expertise, and logistical assistance. By pursuing these strategies, partners can persuade states both to allow IO officials to set the policy agenda and to employ majoritarian decision-making procedures in IO governing bodies. Second, partners can provide material assistance to IOs – in the form of direct contributions as well as payments for their products and services – and thus prevent states from monopolizing funding.

Yet institutionalized alliances do not always lead to high levels of *de facto* policy autonomy. Rather, the willingness and ability of partners to defend IOs against state interference vary as a function of three factors. The first is the depth of cooperation between IOs and partners. While most IOs can list a host of partnerships, these are often symbolic arrangements that pay lip service to norms of stakeholder engagement while providing partners with a modicum of legitimacy. Only when alliances involve a meaningful exchange of resources or services will partners possess incentives to safeguard policy autonomy. The second factor is the alignment between partner and IO policy preferences. While IOs will not enlist actors with sharply conflicting preferences, some partners benefit more from the achievement of organizational goals than others and thus have stronger incentives to prevent state interference. Preference misalignment is particularly common when there are differences in the geographical

³⁸ See Dahl 1982; McCubbins and Schwartz 1984; and Tallberg et al. 2013.

focus and issue area of IOs and partners. The third factor is the complementarity between IO and partner capabilities. Partners with ample material and nonmaterial resources are naturally more effective in protecting policy autonomy. Wealthy partners, for instance, can hire more lobbyists to influence policymakers at the domestic level; launch more extensive publicity campaigns about the harmful effects of state interference; provide more funds to counter-coalitions of nonstate actors; and make more generous donations to IOs themselves. Importantly, however, since the gains from pursuing national interests – and thus the costs of deterring interference – increase with an IO's own level of resources, complementarity depends not on the *absolute* level of partner resources but on the *ratio* of partner to IO resources.

Technical complexity. The second determinant of *de facto* policy autonomy is the technical complexity of IO activities, which I define in terms of the length and specificity of the training needed to develop policies. In highly technical domains, policymakers are typically required to spend several years acquiring appropriate formal (graduate-level) qualifications and professional experience. Only a narrow range of formal certification and employment roles are regarded as signaling possession of the specialized knowledge needed to participate in the policymaking process. Conversely, in nontechnical domains little specialized training (formal or informal) is needed to formulate policies. Policymakers can thus be recruited from a wide variety of professional, educational, and institutional backgrounds.

Technical complexity protects IOs against state interference by generating information asymmetries between IO officials and states, which, as implied by principal-agent theory, reduces the latter's control over the former. Specifically, such asymmetries have two key consequences. First, they make it difficult for states to *propose* new policies, forcing them to cede agenda-setting powers to officials. Second, they prevent states from fully grasping the distributional implications of policies and thus effectively *opposing* proposals that conflict with national interests. Uncertainty about distributional consequences also makes states less likely to withhold financial contributions because they cannot anticipate how they will be affected by a given policy. Perhaps even more important in this regard, technical knowledge provides opportunities for IOs to earn independent revenue through the sale of specialized products and publications and the provision of expert advice.

It is worth noting the contrast between this analysis and the public choice perspective. According to the latter, information asymmetries are a source of weak institutional performance because they reduce the capacity of state principals to monitor and control IO agents. My analysis agrees that such asymmetries enhance the autonomy of IOs but posits that the consequences for performance are *positive*. This difference, of course, stems from the key insight that in the context of international delegation it is the principal – not the agent – who suffers from a moral hazard problem.

In sum, I propose two hypotheses regarding the determinants of *de facto* policy autonomy:

H4: An IO's ability to forge institutionalized alliances characterized by deep cooperation, aligned policy preferences, and complementary capabilities is positively and strongly related to its de facto policy autonomy.

H5: The technical complexity of an IO's activities is positively and strongly related to its de facto policy autonomy.

Data and Statistical Analysis

In this section, I subject my hypotheses to rigorous empirical tests based on the first quantitative dataset on IO performance. The first part of the section introduces the dataset, describing in detail the operationalization and measurement of IO performance as well as the explanatory and control variables. The second part presents the results of a series of regression analyses in which the dependent variable is first IO performance and then *de facto* policy autonomy.

Operationalizing and Measuring IO Performance

As defined in the introduction, the concept of IO performance has three distinct dimensions: (1) the achievement of stated objectives; (2) cost-effectiveness; and (3) responsiveness to a wide range of (public and private) stakeholders. This definition bridges two traditions of conceptualizing institutional performance: the organizational theory tradition, which emphasizes goal attainment, and the public administration tradition, which emphasizes efficiency and responsiveness.³⁹ Each tradition captures

³⁹ On the organizational theory tradition, see Etzioni 1964; and Price 1972. On the public administration tradition, see Cameron 1978; Boyne 2002; and Boyne et al. 2006.

an essential part of the concept. Institutions that achieve their objectives but are wasteful with resources and ignore the concerns of key stakeholders cannot be seen to be performing at a high level. Nor, conversely, can institutions that use resources efficiently and are sensitive to the interests of diverse stakeholders but make little progress in realizing their goals. Goal attainment, cost-effectiveness, and responsiveness to a wide range of stakeholders are thus all integral elements of a plausible and comprehensive definition of IO performance.

In seeking to measure and operationalize the concept of IO performance, we confront three challenges. The first is what Gutner and Thompson refer to as the “eye of the beholder” problem: the performance of IOs is evaluated by a diverse array of stakeholders – including governments, civil society groups, and IO officials – whose assessments may vary significantly.⁴⁰ To deal with this problem, any measure of IO performance must meet two requirements. First, since there are no grounds for privileging any one set of stakeholders, it should incorporate the views of a variety of groups. Second, it should supplement these views with more “objective” sources of evidence, such as quantitative data. I call these the requirements of *inclusiveness* and *objectivity*, respectively. The more consistent the views of stakeholders with each other and with less subjective evidence, of course, the more confident we can be that our measure represents an accurate assessment of IO performance.

The second challenge, which has been much discussed in the public administration literature, is that institutional performance is a complex, multidimensional concept that cannot be fully captured by any one indicator alone.⁴¹ To deal with this challenge, public administration scholars argue, we must develop measures of the concept that reflect all of its theoretically relevant dimensions. A valid measure of IO performance must therefore encompass all three dimensions discussed earlier, ideally by combining indicators of each dimension into an aggregate index. In other words, it must meet the requirement of *multidimensionality*.

The third challenge, which follows naturally from the second, is that the concept of IO performance may not be coherent from an empirical perspective. It is conceivable that IOs simply excel on different dimensions, with some being the most effective at

⁴⁰ Gutner and Thompson 2010, 233.

⁴¹ See Cameron 1978; Quinn and Rohrbaugh 1981; and Boyne 2002.

realizing their stated objectives, others at operating efficiently, and still others at addressing the concerns of diverse stakeholders. In this situation, it would make little sense to treat IO performance as a single latent concept; instead, each dimension should be treated a distinct variable. This suggests a fourth requirement for a valid measure of IO performance, namely, *coherence*: its constituent indicators must have a strong positive correlation with each other.⁴²

<< **Table 1 around here** >>

Data from the recent wave of official government assessments mentioned in the introduction enable us to develop measures of IO performance that satisfy the requirements of inclusiveness, objectivity, multidimensionality, and coherence. My main measures are based on performance ratings from the assessments conducted by DFID and AusAID, which fulfill the four requirements to the greatest extent and have the widest coverage of IOs. However, I also employ data from assessments carried out by the governments of Denmark, the Netherlands, and Sweden to test for possible national bias in the DFID and AusAID ratings and to check the robustness of statistical results based on the latter.⁴³ Table 1 provides more detailed information on the five assessments and the extent to which they fulfill the four requirements.

