Beyond Institutional Design: Explaining the Performance of International Organizations

Ranjit Lall

Abstract   International organizations (IOs) have long been a central focus of scholarship in international relations, yet we know remarkably little about their performance. This article offers an explanation for differences in the performance of IOs and tests it using the first quantitative data set on the topic. I argue that the primary obstacle to effective institutional performance is not deviant behavior by IO officials—as conventional "rogue-agency" analyses suggest—but the propensity of states to use IOs to promote narrow national interests rather than broader organizational objectives. IOs that enjoy policy autonomy vis-à-vis states will thus exhibit higher levels of performance. However, in the international context policy autonomy cannot be guaranteed by institutional design. Instead, it is a function of (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 data set on IO performance—based in part on a new wave of official government evaluations of IOs and in part on an original survey of IO staff—and conducting a comparative case study in the realm of global food security.

Do international organizations (IOs) do what they are meant to do? IOs have long been a central focus of scholarship in international relations (IR), yet we know remarkably little about their performance—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

1. Only a handful of IR studies 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.

from asymmetric information, transaction costs, moral hazard, and other sources. The implicit assumption is that institutions will succeed in providing the benefits their creators desired. 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 IO performance. 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 interorganizational 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 fifty IOs based on numerical ratings from the two most widely cited evaluations, namely, those conducted by the British Department for International Department (DFID) and the Australian Agency for International Development (AusAID).[4] The axes correspond to the three dimensions of IO performance highlighted here: the achievement of stated objectives (*z*-axis), cost-effectiveness (*y*-axis), and responsiveness to a wide range of stakeholders (*x*-axis).

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 seventy countries every year for the past decade.[5] The World Bank provided half a billion people with access to health services and 64 million people with access to improved water sources between 2011 and 2016.[6] The GAVI Alliance has helped to immunize 580 million children against vaccine-preventable diseases since 2000, averting more than 8 million deaths.[7] The low-

---

2. See Australian Agency for International Development 2012; Department for International Development (United Kingdom) 2011; Ministry of Foreign Affairs (Denmark) 2012; Ministry of Foreign Affairs (Netherlands) 2013 and Ministry for Foreign Affairs (Sweden) 2013. (The Swedish assessments were published between 2008 and 2011, the Dutch scorecards between 2011 and 2013.) In interviews with more than 150 officials from IOs based in Washington DC, 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. 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.

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 A1 in the online appendix.

5. World Food Programme Annual Reports 2006–15, <http://www.wfp.org/policy-resources/corporate?type=37&tid_2=All&tid_4=All>, accessed 14 November 2016.

6. World Bank International Development Association website, <http://ida.worldbank.org/results>, accessed 14 November 2016.

7. GAVI Alliance website, <http://www.gavi.org/about/mission/facts-and-figures/>, accessed 14 November 2016.

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 as a result of its failure to achieve results on the ground.
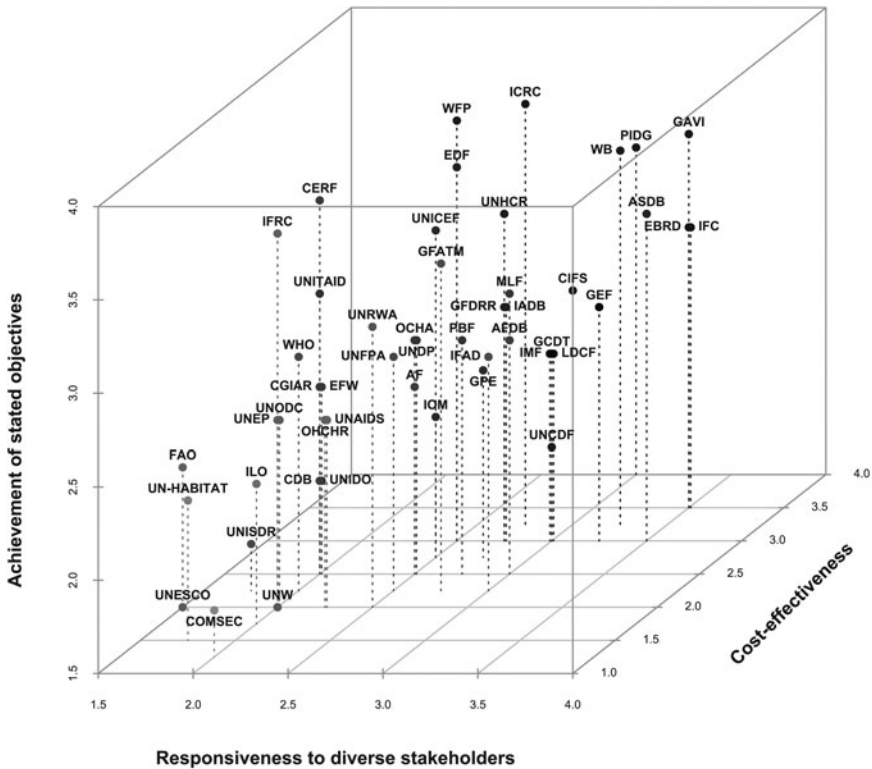
Notes: Ratings are from Department for International Development (United kingdom) 2011; and Australian Agency for International Development 2012. For IOs included in both assessments, an average score is displayed.

FIGURE 1. *Official government ratings of IO performance*

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

8. International Labour Organization 2000.

levels. The WFP and the Food and Agricultural Organization (FAO), for instance, both have a universal membership that includes almost every state; a mandate to 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 sixty-five 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 fifty 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 fifty 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 is why only *some* institutions succeed in realizing this potential.

I argue that the primary obstacle to effective IO performance is not deviant behavior by international bureaucrats—as suggested by conventional "rogue-agency" analyses—but the propensity of *states* to use IOs to advance narrow national interests rather than broader organizational objectives. In 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, in the international context policy autonomy cannot be guaranteed by institutional design. Thus, 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

---

9. See Pevehouse, Nordstrom, and Warnke 2004; and Union of International Associations 2015.

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 data set on IO performance, which covers fifty-three 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 fifty-three 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, namely, the FAO and the WFP. This examination, which draws on archival data and extensive interviews with IO officials, complements the statistical tests by providing evidence for my argument's posited causal mechanisms and shedding light on temporal as well as cross-sectional patterns in IO performance.

