When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation

Dan Honig

January 29, 2018

This is the author's final version of this article, save the addition of these lines. Please cite this article as:

Honig, Dan. 2019. When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation. *International Organization*, Forthcoming Winter 2019.

Acknowledgments

I thank the National Science Foundation Graduate Research Fellowship for its support under grant# DGE-1144152. Matt Andrews, Sam Asher, Nancy Birdsall, Mark Buntaine, Andreas Fuchs, Peter Hall, Conor Hartman, Bob Keohane, Chris Kilby, Steve Knack, Aart Kraay, Jane Mansbridge, Sheila Page, Woody Powell, Lant Pritchett, Simon Quinn, Steve Radelet, Tristan Reed, Alasdair Roberts, Bill Savedoff, Evan Schofer, Ryan Sheely, Beth Simmons, Vivek Srivastava, Martin Steinwand, Mike Tierney, Dustin Tingley, Eric Werker, Michael Woolcock, anonymous reviewers at IO, and many others have all provided helpful comments on earlier versions of these ideas and/or this work. Many thanks to Yi Yan, Smriti Sakhamuri, Grace Chao, Kyle Kessler, and the World Bank Archives staff (particularly Sherrine Thompson) for their research assistance. I thank the European Commission, the UK's Department for International Development, the Asian Development Bank, the Global Fund for AIDS, Tuberculosis, and Malaria, and the German Development Bank for providing data. World Bank data used in this analysis are publicly available; as such the World Bank deserves perhaps the greatest thanks. Data for the Japanese International Cooperation Agency (JICA), the German Society for International Cooperation (GiZ), and the International Fund for Agricultural Development were assembled from individual project completion reports by Odesk-contracted research assistants under my supervision, with the compiled data then sent back to the originating agency for comment and/or correction. GiZ was kind enough to respond with corrections, which were incorporated; JICA wished it to be made clear that these data were generated by me rather than by JICA and that it is not responsible for them. The case studies would not have been possible without the hundreds of hours spent by a diverse group of interviewees listed in the online appendix (available at danhonig.info) talking to me and answering odd questions, for which I am very grateful. I can be reached at dhonig@jhu.edu. Replication materials for this article can be found on the IO replication archive, iojournal.org/data-archive. A fuller Project Performance Database (PPD) listing all publicly disclosable development projects available (the world's largest database of development projects at present) is also available on danhonig.info. This article has overlap (in empirics and text) with Honig 2018 (*Navigation by Judgment: Why and When Top Down Management of Foreign Aid Doesn't Work* from Oxford University Press). Both the book press and journal are aware of this overlap.

When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation

Dan Honig

Abstract

Bureaucracies with field operations that cannot be easily supervised and monitored by managers are caught between two sources of dysfunction that may harm performance. The first source of dysfunction is straightforward: field workers may use operating slack and asymmetric information to their own advantage, thwarting an organization's objectives. The second source of dysfunction is often overlooked: attempts to limit workers' autonomy may have deleterious effects, curbing agents' ability to respond efficaciously to the environment. I find that the parliaments and executive boards to whom International Development Organizations (IDOs) are accountable differentially constrain IDO organizational autonomy, which in turn affects management's control of field agents. Tight management control of field agents has negative effects, particularly in more unpredictable environments. Attempts by politicians to constrain organizations in an effort to improve performance may sometimes be self-undermining, having net effects opposite those intended.

Introduction

Bureaucrats, and employees more broadly, often decry the "red tape", controls, and reporting requirements their organizations place on them. This paper investigates whether these controls are beneficial or counterproductive for International Development Organizations (IDOs, e.g. the World Bank and US Agency for International Development) that deliver foreign aid. There is a real trade-off in allowing field worker judgment to guide foreign aid project implementation; following Aghion and Tirole, letting field workers (agents) take more initiative requires circumscribing managerial (principal) control.[1] This paper, then, compares the benefits of greater agent autonomy to its costs. Is more reliance on the perceptions and judgment of field agents associated with better foreign aid project performance, or do greater top-down controls and orientation of agents towards measureable targets yield better results?

To explore these questions I examine variation in the constraints placed by political authorizing environment on IDOs, arguing that politically constrained organizational autonomy induces tight control of field agents by an organization's managers. When IDO managers need to report measurable success to legislators and executive boards, they need to manage field staff in such a way as to generate numbers. Where agent autonomy is important, then, I argue that a reform-minded politician's desire to improve aid project performance by requiring measurable short-term results may undermine the success of the very aid projects the politician wishes to see perform well. Accounting for results and actually succeeding in delivering results are sometimes in tension with one another.

[1] Aghion and Tirole 1997.

This does not mean top down controls are necessarily counterproductive; where rules and targets are well aligned with an organization's objectives tight control may improve performance. However tight control also reduces agent flexibility and risks orienting agents towards meeting targets at the expense of delivering on broader organizational goals. As such, the costs of control need to be weighed against its benefits. This paper uses variation in recipient country environments as a source of exogenous variation in the net effects of tight principal control. Some recipient country environments are more unstable, and thus there are greater returns to the agent flexibility and use of judgment that tight principal control precludes.

IDOs are well suited for exploring how environments and organizational control practices jointly impact organizational success. Individual IDOs work in a wide range of country environments across a range of tasks from road construction to anti-corruption efforts, with limited ability to exit contexts or tasks in response to poor performance. This paper provides novel evidence connecting political authorizing environment constraint and agency performance via management practice. The empirical findings strongly suggest that more politically constrained IDOs do indeed differentially engage in management practices that increasingly undermine their own performance as environments become more unpredictable.

Examinations of IDO effectiveness, and international organization behavior more broadly, have recently taken what Gulrajani calls the "bureaucratic turn".[2] This work

[2] Gulrajani 2017, 375. Recent works to bring serious empirical study of organization to bear on foreign aid include Buntaine, Parks, and Buch 2017; Buntaine 2016; Bush 2015; Swedlund 2017. On the "bureaucratic turn" in IO more generally I have in mind e.g. Johns 2007; Johnson and Urpelainen 2014.

contributes to that turn by focusing on the under-explored actor in principal-agent models of IO bureaucracy, the agent. As Hawkins & Jacoby put it, "In spite of the growing sophistication of the principal-agent literature, it still contains a remarkably thin view of agent behavior."[3] This paper connects the study of IOs to the rich literature on delegation and autonomy[4] by describing the costs and benefits of principal control for IDOs.

After discussing the costs and benefits of principal vs. agent control, this paper discusses how environmental unpredictability impacts the balance of costs and benefits. It then argues why we should expect IDOs with insecure political authorizing environments to more tightly control field staff, even where such control is unlikely to augur for greater organizational success. The paper then turns to operationalizing key variables and formalizing the hypothesis. Quantitative analysis then brings to bear an original cross-IDO dataset composed of over 9,000 unique development projects, allowing us to test whether the impact of constrained autonomy is as hypothesized in the world's largest cross-organizational database to incorporate development outcomes.[5] The quantitative empirics are then complemented by qualitative interview data illustrating the central mechanism the paper theorizes.

Theory

[3] Hawkins and Jacoby 2006, 199.

[4] Calvert, McCubbins, and Weingast 1989; Carpenter 2001; Carpenter 2010; Nielson and Tierney 2003; Hawkins et al. 2006; Huber and McCarty 2004; Huber and Shipan 2006; Huber and Shipan 2002; Rasul and Rogger 2013; Keiser 1999; Lipsky 1980; Hupe and Hill 2007; Pepinsky, Pierskalla, and Sacks 2016; Kauppi and Van Raaij 2015.

[5] While the movement for aid information transparency has made impressive strides in the past few years, most of the progress to data has been on inputs – on spending data and financial flows. No other source (including the International Aid Transparency Initiative, the OECD Development Assistance Committee's Creditor Reporting System, and the AidData archive) includes systematic information on the results of projects in a way tractable to quantitative analysis for any donor other than the World Bank. The World Bank also makes these data public and easily accessible from the World Bank's website (the only donor to do so).

How Principal Control Can Undermine Organizational Success

Principal-agent models have long wrestled with the reality that agents have asymmetric information – access to local knowledge that distant principals lack. Conventional applications of principal-agent models tend to focus on the risk this asymmetric information poses.[6] In a recent piece entitled "Why Organizations Fail", two distinguished scholars write "Incentive problems arise due to the presence of asymmetric information or imperfect commitment, which lead agents to act according to their own biases or preferences rather than in the interest of the organization."[7]

Attempts at strengthening principal control often take the form of costly monitoring technology. Principals can also induce agents to do what principals want by tying compensation, promotion, etc. to outcomes the principal can observe. These attempts at control have costs as well, however. First, controls may induce agents to focus on meeting formal requirements rather than the service they are meant to deliver. Second, controls may preclude agent initiative and productive use of the asymmetric information to which agents uniquely have access. This section describes each problem in turn.

- What Principal Control May Unproductively Induce

Given the difficulty of directly observing agent action, IDOs' primary tool of agent control is the setting of performance targets and requiring reporting against them.[8] A recent OECD review of the US Agency from International Development (USAID) finds

[6] E.g. Barnett and Finnemore 2003 conceive of IO asymmetric information as a source of dysfunction, of unresponsiveness to the desires of political principals.
[7] Garicano and Rayo 2016, 138–9. The authors in turn cite Hölmstrom 1979 and Shavell 1979 on this point.
[8] See e.g. Natsios 2010; Gulrajani 2011.

USAID uses "approximately 200 standard indicators (recently reduced from 500), and many more custom indicators" in their monitoring and evaluation of projects.[9] These indicators orient agent action, thus acting as a *de facto* management tool irrespective of whether their intent was in fact to centralize control with the principal. Targets can orient field staff towards clear results and hold staff accountable if targets are not reached.