The assessments cover a total of 53 IOs, a list of which can be found in Table 2.⁴⁴ The IOs span a wide range of issue areas, though more than three-quarters operate in at least one of the following five: economic development, education, the environment, humanitarian aid, and public health. There is also some variation in their geographical scope, with six IOs possessing a purely regional membership (five of which are development banks). Finally, note that in addition to IGOs the dataset includes a small number of PPPs (five) and INGOs (two). These institutions closely resemble IGOs – they are often labeled “quasi-IGOs” – and thus fall within my argument’s analytical

⁴² This is similar to Putnam’s requirement of “internal consistency.” See Putnam 1993, 64.

⁴³ Ministry for Foreign Affairs (Sweden) 2011; Ministry of Foreign Affairs (Denmark) 2012; and Ministry of Foreign Affairs (Netherlands) 2013.

⁴⁴ The DFID assessment includes a small number of institutions that are not IOs (mostly divisions and departments of IOs), which are excluded from the dataset.

scope.⁴⁵ Indeed, states were closely involved in their creation, provide the vast majority of their contributions (84 percent on average in the period 2009-11 inclusive), and are formally represented in almost all of their governing bodies.⁴⁶ Nevertheless, as discussed below, the inclusion of these IOs in the dataset does not affect the strength of the empirical support for the argument.

<< **Table 2 around here** >>

All the assessments are informed by a diverse set of stakeholders and thus satisfy the requirement of inclusiveness. DFID and AusAID solicited the views of governments, civil society, businesses, and IO officials through a multiyear series of consultations, workshops, interviews, and written submissions. In addition, they incorporated data from cross-national stakeholder surveys such as Multilateral Organization Performance Assessment Network (MOPAN) Common Approach and the Survey on Monitoring the Paris Declaration. The Danish, Dutch, and Swedish governments also drew on the MOPAN data but did not directly solicit the views of stakeholders, relying instead on indirect feedback from relevant ministries, overseas missions, and embassies. In terms of objectivity, DFID and AusAID crosschecked stakeholder evaluations against quantitative performance data from multiple sources, including the Quality of Official Development Assistance Assessment (QuODA), the Publish What You Fund Aid Transparency Index (ATI), the Heavily-Indebted Poor Countries Capacity Building Project (HIPC CBP), and the Common Performance Assessment System (COMPAS). The other assessments primarily used external audits, independent evaluations, and IO reporting to validate subjective sources of evidence.

Might the assessments still reflect some “national bias” toward IOs that promote the interests of each government? Importantly, the five sets of ratings are strongly correlated with each other. As shown in Table 1, the average correlation between each set of ratings and the remaining four sets is $r = 0.49$ (seven of the 10 coefficients are

⁴⁵ Author interview with Sarah Boulton, Head of DFID’s International Directors’ Office and Lead Coordinator of the UK Multilateral Aid Review, 29 June 2012, London, United Kingdom.

⁴⁶ Only the governing bodies of the International Committee of the Red Cross (ICRC), the International Federation of the Red Cross (IFRC), and the CGIAR Consortium do not include state representatives.

positive and statistically significant at the 10 percent level). Rather than any indication of national bias, therefore, we observe a high degree of *consensus* among states about which IOs are performing well and which IOs are performing poorly. Nor is there any evidence of a broader “advanced industrialized country” bias in the ratings. The average correlation between each set of ratings and the results of the latest round of MOPAN surveys, which capture the views of governments and stakeholders in developing nations, is $r = 0.44$ (three of the five coefficients are positive and significant).⁴⁷ This suggests that the consensus about the relative performance of IOs extends across both developed and developing countries.

Most of the assessments also satisfy the requirement of multidimensionality. The DFID and AusAID assessments contain six performance indicators: (1) delivery of results; (2) contribution to meeting the international community’s objectives; (3) cost and value consciousness; (4) financial resources management; (5) accountability and transparency; and (6) strategic/performance management.⁴⁸ The first two indicators correspond to the first dimension of IO performance (goal attainment); the third and fourth to the second dimension (cost-effectiveness); and the fifth and sixth to the third dimension (responsiveness to diverse stakeholders). The Dutch assessment includes four performance indicators that correspond to the first, second, fourth, and sixth indicators in the DFID and AusAID assessments. The Swedish assessment contains only two indicators, “internal effectiveness” and “external effectiveness,” the first capturing a combination of cost-effectiveness and responsiveness and the second goal attainment. Finally, Denmark’s assessment contains just one overall performance measure and thus does not satisfy the requirement of multidimensionality.

In the four assessments that do satisfy the requirement, indicators of performance have a strong positive correlation with each other and thus pass the final test of coherence. The average correlation among indicators in the DFID, AusAID, Dutch, and Swedish assessments is, respectively, $r = 0.57, 0.49, 0.50,$ and 0.90 (76 of the 92 individual coefficients are positive and significant at the 10 percent level). In addition, the Cronbach’s alpha among the four sets of indicators is $\alpha = 0.85, 0.79, 0.66,$ and

⁴⁷ The MOPAN surveys contain almost 20 individual indicators that are reduced to an aggregate index using principal component analysis.

⁴⁸ Detailed criteria for each indicator can be found in the assessments. The sixth indicator measures whether IOs have an effective evaluation function and use evidence on past performance – from stakeholders and other sources – as a basis for improvement.

0.88, far exceeding the generally accepted minimum consistency threshold of 0.5. This evidence suggests that in each case the indicators conceal a single latent variable.

I combine each set of indicators into a composite index of IO performance using principal component analysis, a standard technique for reducing dimensionality in multivariate data. Specifically, the index is represented by scores on the first principal component in the analysis, which accounts for the maximum possible variance in the data.⁴⁹ The high proportion of variance explained by the first component in each case – 62 percent on average, compared with 16 percent for the second and 13 percent for the third – indicates that the data can plausibly be reduced to a single variable. Nevertheless, to explore patterns within different dimensions of performance I employ both individual indicators and aggregate indices in the statistical tests below.

Explanatory and Control Variables

Data on the key explanatory variables are drawn from a number of original sources. To gather information on *de facto* policy autonomy, I conducted an online survey of the heads of all 53 IOs in the dataset.⁵⁰ The survey was launched in September 2013 and responses were received from every IO by January 2015. To verify the reliability of the responses, in 20 percent of cases the survey was sent to another senior IO official (usually a division or department head). In no instances were there discrepancies between the two sets of answers, indicating a high degree of reliability.

DE FACTO POLICY AUTONOMY is an index that sums the values of six indicators measuring the three characteristics of the concept described in the third section: IO agenda-setting powers, decision-making procedures, and access to nonstate sources of financing. The first three indicators are based on responses to the following questions, which capture the three specific agenda-setting abilities outlined earlier:

⁴⁹ Table 2 displays an average of the four indices and the Danish ratings for each IO (with all five variables rescaled between 0 and 1).

⁵⁰ The survey, which was implemented using the Qualtrics Survey Software, was sent to participants via a link contained in an email. In most cases, 2-3 reminder messages were sent before the response was submitted. The survey contained 31 questions – 26 multiple choice and five write-in – and the average completion time was 37 minutes (excluding one outlier response that was completed over several days). In five cases, participants preferred to provide their responses either verbally (during a telephone interview) or in written form.