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 especially public administration scholars have made considerable progress in measuring and operationalizing institutional performance at the domestic level—progress that I later build on. On the explanatory front, theories of domestic institutional performance have generally focused on how population characteristics (such as size, cohesion, and diversity) affect whether citizens mobilize to demand 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. 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

---

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

problems among states by reducing transaction costs, providing information, and establishing a form of legal liability. Yet implicit in this functionalist analysis is essentially the same assumption about performance: institutions *do* provide the benefits desired by their creators. 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.[11]

Only a small number of studies have questioned this assumption. Drawing on public choice theory, Vaubel and Dreher highlight the performance problems caused by IO officials' rent-seeking behavior.[12] This perspective views states as a collective "principal" who strategically delegates authority to an IO "agent" to pursue a set of common objectives. Because of the costliness of monitoring and writing complete contracts, it suggests, IOs are able to acquire sufficient autonomy to advance their *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."[13] 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: IOs that enjoy a high level of autonomy are likely to perform worse than IOs that are tightly controlled by states.

By drawing attention to the possibility of institutional failure, these studies have significantly advanced our theoretical understanding of IO performance. However, 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. Moreover, 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 been examined through (qualitative and quantitative) studies of only 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

---

11. Key works in these literatures include Abbott and Snidal 2000; Hawkins et al. 2006; and Koremenos, Lipson, and Snidal 2001.

12. See Dreher and Vaubel 2004; and Vaubel 1986.

13. Barnett and Finnemore 1999, 716.

words of its editors, "[IO] bureaucracies will perform their functions … better if they are given some degree of autonomy."[14]

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 by states (and other potential donors) to be performing effectively. The same is true of officials seeking to maximize their policy influence: IOs typically have to rely on moral suasion to secure state compliance, and since they are not accountable to citizens in the same way as national governments, their surest path to legitimacy is to "effectively perform the functions invoked to justify [their] existence."[15] In addition, there is evidence from surveys and in-depth ethnographic studies that officials' goals often *encompass* the stated aims of their IOs.[16] 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.

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. Subsequent studies of

---

14. 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—as well as usefully highlighting some of the methodological challenges of investigating the topic.

15. Buchanan and Keohane 2006, 422.

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

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 bureaucratic agencies' efficiency as a result (in part) of fears that political turnover will enable opponents to use such institutions to pursue different objectives in the future.[17] More recently, building on the economics literature on time inconsistency in monetary policy, Miller and Whitford argue that changes in political conditions can cause the *same* politician to favor different bureaucratic policies at different points in time.[18]

Do analogous situations arise in the international context? Recent work on informal governance in IOs emphasizes how states are frequently prepared to interfere with IOs in ways that jeopardize institutional performance to safeguard core foreign policy interests.[19] The implication is that the delegation of authority to IOs is characterized by a time-inconsistency problem similar 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—because of 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 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 policy-making authority to an *independent* agent whose preferences are aligned with the desired policy outcome in both periods.[20] 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 policy-making

---

17. Moe 1990.
18. Miller and Whitford 2016.
19. Stone 2011.
20. The delegation solution was first proposed by Rogoff 1985.

functions to IOs.[21] 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, because of collective action problems among states, is any effective system of decentralized voluntary enforcement likely to arise.

Circumventing formal safeguards on policy autonomy 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). 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. 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." As a result, 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") as a result of 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.[22]

The emergence of informal policy-making practices that are at odds with formal rules entails that there will be only a weak correlation between an IO's de jure policy autonomy, or the policy autonomy its creators designed it to possess, and its de facto policy autonomy, or the policy autonomy 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. In other words, in the international context institutional design will not be sufficient to guarantee policy autonomy and thus effective performance.

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

---

21. Stone 2011, 29–30.
22. Koremenos, Lipson, and Snidal 2001, 794.

take to address such problems in the absence of interference from states.[23] This definition suggests that policy autonomy has three characteristics. An IO can be classified as 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 IO officials' ability 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. 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).[24] 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 make decisions by consensus, by contrast, every state effectively possesses a veto. 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 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 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.

---

23. For similar definitions, see Abbott and Snidal 1998; Barnett and Finnemore 1999; and Haftel and Thompson 2006. Stone 2011 has a more expansive conception of IO autonomy that encompasses all executive, legislative, and judicial functions that can be delegated to such institutions.

24. Johnson 2014.

Sources of De Facto Policy Autonomy

Where does de facto policy autonomy come from, if not institutional design? While the sources of de jure policy autonomy have received considerable attention from IR scholars, the factors affecting de facto policy autonomy have yet to be systematically investigated.[25] I highlight two factors: (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 objectives and possess complementary material, informational, organizational, and other resources. Such collaboration is mutually beneficial: in return for their assistance in developing, implementing, and enforcing policies, IOs provide partners with access to contacts and networks, normative guidance, and the legitimacy that comes with endorsement by a multilateral body.[26]

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 IOs against state interference.[27] Since partners derive benefits from collaboration and have similar goals to IOs—a key reason they are enlisted in the first place—they have incentives to assist IOs in addressing not only the *practical* challenge of policy formulation and implementation but also the *political* challenge of state interference. Partners can protect IO policy autonomy in two ways. First, they can alter the payoff to states of pursuing national interests, for instance, by lobbying policy-makers 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, partners' willingness and ability 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

25. See Haftel and Thompson 2006; Hooghe et al. 2017; and Johnson 2014.
26. Abbott and Snidal 1998.
27. Abbott et al. 2015.

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 conflicting preferences, some partners benefit more from achieving 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 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 policy-makers 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, policy-makers 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 the specialized knowledge needed to participate in the policy-making process. Conversely, in non-technical domains little specialized training (formal or informal) is needed to formulate policies. Policy-makers can thus be recruited from a wide variety of professional, educational, and institutional backgrounds.

Technical complexity protects IOs against interference by generating information asymmetries between IO officials and states regarding the content of policies. Such asymmetries make it difficult for states to propose new policies, forcing them to cede agenda-setting powers to officials. In addition, they prevent states from fully grasping the distributional implications of policies and thus effectively opposing proposals that conflict with national interests.[28] Uncertainty about distributional consequences also makes states less likely to withhold financial contributions from IOs because they cannot anticipate how they will be affected by a given policy. Perhaps even more important, technical knowledge enables IOs to earn independent revenue by selling specialized products and publications and providing expert advice.