However, target-setting may also induce distortions, e.g. inducing agents to focus on producing what can be measured and reported upon.[10] Due to their difficult monitoring environment, IDOs often measure short-term outputs to proxy longer term outcomes; agents may achieve these outputs but without actually forwarding the IDOs' goals.[11] As Kerr put it over forty years ago, there is potential for IDOs to engage in "the folly of rewarding A while hoping for B."[12]

- What Principal Control May Unproductively Preclude

The danger to principals of agent asymmetric information and hidden action are well explored in applications of the principal-agent model to international organizations. While less commonly explored in conventional applications of principal-agent models, there have been scholars who conceive of agents' private information as valuable for good organizational performance. Most directly, Lisa Martin has theorized that IMF staff

[9] OECD 2016, 82.
[10] Holmstrom and Milgrom 1991; Wilson 1989.
[11] Clist 2016; Gelb and Hashmi 2014; Perakis and Savedoff 2015.
[12] Kerr 1975.

members have private information about borrowers' contexts that are important for IMF loan performance.[13]

Like IMF loan performance, foreign aid project success also depends on asymmetric information held by agents. Attempts at control inevitably produce rules, targets, and other accountability measures that purposefully constrain agents. But the very constraint that precludes bad behavior by agents may also unintentionally preclude behaviors that *are* in service of the organization's mission. Sometimes good organizational performance depends on the gathering and use of asymmetric information, such as local contextual knowledge. When there are important things agents can know and their supervisors cannot (asymmetric information), Aghion and Tirole argue it is critical that agents have not just formal but "real" authority.[14] This means agents are not just given formal ability to make judgments but also that the organizational incentives they face encourage the use of their judgment. Aghion and Tirole argue that agents who do not have an incentive to gather asymmetric information will not do so, framing this as the tradeoff between agent initiative and principal control.[15]

Asymmetric information can also include soft information. Soft information is defined as information that a skilled observer might use to inform his or her decisions, but cannot be proven or "cannot be directly verified by anyone other than the agent who

[13] Martin 2006. Johns 2007 also conceives of agent private information as critical to implementation in the context of IOs.
[14] Aghion and Tirole 1997.
[15] Aghion and Tirole's tradeoff has a number of cousins; E.g. James Scott's claim that top-down planning precludes "metis", or "knowledge embedded in local experience", and Gailmard & Patty's notion that inducing agent investment in expertise requires a loosening of principal control, as only agents with the ability to make use of expertise will work to cultivate it. Scott 1998; Gailmard and Patty 2007.

produces it."[16] Tight principal control means agents will not gather asymmetric information, including soft information; as such organizations are left with a poorer knowledge base.

Principal control may also impede organizational response to changing contexts. Putting more control in the hands of agents empowers actors who are better placed to rapidly respond when flexibility and adaptation is needed, while simultaneously reducing the control mechanisms (review procedures, approval processes, etc.) that might impede rapid response. Flexibility is complementary, but distinct, from the asymmetric (soft) information channel; flexibility is in greater demand when contexts change more rapidly, whereas the direct returns to soft information may persist irrespective of the rate of environmental change.

The Benefits of Principal Control

These benefits of agent autonomy must be balanced against their costs. Putting more control in the hands of field agents also means those agents will find it easier to engage in a range of actions, including those that may be illegal or undesired.[17] Agents may be "captured" in their time away from headquarters, maximizing private benefits or simply implementing their own plans even when those plans do not serve organizational best interest.[18]

[16] Stein 2002, 1892. Soft information is perhaps most easily understood as the informational cousin of tacit knowledge (Polanyi 1966), inasmuch as a key feature of both is the difficulty of codification and transmission.
[17] Tirole 1994.
[18] For this discussion in IDOs see e.g. Woods 2006, 56.

Agents will be more susceptible to capture with less principal control. In addition, agent judgment can simply be wrong even when well-intentioned. An IDO that gives agents more control will have more to fear from fallible agent judgments.

Principal control also produces more standardized behavior. By shifting control to agents, an organization may allow more scope for bias and prejudice.[19] Where standardization is critical to good outcomes – the organizational equivalent of baking a cake, where following a precise recipe is likely to yield the best results – less principal control will likely induce variation that will be detrimental to organizational performance.[20]

There are potential costs to organizations in giving agents more autonomy and thus relying on their initiative, just as there are costs to tighter principal control and less autonomy. Aghion & Tirole frame this as the tension between principal control and agent initiative. The following section explores how the level of environmental unpredictability affects the relative value of agent asymmetric information and flexibility, as well as the potential for outputs to induce distortions in agent performance unhelpful to the principal. It argues that environmental unpredictability plays an important role in determining how IDOs ought best resolve the tension between principal control and agent initiative in a given context.

Environmental Unpredictability, Principal Control, and IDO Success

[19] Policing is perhaps the context in which this issue is most contested. Relying on police judgment clearly gives rise to disparate treatment e.g. by race; however it is unclear if tight control of police officers is likely to lead to better outcomes. See e.g. Brehm and Gates 1999; Prendergast 2001.
[20] See e.g. Chandler 1977; Williamson 1983 on the tension between flexibility and standardization on organizational management and structure.

Whether more or less principal control augurs for greater success in delivering foreign aid depends on a number of situational factors; one critical systematic source of variation is environmental unpredictability. Rapidly changing contexts both increase the chances that targets will induce distortion and raises the value of what controls can preclude. More unpredictable environments require more flexibility and more use of asymmetric (soft) information.

As IDO project implementation occurs, there are many things that may impact how interventions ought best proceed. Some changes are foreseeable, and thus a smart project plan could account for these contingencies. However, there are frequently what then-US Secretary of Defense Donald Rumsfeld once referred to as "unknown unknowns".[21]

Some "unknown unknowns" may be unforeseeable when a project commences, but nonetheless predictable at some time before the risk occurs. When a hypothetical project to provide youth vocational skills in collaboration with the Ministry of Youth and Sports begins, the recipient country's political environment may appear stable. However a year into implementation, the Minister falls out of favor with the Prime Minister and is likely to lose his job; the current Minister's successor is likely to marginalize a program closely associated with her predecessor. A wise and well-informed IDO field agent, foreseeing this possibility, may begin to include more career civil servants in the steering committee of the project and consult the Minister himself less. Such a decision requires agent freedom of action and use of asymmetric soft information. An IDO with tight principal control will have more poorly informed agents who would in any case be less able to act on their own

[21] This was at a press conference on February 12, 2002 regarding the link between the Government of Iraq and weapons of mass destruction.

unverifiable judgments of changing context to respond strategically to changing circumstances.[22]

Additionally, environments vary with regards to legibility - the extent to which they can be understood from a distance.[23] More unpredictable environments are also likely to be less legible. In the context of international development this might be understood as the correlation of *de jure* structures with *de facto* reality. Formal structures and hierarchy vary with regards to whether they are good indicators, for example, of whose approval is needed in practice to ensure a project will proceed. The greater the gap between structures and reality, the greater the returns to soft information and thus agent autonomy. In less legible environments it will be hard for anyone other than field agents to make judgments about how to proceed in designing and implementing projects.

Deviations from Equilibrium: Heterogeneous Political Authorizing Environments

The discussion thus far perhaps begs the question why IDOs, and indeed all organizations, would not simply adapt principal control appropriately to differences in the environment. Just as field operatives report to IDO headquarters, IDOs themselves are agents reporting to authorizing environments, the collection of actors to whom organizations are accountable (e.g. their political principals). IDOs respond to the shadow of their authorizing environments. By "shadow" I mean the threat of possible future

[22] Another way of framing this point is via the economics literature on incomplete contracting (e.g. Grossman and Hart 1986; Hart and Moore 1988; Hart and Moore 1990), which argues that decision rights for unforeseen contingencies need to rest with the party who needs to make uncontractible investments. In the sense of Gailmard and Patty 2012, it is the agents who need to invest in their expertise; in the sense of Aghion and Tirole 1997, these agents need to invest in gathering contextual asymmetric information.

[23] Legibility is a key element of James Scott's framework for thinking about top-down planning in Seeing Like a State (Scott 1998).

authorizer sanction, which in turn affects management's actions and degree of principal control.[24] For public organizations political authorizing environments are critical gatekeepers to resources, controlling the funding, mandate, and ultimately survival of public agencies.[25]

Different IDOs have very different relationships with their authorizing environments. The expected probability of sanctions for e.g. failure or reputation-damaging cases of corruption and fraud varies. As such the "length" of an authorizing environment's shadow varies; some IDOs worry about how their performance will be viewed by authorizers to a much a greater degree than do others. Exploring the reasons for authorizing environment differences are beyond the scope of this work, which will largely take authorizing environments as given. I focus instead on the consequences of differential authorizing environment insecurity.

Insecure agencies will take fewer risks than they otherwise would.[26] If an organization needs to meet measures in the short term in order to e.g. receive continued funding, the organization may not take the risks necessary to achieve long term ends—an organization's "risk appetite" may be inefficiently constrained. Insecure agencies are less likely to take smart risks, where the expected probability-weighted value of benefits exceeds costs.

[24] This thus takes the intuition of Axelrod's famous "shadow of the future" (Axelrod 1984) but operates not through a probabilistic chance of repeated interaction but rather a probabilistic chance of future sanction.
[25] Organizational behavior scholarship has long accepted that organizational strategies are determined in part by the need to access critical resources. See e.g. Resource Dependence Theory, and the long literature following Pfeffer and Salancik 1978.
[26] Bozeman and Kingsley 1998; Singh 1986.

Insecure agencies are also much more likely to be concerned with reporting success to authorizers. The greater the pressure to report organizational results, the greater the need for senior management to manage via measurement and target setting inside an organization, in order to generate the data which can then be reported to authorizers.[27] Target setting does more than simply add an additional reporting step to agents' workload; when pressure is put upon these measures for control purposes, measures change what agents and organizations actually do. While this is true of all kinds of measures, there are particularly large reasons to worry when management by measurement is employed for legitimacy-seeking reasons.

As seen from the perspective of an agency in need of justifying itself one attractive feature of measurement and reporting is measurement's role in making the activities of the organization seem legitimate.[28] In the public sector measurement has increasingly become critical to justifying continued funding and building legitimacy as part of a broader discourse on accountability and control; the spread of performance measurement is often linked with legitimacy seeking.[29] The reason for measurement is then to *appear* successful; where appearing successful and actually accomplishing the organization's objectives are in tension, the latter is likely to be sacrificed in favor of the former. By creating metrics and meeting targets, even when those targets are not well linked to

[27] E.g. Wynen and Verhoest 2016.
[28] Meyer and Rowan 1977.
[29] Dunleavy and Hood 1994; Hood 2004; Lynn Jr 1998; Modell 2004; Oliver 1991.

ultimate organizational goals, organizations can appear to be performing well to political authorizers.[30]

Insecure agencies are likely to engage in greater principal control at the expense of agent initiative even where that principal control may undermine the success of interventions. This is both because principal control is likely to better generate standardized data that can be used for legitimacy-seeking purposes and because tight control limits opportunities for agent malfeasance or bad action that might serve as a reputational risk for agencies.

Hypotheses & Operationalization

This work conceptualizes letting go of principal control and thus giving agents greater autonomy as a 2nd-best strategy; a strategy to employ when it is less bad than the distortions and constraints of top-down control. In some contexts, tight principal control is clearly superior to relying on fallible agent judgment. In other contexts the gains of being able to respond more flexibly and better utilize asymmetric (soft) information are superior to distortionary tight principal control. IDOs with greater political authorizing constraint will be less likely to give up principal control when appropriate, as environments become more unpredictable. More constrained IDOs will thus be less able to cope with unpredictability than will their less constrained peers. Thus

[30] Authorizing environments are not the only source of legitimacy-seeking measurement; Buntaine, Parks, and Buch 2017 suggest that recipient countries play a key role in choosing less substantial, easier to monitor targets. This effect as they theorize it should not differentially affect different IDOs after controlling for recipient country fixed effects (as the empirical models below do), however.

IDOs with more stable political authorizing environments will see less of a decline in performance in response to increasing environmental unpredictability than their more constrained peers.