1. *Does the permanent staff propose new policies (for instance, in the form of programs or projects) or issue draft rules for your organization?* (“Yes” = 1; “No” = 0)
2. *Does the permanent staff draft your organization’s annual budget?* (“Yes” = 1; “No” = 0)
3. *Does the permanent staff prepare the agenda for your organization’s governing bodies?* (“Yes” = 1; “No” = 0)

The fourth indicator, measuring decision-making procedures, is based on responses to the question:

4. *How are policy decisions typically made in your organization’s governing body?* (“Majority voting” = 1; “Supermajority voting” = 0.5; “Unanimity” = 0; scores are averaged across all IO governing bodies)

The last two indicators, which are based on data from IO financial statements for the period 2009-11 inclusive, cover access to nonstate funding. The first measures the proportion of an IO’s financial contributions that are received from nonstate actors, while the second measures the proportion of an IO’s total income that is independently earned from investments, interest on loans, sales, fees, and other sources (both are rescaled between 0 and 1).⁵¹

DE JURE POLICY AUTONOMY has the same six indicators as DE FACTO POLICY AUTONOMY, with the sole difference that they are based not on survey data but on formal rules set out in an IO’s constitution, charter, treaty, or rules of procedure.⁵² As shown in the fourth and fifth columns of Table 2, the two variables have a positive but modest association, with a correlation coefficient of $r = 0.35$ (which is not statistically significant at the 10 percent level). In almost four-fifths of IOs, the rounded values of the variables diverge; in around 40 percent of these cases, the discrepancy is at least two points (approximately two standard deviations of each variable). This evidence confirms that formal rules alone do not provide the basis for an accurate assessment of an IO’s real level of policy autonomy and thus provides strong support for H2.

⁵¹ While the majority of IOs receive only a small fraction of their contributions from nonstate sources, this proportion exceeds 20 percent for approximately one-quarter of IOs (and one-fifth of IGOs) and 50 percent in one case (the World Health Organization).

⁵² A corollary of this difference is that the fifth and six indicators are no longer proportions but binary variables measuring whether an IO is permitted by its formal rules to receive donations from nonstate actors (fifth indicator) and to independently earn revenue (sixth indicator).

Turning to the other key explanatory variables, INSTITUTIONALIZED ALLIANCES is a summed index of four indicators measuring the three dimensions of variation in such arrangements discussed earlier. All indicators are based on information from official IO websites, most of which have a section devoted to “partnerships.”⁵³ The first indicator measures the depth of cooperation between IOs and partners. IOs receive a score between 0 and 1 based on the proportion of partnerships that involve substantive collaboration in the formulation, monitoring, implementation, or enforcement of policies (as opposed to a purely symbolic affiliation). The second and third indicators measure policy preference alignment. IOs are assigned a score from 0 to 1 reflecting the proportion of (non-symbolic) partners that share their geographical scope (second indicator) and issue area (third indicator).⁵⁴ The fourth indicator measures complementarity in capabilities, taking a value of 1 if the ratio of (non-symbolic) partner to IO expenditures in the period 2009-11 inclusive exceeds 10 and 0 otherwise.

I employ two measures of technical complexity. The first, TECHNICAL COMPLEXITY (SURVEY), is based on responses to the following survey question: “Does the permanent staff conduct independent (systematic empirical or theoretical) research in support of policymaking?” (“Yes” = 1; “No” = 0). The second, TECHNICAL COMPLEXITY (QUALIFICATIONS), is based on officials’ formal qualifications and thus adheres more closely to the definition of the concept provided in the third section. IOs are awarded a score between 0 and 1 based on the proportion of their senior management that possesses a graduate-level educational or professional qualification in a field directly related to their issue area (for instance, a medical doctorate for IOs in the issue area of public health).⁵⁵ This information is obtained from staff biographies on the websites of IOs and former employers as well as résumés posted on professional networking websites (such as LinkedIn).

<< **Table 3 around here** >>

⁵³ IOs that have no listed partnerships are assigned a score of 0 on every indicator.

⁵⁴ I distinguish between three levels of geographical focus – global, regional, and national – and 25 issue areas based on a list constructed by Hooghe et al. forthcoming.

⁵⁵ An IO’s senior management is defined as its head, deputy head, and the directors of its non-administrative departments.

Finally, the dataset includes a small number of control variables, some of which are suggested by the literature on institutional performance reviewed in the second section. NUMBER OF STAFF is the logged number of full-time staff employed by an IO. AGE is the logged number of years since an IO's establishment. FIELD PRESENCE is a dummy variable measuring whether an IO has any offices outside the country in which it is headquartered.⁵⁶ Unless otherwise specified, all explanatory and control variables are measured as of the end of 2011. Summary statistics on all variables in the dataset are displayed in Table 3.

Statistical Analysis

<< **Table 4 and Figure 2 around here** >>

I begin by estimating a series of ordinary least squares regressions with heteroskedasticity-robust standard errors. In Tables 4 and 5, the dependent variable is IO performance, measured by the four indices of government ratings in Models 1-4, the Danish scores in Model 5, and an average of the previous five measures – rescaled between 0 and 1, with missing values filled in using multiple imputation – in Model 6.⁵⁷ In Table 4, the key explanatory variable is DE FACTO POLICY AUTONOMY. In line with H1, the coefficient on this variable is positive and statistically significant at the one percent level in five of the six models – including, crucially, Models 1 and 2, in which the DFID and AusAID indices are the respective dependent variables. This relationship is visualized in Figure 2, which displays a scatterplot of DE FACTO POLICY AUTONOMY against the average measure of IO performance with the estimated regression line from Model 6 (data points represent averages across the imputed datasets). In Table 5, the main explanatory variable is DE JURE POLICY

⁵⁶ Data on the control variables come from IO websites, annual reports, and brochures as well as external databases such as the *Yearbook of International Organizations*.

⁵⁷ I use Honaker, King, and Blackwell's *Amelia II* program in R to implement multiple imputation, creating 10 complete datasets. The imputation model contains every variable in the dataset and a ridge prior of 0.5 percent of the number of observations.

AUTONOMY. Consistent with H3, the coefficient on this variable is nonsignificant in all six models and positive in five (including the DFID and AusAID models).

<< **Table 5 around here** >>

To investigate patterns within different dimensions of IO performance, I disaggregate the four indices of government ratings into their 18 constituent indicators (while maintaining the same two sets of explanatory variables). The key results are summarized in Table 6, which reports the sign and significance level of the coefficients on DE FACTO POLICY AUTONOMY and DE JURE POLICY AUTONOMY in the 36 analyses (full regression results are shown in Tables A1-A8 in the online appendix). H1 and H3 receive consistent support across all three dimensions. The coefficient on DE FACTO POLICY AUTONOMY is positive in all 18 models and significant at the five percent level in 16. Importantly, it is both positive and significant in 11 of the 12 models based on the DFID and AusAID indicators, which measure the three dimensions most directly and precisely. The coefficient on DE JURE POLICY AUTONOMY, meanwhile, is positive in 13 of the 18 models and nonsignificant in 16. Similarly to above, it is both positive and nonsignificant in 9 of the 12 DFID and AusAID models.

<< **Table 6 around here** >>

Finally, in Table 7 the dependent variable is DE FACTO POLICY AUTONOMY. As predicted by H4 and H5, the coefficients on INSTITUTIONALIZED ALLIANCES, TECHNICAL COMPLEXITY (SURVEY), and TECHNICAL COMPLEXITY (QUALIFICATIONS) are all large, positive, and significant at the five percent level.⁵⁸ By contrast, in accordance with H2, the coefficient on DE JURE POLICY AUTONOMY is positive but nonsignificant in both models. Thus, strikingly, the

⁵⁸ Note, however, that the latter two coefficients are markedly larger than the former.

degree of policy autonomy that an IO is designed to possess has relatively little bearing on the degree of policy autonomy that it possesses in practice.