It is worth noting the striking contrast between this analysis and the public choice perspective, which views information asymmetries as a source of weak institutional

---

28. Mattli and Seddon 2015.

performance because they reduce the capacity of state principals to monitor and control IO agents. My analysis agrees that such asymmetries enhance IO autonomy but posits that the consequences for performance are *positive*. This difference 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

Operationalizing and Measuring IO Performance

As defined earlier, 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.[29] Each tradition captures 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 operationalize and measure 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.[30] 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

---

29. On the organizational theory tradition, see Etzioni 1964; and Price 1972. On the public administration tradition, see Boyne 2002; Boyne et al. 2006; and Cameron 1978.
30. Gutner and Thompson 2010, 233.

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.[31] 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 excel on different dimensions, with some being the most effective at 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 as 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.[32]

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. 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 fifty-three IOs, a list of which can be found in Table A1 in the online appendix.[33] The IOs span a wide range of issue areas, though more than three-quarters operate in at least one of the following five domains: economic development, education, environmental protection, humanitarian aid, and public health. There is also some variation in their geographical scope, with

---

31. Boyne 2002; Cameron 1978; Quinn and Rohrbaugh 1981.

32. This is similar to Putnam's requirement of "internal consistency." Putnam 1993, 64.

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

TABLE 1. *Government assessments of IO performance*

Assessment author	Year	IOs included	Main data sources	Performance indicators	r with others	r with MOPAN
Department for International Development (DFID), United Kingdom	2011	41 (36 IGOs, 2 INGOs, 3 PPPs)	*Subjective*: MOPAN surveys; Paris Aid Survey; HIPC CBP evaluations; stakeholder consultations, workshops, interviews, and written submissions	Delivery of results; contribution to international objectives; cost consciousness; financial management; transparency and accountability; strategic management	0.67	0.53
Australian Agency for International Development (AusAID), Australia	2012	41 (35 IGOs, 1 INGO, 4 PPPs)	*Objective*: QuODA data; Publish What You Fund Aid Transparency Index; COMPAS indicators; analysis of IO reports		0.62	0.63
Ministry for Foreign Affairs, Sweden	2008–11	23 (21 IGOs, 2 PPPs)		Internal and external effectiveness	0.41	0.19
Ministry of Foreign Affairs, Netherlands	2011–13	31 (29 IGOs, 2 PPPs)	*Subjective*: MOPAN surveys; reporting from government ministries, overseas missions, and embassies	Implementation; focus on core mandate; financial management; effectiveness of administration and management	0.44	0.59
Ministry of Foreign Affairs, Denmark	2012	17 (15 IGOs, 1 INGO,1 PPP)	*Objective*: analysis of IO reports; external audits; independent evaluations and reviews	One overall performance measure	0.30	0.25

Notes: IGO = intergovernmental organization; INGO = international nongovernmental organization; PPP = public-private partnership. See text for full names of data sources in fourth column.

six IOs possessing a purely regional membership (five of which are development banks). Finally, in addition to IGOs, the data set includes a small number of PPPs (five) and INGOs (two). While not the direct focus of my argument, these institutions closely resemble IGOs and thus fall within its analytical scope. Indeed, states were closely involved in their creation, provide the vast majority of their contributions (84 percent on average in the period 2009–11), and are formally represented in almost all of their governing bodies. Nevertheless, the inclusion of these IOs in the data set does not affect the strength of empirical support for the argument.

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 stakeholders' views, relying instead on indirect feedback via relevant ministries, overseas missions, and embassies. In terms of objectivity, DFID and AusAID cross-checked 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 Table 1 shows, the average correlation between each set of ratings and the remaining four sets is $r = 0.49$ (seven of the ten coefficients are 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 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).[34] This suggests that the consensus about the relative performance of IOs extends across both developed and developing countries.

---

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

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.[35] The first two 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). The Cronbach's alpha among the four sets of indicators is $\alpha = 0.85, 0.79, 0.66$, and 0.88, well above 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. 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.[36] 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 subsequent statistical analyses.

Explanatory and Control Variables

DE FACTO POLICY AUTONOMY is an index that sums the values of six indicators measuring the three characteristics of the concept described earlier: IO agenda-setting powers, decision-making procedures, and access to nonstate sources of financing.

---

35. 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.

36. The third column of Table A1 displays an average of the four indices and the Danish ratings for each IO (with all five variables rescaled between 0 and 1).

The first three indicators are based on responses to an online survey of the heads of all fifty-three IOs that I conducted between September 2013 and January 2015.[37] Specifically, respondents were asked the following questions, which capture the three distinct agenda-setting abilities outlined earlier:

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, 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).

DE JURE POLICY AUTONOMY has the same six indictors 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.[38] As shown in the fourth and fifth columns of Table A1, 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 standards deviations of each variable). This evidence confirms that formal rules alone do not provide the basis for an accurate assessment of an IO's true level of policy autonomy and thus provides strong support for H2.

---

37. The survey, implemented using Qualtrics Survey Software, was sent to participants via a link in an email. In most instances, two to three reminder messages were sent before the response was submitted. In a small number of cases, participants preferred to provide their responses verbally (during a telephone interview) or in written form.

38. 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 the official websites of IOs and partners.[39] The first 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 (nonsymbolic) partners that share their geographical scope (second indicator) and issue area (third indicator).[40] The fourth measures complementarity in capabilities, taking a value of 1 if the ratio of (nonsymbolic) partner to IO expenditures in the 2009–11 period exceeds 10 and 0 otherwise.

I employ two measures of technical complexity. 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). TECHNICAL COMPLEXITY (QUALIFICATIONS) is based on IO officials' formal qualifications and thus adheres more closely to the definition of the concept provided earlier. 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).[41] This information is obtained from staff biographies on the websites of IOs and former employers as well as resumés posted on professional networking websites (such as LinkedIn).

Finally, the data set includes a small number of control variables. 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 operational offices outside the country in which it is headquartered.[42] Unless otherwise specified, all explanatory and control variables are measured as of the end of 2011. Summary statistics on all variables in the data set appear in Table A2 in the online appendix.