The claim is not, then, that tight principal control is always inferior; nor is it that agent initiative allows IDOs to improve their absolute level of performance as environments become less predictable. It is simply that less principal control and greater agent initiative will be more helpful in more unpredictable contexts. This is because the costs imposed by principal control will go up as unpredictability rises, as will the benefits of relying on agent initiative and judgment. Insecure political authorizing environments will preclude IDOs from giving up principal control when otherwise appropriate.

Operationalizing Success

In implementing their work IDOs structure their activities through projects. Projects are discrete, time- and place-bound activities implemented after careful planning and preparation. These projects can vary widely in location, sector, and purpose; World Bank projects approved by the World Bank's board in April 2017 include projects in the Dominican Republic focused on the quality of educational statistics, in Bosnia on public health behavior, in Benin on enhancing agricultural productivity, and in India on state-level urban development.[31]

The empirics below employ a novel dataset consisting of over 9,000 unique projects in 140 countries carried out by nine donor agencies from 1994-2013, the product of many

[31] World Bank Project #s P163049, P160512, P160029, and P155303, respectively. April 2017 chosen as most recent available data from the public World Bank Projects & Operations database on date of access. World Bank 2017.

months of labor; this dataset is unique in including project performance data for multiple IDOs. More detail on the data collection, cleaning, and coding process can be found in the online appendix.[32]

Figure 1 shows the distribution of projects across countries, demonstrating the wide range of countries in which projects occur.

[Figure 1 Here]

Project success ratings are assigned on a six point likert-type scale, with e.g 6 as "highly satisfactory" and 1 as "highly unsatisfactory".[33] The underlying construct employed by different IDOs for measuring the success of projects is relatively consistent, with an OECD-wide standard in place. A given project's rating is intended to incorporate a project's relevance, effectiveness, efficiency, sustainability, and impact.[34] Holistic success ratings are variously calculated by IDO staff, external evaluation departments, or independent evaluators.

This is, of course, a less than fully precise standard as to what constitutes success. Success may be defined differently for different IDOs, or in different sectors. Fixed effects by IDO, sector, and recipient country partially help control for these potential sources of

[32] The online appendix is on my personal website, danhonig.info. The fuller Project Performance Database (PPD), containing the publicly disclosable projects employed in this work as well as others (over 14,000 total projects), is also on the website.

[33] This example is drawn from the World Bank's six-point rating system, as it is perhaps the best known. Some organizations evaluate projects on alternative likert-type scales (such as a four-point scale, with 4 being best); I transform all scales to be on a consistent six-point scale and employ IDO fixed effects in all models that use this six-point scale. I also employ a z-transformed version of this variable in the analysis when IDO fixed effects are absent. This process effectively de-means overall project success, just as employing IDO fixed effects would do.

[34] For more on these terms see OECD 2000; OECD 1991.

bias. Poor data quality and evaluation bias are potential threats to validity that are discussed below and treated in robustness checks in the online appendix. It is also possible for the data to be accurate in the sense of correctly reflecting an organization's assessment, but for that assessment to bear little connection to the actual performance of the project. To the extent possible, I have also attempted to validate these evaluations by turning to primary documentation; the online appendix describes this archival work, which broadly supports the conclusion that organizational assessments mapped "real" success and failure in the projects examined.

IDO Autonomy and Constraint: Authorizing Environment Insecurity and Propensity to Give Up Principal Control

In 2005 IDOs and recipient countries came together to agree to the Paris Declaration, a set of principles for achieving more effective aid tied to measurable targets.[35] Follow up Paris Declaration monitoring surveys focused on various elements of aid delivery. The monitoring surveys asked both donors and recipient countries for reports on their own, and each other's, practices (i.e. recipient countries also reported on donor behavior).

From the quantitative indicators that formed part of the monitoring reports I construct IDO proxies for "propensity to devolve control" and "authorizing environment insecurity".[36] The online appendix describes the construction of these scales in substantial

[35] The Paris Declaration on Aid Effectiveness 2005.

[36] These data were coded from appendix C of the published 2011 monitoring survey, which summarized performance on all three waves. They are indicators 5a (PFM), 5b (procurement), 6 (PIU), 7 (predictability), and 8 (tied aid). OECD 2012. The three waves of Paris Declaration surveys (2005, 2007, 2010) are averaged here, in keeping with expert advice that these were effectively multiple mappings of the same facts, with insufficient time for organizations to change significantly

detail. The devolution propensity and authorizing environment constraint measures are reasonably well correlated (.41). I take the simple average of the indicators to form a simple scale of autonomy ranging from 0 to 1, coded so that higher scores on the scale represent lower levels of political authorizing insecurity and higher IDO propensity to devolve control.[37] The overall scale has a Cronbach's alpha of .825.[38] This provides reasonable confidence that these measures and the two subscales map the same essential facts regarding IDOs and thus provide suggestive evidence for my conjecture that political constraints do in fact trickle down to IDOs' management practices.

A principal components analysis suggests this simple average is a more intuitive solution that will yield similar results to formal use of principal components; in any case results are robust to using a principal components approach. The online appendix presents the relevant technical information (e.g. eigenvector scree plots and component loading tables), as well as robustness checks employing principal components.

Given the critical role measurement of politically constrained autonomy plays for the empirical strategy, I validated the Paris Declaration scale with more direct measurement. I conducted a small-scale direct field survey of aid experts—individuals who have substantial development experience or whose jobs bring them into contact with a wide variety of donors. The online appendix contains a fuller explanation of this field survey measure.

---

between the first wave in 2005 and the last wave in 2010. Results are robust to using any wave and dropping any wave of the survey.

[37] For multilaterals (AsDB, WB, IFAD, EC) tied aid is not reported in the Paris Declaration monitoring surveys; in these cases the scale is an average of the remaining four measures.

[38] This is for the full scale with all IDOs.

Both the field survey measure and the Paris Declaration measure of IDO autonomy are time-invariant. In employing a time-invariant measure of IDO autonomy I do not mean to imply that IDO autonomy does not, in fact, vary across time – it certainly does, as agencies' relationships with their political authorizing environments change. While data limitations preclude modeling this source of variation, I do attempt to control for these dynamics to the extent possible. I employ year-by-IDO fixed effects as a robustness check in Table 3, which will absorb any changes in autonomy for a given IDO (by absorbing any IDO-specific changes in performance dynamics where they differ from the general pattern). The results below are robust to using any of the (differently timed) waves of the Paris Declaration monitoring reports (and thus using only the most recent, or least recent, wave or waves), as well as employing the (even more recent) direct field survey measure of autonomy. This field survey is well correlated with the Paris Declaration-derived politically constrained autonomy scale (.73), providing both an additional level of confidence in the accuracy of the Paris Declaration-based measure and suggestive evidence that IDO autonomy has not changed so greatly within-IDO over the period of the data so as to make the time-invariant measure uninformative.

Environmental Unpredictability

This paper operationalizes unpredictability by focusing on differential state fragility. Predictability and fragility are often linked explicitly in development practice, with practitioners speaking about the difficult and unpredictable nature of fragile state environments.[39] Fragility is in some sense the likelihood that the current equilibrium will break down or change rapidly. As the World Bank puts it, fragile states are "more unstable

[39] Weijer 2012; Institute of Development Studies 2014; Ghani, Lockhart, and Carnahan 2005.

and unpredictable" than their less fragile peers.[40] The focus of this work is not on fragile states as a class; on those at the very extreme of the state fragility measure. The theory above is intended to apply to the entire range of state fragility, and thus comparisons will be made across the entire universe of developing countries.

Environmental unpredictability is measured via the Polity IV State Fragility Index (SFI).[41] This index incorporates security, governance, economic development, and social development measures and has two subscales: effectiveness and legitimacy. The two subscales are highly correlated (.66) and Cronbach's alpha (.78) suggests that they map the same underlying construct.[42] The SFI varies at the country-year level, with every country holding an annual SFI score from 1994 to present.

Does Political Constraint on Autonomy Have Differential Effects Across Environment and Task?

The following section explores the relationship between IDO politically constrained autonomy, environmental unpredictability, and project success.

Summary Statistics of Key Variables

Table 1 presents summary statistics.

[Table 1 Here]

The online appendix provides additional summary statistics by IDO regarding the key dependent variable, project success. A key weakness of these data is the modest number of IDOs in the sample. Throughout the analysis below I will take care to ensure this small "2nd-level N" is not leading to spurious conclusions. In particular, I employ quite

[40] World Bank 2006, 55.
[41] Center for Systemic Peace 2014.
[42] In the sample data.

simple and straightforward econometric models to minimize the chance that these models are "overfit," with results driven by the relative lack of variation in outcome data as compared to the number of explanatory variables.

Quantitative Results: Politically Constrained Autonomy and Project Success

This section lays out the primary findings then addresses potential econometric concerns. The model for project *i* in recipient country *j* implemented by IDO *k* generalizes to

$$\text{Project Success}_{i,j,k} = \beta_1\text{*Environmental Unpredictability}_j + \beta_2\text{*Environmental Unpredictability}_j\text{*IDO Autonomy}_k + \beta_3\text{*Controls}_i + \text{Fixed Effects}_j + \text{Fixed Effects}_k + \varepsilon_i.$$

The key empirical prediction is that the coefficient on β_2 will be positive and statistically significant; that $\text{Project Success}_{i,j,k}$ will increasingly benefit from greater IDO Autonomy_k as $\text{Environmental Unpredictability}_j$ rises.

One key shortcoming of the dependent variable, project success, is that it is not amenable to direct inter-organizational comparisons; there is no reason to believe that one IDO's rating of "4" is in fact more successful than another IDO's rating of "3". Any (constant) systematic differences amongst IDO evaluation criteria or measurement standards are addressed in two ways: by including IDO *k* fixed effects in the models below (thus generating results which leverage intra-IDO comparisons across projects) and by normalizing project ratings using IDO-specific z-scores where fixed effects are not employed. As noted above, the measure of IDO politically constrained autonomy varies at the IDO *k* level and is time-invariant. This means that the measure is collinear to IDO *k* fixed effects. As a result quantitative analysis cannot directly compare IDOs' performance – it cannot say that e.g. KfW projects were more successful than IFAD in country X while IFAD projects were more successful than KfW in country Y. The interaction of IDO autonomy and a given country's level of environmental predictability does vary at the *j,k* level. The

interaction term can thus still be informative as to how within-IDO performance varies over recipient country and across time, though the absence of an autonomy base term precludes a direct comparison of two different IDOs' project success in a given country-year. Some models also use recipient country *j* fixed effects, thus ensuring any fixed country-specific features are not driving results.

In a literal sense, using IDO fixed effects removes the mean of the dependent variable – project success - for each IDO. This also means we need not trust that projects are as successful as donors say they are to believe the results of this model. Table 1 indicates than the average project scores a 4.3 on a six point scale. It seems possible, even highly likely, that the average project is not in fact a clear success; that these ratings are biased upwards. This will not bias the results so long as for a given IDO higher numbers are still associated with greater success; so long as a project scoring a 6 is more successful than a 4, a 4 more than a 2, etc. By de-meaning project success we also avoid spurious conclusions about absolute levels of successfulness.