<< **Table 7 around here** >>

Robustness Checks

I subjected the preceding analyses to a variety of robustness checks, the results of which are reported in the online appendix. To correct for possible omitted variable bias, I added a number of control variables to the baseline model in Tables 4 and 5: (1) dummy variables for the five most common issue areas in the dataset (listed above), since some types of tasks may be inherently more challenging than others; (2) a variable measuring the degree of asymmetry in the material capabilities of member/donor states, as IOs may be more susceptible to inference from powerful nations;⁵⁹ and (3) a variable measuring the similarity of member/donor states' voting patterns in the UN General Assembly, on the intuition that homogeneous groups can more easily overcome collective action problems;⁶⁰ and (4) a variable measuring the geographical diversity of member/donor states, another proxy for group heterogeneity.⁶¹ Since all of these variables could conceivably affect *de facto* policy autonomy as well as performance, I also included them in the baseline model in Table 7. As shown in Tables A9-A11, these changes did not alter the key results.

I also examined whether the results were robust to the exclusion of different sets of observations. First, I omitted all INGOs and PPPs from the baseline models in Tables 4 and 5 to address the possibility that such institutions systematically differ from IGOs on any variable of interest (Tables A12-A14). Second, given the small size of the dataset, I tested whether the results were driven by a handful of influential

⁵⁹ This is operationalized as the ratio of the material capabilities of the most powerful state to those of the remaining states. Capabilities are measured using the Correlates of War Project's *Composite Index of National Material Capability* (version 4.0), available at <<http://cow.dss.ucdavis.edu/datasets/national-material-capabilities/national-material-capabilities-v4-0>>, accessed 12 December 2013.

⁶⁰ Data come from Gartzke's *Affinity of Nations Index* (version 4.0), available at <<http://pages.ucsd.edu/~egartzke/datasets.htm>>, accessed 12 December 2013. Index scores are averaged across all dyads of member/donor states.

⁶¹ This is measured as $1 - \sum_{i=1}^m s_i^2$, where s is the share of a given geographical region among states and m is the total number of regions. Regions are defined according to the UN geoscheme.

observations, that is, observations whose exclusion from the analysis makes a significant difference to parameter estimates. Specifically, for each of the baseline models I computed the three most widely used measures of influence in linear regression – Cook’s Distance, DFFITS, and DFBETAS – and excluded all observations that exceeded at least one of their usual cutoff points (Tables A15-A17).⁶² Again, the results were not materially affected by any of these changes.

Another potential issue with the results is endogeneity in Table 3: DE FACTO POLICY AUTONOMY may be positively affected by IO performance, for instance, if successful IOs are “rewarded” by states with more independence or find it easier to attract partners.⁶³ One way to address this issue is to employ an instrumental variables approach, which involves replacing the endogenous variable with an “instrument” that is correlated with it but uncorrelated with the error term (and thus does not directly affect the dependent variable). This approach is well suited to the two-stage structure of my argument, which naturally suggests two instruments: institutionalized alliances and technical complexity. Technical complexity is the more promising candidate since, as suggested earlier, alliances could directly enhance performance by filling gaps in IO capacities.⁶⁴ It is not obvious, by contrast, why technical complexity would have such an effect; indeed, if anything, we should expect it be *negatively* associated with performance. Technical complexity is also a strong candidate because both of its measures are powerful predictors of DE FACTO POLICY AUTONOMY, easily exceeding the standard F-test threshold for a “weak instrument.”⁶⁵ I thus used the measures as separate instruments in a series of two-stage least squares regressions. The results, reported in Tables A18 and A19, are consistent with those in Table 4.

It is important to stress, however, that these results should be treated with caution because the absence of a direct link between technical complexity and IO performance can never be verified empirically. It is conceivable, for instance, that

⁶² See the tables for details on the cutoff points, relevant DFBETAS coefficients, and excluded IOs (which mostly match those that appear to be outliers in Figure 2). Table A11 does not include Model 6 in Table 4 because the three measures cannot be calculated for models based on multiple imputation.

⁶³ It is also possible that INSTITUTIONALIZED ALLIANCES is endogenous to DE FACTO POLICY AUTONOMY in Table 4. If this were the case, however, we would expect a *negative* relationship between the two variables: highly autonomous IOs should be less likely to forge alliances because they do not require further protection against state interference (and are typically already performing well).

⁶⁴ If, as suggested above, high-performing IOs can more easily find partners, alliances would also be directly affected by performance.

⁶⁵ TECHNICAL COMPLEXITY (SURVEY) and TECHNICAL COMPLEXITY (QUALIFICATIONS) have F-statistics of 12.33 and 14.48, respectively, compared with the standard threshold of 10.

progress in performing technically complex tasks is easier to demonstrate in a measurable way. While such a scenario seems unlikely to fully explain the results of the instrumental variables analysis, the possibility of some reverse causation between performance and *de facto* policy autonomy resulting from high-performing IOs being granted more independence or enjoying greater success in attracting partners should not be ignored (and is politically interesting). Indeed, the qualitative evidence provided in the next section suggests that both of these trends *do* occur and that they complement and reinforce the causal mechanisms posited by my argument.

Case Study: A Tale of Two Food Agencies

This section presents a short comparative case study of the FAO and the WFP. This examination contributes to the hypothesis testing by providing three types of evidence that could not be obtained from the quantitative analysis: (1) evidence that the argument's causal mechanisms operate as stipulated; (2) evidence that the argument can shed light on temporal as well as cross-sectional patterns in IO performance; and (3) evidence regarding the direction of causation between variables, in particular performance and *de facto* policy autonomy. At the same time, since the FAO and the WFP are similar in several respects *other* than their values of the dependent variable – including, as noted earlier, their mandate, membership size, organizational structure, and number of staff – the research design retains a key advantage of the quantitative approach, namely, the ability to control for potentially confounding factors.

The case study, which draws on archival data and a large number of interviews with IO officials, is divided into two parts, the first focusing on the FAO and the second on the WFP. In each part, I begin with a detailed comparison of the IO's *de jure* and *de facto* policy autonomy based on the six indicators of each variable outlined in the previous section. I then explore how the IO's degree of *de facto* policy autonomy has shaped and interacted with its performance over time, paying particular attention to its success in resisting state attempts to interfere in the policymaking process to promote narrow national interests – a key causal mechanism highlighted by my argument. Finally, I turn to the sources of the IO's level of *de facto* policy autonomy, examining whether (1) the technical complexity of its activities and (2) the number and type of institutionalized alliances it has formed have influenced this variable in a manner consistent with the argument's predictions and posited causal processes.