Statistical Analysis

I begin by estimating a series of ordinary least squares regressions with heteroskedasticity-robust standard errors. In Tables 2 and 3, the dependent variable is IO

---

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

40. I distinguish between three levels of geographical focus—global, regional, and national—and twenty-five issue areas based on a list constructed by Hooghe et al. 2017.

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

42. 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*.

TABLE 2. *Relationship between IO performance and de facto policy autonomy*

	Dependent variable: Government ratings of IO performance					
	DFID INDEX (1)	AUSAID INDEX (2)	DUTCH INDEX (3)	SWEDISH INDEX (4)	DANISH RATINGS (5)	AVERAGE OF MEASURES (6)
DE FACTO POLICY AUTONOMY	1.070***	1.210***	0.557***	0.685***	0.156	0.111***
	(0.207)	(0.195)	(0.168)	(0.155)	(0.263)	(0.032)
NUMBER OF STAFF (*log*)	0.209	−0.086	0.049	−0.048	0.027	0.007
	(0.167)	(0.114)	(0.156)	(0.115)	(0.258)	(0.017)
FIELD PRESENCE	−1.315	−0.145	−0.288	−0.78	−1.296***	−0.053
	(1.144)	(0.898)	(0.979)	(1.306)	(0.367)	(0.116)
AGE (*log*)	−0.368	−0.0004	−0.774*	0.338	1.226***	−0.035
	(0.413)	(0.428)	(0.411)	(0.410)	(0.346)	(0.046)
Constant	−2.687***	−3.420***	0.84	−2.384	−1.434	0.274**
	(0.849)	(0.947)	(0.795)	(1.653)	(1.312)	(0.126)
Observations	41	41	31	23	17	53
R^2	0.402	0.353	0.352	0.387	0.549	0.224
Adjusted R^2	0.336	0.281	0.253	0.251	0.399	0.159

Notes: Ordinary least squares regressions with robust standard errors in parentheses. In Model 6, the five measures of IO performance are rescaled between 0 and 1. *$p < .1$; **$p < .05$; ***$p < .01$.

TABLE 3. *Relationship between IO performance and de jure policy autonomy*

	Dependent variable: Government ratings of IO performance					
	DFID INDEX (1)	AUSAID INDEX (2)	DUTCH INDEX (3)	SWEDISH INDEX (4)	DANISH RATINGS (5)	AVERAGE OF MEASURES (6)
DE JURE POLICY autonomy	0.137	0.167	−0.043	0.034	0.112	0.018
	(0.165)	(0.205)	(0.179)	(0.125)	(0.124)	(0.022)
NUMBER OF STAFF (*log*)	0.439**	−0.065	0.146	0.056	0.046	0.005
	(0.221)	(0.131)	(0.164)	(0.153)	(0.231)	(0.017)
FIELD PRESENCE	−2.193*	−0.376	−0.029	−1.144	−1.257***	−0.065
	(1.188)	(0.974)	(1.013)	(1.431)	(0.386)	(0.110)
AGE (*log*)	−0.424	0.042	−0.907*	0.346	1.153***	−0.014
	(0.493)	(0.470)	(0.482)	(0.634)	(0.321)	(0.048)
Constant	−0.352	−0.046	2.321**	−0.751	−1.162	0.522***
	(0.933)	(1.063)	(1.051)	(1.982)	(0.994)	(0.131)
Observations	41	41	31	23	17	53
R^2	0.148	0.036	0.162	0.05	0.563	0.029
Adjusted R^2	0.053	−0.071	0.034	−0.161	0.417	−0.052

Notes: Ordinary least squares regressions with robust standard errors in parentheses. In Model 6, the five measures of IO performance are rescaled between 0 and 1. *$p < .1$; **$p < .05$; ***$p < .01$.

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) in Model 6. In Table 2, 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 1 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. In Table 3, DE JURE POLICY AUTONOMY is substituted for DE FACTO POLICY AUTONOMY. Consistent with H3, the coefficient becomes nonsignificant in all six models but remains positive in five (including the DFID and AusAID models). This striking discrepancy is illustrated in Figures 2 and 3, which display scatterplots of the average measure of IO performance against DE FACTO POLICY AUTONOMY and DE JURE POLICY AUTONOMY, respectively, with the estimated regression line from Model 6 in each table.
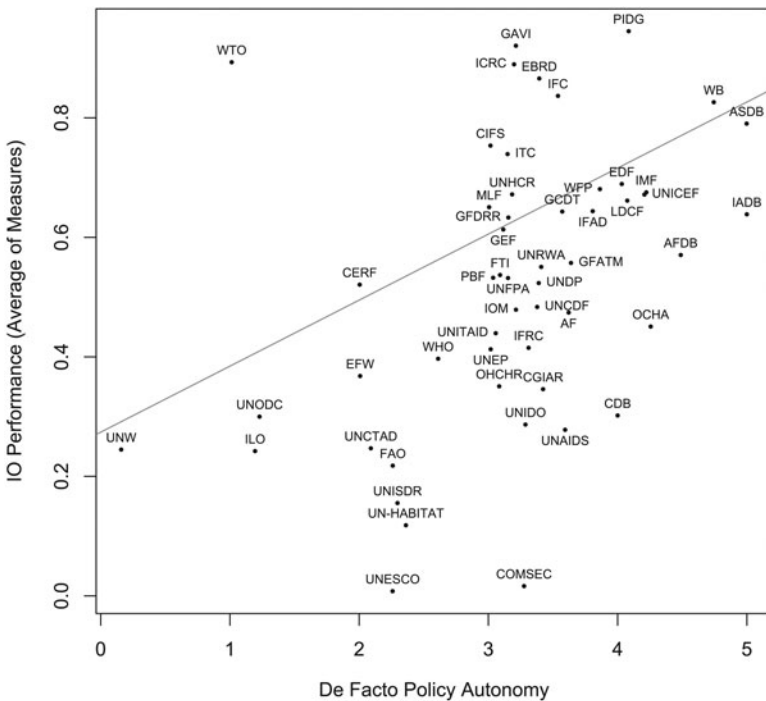
Note: The solid line represents the estimated regression line for DE FACTO POLICY AUTONOMY from Model 6, Table 2.