The quantitative analysis instead focuses on the differential performance of IDOs with varying levels of politically constrained autonomy in interaction with other explanatory variables. This takes advantage of the fact that a rating of 4 given by KfW means a project was more successful than a project assigned a 3 by KfW, while a 2 given by IFAD means a project was less successful than one given a 3 by IFAD. It is possible, then, for the quantitative analysis to yield conclusions of the type "KfW projects are more successful in country X than country Y, while IFAD projects are more successful in country Y than country X". In this way inter-IDO comparisons can be made by comparing intra-IDO variation in project success.

To adjust for the possibility that project success may be correlated within a given recipient country, the main analyses report standard errors clustered at the recipient country level. It is also possible that project success is correlated within IDOs (even with fixed effects, errors may be correlated if the assignment of independent variables are clustered and there are heterogeneous treatment effects). Online appendix Tables I.10 and I.11 suggest that clustering by IDO (or, when practicable, double-clustering on IDO and recipient country) does not alter the substantive findings presented below.

Politically Constrained Autonomy and Environmental Unpredictability

Table 2 reports the core findings.

[Table 2 Here]

There is a robust and statistically significant negative relationship between environmental unpredictability and overall project success. Environmental unpredictability is associated with less successful project evaluations for IDOs, on average. The key explanatory variable, the interaction of IDO autonomy with environmental unpredictability, indicates that autonomy mediates the effect of environmental unpredictability on project success. While all IDOs see a decline in project success as environmental predictability falls, for more autonomous IDOs this decline is much less steep.

All models include IDO fixed effects. The model's comparison is being made within each IDO's projects, comparing whether a given IDO – e.g. the Asian Development Bank – sees more successful projects on average in more or less unpredictable environments (as measured by the State Fragility Index). Models 3 and 4 in Table 2 incorporate recipient

country fixed effects, thus focusing only on changes within recipient countries over time. Models 5 and 6 incorporate sector fixed effects, controlling for sectors at the most fine-grained level available, the 222 unique five-digit OECD Development Assistance Committee Creditor Reporting System (CRS) purpose sectors. Findings are robust to focusing on differences in performance within sectors as well.

If project success ratings were simply arbitrarily assigned, we would expect no relationship between project success and environmental unpredictability. If in harder to monitor unpredictable environments all projects were declared more successful, we would expect environmental unpredictability to be associated with higher success ratings. But instead here we see the relationship theory, and arguably intuition, would predict: consistent with my theory, as environments become more unpredictable project success falls.

As noted above, the key empirical prediction regards the interaction between IDO autonomy and environmental unpredictability. This interaction term is robustly positive and statistically significant, suggesting that autonomy does indeed play an important role in allowing an IDO to respond to greater environmental unpredictability. Once again this result holds when focusing on within-sector or within-recipient country data.

[Figure 2 Here]

Figure 2 draws from Model 1 of Table 2 to graphically represent differential performance by level of politically constrained autonomy. Note that the y-intercepts, and thus the relative level of the two lines in Figure 2, do not contain useful information. As described above the direct effect of autonomy is absorbed by IDO fixed effects, making the vertical positions of the two lines arbitrary. What *is* informative is the differential slopes of

the two lines – the differential success of IDOs of varying levels of autonomy in response to varying levels of environmental unpredictability.

All IDOs perform better in more predictable, stable contexts than they do in less predictable environments. More autonomous (less constrained) IDOs see a much smaller deline in performance than than their less autonomous (more constrained) peers as unpredictability rises. While an IDO with the lowest observed level of autonomy is predicted to have over half a point (.5) of difference between its performance in a state like Armenia (SFI=7 in 2014, or one standard deviation more stable than the mean) and its performance in a state like Nigeria (SFI=17 in 2014, or one standard deviation below the mean), an IDO with the highest observed level of autonomy is predicted to have about .06 of a point, or one tenth as much, performance differential.[43]

Table 3 adds a series of fixed effects to the main findings. Inclusion of time fixed effects (either yearly or in five-year periods) does nothing to diminish the association between autonomy and recipient unpredictability. The result remains robust to including time*IDO fixed effects and time*recipient fixed effects. These results should allay any concerns that the primary results are driven by heterogeneous IDO project performance over time or by heterogeneous entry of IDOs into and out of recipient countries over time.

[Table 3 here]

[43] While choosing the extremes for this graphical representation, relatively high and relatively low autonomy IDOs are differentiable at more modest degrees of differences as well. Online appendix Figure 3 demonstrates that using the 25th and 75th percentile also yields statistically significant differences in performance, though with smaller realized differences in predicted performance, of course.

Robustness

The online appendix outlines a series of robustness checks that speak to the validity of the autonomy measure, as well as the sensitivity of the analysis to how the interaction term is modeled, outliers, quirks in outcome variance, and a number of other potential threats to validity. Two primary concerns will be addressed here: the quality of the evaluations that form the core of the analysis and differential selection of IDOs into sectors or environments.

Evaluation Bias

These data rely on evaluations of project success made by the agencies themselves. One might worry that an agency with a fragile relationship with its political authorizing environment would, in addition to being less autonomous, have a greater incentive to self-evaluate projects to have been successes. This is not a threat to validity, inasmuch as a consistent bias would be absorbed by the IDO fixed effect; of greater concern would be bias that moves with the interaction of autonomy and environmental predictability. If, for example, more autonomous IDOs give their agents more leeway in self-evaluations, which those agents differentially exercise to a greater degree as environmental unpredictability rises, this would be a threat to the validity of the main findings.

The involvement of independent evaluation units provides suggestive insight into this problem, as independent evaluation units should not have the same degree of incentive as agents themselves to give favorable evaluations. Table 4 controls for the type of evaluation; that is, whether the data source is an internal review by project staff, a review conducted by an IDO's own independent evaluation unit, or a review conducted by an externally contracted evaluator.

[Table 4 Here]

The relationship between autonomy and environmental unpredictability remains unchanged, suggesting that differential evaluation bias by different IDOs is not driving the results.

Selection

One natural concern might be that controlling for recipient and sector fixed effects does not account for the fact that different IDOs may make decisions about what projects to pursue in light of where projects would be more successful. Perhaps more autonomous IDOs engage in greater strategic selection of recipient countries and sectors, placing themselves in a better position to succeed. While this selection effect would be a channel from autonomy to differential IDO project success, it would be one that meant more autonomous IDOs were not in fact more successful than their less autonomous peers in actually delivering projects in more unpredictable environments relative to their own performance in more predictable environments.

To explore selection I construct a parallel dataset with the number of observations from each IDO in each country in each sector in each year. If indeed differential IDO autonomy is working via selection, we should see differential presence or absence of projects by level of environmental unpredictability. Table 5 replicates Table 2's regression model, but substitutes the number of projects completed in each IDO-country-sector-year as the dependent variable.[44] There are over 900,000 unique IDO-country-sector-years, allowing quite a bit of precision in this selection estimate.

[44] Due to computational limitations, "sector" for the purposes of Table 5 are the 3-digit, rather than 5-digit, CRS sector codes.

[Table 5 Here]

Table 5 finds no selection along the main dimension of inquiry, the interaction between environmental unpredictability and IDO autonomy. This suggests that IDO selection of sectors and/or countries is not a systematic problem for this analysis.

The online appendix provides a range of additional tests. To partially summarize, it does not seem to be quirks of measurement or subtle features of the construction of any of the key measures that are driving results. Using the survey measure of IDO autonomy or a principal components approach does not alter findings. The appendix also presents the IDO-by-IDO statistics on the relationship between environmental unpredictability and project success, explores variance in project success and how this varies with level of environmental unpredictability and IDO autonomy, and presents IDO-by-IDO summary statistics.

Qualitative Illustrations: Comparing USAID and DFID Authorizing Environments and Their Impact

To further investigate the relationship between political authorizing environments, IDO autonomy, environmental unpredictability, and project success I conducted eight case studies examining US Agency for International Development (USAID) and UK Department for International Development (DFID) projects in Liberia and South Africa.[45] These case studies allow a direct comparison of IDO performance, going beyond the intra-agency comparisons to which the quantitative analysis is limited. While I cannot do justice to the richness of the case study data here, I believe these data can help illustrate the mechanisms theorized above. This section discusses differences in USAID and DFID authorizing

[45] This was done as part of the research for Honig 2018.

environments, then turns to illustrating the differences in level of principal control engaged in by USAID and DFID in a representative case in which both agencies were pursuing a similar goal (building municipal government capacity) in South Africa.

Authorizing Environments

USAID and DFID are IDOs with a clear difference in authorizing environment insecurity and constraint. The formal status of an agency and its level of formal independence have long been thought of as important features of agency independence and insulation from political oversight.[46] DFID is a separate ministry run by a cabinet-level minister, while USAID is subordinate to a cabinet secretary, reporting to the US Secretary of State. DFID has power and access that USAID does not, a sign of the relative importance and power of DFID vis-à-vis USAID.

DFID has stable budgets and strong parliamentary support for foreign aid. USAID's budget is quite unstable, with no long-term budgetary commitments and a much lower level of funding as a share of government spending. USAID's budget, unlike DFID's, involves heavy use of "earmarks" which pre-specify what funds must be used for; by one Congressional Research Service estimate earmarks comprised almost 75% of USAID's budget.[47] While tight control cannot be measured merely by the number of words devoted

[46] Huber and Shipan 2002; Carpenter 2001; Gilardi 2002.

[47] Congressional Research Service 2006, 19. This 2006 survey is the most recent comprehensive review of earmarks; this figure is an estimate of earmarks in the Foreign Operations Appropriations Act of 2005, and combines 'soft' (19.8%) and 'hard' (53.4%) earmarks. Foreign Operations budgets are put forward by a number of entities, including notably the State Department. However as USAID tends to have more earmarks than others (interviews) this estimate is more likely to be low than high.

to legislation, it is suggestive of the difference that the US foreign assistance act runs over 300 printed pages; comparable UK legislation runs fewer than 40 pages.[48]

USAID and DFID are by their own admission very different organizations as regards risk-taking. USAID describes itself as having a "conservative risk appetite"; by contrast DFID describes itself as having "a relatively high risk appetite, and [DFID] is often willing to tolerate high levels of risk where there are substantial potential benefits."[49] USAID also stands out with regards to flexibility and the use of measurement. A recent OECD peer review of USAID – essentially a report written by other IDOs regarding USAID's systems – finds that USAID's need for authorization from Washington constrains its operating flexibility.[50]

The IDO autonomy score for each IDO supports the view that USAID and DFID's differences in authorizing environment do indeed lead to different behavior. USAID has a score of .36, 30th amongst the 33 IDOs for whom Paris Declaration monitoring surveys allow the calculation of scores. DFID, on the other hand, has a score of .80, 2nd amongst the 33 IDOs.[51]

Politics and authorizing environments are much more salient for those discussing USAID than DFID government interventions. This is suggestive evidence of greater relative organizational focus on and preoccupation with its authorizing environment for USAID. In

[48] US aid expenditures still flow through the Foreign Assistance Act of 1961. This act, including amendments, runs 384 pages. US Congress Committee on International Relations and US Congress Committee on Foreign Relations 2003. For the UK there are a number of short pieces of legislation; these are Government of the United Kingdom 2002; Government of the United Kingdom 2006; Government of the United Kingdom 2014; Government of the United Kingdom 2015.
[49] DFID 2010, 6; USAID 2014, 16.
[50] OECD 2016, 59.
[51] See the appendix for a full list of scores by IDO.

the interview data from South Africa, the word "Congress" appears thirteen times; "Parliament" is mentioned only once, and by way of *contrasting* DFID with USAID.