The Food and Agriculture Organization

The oldest permanent specialized agency of the UN, the FAO was created in 1945 with the objectives of eradicating hunger, raising nutrition levels, and improving rural living standards. Conscious of the need for a strong autonomous secretariat to effectively carry out this mission – and assuming that serious threats to national interests would not arise in the issue area of food security – states designed the FAO to possess substantial policy discretion. The FAO Constitution requires its Director-General (i.e., chief of staff) to formulate “proposals for appropriate action”; draft its Programme of Work and Budget; and set the agenda for its two governing bodies: the Conference, which meets every two years and includes all 197 member states; and the Council, which meets 1-3 times per year and includes 49 (rotating) nations.⁶⁶

Governing body decisions are to be made by majority voting, with the exception of those concerning constitutional amendments, the admission of new members, the size of the budget, mid-session changes to the agenda, and the adoption of conventions and recommendations, which require a two-thirds majority vote.⁶⁷ While not explicitly authorized to accept donations from nonstate actors, the FAO is permitted to independently earn revenue from selling and licensing its products and services.⁶⁸

In practice, however, the FAO’s policy autonomy is limited. As one senior official remarked, “The rules say one thing, but member states do something else – they dominate every aspect of the policymaking process.”⁶⁹ The FAO’s key policies and programs, including the Programme of Work and Budget, are proposed not by the Director-General but by Council committees composed entirely of state delegates, such as the Programme Committee and the Finance Committee. These committees are also responsible for the vast majority of reports, discussion topics, and policy proposals placed on the governing bodies’ agendas, with the Director-General rarely exercising the right to add, modify, or remove items.⁷⁰ In direct violation of the constitution, governing body decisions are almost always made by consensus, even on some of the issues that require supermajority consent, such as the size of the budget and the adoption of conventions and recommendations. Finally, the FAO is heavily

⁶⁶ Food and Agriculture Organization 2011, Volume I (A), Article VII.

⁶⁷ Ibid, Volume I (B), Rule VII, fn. 1.

⁶⁸ Ibid, Volume I (C), Regulation VII.

⁶⁹ Author interview with FAO department director, 21 January 2015, Rome, Italy.

⁷⁰ Author interview with FAO office director, 23 January 2015, Rome, Italy.

reliant on funding from member states: nonstate actors have provided an average of 12 percent of contributions during the past decade, while independent earnings have accounted for just three percent of total income.⁷¹

Despite its high level of *de jure* policy autonomy, therefore, the FAO has been highly vulnerable to state interference throughout its history. In its first three decades, the FAO was dominated by the major Western food producers – in particular the United States – whose initial enthusiasm for its objectives soon gave way to narrower concerns that assisting food-insecure countries could reduce their commercial exports. To prevent this possibility, in 1954 they secured the Council’s endorsement of a set of guidelines that prohibited bilateral food aid from displacing imports in recipient nations – rules that, critics argued, defeated the very purpose of such assistance.⁷² In addition, they vetoed all proposals by officials to create a global food reserve for emergencies – most notably the 1946 “World Food Board” plan – fearing that it would lower agricultural prices.⁷³ Decolonization and the ensuing growth in the FAO’s membership in the 1960s and 1970s increased the influence of developing nations, which demanded an expansion of its field programs and an increase in levels of food aid. Unwilling to meet these demands, developed states resorted to alternative strategies of interference, reducing, delaying, and withholding contributions and sporadically issuing exit threats from 1980 onward. The upshot was a state of financial peril and institutional gridlock that has largely persisted to the present day.⁷⁴

<<

Figure 3 around here

>>

In line with my argument, these problems have resulted in low levels of performance on all three dimensions. Deprived of funds and marginalized in the policymaking process, FAO officials have been unable to establish the deep field presence or bring

⁷¹ FAO audited accounts 2004-13, David Lubin Memorial Library, FAO, Rome, Italy.

⁷² See Hopkins 1992.

⁷³ Hambidge 1955, 68-69.

⁷⁴ A widely cited independent evaluation of the FAO conducted in 2007 found that “low levels of trust and mutual understanding between Member Nations” had brought about “a financial and programme crisis that imperils the Organization’s future.” Food and Agriculture Organization 2007, 3.

about the domestic agricultural reforms required to enhance global food security.⁷⁵ These failings have, in turn, further undermined support for a powerful independent secretariat among states, who have instead restricted it to carrying out limited tasks such as compiling agricultural data, offering policy advice, and developing voluntary codes and guidelines – evidence of the two-way relationship between performance and *de facto* policy autonomy suggested above. This vicious cycle has ensured that the FAO has made few inroads into the core problem that motivated its creation: as illustrated in Figure 3, the number of undernourished people in the world has barely changed over the past 40 years, even as the number of people living in extreme poverty has sharply declined (the latter is now lower than the former).⁷⁶ Nor has it performed its narrow range of tasks in a cost-effective manner: archival financial records indicate that administrative costs have consumed around a one-fifth of its annual expenditures for most of its 70-year history, with the figure even higher during its early years (see Figure 4).⁷⁷ Finally, as discussed earlier, it has generally been responsive to a small group of (entirely governmental) stakeholders.

<< **Figure 4 around here** >>

The argument also helps to explain the FAO's failure to safeguard the high level of policy autonomy enshrined in its constitution. While the FAO's in-country advisory work can be relatively technical, its core information-gathering and norm-building activities do not require extensive specialized knowledge. This is reflected in the wide range of educational and professional backgrounds from which officials are recruited: as of 2011, only one member of the FAO's senior management (the Director-General Jacques Diouf) possessed a graduate-level qualification in the field of agricultural science or public health. It is also illustrated by the highly politicized nature of deliberations in its governing bodies, which are frequently characterized by lengthy

⁷⁵ Only 30-40 percent of FAO employees have been based in the field in recent decades. Ibid, 253.

⁷⁶ The undernourishment data come from Food and Agriculture Organization 2010 (1969-89); and FAO Hunger Statistics (1990-2014), available at <<http://www.fao.org/hunger/en/>>, accessed 4 October 2014. The poverty data come from Bourguignon and Morrisson 2002 (1970-80); and the World Bank's PovcalNet database (1981-2011), available at <<http://iresearch.worldbank.org/PovcalNet/index.htm>>, accessed 4 December 2014.

⁷⁷ FAO audited accounts 1945-2013, David Lubin Memorial Library, FAO, Rome, Italy.

and acrimonious exchanges in which delegates repeatedly condemn each other's national policies.⁷⁸ In sum, there is little scope for officials to exploit information asymmetries to protect themselves against state interference.

Equally important, officials have failed to forge the type of institutionalized alliances that are conducive to a high level of *de facto* policy autonomy. The FAO has long suffered from a “bad name” as a partner due to its perceived territoriality and failure to deliver results – an example of reverse causation between performance and alliances – and recent efforts to shed this reputation have enjoyed little success.⁷⁹ Data on the FAO's website indicate that 58 percent of its partnerships – which are divided fairly evenly among IGOs, NGOs, and businesses – consist of a purely symbolic affiliation.⁸⁰ The rest have below-average levels of policy preference alignment for IOs in the dataset, with only 57 percent of partners possessing a global scope and the same proportion operating in the issue area of agriculture or public health. Complementarity in capabilities is even lower: the ratio of partner to FAO expenditures in the period 2009-11 inclusive is just 7.01, the third lowest in the dataset. Consequently, as one official lamented, “Partners have neither the desire nor the ability to prevent states from pursuing their own agendas. Most of them either have no meaningful relationship with the FAO or possess very different goals from it – and the remainder simply lack the resources to alter state behavior.”⁸¹

The World Food Programme

The WFP was jointly established by the UN and the FAO in 1962 with a mandate to use food aid to promote development and provide relief in emergencies. While the FAO already played a role in regulating bilateral food aid, as noted earlier, the WFP represented the first attempt to deliver nutritional assistance on a multilateral basis. For similar reasons to the FAO – namely, a belief among states that a strong impartial secretariat was needed to effectively administer food aid and that the multilateral character of such aid would render it relatively harmless to national interests – the WFP was designed to possess extensive policy autonomy. According to its General Regulations, the Executive Director (i.e., chief of staff) is responsible for preparing its

⁷⁸ Author interview with FAO deputy director-general, 19 January 2015, Rome, Italy.