FIGURE 2. *Relationship between IO performance and de facto policy autonomy*

To investigate patterns within different dimensions of IO performance, I disaggregate the four indices of government ratings into their eighteen constituent indicators (while maintaining the same two sets of explanatory variables). Table 4 summarizes the key results, reporting the sign and significance level of the coefficients on DE FACTO POLICY AUTONOMY and DE JURE POLICY AUTONOMY in the thirty-six analyses (full regression results are shown in Tables A3–A10 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 eighteen models and significant at the 5 percent level in sixteen. Importantly, it is both positive and significant in eleven of the twelve 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 thirteen of the eighteen models and nonsignificant in sixteen. It is both positive and nonsignificant in nine of the twelve DFID and AusAID models.

Note: The solid line represents the estimated regression line for DE JURE POLICY AUTONOMY from Model 6, Table 3.

FIGURE 3. *Relationship between IO performance and de jure policy autonomy*

Finally, in Table 5 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 5 percent level. By contrast, in accordance with H2, the coefficient on DE JURE POLICY AUTONOMY is positive but nonsignificant in both models. Thus, strikingly, the degree of policy autonomy an IO is designed to possess has relatively little bearing on the degree of policy autonomy it possesses in practice.

TABLE 4. *Summary of results for constituent performance indicators*

Performance Indicators	DFID		AUSAID		DUTCH		SWEDISH	
	DFPA	DJPA	DFPA	DJPA	DFPA	DJPA	DFPA	DJPA
Delivery of Results	+ ***	+ *	+ ***	+	+	−		
Contribution to Objectives	+ ***	+	+	−	+ **	− **		
Cost Awareness	+ ***	+	+ ***	+				
Financial Management	+ ***	+	+ ***	+	+ ***	+		
Accountability and Transparency	+ ***	−	− ***	+				
Strategic Management	+ **	+	+ ***	+	+ ***	+		
Internal Effectiveness							+ ***	−
External Effectiveness							+ ***	+

Notes: The table reports the sign and significance level of coefficients on DE FACTO POLICY AUTONOMY (DFPA) and DE JURE POLICY AUTONOMY (DJPA) in regressions of each performance indicator in the left-hand column on the independent variables in Tables 2 and 3. *$p < 0.1$; **$p < 0.05$; ***$p < .01$.

TABLE 5. *Determinants of de facto policy autonomy*

	Dependent variable: DE FACTO POLICY AUTONOMY	
	(1)	(2)
INSTITUTIONALIZED ALLIANCES	0.236***	0.200**
	(0.079)	(0.094)
TECHNICAL COMPLEXITY (SURVEY)	0.691***	
	(0.240)	
TECHNICAL COMPLEXITY (QUALIFICATIONS)		0.914**
		(0.454)
DE JURE POLICY AUTONOMY	0.122	0.115
	(0.077)	(0.076)
NUMBER OF STAFF (*log*)	−0.085	−0.1
	(0.059)	(0.066)
FIELD PRESENCE	0.271	0.122
	(0.293)	(0.342)
AGE (*log*)	0.099	0.215
	(0.170)	(0.187)
Constant	1.778***	1.796***
	(0.372)	(0.391)
Observations	53	53
R^2	0.365	0.335
Adjusted R^2	0.282	0.248

Notes: Ordinary least squares regressions with robust standard errors in parentheses. *$p < .1$; **$p < .05$; ***$p < .01$.

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 2 and 3: (1) dummy variables for the five most common issue areas in the data set, 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 because IOs may be more susceptible to interference from powerful nations;[43] (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;[44] and (4) a variable measuring the geographical diversity of member/donor states, another proxy for group heterogeneity.[45] 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 5. As Tables A11–A13 show, 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 2 and 3 to address the possibility that such institutions systematically differ from IGOs on any variable of interest (Tables A14–A16). Second, given the data set's small size, I tested whether the results were driven by influential observations. 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 exceeding at least one of their standard cut-off points (Tables A17–A19).[46] Again, the results were not materially affected by any of these changes.

Another potential issue with the results is endogeneity in Table 2: DE FACTO POLICY AUTONOMY may be positively affected by IO performance, for instance, if successful IOs are "rewarded" by states with more independence. 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 my argument's two-stage structure, which naturally suggests two instruments: institutionalized alliances and technical complexity. The latter is the

---

43. This is measured as the ratio of the material capabilities of the most powerful state to those of the remaining states. Material 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/data-sets/national-material-capabilities-national-material-capabilities-v4-0>, accessed 12 December 2013.

44. 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.

45. This is defined as $1 - \sum_{i=1}^{m} s^2$, 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.

46. See the tables for details on the cutoff points, relevant DFBETAS coefficients, and excluded IOs.

more promising candidate since institutionalized alliances could directly enhance performance by filling gaps in IO capacities. Both measures of technical complexity are strong predictors of DE FACTO POLICY AUTONOMY, comfortably exceeding the standard F-test threshold for a "weak instrument."[47] I thus used the measures as separate instruments in a series of two-stage least squares regressions. The results, reported in Tables A20 and A21, are consistent with those in Table 2. While these results should be treated with caution—the absence of a direct link between technical complexity and performance cannot be verified empirically—they suggest that the positive relationship between performance and de facto policy autonomy is not solely the product of reverse causation.

Case Study: A Tale of Two Food Agencies

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: (1) the conference, which meets every two years and includes all 197 member states; and (2) the council, which meets one to three times per year and includes forty-nine (rotating) nations.[48] 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.[49] 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.[50]

In practice, however, the FAO's policy autonomy is very limited. As one senior official remarked, "the rules say one thing, but member states do something else— they dominate every aspect of the policy-making process."[51] 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

47. TECHNICAL COMPLEXITY (SURVEY) and TECHNICAL COMPLEXITY (QUALIFICATIONS) have F-statistics of 12.33 and 14.48, respectively, compared with the standard threshold of 10.
48. Food and Agriculture Organization 2011, Volume 1 (A), Article 7.
49. Ibid., Volume 1(B), Rule VII, n1.
50. Ibid., Volume 1(C), Regulation VII.
51. Author interview with FAO department director, 21 January 2015, Rome.