USAID's belief that Congress does not trust the organization and fear of being on the proverbial chopping block came up frequently in interviews. This insecurity was evoked with regards to USAID's hiring practices, limiting USAID's ability to hire full time staff and thus the use of contractors in project supervisory roles.[52] It was used to explain budget unpredictability and the constant need to fight for funding.[53] In perhaps the most vivid depiction, one senior official described USAID as "under siege" from Congress, saying

> When you [USAID] have a hostile Congress and an ineffectual president which was the case pretty much since, I think the last time USAID had any true swagger was under Reagan, it has been an agency under siege for, I guess it would be going on for over thirty years now.[54]

The need for reporting looms large for USAID. As one staffer of a development contractor implementing a USAID project put it, "USAID wanted reports. USAID pushed management, and management pushed us."[55] An individual with experience at a number of development contractors described this as the pressure from USAID to "document more than do," suggesting that such tends to leave projects with "some really beautiful reports" to please funders and authorizers but less impact on the ground than might have occurred

[52] Interview 87, 9/20/13. This interview numbering scheme anonymizes specific remarks per the confidentiality offered interviewees; greater detail in the appendix.
[53] "It is all about the budget, right, the budget battle and the dream of federal agencies is that they could get, put together projects that year in and year out, programmes that year in and year out would have predictable funding, but every year the budget fight is a new adventure." (ibid)
[54] ibid.
[55] Interview 44, 6/6/13

if attention had instead focused on implementation.[56] A former senior manager of a USAID project summarized a widely echoed view in saying "USAID's focus was around meeting numbers as opposed to the impact."[57]

Key to the Aghion & Tirole model is the insight that if agents cannot use asymmetric information there is no point in gathering it. One USAID interviewee suggested that very few of her colleagues ever learned much about the world just beyond the embassy walls, as there was no way to make use of that information.[58] As one USAID official put it, the effect of the restrictions and constraint from above is to "make you cautious."[59] Insecurity breeds conservatism, the need to ensure that any action taken can be defended.

The Impact of Authorizing Environment Constraint: Differential Principal Control, and Project Success, in South Africa

DFID and USAID's projects in South African municipal governance in the mid-2000's illustrate how legitimacy-seeking reporting induced by political authorizing environments can affect IDO projects in practice in a relatively predictable environment (and thus a 'hard case' for this paper's theory). DFID's Consolidated Municipal Transformation Programme[60] and USAID's Local Governance Support Program, Phase 2 [61] both aimed to help municipalities efficiently and effectively deliver services.[62] Both focused on making local

[56] Interview 49, 5/30/13
[57] Interview 81, 8/1/13
[58] Interview 12, 6/4/13.
[59] Interview 109, 9/12/13
[60] DFID Project Reference #104886
[61] USAID Contract #674-C-00-05-00001-00
[62] DFID's CMTP stated its goal as to "consolidate accountable local democracy and pro-poor service delivery". The purpose is to promote the development of effective and efficient municipalities". (DFID 2003, p. 1) USAID's Phase 2 focused on efforts to "improve municipal planning skills and operating systems to increase effectiveness, transparency and accountability; and increase the revenue stream of selected municipalities ensuring that citizens meet their obligations." (LGSP/RTI 2006)

government more effective via capacity building by transferring knowledge to municipal staff. This capacity building was of both a management and a financial nature; both projects aimed to improve municipal accounting and billing systems and municipal debt management. How USAID and DFID delivered their interventions was quite different, however, as was their management, reporting, and design processes.

USAID settled on an initial model that delivered monitorable and measurable training to municipalities.[63] On a pre-arranged day, a trainer would arrive and hold a session, often in a conference room, for part or all of the day on the pre-arranged topic. Many municipalities were served by the project and many training sessions were delivered. Following the trainings, trainers verified that trainings had occurred and tracked how many individuals were trained.[64]

DFID's project shared the broad focus of USAID's on improving municipal functioning via skills transfer and systems building. Contractors implemented DFID's project, as they did USAID's. Unlike USAID's project, DFID's project worked primarily by embedding advisors in local municipalities. Advisors resided in the municipalities for extended periods of time to build skills and systems on an ongoing basis. DFID's advisers were ultimately in charge of project direction; *they* set the specific goals against which they

[63] USAID's intervention eventually did place advisers briefly in municipalities; on some accounts this was in response to pressure from South African officials observing the relative success of the two projects. (Interview 97, 8/15/13) These USAID advisers still faced quite specific reporting regimes based on externally verifiable data and were far less able to use soft information than their DFID project cousins.

[64] Reports from contractors (various), acquired via anonymous source but producible (with redacted names) on request. Confirmed via interview 88, 8/1/13.

were reporting. As one implementer put it, DFID's reporting was "more content-rich; it was not a numbers game."[65]

DFID and USAID both had reporting requirements for their respective projects. However DFID did not rely primarily on externally verifiable data in reporting, unlike USAID. DFID effectively put resident advisers and their soft information-laden judgments in control, something DFID not only condoned but actually explicitly designed into the project. The "price" of this greater degree of agent initiative was a lesser degree of principal control.

Meeting targets clearly served as a control in USAID's municipal governance project. Michelle Layte, the head of project implementation towards the end of the project, said indicators were chosen "because it was easier to count… but the numbers didn't tell about the impact". Layte went on to say that, while USAID had been better earlier in the project, "It was more a number chasing towards the end especially because we needed to reach our target."[66] Another interviewee described implementing the USAID project as "a numbers game… [USAID would say] we want the numbers, we want information."[67]

One USAID implementer described a clear sense inside the project implementation team that the trainings were failing.[68] The correlation between measures and ultimate outcomes broke down in USAID's municipal governance project. The training numbers weren't fabricated; trainings were occurring and individuals were attending. One USAID

[65] Interview 82, 8/1/13. The interviewee was contrasting their experiences with DFID's municipal governance project with their experiences working for a variety of other organizations, not USAID specifically.
[66] This continued in the later phases of LGSP, which did involve placing some trainers in municipalities after many years of largely fruitless training. In the later phases of the project, resident advisors would report on meetings held, guidelines drafted, and other such externally observable and verifiable indicators. Interview 82, 8/1/13
[67] Interview 82, 8/1/13
[68] Interview 88, 8/1/13.

actor described this as counting "bums on seats."[69] These targets were simply disconnected with the actual broader purpose they had been designed to serve. These measures may have had little connection to impact, but they certainly affected implementation. Target-setting constrained the behavior of field agents and their managers, precluding alteration of the project.

There is very little evidence that training under the USAID project was effective. This was the view not just of observers but also of those who actually worked to implement the project.[70] As one team member put it, "I don't think [training under the project] contributed much… because you go there, you don't have any authority over the people that you training, so if they don't cooperate you cannot say anything, you go there sometimes, they tell you that we have other priorities, we don't have time now, those kinds of things."[71]

DFID's project was by no means an overwhelming success; that said, it was substantially more successful than was USAID's. Being full time resident for the long term (2-3 years), DFID project advisors were often – though not always - able to find a way to positively influence municipal systems. Both beneficiaries and project staff reported that advisors achieved some shifts in municipal practices.[72] Multiple actors noted the

[69] Interview 86, 7/30/13
[70] e.g. interview 93, 7/18/13; interview 73, 7/19/13
[71] Interview 88, 8/1/13
[72] Interview 103, 7/30/13; interview 72, 7/31/13; interview 98, 8/14/13; interview 75, 7/25/13; interview 76, 7/29/13

permanent status of advisors in the municipality prevented the program from being "sidelined" in the way USAID's project seems to have been.[73]

The South African municipal governance case allows us to see what different levels of IDO autonomy look like in practice. USAID and DFID implemented programs with quite similar goals. They did so through rather similar contracting structures. But DFID's project exhibited far greater flexibility and use of agent initiative. Measurement and reporting via pre-specified targets played a substantial role in USAID's intervention but little in DFID's. USAID was more rule-bound, with substantial process controls and an orientation towards satisfying bureaucratic requirements.[74] DFID, by contrast, placed resident advisers in municipalities, and designed its project in a manner less tractable to control from above. DFID's project created reporting requirements that did not rely on externally verifiable and observable information. DFID's project was more successful than USAID's, a success clearly linked to the differences in how the projects were implemented and managed.

Conclusion

IDO project success is negatively impacted by environmental unpredictability. However less politically constrained IDOs see systematically lower performance declines in more unpredictable contexts than do their less constrained peers. The South Africa case study comparison provides suggestive evidence that what is true intra-organizationally is also true when comparing across organizations; constraints induced by political authorizing environment insecurity sometimes undermine comparative project success.

[73] Interview 74, 7/30/13; interview 72, 7/31/13. These actors didn't make explicit comparison to LGSP.
[74] Interview 103, 7/30/13

Variation in political authorizing environments has quite substantial potential impacts on development outcomes and consequently on developmental trajectories and conflicts. Effective delivery for a range of IDO projects is for some, but not all, IDOs precluded by political authorizing environments and the measurement and control systems to which they give rise. Constraints on agents that flow from an understandable, even laudable, desire to demonstrate results and accountability to politicians and citizens can undermine IDO performance.

In some instances output measurement and reporting may well improve organizational performance; when working in relatively predictable environments this may well be the superior strategy. However in less predictable environments, this reporting and tight principal control crowds out the organization's ability to serve its ends. The more unpredictable the environment, the more important it is for power and decision-making to sit with field agents.

From public schools to multinational firms, many organizations struggle with Aghion & Tirole's tension between principal control and agent initiative. Philippe Aghion himself, in collaboration with a number of illustrious coauthors, has recently applied his model to private firms during the Great Recession.[75] Using data from 11 OECD countries they find that private firms with more local plant manager control out-perform more centralized firms in the sectors hardest hit by the crisis. As they put it, "Higher turbulence benefits decentralized firms because the value of local information and urgent action

[75] Aghion et al. 2017.

increases."[76] The usefulness of agent initiative and ability to gather and use asymmetric (soft) information are far from an IDO-only, or even public sector, phenomenon.