⁷⁹ Food and Agriculture Organization 2007, 215.

⁸⁰ See <<http://www.fao.org/partnerships/en/>>, accessed 14 December 2014.

⁸¹ Author interview with FAO partnerships officer, 22 January 2015, Rome, Italy.

Management Plan (a “comprehensive plan of work”) and Strategic Plan (a description of “strategic objectives and their implications for the proposed programme of work”); drafting its biennial budget; and setting the agenda for its governing body, the Executive Board, which includes 36 (rotating) nations.⁸² Executive Board decisions are to be made unanimously, although majority voting is required if consensus cannot be reached. Similarly to before, decisions about amendments to the General Regulations and mid-session changes to the agenda require a two-thirds majority vote. With respect to nonstate funding, the WFP is permitted both to accept donations from “intergovernmental bodies, other public and appropriate non-governmental, including private, sources” and to independently earn income from investing funds that are not “immediately required.”⁸³

In contrast to the FAO, however, the WFP possesses a high level of policy autonomy *in practice* as well as in design. WFP officials working in the Resource Management and Accountability Department and the Operations Services Department are the only actors involved in preparing the Management Plan, Strategic Plan, and budget. They are also the source of almost every report and project proposal included on the Executive Board’s agenda, with states largely restricted to a “rubber-stamping” role.⁸⁴ Executive Board decisions are typically made by consensus, as per the General Regulations, although majority voting is common on politically sensitive matters.⁸⁵ Unlike in the FAO, moreover, rules regarding issues that require supermajority consent are always honored. Lastly, while most of the WFP’s funding is provided by member states, it derives a higher proportion from nonstate sources than the FAO.⁸⁶

The WFP’s high level of *de facto* policy autonomy has enabled it to consistently resist state attempts to promote national interests – attempts that have circumvented formal restrictions on interference in policymaking – earning it a widespread reputation for political neutrality.⁸⁷ During the Cold War, for instance, WFP officials used their

⁸² World Food Programme 2010, 39 and 9. Some of these responsibilities were originally assigned to the Director-General of the FAO – a source of tension between the two institutions – but were transferred to the Executive Director in the early 1990s.

⁸³ *Ibid.*, 21 and 47.

⁸⁴ Author interview with WFP office director, 28 January 2015, Rome, Italy.

⁸⁵ Author interview with WFP official, 22 January 2015, Rome, Italy.

⁸⁶ On average, 15 percent of contributions have come from nonstate actors during the past decade, while five percent of total income has been independently earned. WFP annual reports (nonstate contributions) and audited accounts (independent earnings) 2004-13, WFP Library, Rome, Italy.

⁸⁷ See Barrett and Maxwell 2005.

policymaking influence to push through – in the face of informal opposition from the United States, historically the largest donor – much-needed projects in Soviet allies such as Cuba, Ethiopia, Nicaragua, and Vietnam.⁸⁸ More generally, econometric studies have found that levels of WFP assistance – unlike those of bilateral food aid – are highly correlated with indicators of recipient country need (such as income and food availability) and largely uncorrelated with proxies for the political interests of donors (such as their geographical distance from recipients).⁸⁹ The WFP has also been responsive to the needs of recipients in *how* it procures food. The traditional model of procurement, in which food is sourced from donor countries, has been widely criticized for its slow speed, high transport costs, and tendency to lower agricultural prices in recipient markets.⁹⁰ In response, despite resistance from donor countries that benefit from the surplus-disposal outlet provided by the traditional model (in particular the United States), the WFP has increasingly turned to alternative forms of procurement such as local and regional purchases (which now account for more than 75 percent of its food) and the provision of cash and vouchers to beneficiaries.⁹¹

<<

Figure 5 around here

>>

Consistent with theoretical expectations, the WFP's capacity to resist state interference has enabled it to achieve high levels of performance on all three dimensions. As shown in Figure 5, which is based partly on archival distribution records, WFP food deliveries have continuously increased since its establishment, with particularly rapid growth in the last 30 years: the number of beneficiaries of WFP aid soared from 15 million in 1980 to an average of 97 million over the past decade, while the total metric tonnage of food distributed rose from 1.5 million to four million.⁹² During this period, the WFP has helped to avert several possible famines – including in southern Africa (1991-92 and 2000-01), Afghanistan (2001),

⁸⁸ Interestingly, in most of these cases the United States chose not to block consensus in the governing body, anticipating that it would be defeated in a majority vote.

⁸⁹ Barrett and Heisey 2002; and Neumayer 2005. For the WFP, donors' political interests are measured using a weighted average of individual scores for the United States and European Union (its two largest donors).

⁹⁰ Barrett and Maxwell 2005.

⁹¹ See <<http://www.wfp.org/purchase-progress/overview>>, accessed 15 December 2014.

⁹² WFP annual statements of the Executive Director 1963-2014, WFP Library, Rome, Italy.

and western Africa (2012) – by building up a vast logistical network that operates 70 aircraft, 20 ships, and 5,000 trucks on any given day.⁹³ These feats have, in turn, strengthened state support for organizational autonomy – the reverse of what occurred in the FAO’s case – helping to explain the notable lack of governing body opposition to staff-initiated projects since the end of the Cold War.⁹⁴ State backing for the organization has also been boosted by its consistently high levels of efficiency: archival financial data show that administrative costs have consumed an average of seven percent of annual expenditures since its creation (see Figure 4), a figure that is likely to be among the lowest of any multilateral development agency.⁹⁵ Finally, as discussed earlier, the WFP has responded to the demands of a wide range of (state and nonstate) stakeholders in both allocating and procuring food aid.

Effective performance is not the only source of the WFP’s high level of *de facto* policy autonomy. No less important is the technically challenging nature of delivering emergency food aid on a global scale. The WFP manages a lengthy and complex supply chain in which, within a matter of days, food is procured from governments and commercial producers in multiple countries; subjected to quality control, processed, fortified, packaged, and labeled; transported by some combination of air, sea, and land to the crisis-hit area; deposited in existing or mobile warehouses; and finally distributed to beneficiaries via a vast network of intermediary organizations (discussed below).⁹⁶ Officials thus require considerable expertise in logistics and supply chain management, as illustrated by the fact that more than half of the WFP’s senior management possessed a graduate-level qualification in engineering or business administration as of 2011. The technical character of the WFP’s activities is also reflected in the Executive Board’s willingness to allow officials to set its agenda and to approve almost all of their policy proposals without modification, a practice that results in short meetings (usually 1-2 hours) and low levels of internal conflict.⁹⁷

Perhaps the most important source of the WFP’s *de facto* policy autonomy, however, is its institutionalized alliances. To help it distribute food at the final stage of its

⁹³ See <<https://www.wfp.org/logistics>>, accessed 14 January 2015.

⁹⁴ Author interview with WFP office director, 28 January 2015, Rome, Italy.

⁹⁵ WFP audited accounts 1962-2013, WFP Library, Rome, Italy. While there are no comparable historical data for other such agencies, only four currently have a ratio of administrative costs to development assistance lower than 0.07. Easterly and Williamson 2011.

⁹⁶ Author interview with WFP department director, 27 January 2015, Rome, Italy.