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.[52] 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 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 3 percent of total income.[53]

Despite its high level of de jure policy autonomy, therefore, the FAO has been consistently vulnerable to state interference. In its first three decades, the FAO was dominated by 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.[54] In addition, they vetoed staff proposals to create a global food reserve for emergencies—most notably the 1946 "World Food Board" plan—fearing that it would lower agricultural prices. 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 today.[55]

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 policy-making process, FAO officials have been unable to establish the deep field presence or bring about the domestic agricultural reforms required to enhance global food security.[56] Instead, it has been reduced to carrying out a narrow range of tasks that pose little threat to national interests, such as compiling agricultural data, offering policy advice, and developing voluntary codes and guidelines. As a result, it has made few inroads into the core problem that motivated its creation: the number of undernourished people in the world has barely changed over the past forty years,

---

52. Author interview with FAO office director, 23 January 2015, Rome.
53. FAO audited accounts 2004–13, David Lubin Memorial Library, FAO, Rome.
54. Hopkins 1992.
55. 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.
56. Ibid., 253. Only 30 to 40 percent of FAO employees have been based in the field in recent decades.
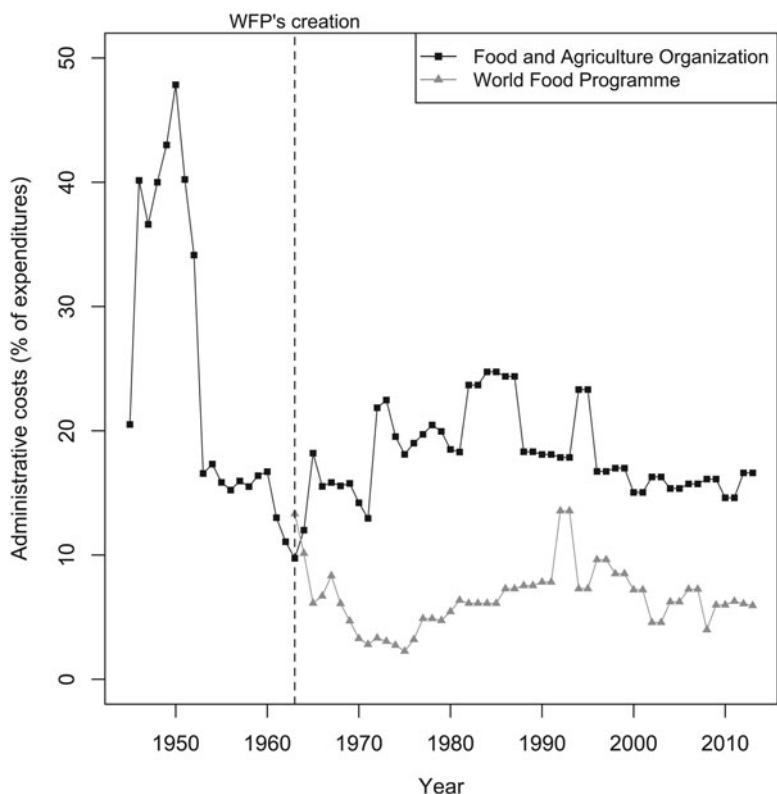
Sources: FAO archival audited accounts 1945–2013, David Lubin Memorial Library, FAO, Rome; WFP archival audited accounts 1962–2013, WFP Library, Rome.

FIGURE 4. *FAO and WFP administrative costs, 1945–2013*

even as the number of people living in extreme poverty has sharply declined (see Figure A1 in the online appendix). Nor has it performed its limited set of tasks in a cost-effective manner: as Figure 4 illustrates using archival financial data, administrative costs have consumed around one-fifth of its annual expenditures for most of its seventy-year history, with the proportion even higher during its early years.[57] Finally, as discussed earlier, it has generally been responsive to a small group of (entirely governmental) stakeholders.

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

---

57. FAO audited accounts 1945–2013, David Lubin Memorial Library, FAO, Rome.

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 (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 and acrimonious exchanges in which delegates repeatedly condemn each other's national policies.[58] 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 because of its perceived territoriality, and recent efforts to shed this reputation have enjoyed little success.[59] An examination of the organization's partnerships—which are divided fairly evenly among IGOs, NGOs, and businesses—reveals that 58 percent consist of a purely symbolic affiliation.[60] The rest have below-average levels of policy preference alignment for IOs in the data set, 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 2009–11 period is just 7.01, the third lowest in the data set. 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."[61]

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, the WFP represented the first attempt to deliver nutritional assistance multilaterally. For similar reasons to the FAO—namely, a belief among states that a strong independent secretariat was needed to effectively administer food aid and that the multilateral character of such aid would render it politically harmless—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 management plan (a "comprehensive plan of work") and strategic plan (a description of "strategic objectives

---

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

59. Food and Agriculture Organization 2007, 215.

60. All data on partnership characteristics reported in this section are based on the variable INSTITUTIONALIZED ALLIANCES and thus drawn from the official websites of the FAO, the WFP, and their respective partners.

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

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 thirty-six (rotating) nations.[62] Executive board decisions are to be made unanimously, although majority voting is required if consensus cannot be reached. As before, decisions about amendments to the General Regulations and mid-session changes to the agenda require a two-thirds majority vote. With respect to non-state 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."[63]

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.[64] Executive board decisions are typically made by consensus, although majority voting is common on politically sensitive matters.[65] Unlike in the FAO, rules regarding issues that require supermajority consent are almost 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.[66]

The WFP's high level of de facto policy autonomy has allowed it to consistently resist state attempts to promote parochial national interests—attempts that have circumvented formal restrictions on interference in policy-making—earning it a widespread reputation for political neutrality.[67] During the Cold War, for instance, WFP officials used their policy-making 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.[68] 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

62. 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.

63. Ibid., 21 and 47.

64. Interview with WFP office director, 28 January 2015, Rome.

65. Interview with WFP official, 22 January 2015, Rome.

66. On average, 15 percent of contributions have come from nonstate actors during the past decade, while 5 percent of total income has been independently earned. WFP annual reports (nonstate contributions) and audited accounts (independent earnings) 2004–13, WFP Library, Rome.