This paper's analysis suggests limits to the range of where external monitoring – the workhorse solution of applications of principal-agent theory to public bureaucracy – may indeed be a workable solution. For some tasks, in some environments, it is not just that the monitoring is costly – the monitoring *itself* may have deleterious effects. If indeed it is true that tight oversight is detrimental to performance in some circumstances, public accountability as conventionally conceived may sometimes come at the expense of desired performance outcomes. Reporting may be a façade, with reporting requirements inducing agents to produce numbers at the expense of actually forwarding the broader goals of their organizations.

IDOs operate in difficult contexts, and attempt to do difficult things. They are, perhaps unsurprisingly, often unsuccessful. In some of the domains where foreign aid has the potential to make the most difference, e.g. in unpredictable fragile states, politically induced constraints on IDO autonomy make project success even less likely. This paper's findings suggest not only that we could do more to improve aid delivery, but that the move towards measurement and control in foreign aid in recent years may in some cases actually be hindering progress. The drive for measurement and quantitative results is usually framed around efficacy and value for money. If this encourages political authorizers to constrain IDOs' ability to engage in more flexible, autonomous operational strategies, well-intentioned authorizers may end up accomplishing precisely the opposite of what they intend.

[76] Ibid., abstract.

References

Aghion, Philippe, Nicholas Bloom, Brian Lucking, Raffaella Sadun, and John Van Reenen. 2017. *Turbulence, Firm Decentralization and Growth in Bad Times*.

Aghion, Philippe, and J Tirole. 1997. Formal and real authority in organizations. *Journal of political economy* 105 (1): 1–29.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Barnett, Michael N., and Martha Finnemore. 2003. The Politics, Power, and Pathologies of International Organizations. *International Organization* 53 (4): 699–732.

Bozeman, B, and G Kingsley. 1998. Risk Culture in Public and Private Organizafions. *Public Administration Review* 58 (2): 109–118.

Brehm, John O., and Scott Gates. 1999. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. University of Michigan Press.

Buntaine, Mark T. 2016. *Giving Aid Effectively*. Oxford University Press.

Buntaine, Mark T., Bradley C. Parks, and Benjamin P. Buch. 2017. Aiming at the Wrong Targets: The Domestic Consequences of International Efforts to Build Institutions. *International Studies Quarterly*: 1–60.

Bush, Sarah Sunn. 2015. *The Taming of Democracy Assistance: Why Democracy Promotion Does Not Confront Dictators*. Cambridge University Press.

Calvert, R, M McCubbins, and B Weingast. 1989. A Theory of political Control and Agency Discretion. *American Journal of Political Science* 33 (3): 588–611.

Cameron, AC, J Gelbach, and DL Miller. 2006. *Robust Inference With Multi-Way Clustering*. NBER Technical Working Paper. Cambridge, MA.

Carpenter, Daniel P. 2010. *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton University Press.

Carpenter, Daniel P. 2001. *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862-1928*. Princeton University Press.

Celasun, Oya, and Jan Walliser. 2008. Predictability of aid: Do fickle donors undermine aid effectiveness? *Economic Policy* 23 (55): 545–594.

Center for Systemic Peace. 2014. State Fragility Index.

Chandler, Alfred D. 1977. *The Visible Hand: the Managerial Revolution in American Business*. Cambridge, MA: Belknap Press.

Clist, Paul. 2016. Payment by Results in Development Aid: All that Glitters is not Gold. *The World Bank Research Observer*: lkw005.

Desai, Raj, and Homi Kharas. 2010. *The Determinants of Aid Volatility*.

DFID. 2010. *Working Effectively in Conflict-affected and Fragile Situations (Briefing Paper H: Risk Management)*.

Dunleavy, P, and C Hood. 1994. From old public administration to new public management. *Public money & management* 14 (3): 9–16.

Gailmard, Sean, and John W. Patty. 2012. *Learning While Governing: Expertise and Accountability in the Executive Branch*. University of Chicago Press.

Gailmard, Sean, and John W. Patty. 2007. Slackers and zealots: Civil service, policy discretion, and bureaucratic expertise. *American Journal of Political Science* 51 (4): 873–889.

Garicano, Luis, and Luis Rayo. 2016. Why Organizations Fail: Models and Cases. *Journal of Economic Literature* 54 (1): 137–192.

Gelb, Alan, and Nabil Hashmi. 2014. *The Anatomy of Program-for-Results: An Approach to Results-Based Aid*. Working Paper.
Ghani, Ashraf, Clare Lockhart, and M Carnahan. 2005. *Closing the Sovereignty Gap : an Approach to State-Building*.
Gilardi, Fabrizio. 2002. Policy Credibility and Delegation to Independent Regulatory Agencies: a comparative empirical analysis. *Journal of European Public Policy* 9 (6): 873–893.
Government of the United Kingdom. 2014. *International Development (Gender Equality) Act 2014*.
Government of the United Kingdom. 2015. International Development (Official Development Assistance Target) Act 2015.
Government of the United Kingdom. 2006. *International Development (Reporting and Transparency) Act 2006*.
Government of the United Kingdom. 2002. *International Development Act 2002*.
Grossman, Sanford J., and Oliver D. Hart. 1986. The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy* 94 (4): 691.
Gulrajani, Nilima. 2017. Bilateral Donors and the Age of the National Interest: What Prospects for Challenge by Development Agencies? *World Development* 96. Elsevier Ltd: 375–389.
Gulrajani, Nilima. 2011. Transcending the Great Foreign Aid Debate: managerialism, radicalism and the search for aid effectiveness. *Third World Quarterly* 32 (2): 199–216.
Hart, O, and John Moore. 1988. Incomplete contracts and renegotiation. *Econometrica: Journal of the Econometric Society* 56 (4): 755–785.
Hart, Oliver, and John Moore. 1990. Property Rights and the Nature of the Firm. *Journal of Political Economy* 98 (6): 1119–1158.
Hawkins, Darren G., and Wade Jacoby. 2006. How Agents Matter. In *Delegation and Agency in International Organizations*, 199–228.
Hawkins, Darren G., David A. Lake, Daniel L. Nielson, and Michael J. Tierney. 2006. Delegation under anarchy: states, international organizations, and principal-agent theory. In *Delegation and Agency in International Organizations*, edited by Darren G. Hawkins, David A. Lake, Daniel L. Nielson, and Michael J. Tierney, 3–38.
Holmstrom, B., and P. Milgrom. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization* 7: 24.
Hölmstrom, Bengt. 1979. Moral hazard and observability. *Bell Journal of Economics* 10, (1): 74–91.
Honig, Dan. 2018. *Navigation by Judgment: Why and When Top-Down Control of Foreign Aid Doesn't Work*. New York: Oxford University Press.
Hood, Christopher. 2004. The Middle Aging of New Public Management: Into the Age of Paradox? *Journal of Public Administration Research and Theory* 14 (3): 267–282.
Huber, John D., and Nolan McCarty. 2004. Bureaucratic capacity, delegation, and political reform. *American Political Science Review* 98 (3): 481–494.
Huber, John D., and Charles R Shipan. 2006. Politics, Delegation, and Bureaucracy. In *The Oxford Handbook of Political Economy*, edited by Robert E. Goodin, 256–272. Oxford University Press.
Huber, John D., and Charles R. Shipan. 2002. *Deliberate Discretion: The Institutional*

Foundations of Bureaucratic Autonomy. Cambridge University Press.
Hupe, Peter, and Michael Hill. 2007. Street-Level Bureaucracy and Public Accountability. *Public Administration* 85 (2): 279–299.
Institute of Development Studies. 2014. Conflict and Fragility. Available from <http://www.ids.ac.uk/idsresearch/conflict-and-fragility>.
Johns, Leslie. 2007. A Servant of Two Masters: Communication and the Selection of International Bureaucrats. *International Organization* 61 (2): 245–275.
Johnson, Tana, and Johannes Urpelainen. 2014. International Bureaucrats and the Formation of Intergovernmental Organizations: Institutional Design Discretion Sweetens the Pot. *International Organization* 68 (1): 177–209.
Kauppi, Katri, and Erik M. Van Raaij. 2015. Opportunism and Honest Incompetence - Seeking Explanations for Noncompliance in Public Procurement. *Journal of Public Administration Research and Theory* 25 (3): 953–979.
Keiser, Lael R. 1999. State Bureaucratic Discretion and the Administration of Social Welfare Programs: The Case of Social Security Disability. *Journal of Public Administration Research and Theory* 9 (1): 87–106.
Kerr, S. 1975. On the Folly of Rewarding A, While Hoping for B. *Academy of Management journal* 18 (4): 769–783.
Leuffen, Dirk, S Shikano, and S Walter. 2012. Measurement and data aggregation in small-n social scientific research. *European Political Science*: 1–20.
Lipsky, Michael. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. Russell Sage Foundation.
Lynn Jr, Laurence E. 1998. The New Public Management: How to Transform a Theme into a Legacy. *Public Administration Review* 58 (3): 231–237.
Martin, Lisa. 2006. Distribution, Information, and Delegation to International Organizations: The Case of IMF Conditionality. In *Delegation and Agency in International Organizations*, edited by Darren G. Hawkins, David A. Lake, Daniel L. Nielson, and Michael J. Tierney, 140–164. New York: Cambridge University Press.
Meyer, JW, and Brian Rowan. 1977. Institutionalized organizations: Formal structure as myth and ceremony. *American journal of sociology* 83 (2): 340–363.
Modell, S. 2004. Performance measurement myths in the public sector: a research note. *Financial Accountability & Management* 20 (February): 39–56.
Natsios, A. 2010. *The Clash of the Counter-bureaucracy and Development*. Washington, DC.
Nielson, Daniel L., and Michael J. Tierney. 2003. Delegation to International Organizations: Agency Theory and World Bank Environmental Reform. *International Organization* 57 (2): 241–276.
OECD. 2012. *Aid Effectiveness 2011: Progress in Implementing the Paris Declaration*. OECD Publishing.
OECD. 2000. DAC Criteria for Evaluating Development Assistance Factsheet: 2.
OECD. 2016. *OECD Development Co-operation Peer Reviews: United States 2016*. Paris: OECD Publishing.
OECD. 1991. *The DAC Principles for the Evaluation of Development Assistance*.
Oliver, C. 1991. Strategic Responses to Institutional Processes. *Academy of management review* 16 (1): 145–179.
Pepinsky, Thomas B, Jan H Pierskalla, and Audrey Sacks. 2016. Bureaucracy and Service Delivery. *Available at SSRN: http://ssrn.com/abstract=2798912*: 1–47.