⁹⁷ Author interview with WFP division director, 23 January 2015, Rome, Italy.

supply chain, the WFP has built up the largest partnership network of any IO, comprising around 1,250 NGOs, 50 businesses, and 15 IGOs. Most of this network was formed in the 1990s – archival records show that the number of NGO partners grew from 170 in 1988 to 1,120 in 1998 – suggesting that its growth both fuelled and fed off the sharp rise in WFP deliveries during this period (another instance of two-way causation between performance and alliances).⁹⁸ Critically, partnerships are characterized by high levels of depth, preference alignment, and complementarity: online information indicates that all of them involve substantive collaboration in the distribution of food; 84 percent of partners operate in the issue area of agriculture, public health, or humanitarian aid (although only 15 percent possess a global scope); and the ratio of partner to WFP expenditures in the period 2009-11 inclusive is 82.42, more than 10 times the FAO’s figure.⁹⁹ As expected, therefore, partners have been both willing and able to defend the WFP against state interference. As one official explained, “When states try to derail WFP programs for the sake of national interests, partners step in with vital financial support and launch intense lobbying campaigns at the domestic level. Pressure from NGOs such as CARE and Save the Children, for instance, has been instrumental in persuading the United States to reverse its long-standing opposition to local and regional purchases in recent years.”¹⁰⁰

Conclusion

Two key conclusions about the sources of variation in IO performance emerge from this article. First, since states possess incentives to pursue collective interests before creating IOs but individual interests afterward, the level of policy autonomy enjoyed by such institutions is a critical determinant of their performance. Second, and most strikingly, high levels of policy autonomy cannot be secured through institutional design: due to the difficulty of enforcing formal rules at the international level, *de jure* policy autonomy does not translate into *de facto* policy autonomy. Instead, the latter arises when IOs (1) forge institutionalized alliances with actors above and below the state – specifically, alliances characterized by deep cooperation, aligned preferences, and complementary capabilities – or (2) engage in technically complex activities.

⁹⁸ WFP Committee on Food Aid Policies and Programs/Executive Board reports 1988-98, WFP Library, Rome, Italy.

⁹⁹ See <<http://www.wfp.org/partners>>, accessed 12 October 2014.

¹⁰⁰ Author interview with WFP partnerships coordinator, 23 January 2015, Rome, Italy.

By operationalizing, measuring, and analyzing the relationships between these variables using a variety of new data sources, the article enhances our empirical knowledge of IOs. In addition, it contributes to a number of theoretical literatures in IR. First, it provides a corrective to the burgeoning literature on institutional design – and complements recent scholarship on informal governance – by showing that we cannot fully understand the performance of IOs by focusing solely on their formal rules. Second, it extends scholarship on IO autonomy, nonstate actors, and epistemic communities by highlighting the critical role of institutionalized alliances and technical complexity in protecting IOs against state interference. Third, it contributes to the IR literatures on delegation and principal-agent theory by drawing attention to the moral hazard problem experienced by state principals rather than IO agents, thereby raising questions about the validity of public choice analyses of international institutions. Finally, it joins a growing number of studies in seeking to move beyond the traditional functionalist assumption of institutional “efficiency” by analyzing the behavior and impact of IOs *after* they have been created.

Beyond these contributions, the article has important policy implications. Perhaps the most surprising is that efforts to improve the performance of IOs by strengthening formal safeguards against state interference in the policymaking process are unlikely to be successful. While *de jure* policy autonomy may provide IOs with a veneer of independence, it does not guarantee *de facto* policy autonomy and thus real protection against interference. This does not entail, however, that nothing can be done to make IOs more effective. To the contrary, the analysis offers valuable practical lessons for actors seeking to enhance IO performance. It suggests, for example, that IO officials should forge deep institutionalized alliances with subnational and supranational actors who share their policy preferences and have complementary capabilities; and other stakeholders – whether allied to IOs or not – should seek to mitigate state interference by placing pressure on policymakers in the domestic arena and providing IOs with the resources they need to pursue policies that advance organizational objectives.

References

Abbott, Kenneth W., and Duncan Snidal. 1998. Why States Act through Formal International Organizations. *Journal of Conflict Resolution* 42 (1):3-32.

Abbott, Kenneth W., and Duncan Snidal. 2000. Hard and Soft Law in International Governance. *International Organization* 54 (3):421-456.

- Abbott, Kenneth W., Philipp Genschel, Duncan Snidal, and Bernhard Zangl, eds. 2015. *International Organizations as Orchestrators*. Cambridge, UK: Cambridge University Press.
- Alesina, Alberto, Reza Baqir, and William Easterly. 1999. Public Goods and Ethnic Divisions. *The Quarterly Journal of Economics* 114 (4):1243-1284.
- Andrews, Rhys, George A. Boyne, Jennifer Law, and Richard M. Walker. 2005. External Constraints on Local Service Standards: The Case of Comprehensive Performance Assessment in English Local Government. *Public Administration* 83 (3):639-656.
- Ashworth, Rachel E., George A. Boyne, and Tom Entwistle, eds. 2010. *Public Service Improvement: Theories and Evidence*. New York, NY: Oxford University Press.
- Ascher, William. 1983. New Development Approaches and the Adaptability of International Agencies: The Case of the World Bank. *International Organization* 37 (3):415-439.
- Australian Agency for International Development. 2012. *Australian Multilateral Assessment*. Canberra.
- Barnett, Michael N. 1997. The UN Security Council, Indifference, and Genocide in Rwanda. *Cultural Anthropology* 12 (4):551-578.
- Barnett, Michael N., and Martha Finnemore. 1999. The Politics, Power, and Pathologies of International Organizations. *International Organization* 53 (4):699-732.
- Barrett, Christopher B., and Kevin C. Heisey. How effectively does multilateral food aid respond to fluctuating needs? *Food Policy* 27 (5):477-491.
- Barrett, Christopher B., and Daniel G. Maxwell. 2006. *Food Aid After Fifty Years: Recasting its Role*. Oxford, UK: Routledge.
- Bourguignon, François, and Christian Morrisson. 2002. Inequality Among World Citizens: 1820-1992. *American Economic Review* 92 (4):727-744.
- Boyne, George A. 2002. Concepts and Indicators of Local Authority Performance: An Evaluation of the Statutory Frameworks in England and Wales. *Public Money and Management* 22 (2):17-24.
- Boyne, George A., Kenneth J. Meier, Laurence J. O'Toole Jr., and Richard M. Walker eds. 2006. *Public Service Performance: Perspectives on Measurement and Management*. Cambridge, UK: Cambridge University Press.
- Buchanan, Allen, and Robert O. Keohane. 2006. The Legitimacy of Global Governance Institutions. *Ethics & International Affairs* 20 (4):405-437.
- Cameron, Kim S. 1978. Measuring Organizational Effectiveness in Institutions of Higher Education. *Administrative Science Quarterly* 23 (4):604-632.
- Copelovitch, Mark S. 2010. *The International Monetary Fund in the Global Economy: Banks, Bonds, and Bailouts*. Cambridge, UK: Cambridge University Press.
- Cusack, Thomas R. 1999. Social capital, institutional structures, and democratic performance: A comparative study of German local governments. *European Journal of Political Research* 35 (1):1-34.

Dahl, Robert A. 1982. *Dilemmas of Pluralist Democracy: Autonomy vs. Control*. New Haven, CT: Yale University Press.

Department for International Development (United Kingdom). 2011. *Multilateral Aid Review: Ensuring Maximum Value for Money for UK Aid through Multilateral Organisations*. London.

Dreher, Axel, and Roland Vaubel. 2004. The Causes and Consequences of IMF Conditionality. *Emerging Markets Finance and Trade* 40 (3):26-54.

Etzioni, A. 1964. *Modern Organizations*. Englewood Cliffs, NJ: Prentice-Hill.

Food and Agriculture Organization. 2007. *Report of the Independent External Evaluation of the Food and Agriculture Organization of the United Nations (FAO)*. Conference document C 2007/7A.1-Rev.1, 34th Session. Rome.