67. Barrett and Maxwell 2005.

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

from recipients).[69] 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.[70] In response, despite resistance from donor countries that benefit from the surplus-disposal outlet provided by the traditional model (in particular the US), 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 providing cash and vouchers to beneficiaries.[71]

Consistent with theoretical expectations, the WFP's capacity to resist state interference has enabled it to excel on all three dimensions of performance. 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 thirty 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.[72] During this period, the WFP has helped to avert several possible famines—including in southern Africa (1991–92 and 2000–01), Afghanistan (2001), and western Africa (2012)—by building up a vast logistical network that operates seventy aircraft, twenty ships, and 5,000 trucks on any given day.[73] It has achieved these feats while maintaining a consistently high level of efficiency: archival financial data show that administrative costs have consumed an average of just 7 percent of annual expenditures since its creation, a figure that is likely to be among the lowest of any multilateral development agency (see Figure 4).[74] Finally, as discussed earlier, the WFP has responded to the demands of a wide range of (governmental and nongovernmental) stakeholders in both allocating and procuring food aid.

How has the WFP succeeded—where the FAO failed—in protecting the substantial autonomy with which it was designed? The technically challenging nature of delivering food aid on a global scale is part of the answer. 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

69. See Barrett and Heisey 2002; and Neumayer 2005.

70. Barrett and Maxwell 2005.

71. World Food Programme website, available at <http://www.wfp.org/purchase-progress/overview>, accessed 15 December 2014.

72. WFP annual statements of the Executive Director 1963–2014, WFP Library, Rome.

73. World Food Programme Logistics, available at <https://www.wfp.org/logistics>, accessed 14 January 2015.

74. WFP audited accounts 1962–2013, WFP Library, Rome. 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.

organizations.[75] Officials thus require considerable expertise in logistics and supply chain management: 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 one to two hours) and low levels of internal conflict.[76]
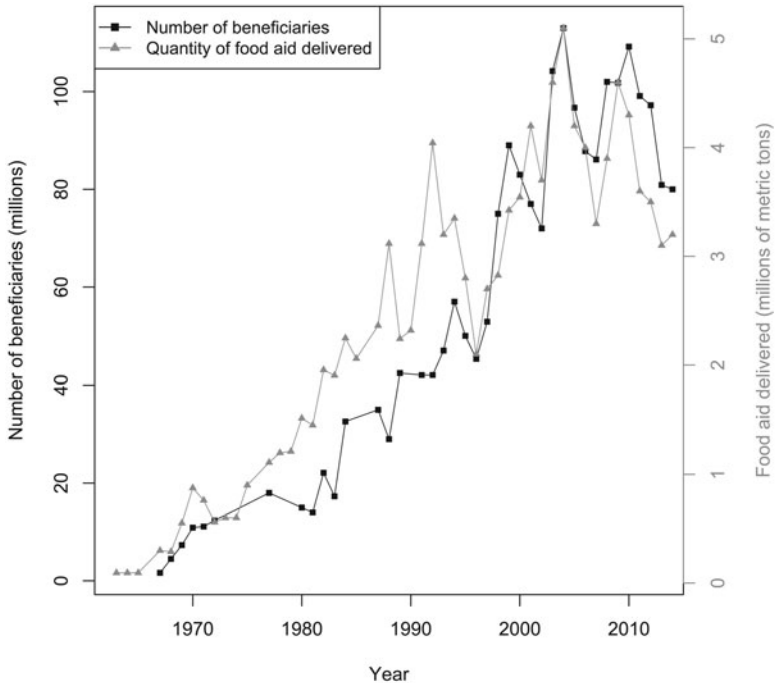
Source: WFP archival annual statements of the Exevutive Director 1963–2014, WFP Library, Rome.

FIGURE 5. *WFP food aid deliveries, 1963–2014*

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 supply chain, the WFP has built up the largest partnership network of any IO, comprising around 1,250 NGOs, fifty businesses, and fifteen IGOs. Most

75. Interview with WFP department director, 27 January 2015, Rome.
76. Interview with WFP division director, 23 January 2015, Rome.

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—helping to explain the sharp rise in WFP deliveries during this period.[77] Critically, partnerships are characterized by high levels of depth, preference alignment, and complementarity: all of them involve substantive collaboration in food distribution; 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 2009–11 period is 82.42, more than ten times the FAO's figure. As expected, partners have therefore 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."[78]

Conclusion

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. Critically, however, institutional design does not guarantee a high level of policy autonomy: because of the difficulty of both anticipating future threats to national interests and enforcing formal rules at the international level, de jure policy autonomy does not translate into de facto policy autonomy. Instead, de facto autonomy arises when IOs (1) forge institutionalized alliances with actors above and below the state—specifically, alliances characterized by deep cooperation, aligned policy preferences, and complementary capabilities—or (2) engage in technically complex activities.

By operationalizing, measuring, and analyzing the relationships between these variables using a variety of new data sources, the study enriches our empirical knowledge of IOs. In addition, it contributes to several areas of IR theory. It challenges a central premise of the influential literature on institutional design by showing that the formal rules with which IOs are created often have little bearing on how they function in reality. This surprising finding calls for both a more sustained theoretical and empirical focus on de facto rather than de jure characteristics of IOs and for further research on the conditions under which the two coincide. More generally, it underlines the need to move beyond the functionalist assumption that design features—and

77. WFP Committee on Food Aid Policies and Programs/Executive Board Reports 1988–98, WFP Library, Rome.
78. Interview with WFP partnerships coordinator, 23 January 2015, Rome.

indeed international institutions themselves—produce "efficient" outcomes by addressing cooperation problems among states.

The article also has important implications for the burgeoning literature on delegation and principal-agent theory. This literature views opportunistic behavior by IO agents as the key strategic problem to be "solved" in international delegation. By highlighting the incentives for state principals to engage in moral hazard—and the incentives for IO agents to resist such behavior—my analysis both reveals another dimension to the strategic structure of delegation and raises questions about the utility of conventional applications of principal-agent theory to IOs. Investigating how the severity of principal's moral hazard problem varies across issue areas and institutional settings—a function, the analysis suggests, of the strength of enforcement mechanisms, the degree of uncertainty about threats to national interests, and the costs of "recontracting" in such contexts—is an important task for future research.