Perakis, Rita, and William Savedoff. 2015. *Does Results-Based Aid Change Anything? Pecuniary Interests, Attention, Accountability and Discretion in Four Case Studies*. CGD Policy Paper.
Pfeffer, Jeffrey, and Gerald R. Salancik. 1978. *The External Control of Organizations: A Resource Dependence Perspective*. Stanford University Press.
Polanyi, Michael. 1966. *The Tacit Dimension*. University of Chicago Press.
Prendergast, C. 2001. *Selection and Oversight in the Public Sector, with the Los Angeles Police Department as an Example*.
Rasul, Imran, and Daniel Rogger. 2013. *Management of Bureaucrats and Public Service Delivery : Evidence from the Nigerian Civil Service*.
Scott, James C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.
Shavell, Steven. 1979. Risk sharing and Incentives in the Principal and Agent Relationship. *Bell Journal of Economics* 10 (1): 55–73.
Singh, JV. 1986. Peformance, Slack, and Risk Taking in Organizational Decision Making. *Academy of management Journal* 29 (3): 562–585.
Stein, JC. 2002. Information production and capital allocation: Decentralized versus hierarchical firms. *The Journal of Finance* LVII (5).
Swedlund, Haley J. 2017. Can foreign aid donors credibly threaten to suspend aid? Evidence from a cross-national survey of donor officials. *Review of International Political Economy*. Taylor & Francis.
Tirole, Jean. 1994. The Internal Organization of Government. *Oxford Economic Papers* 46 (1): 1–29.
US Congress Committee on International Relations, and US Congress Committee on Foreign Relations. 2003. *Legislation on Foreign Relations Through 2002*.
USAID. 2014. *Public Financial Management Risk Assessment Framework (PFMRAF) Manual: A Mandatory Reference for ADS Chapter 220*.
Weijer, Frauke De. 2012. *Rethinking approaches to managing change in fragile states*.
Williamson, Oliver E. 1983. *Markets and Hierarchies*. Free Press.
Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do And Why They Do It*. Basic Books.
Woods, Ngaire. 2006. *The Globalizers: The IMF, the World Bank, and their Borrowers*. Cornell University Press.
World Bank. 2006. *Engaging with Fragile States: An IEG Review of World Bank Support to Low-income Countries Under Stress*. World Bank Publications.
World Bank. 2017. World Bank Projects & Operations Database.
Wynen, Jan, and Koen Verhoest. 2016. Why Do Autonomous Public Agencies Use Performance Management Techniques ? Revisiting the Role of Basic Organizational Characteristics 7494 (June).
2005. The Paris Declaration on Aid Effectiveness. Paris, France: Avaliable Online at http://www.oecd.org/dataoecd/11/41/34428351.pdf.

Tables

Variable	Obs	Mean	Std. Dev.	Min	Max
Overall Project Success (6 pt scale)	9312	4.304	1.139	1	6
Environmental Unpredictability (State Fragility Index)	9312	12.522	4.992	0	25
Project Size (USD Millions)	7247	41.114	102.308	.004	4015
IDO Autonomy (from Paris Declaration monitoring survey)	9312	.659	.075	.559	.799

Table 1: Summary Statistics of Key Variables

DV: Project Success (6-pt scale)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Environmental Unpredictability	-0.170***	-0.171***	-0.149***	-0.147***	-0.112***	-0.107***	-0.0868**
	(0.0307)	(0.0328)	(0.0403)	(0.0414)	(0.0351)	(0.0351)	(0.0391)
Env Unpred*IDO Autonomy	0.205***	0.206***	0.187***	0.180***	0.113**	0.104**	0.107*
	(0.0464)	(0.0467)	(0.0583)	(0.0599)	(0.0547)	(0.0515)	(0.0590)
Project Size (USD Millions)		0.000585***		0.000413***		0.000566**	
		(0.000162)		(0.000135)		(0.000218)	
Constant	4.423***	4.372***	3.807***	3.789***	5.764***	5.780***	4.489***
	(0.118)	(0.125)	(0.204)	(0.224)	(0.115)	(0.120)	(0.215)
IDO Fixed Effects	Y	Y	Y	Y	Y	Y	Y
Recipient Fixed Effects	N	N	Y	Y	N	N	Y
Sector Fixed Effects	N	N	N	N	Y	Y	Y
R^2	0.099	0.114	0.147	0.165	0.154	0.184	0.207
Observations	9312	7247	9312	7247	7370	5446	7370

Standard errors in parentheses, clustered by recipient country

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: IDO Autonomy Mediates the Relationship Between Environmental Predictability and IDO Project Success. While all IDOs see performance decline as environments become less predictable, more autonomous (less constrained) IDOs have substantially smaller declines. Ordinary least squares (OLS) regression.

DV: Project Success (6-pt scale)	(1)	(2)	(3)	(4)
Environmental Unpredictability	-0.169***	-0.167***	-0.0904*	-0.0843*
	(0.0316)	(0.0314)	(0.0482)	(0.0495)
Env Unpred*IDO Autonomy	0.204***	0.200***	0.145**	0.137**
	(0.0475)	(0.0471)	(0.0664)	(0.0681)
Constant	4.401***	4.222***	3.953***	3.783***
	(0.146)	(0.205)	(0.347)	(0.345)
IDO Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	N	N
Year*IDO Fixed Effects	N	Y	N	N
5-yr 'bin' Fixed Effects	N	N	Y	Y
Recipient Fixed Effects	N	N	Y	Y
Recipient* 5-yr bin FEs	N	N	Y	Y
IDO*5-yr bin FEs	N	N	N	Y
R^2	0.101	0.116	0.207	0.211
Observations	9312	9312	9312	9312

Standard errors in parentheses, clustered by recipient country

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Expanding Fixed Effects for Robustness. Controlling for time, or time interacted with recipients or IDOs, does little to change the main effects.

	(1) 6pt scale	(2) Z-score	(3) 6pt scale	(4) Z-score
Environmental Unpredictability	-0.165***	-0.133***	-0.145***	-0.117***
	(0.0323)	(0.0308)	(0.0418)	(0.0390)
Env Unpred*IDO Autonomy	0.199***	0.159***	0.177***	0.138**
	(0.0489)	(0.0479)	(0.0602)	(0.0572)
Internal Evaluator	-9.763**	-13.36	-8.928*	-12.88
	(4.319)	(18660.3)	(5.280)	(3337.9)
Independent Eval Office	3.441***	-2.361	3.883***	-2.244
	(1.284)	(14334.5)	(1.301)	(3218.9)
Internal Eval*IDO Autonomy	14.94**	20.28	13.75*	19.60
	(6.691)	(25267.8)	(8.130)	(3705.2)
Independent Eval*IDO Autonomy	-5.522***	3.169	-6.160***	3.014
	(2.005)	(26193.4)	(2.042)	(5635.5)
IDO Autonomy		-4.665		-5.015
		(18067.1)		(5003.6)
Constant	4.529***	3.679	4.140***	3.647
	(0.110)	(12722.8)	(0.213)	(3073.5)
IDO Fixed Effects	Y	N	Y	N
Recipient Fixed Effects	N	N	Y	Y
R^2	0.103	0.029	0.154	0.084
Observations	7722	7722	7722	7722

Standard errors in parentheses, clustered by recipient country

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Controlling for Evaluation Type. The third, omitted, category is externally contracted evaluators. Models 2 and 4 take an IDO-specific z-score of the dependent variable to allow the IDO fixed effect to be dropped, and thus the base term of IDO Autonomy to be added to the models.

DV: # of observations by IDO-country-sector-year	(1)	(2)	(3)	(4)
Environmental Unpredictability	-0.000102	-0.000847	-0.000102	-0.000847
	(0.000491)	(0.000475)	(0.000491)	(0.000475)
Env Unpred*IDO Autonomy	0.000594	0.000594	0.000594	0.000594
	(0.000666)	(0.000666)	(0.000666)	(0.000666)
Constant	0.00192	0.0174***	-0.00474***	0.0107*
	(0.00129)	(0.00411)	(0.00118)	(0.00416)
IDO Fixed Effects	Y	Y	Y	Y
Recipient Fixed Effects	N	Y	N	Y
Sector Fixed Effects	N	N	Y	Y
R^2	0.007	0.014	0.015	0.021
Observations	957096	957096	957096	957096

Standard errors in parentheses, clustered by recipient country

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: IDO Project Selection. IDOs of different levels of autonomy do not differentially select into more or less unpredictable recipient country environments.

Figures

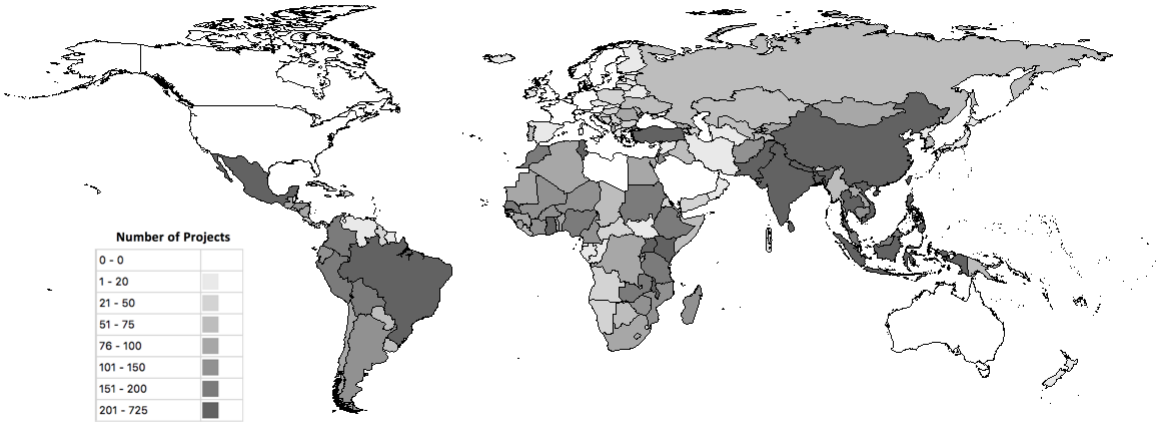

Figure 1: Overview of Projects in Dataset

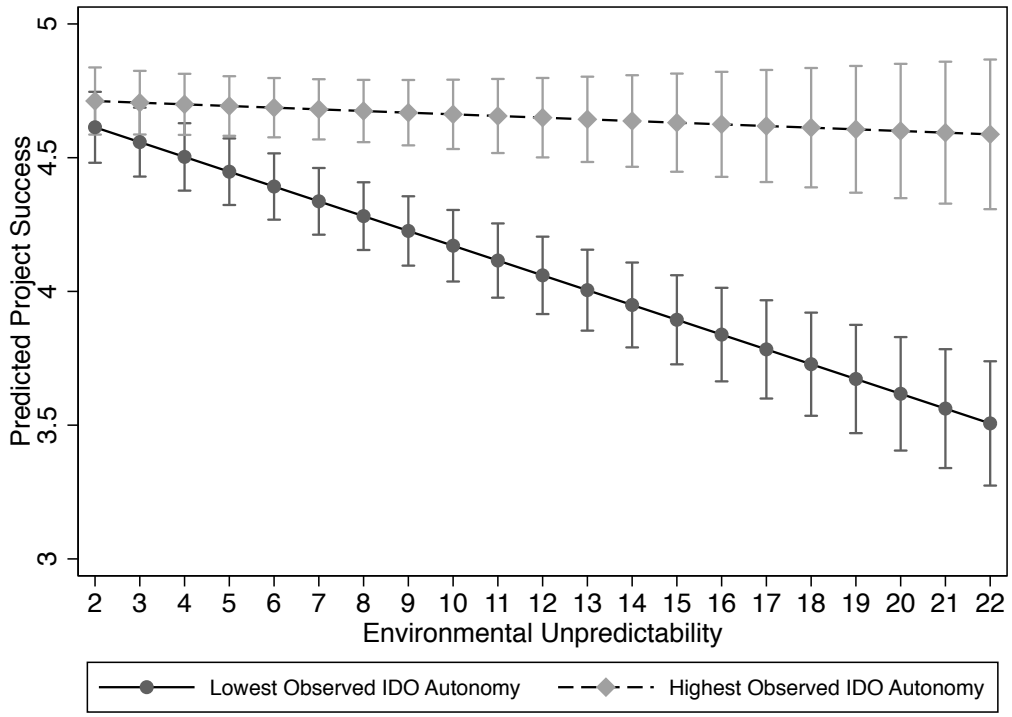

Figure 2: Returns to Autonomy in Countries of Differential Predictability. More autonomous (less constrained) IDOs see much more consistent performance than do less autonomous (more constrained) IDOs as environmental predictability varies. Bars around each point estimate represent 95% confidence intervals. The "lowest observed" autonomy score is .56 on a zero to one scale, the "highest observed" is .80. These represent the extremes of autonomy in the sample data. See the online appendix for a full list of autonomy scores.