Food and Agriculture Organization. 2010. *The State of Food Insecurity in the World: Addressing Food Insecurity in Protracted Crises*. Rome.

Food and Agriculture Organization. 2011. *Basic texts of the Food and Agriculture Organization of the United Nations (2011 edition)*. Available at <<http://www.fao.org/docrep/meeting/022/k8024e.pdf>>, accessed 10 June 2014.

Grieco, Joseph. 1988. Anarchy and the Limits of International Cooperation: A Realist Critique of the Newest Liberal Institutionalism. *International Organization* 42 (3):485-507.

Grossman, Sanford J., and Oliver D. Hart. 1986. The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy* 94 (4):691-719.

Gutner, Tamar, and Alexander Thompson. 2010. The politics of IO performance: A framework. *The Review of International Organizations* 5 (3):227-248.

Haftel, Yoram Z., and Alexander Thompson. 2006. The Independence of International Organizations Concept and Applications. *Journal of Conflict Resolution* 50 (2):253-275.

Hambidge, Gove. 1955. *The Story of FAO*. New York, NY: D. Van Nostrand Company.

Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure*. Oxford, UK: Oxford University Press.

Hart, Oliver, and John Moore. 1990. Property Rights and the Nature of the Firm. *Journal of Political Economy* 98 (6):1119-1158.

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

Hooghe, Liesbet, Gary Marks, Tobias Lenz, Jeanine Bezuijen, Besir Ceka, and Svet Derderyan. Forthcoming. *Scale and Community: The Purpose and Design of International Organizations*. Oxford University Press.

Hopkins, Raymond F. 1992. Reform in the International Food Aid Regime: The Role of Consensual Knowledge. *International Organization* 46 (1):225-264.

- International Labour Organization. 2000. *Possible improvements in the standard-setting activities of the ILO. Addendum: Statistics on the ratification of Conventions*. Governing Body document GB.277.LILS/2(Add.1), 277th Session. Geneva.
- Johnson, Tana. 2014. *Organizational Progeny: Why Governments are Losing Control over the Proliferating Structures of Global Governance*. Oxford, UK: Oxford University Press.
- Keohane, Robert O. 1984 *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, NJ: Princeton University Press.
- Knack, Stephen. 2002. Social Capital and the Quality of Government: Evidence from the States. *American Journal of Political Science* 46 (4):772-785.
- Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. The Rational Design of International Institutions. *International Organization* 55 (4):761-799.
- La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1999. The Quality of Government. *Journal of Law, Economics, and Organization* 15 (1):222-279.
- Mathiason, John. 2007. *Invisible Governance: International Secretariats in Global Politics*. Bloomfield, CT: Kumarian Press.
- McCubbins, Mathew D., and Thomas Schwartz. 1984. Congressional Oversight Overlooked: Police Patrols versus Fire Alarms. *American Journal of Political Science* 28 (1):165-179.
- Mearsheimer, John J. 1994. The False Promise of International Institutions. *International Security* 19 (3):5-49.
- Miguel, Edward, and Mary K. Gugerty. 2005. Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics* 89 (11):2325-2368.
- Miles, Edward L., Arild Underdal, Steinar Andresen, Jørgen Wettestad, Jon B. Skjærseth, and Elaine M. Carlin. 2002. *Environmental Regime Effectiveness: Confronting Theory with Evidence*. Cambridge, MA: MIT Press.
- Miller, Gary J., and Andrew B. Whitford. 2007. The Principal's Moral Hazard: Constraints on the Use of Incentives in Hierarchy. *Journal of Public Administration Research and Theory* 17 (2):213-233.
- Miller, Gary J., and Andrew B. Whitford. 2016. *Above Politics: Bureaucratic Discretion and Credible Commitment*. New York, NY: Cambridge University Press.
- Ministry of Foreign Affairs (Denmark). 2012. *Denmark's Engagement in Multilateral Development and Humanitarian Organizations*. Copenhagen.
- Ministry of Foreign Affairs (Netherlands). 2013. Scorecards of international institutions. Available at <<http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2013/08/30/scorecards-van-internationale-instellingen.html>>. Accessed 15 October 2013.
- Ministry of Foreign Affairs (Sweden). 2011. Swedish assessment of multilateral organisations. Available at <<http://www.regeringen.se>>. Accessed 15 October 2013.

- Moe, Terry M. 1989. The Politics of Bureaucratic Structure. In *Can the Government Govern?*, edited by John E. Chubb and Paul E. Peterson, 267-329. Washington D.C.: The Brookings Institution.
- Moe, Terry M. 1990. The Politics of Structural Choice: Toward a Theory of Public Bureaucracy. In *Organization Theory: From Chester Barnard to the Present and Beyond*, edited by Oliver E. Williamson, 116-53. Oxford, UK: Oxford University Press.
- Neumayer, Eric. Is the Allocation of Food Aid Free from Donor Interest Bias? *The Journal of Development Studies* 41 (3):394-411.
- Pevehouse, Jon, Timothy Nordstrom, and Kevin Warnke. 2004. The Correlates of War 2 International Governmental Organizations Data Version 2.0. *Conflict Management and Peace Science* 21 (2):101-119.
- Pollack, Mark A. 1997. Delegation, agency, and agenda setting in the European Community. *International Organization* 51 (1):99-134.
- Price, James L. 1972. The Study of Organizational Effectiveness. *The Sociological Quarterly* 13 (1):3-15.
- Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
- Quinn, Robert E., and John Rohrbaugh. 1981. A Competing Values Approach to Organizational Effectiveness. *Public Productivity Review* 5 (2):122-140.
- Rogoff, Kenneth. 1985. The Optimal Degree of Commitment to an Intermediate Monetary Target. *The Quarterly Journal of Economics* 100 (4):1169-1189.
- Stone, Randall W. 2011. *Controlling Institutions: International Organizations and the Global Economy*. Cambridge, UK: Cambridge University Press.
- Tallberg, Jonas, Thomas Sommerer, Theresa Squatrito, and Christer Jönsson. 2013. *The Opening Up of International Organizations: Transnational Access in Global Governance*. Cambridge, UK: Cambridge University Press.
- Union of International Associations. 2015. *Yearbook of International Organizations 2015-2016*. Leiden, The Netherlands: Brill.
- Vabulas, Felicity, and Duncan Snidal. 2013. Organization without delegation: Informal intergovernmental organizations (IIGOs) and the spectrum of intergovernmental arrangements. *The Review of International Organizations* 8 (2):193-220.
- Vaubel, Roland. 1986. A public choice approach to international organization. *Public Choice* 51 (1):39-57.
- Vreeland, James Raymond. 2003. *The IMF and Economic Development*. New York, NY: Cambridge University Press.
- Weingast, Barry R., and Mark J. Moran. 1983. Bureaucratic Discretion or Congressional Control? Regulatory Policymaking by the Federal Trade Commission." *The Journal of Political Economy* 91 (5):765-800.

Weingast, Barry R. 1984. The Congressional-Bureaucratic System: A Principal Agent Perspective (with Applications to the SEC). *Public Choice* 44 (1):147-191

World Food Programme. 2010. *General Regulations, General Rules, Financial Regulations, Rules of Procedure of the Executive Board (November 2010 Edition)*. Available at <<http://www.wfp.org/about/general-regulations>>. Accessed 10 June 2014.

Young, Oran R. 1999. *The Effectiveness of International Environmental Regimes: Causal Connections and Behavioral Mechanisms*. Cambridge, MA: MIT Press.