Finally, the article extends and builds bridges between the literatures on IO independence, nonstate actors, and epistemic communities by drawing attention to the key role of institutionalized alliances and technical complexity in protecting IOs against state interference. The notion that an institution's autonomy can be enhanced by links with external stakeholders is particularly surprising from a theoretical perspective because it runs counter to the well-established literatures on pluralism and bureaucratic delegation in American politics, which treat societal influence as a (necessary) constraint on the discretion of policy-makers. Further research is needed both to better understand the differences in the political and strategic context in which domestic and international public institutions operate and to identify the sources of variation in the depth, alignment, and complementarity of alliances across IOs.

Beyond these contributions, the article has significant policy implications. Perhaps the most striking is that efforts to improve the performance of IOs by strengthening institutional safeguards against state interference in the policy-making process are unlikely to be successful. 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 institutional performance. It suggests, for instance, that states should seek—individually or collectively—to counterbalance efforts by other nations to use IOs to advance narrow foreign policy interests; IO officials should forge deep institutionalized alliances with subnational and supranational actors who share their policy preferences and possess complementary capabilities; and other stakeholders—whether allied to IOs or not—should attempt to mitigate state interference by lobbying policy-makers in the domestic arena and providing IOs with the resources they need to pursue policies that advance organizational objectives.

Supplementary Material

Supplementary material for this article is available at https://doi.org/10.1017/S0020818317000066.

References

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

———. 2000. Hard and Soft Law in International Governance. *International Organization* 54 (3):421–56.

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):1243–84.

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):639–56.

Ascher, William. 1983. New Development Approaches and the Adaptability of International Agencies: The Case of the World Bank. *International Organization* 37 (3):415–39.

Australian Agency for International Development. 2012. *Australian Multilateral Assessment*. Canberra, Australia. Available at <http://dfat.gov.au/about-us/publications/Pages/australian-multilateral-assessment-ama-full-report.aspx>. Accessed 18 August 2015.

Barnett, Michael N. 1997. The UN Security Council, Indifference, and Genocide in Rwanda. *Cultural Anthropology* 12 (4):551–78.

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. 2002. How Effectively Does Multilateral Food Aid Respond to Fluctuating Needs? *Food Policy* 27 (5):477–91.

Barrett, Christopher B., and Daniel G. Maxwell. 2005. *Food Aid After Fifty Years: Recasting Its Role*. Oxford, UK: Routledge.

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 and International Affairs* 20 (4):405–37.

Cameron, Kim S. 1978. Measuring Organizational Effectiveness in Institutions of Higher Education. *Administrative Science Quarterly* 23 (4):604–32.

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.

Department for International Development (United Kingdom). 2011. *Multilateral Aid Review: Ensuring Maximum Value for Money for UK Aid through Multilateral Organisations*. London, UK. <https://www.gov.uk/government/publications/multilateral-aid-review>. Accessed 20 May 2013.

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

Easterly, William, and Claudia R. Williamson. 2011. Rhetoric Versus Reality: The Best and Worst of Aid Agency Practices. *World Development* 39 (11):1930–49.

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, Italy. Available at <ftp://ftp.fao.org/docrep/fao/meeting/012/k0827erev1.pdf>. Accessed 14 November 2016.

———. 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 14 January 2016.

Gutner, Tamar, and Alexander Thompson. 2010. The Politics of IO Performance: A Framework. *The Review of International Organizations* 5 (3):227–48.

Haftel, Yoram Z., and Alexander Thompson. 2006. The Independence of International Organizations Concept and Applications. *Journal of Conflict Resolution* 50 (2):253–75.

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. 2017. *Measuring International Authority: A Postfunctionalist Theory of Governance*. Volume 3. Oxford, UK: Oxford University Press.

Hopkins, Raymond F. 1992. Reform in the International Food Aid Regime: The Role of Consensual Knowledge. *International Organization* 46 (1):225–64.

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, Switzerland.

Johnson, Tana. 2014. *Organizational Progeny: Why Governments Are Losing Control over the Proliferating Structures of Global Governance*. Oxford, UK: Oxford University Press.

Katzenstein, Peter J., Robert O. Keohane, and Stephen D. Krasner. 1998. International Organization and the Study of World Politics. *International Organization* 52 (4):645–85.

Knack, Stephen. 2002. Social Capital and the Quality of Government: Evidence from the States. *American Journal of Political Science* 46 (4):772–85.

Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. The Rational Design of International Institutions. *International Organization* 55 (4):761–99.

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):222–79.

Mathiason, John. 2007. *Invisible Governance: International Secretariats in Global Politics*. Bloomfield, CT: Kumarian Press.

Mattli, Walter, and Jack Seddon. 2015. Orchestration Along the Pareto Frontier: Winners and Losers. In *International Organizations as Orchestrators*, edited by Kenneth W. Abbott, Philipp Genschel, Duncan Snidal, and Bernhard Zangl, 315–48. Cambridge, UK: Cambridge University Press.

Miller, Gary J., and Andrew B., Whitford. 2016. *Above Politics: Bureaucratic Discretion and Credible Commitment*. New York: Cambridge University Press.

Ministry of Foreign Affairs (Denmark). 2012. *Denmark's Engagement in Multilateral Development and Humanitarian Organizations*. Copenhagen. Available at <http://um.dk/~/media/UM/Danish-site/Documents/Udenrigspolitik/Nyheder_udenrigspolitik/2012/120709%20Multilateral%20Analysis%202012%20-%20English%20Version.pdf>. Accessed 15 October 2013.

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 for Foreign Affairs (Sweden). 2013. Swedish Assessment of Multilateral Organizations. Available at <http://www.regeringen.se>. Accessed 15 October 2013.

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. 2005. 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):101–19.

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):122–40.

Rogoff, Kenneth. 1985. The Optimal Degree of Commitment to an Intermediate Monetary Target. *The Quarterly Journal of Economics* 100 (4):1169–89.

Stone, Randall W. 2011. *Controlling Institutions: International Organizations and the Global Economy.* Cambridge, UK: Cambridge University Press.

Union of International Associations. 2015. *Yearbook of International Organizations 2015–2016.* Leiden, The Netherlands: Brill. Available at <http://www.uia.org/yearbook>. Accessed 8 October 2015.

Vaubel, Roland. 1986. A Public Choice Approach to International Organization. *Public Choice* 51 (1): 39–57.

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.