Honig When Reporting Undermines Performance Online Appendix: Data Collection Methods and Additional Empirics

Dan Honig

March 10, 2018

This document is the online appendix referred to in Honig, D. (2019). When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation. *International Organization*, Forthcoming Winter 2019.

Data Collection

There is no existing cross-IDO database of project outcome data. The project success data therefore had to be collected from each IDO in the sample individually. I pursued project success data from every OECD bilateral aid agency in the top 10 in terms of the volume of official development assistance aid delivered directly in 2010, the last available data when this research commenced. This includes agencies in the US, Germany, the UK, France, Japan, Canada, Norway, Australia, Sweden, and Denmark. I also pursued data from all of the biggest multilateral aid agencies (the European Commission, UN Development Programme, World Bank, African and Asian Development Banks, and Global Fund), as were other agencies with which I had links (e.g. Irish Aid, International Fund for Agricultural Development, Food and Agriculture Organization, and International Monetary Fund).

There were two basic reasons to exclude IDOs from the sample. First, many IDOs do not in fact assign an overall, holistic success rating to projects ex-post. Second, for some IDOs I could not get access to outcome data that does exist (e.g. the African Development Bank). The IDOs included in this analysis are a convenience sample, raising concerns regarding broader generalizability. To the extent that the willingness to make data public, or the agency's decision to give projects an overall success rating, are plausibly correlated to an agency's autonomy this is a threat to generalizability that must be considered in examining these quantitative results in isolation (that is, without incorporating the case study findings). Table I.12 suggests there is cause for concern, as none of the bottom ten IDOs in autonomy are included in this analysis. It seems plausible that the least autonomous agencies, those with the least stable relationships with their political authorizing environments, are less likely to collect and/or make public information that might cast some of their projects in a less than stellar light.

The most straightforward result of this under-sampling would be to reduce the power of the quantitative tests; it is harder to imagine how this might lead to spurious findings. Spurious findings would result if the "true" shape of the relationship between autonomy and project success were parabolic. This seems most likely in the sense that the most autonomous agencies might engage in "too much" autonomy; however there is a good sampling of the "top" of the IDO distribution as regards autonomy. While seeming unlikely theoretically, if those with modest degrees of autonomy e.g. fared even worse than those with the lowest degrees of autonomy as environmental predictability role, this parabolic relationship (with the extremes of autonomy both faring better than the middle) would be missed due to the lack of data availability for the least autonomous IDOs.

The nine IDOs included are the the World Bank (WB), Asian Development Bank (AsDB), the UK's Department for International Development (DFID), the European Commission (EC), the Global Fund for AIDS, TB, and Malaria (GFATM), the German Agency for Technical Cooperation (GiZ), the German Development Bank (KfW), the International Fund for Agricultural Development (IFAD), and the Japanese International Cooperation Agency (JICA). Of the nine IDOs included only the World Bank's information is publicly accessible. The Asian Development Bank and DFID released data following formal public information requests. The European Commission and KfW released data under confidentiality agreements that limited their disclosure and further use. KfW later waived its confidentiality provision, allowing its data to be included in the publicly posted data that accompanies this paper; the EC declined to do so.

GiZ, IFAD, and JICA all maintain publicly accessible archives of individual project evaluation documents. In converting these individual project documents into a usable database I contracted research assistants using the online job contracting platform Odesk. RAs speaking the appropriate language (English, German, or Japanese as appropriate) extracted the relevant data – project names, performance scores, start and completion dates, budgets, etc. – from source documents, with me selectively double-checking their work (in the case of foreign language documents, with the help of Google translate). After compilation of each IDO's data I sent to each excel spreadsheets containing each agency's data were sent back to the originating agency for comment and/or correction. GiZ was kind enough to respond with a handful of minor corrections, which were incorporated. JICA had no substantive comment on the data itself, but wished it to be made clear that these data were generated by me rather than by JICA, which bears no responsibility for

errors or omissions. I hereby note that is the case, with all JICA data unofficial and unverified. IFAD never responded to multiple queries.

Archival Work on Project Success

As mentioned in the main text, I engaged in archival work to examine the documents underlying project documents. The World Bank uniquely allows access to archived primary project documents.[1] These documents include correspondence between project staff and between World Bank staff and national governments, back-to-office reports and (often handwritten) notes by those monitoring projects, detailed financial and performance indicators, and detailed evaluation reports. For approximately a dozen projects I reviewed archival documents at length, focusing on cases in which similar projects (such as the first and second phases of a particular project in a particular country) received quite different ratings and one might therefore be particularly doubtful about the reliability of ratings. In reviewing the archival documents (which in every case occurred many months after identifying the projects to be reviewed), I intentionally proceeded without knowledge of which projects were more or less successful and attempted to generate my own rating from the primary documentation. I cannot say that my rating on a six-point scale always matched the World Bank's score precisely. Indeed, this would be troubling if true, since evaluators also engage in conversations with project personnel, recipient government officials, and project beneficiaries, transcripts of which are not included in the archived files. However, there were no cases in which my self-generated rating differed by more than one point from the World Bank's official rating on a six point scale. In short, in this small sample success and failure do seem to be different and do map onto real features of the projects.

Construction of the Paris Declaration Monitoring Survey-Derived Measure of IDO Autonomy

As mentioned in the main text, the primary IDO autonomy measure employed in this work is a composite of two scales, one focusing on authorizing environment insecurity and the other on IDO propensity to devolve control over project implementation. The authorizing environment insecurity measure is constructed from two indicators. These indicators are, first, the degree to which aid is untied; that is, the extent to which it is not required that funds be spent on goods and services produced by the donor country. A high level of tied aid is a sign of an IDO's need to build political consensus for aid by serving domestic political constituencies and thus reflects more insecure footing in the IDO's political authorizing environment. The second indicator is the predictability of aid. The Paris

[1] Access to these documents, which require an extended vetting and declassification process, is via the World Bank Group Archives. These documents can be accessed by making requests under the World Bank's Access to Information policy.

Declaration asked donors for the first time to report formal projections of disbursements for future years; the monitoring surveys compare the last (that is, most recent) ex-ante projection of aid spending in a given year to the actual volume of aid disbursed.[2] Previous scholarship suggests that deviations from estimated sums are linked to IDO funding insecurity and political interference in IDO funding levels and direction.[3]

The propensity to devolve control measure is constructed from three indicators examining an IDO's project implementation behaviors. There is no available measure of IDO behavior with regards to their own agents; there are, however, systematic measures of IDO behavior as regards recipient country governments, and the frequency with which IDOs let go of principal control in favor of implementation led by these governments. Many of the same factors that I theorize drive IDOs' inappropriate retention of principal control vis-à-vis their agents – e.g. a worry about reputational risk and a desire to ensure short-term delivery is successful at the expense of long-term development goals – should also reduce an IDOs' propensity to hand over substantive control to developing country governments. I use IDO control tendencies toward recipient governments as a proxy for IDO control tendencies towards their own agents. The specific measures employed are the use of recipient-country public financial management (PFM) systems; the use of recipient-country procurement systems; and the avoidance of parallel implementation units.[4]

Additional Robustness

Additional Summary Statistics: Project Success

Tables II.1 and II.2 provide additional information regarding the key dependent variable, overall project success. Project success is,"inflated" to a six point scale from whatever the likert-type base scale is for each donor. This has no implication for the econometrics so long as IDO fixed effects are employed, but makes interpretation of the results more intuitive.

---

[2] This is a slight simplification; the indicator also penalizes over-disbursement, in fact calculating something like the absolute value of the deviation from prediction. In addition, disbursements are as-reported by partner government, adding inaccuracy borne of partner government data systems. See Ibid., 73–74 for more detail.

[3] Celasun and Walliser 2008; Desai and Kharas 2010.

[4] Parallel implementation units are separate operating units established at donors' insistence. These units use donor standards and thus give donors more control than would the routing of funds fully through recipient country government systems.

IDO	count	mean	sd	min	max
AsianDB	999	4.007508	1.036263	1.5	6
DFID	1795	4.610808	.9684172	1.2	6
EC	586	4.067406	.9810926	1.5	6
GFATM	538	4.750929	1.229771	1.5	6
GiZ	108	4.407407	.9175041	2	6
IFAD	31	4.16129	.7347006	2	5
JICA	672	4.984375	1.188046	1.5	6
KfW	1052	4.223384	1.02328	1	6
WB	3531	4.09544	1.18068	1	6
Total	9312	4.303898	1.138767	1	6

Table I.1: Summary Statistics of Project Success by Donor (6-point scale)

Perhaps an even more intuitive way of thinking about these data is to think of them as z-scores, given that – once IDO fixed effects are taken – each project is essentially being compared to the distribution of a given IDO's other projects. In the robustness checks in the paper and below I also drop the IDO fixed effect from regression models, instead using the z-score as the dependent variable.

IDO	count	mean	sd	min	max
AsianDB	999	.1100482	.9887898	-2.282585	2.011261
DFID	1795	-.0085493	.9993619	-3.528346	1.425033
EC	586	-.0084486	1.004518	-2.637156	1.970289
GFATM	538	.0025107	1.002199	-2.646827	1.020438
GiZ	108	-.0707946	1.050646	-2.827548	1.752903
IFAD	31	-.0282393	1.010707	-3.001466	1.12555
JICA	672	-.0083597	.9893771	-2.910065	.8374288
KfW	1052	.052686	.9197903	-2.8447	1.649623
WB	3531	-.0162114	.9279303	-2.449007	1.480637
Total	9312	.008058	.9637126	-3.528346	2.011261

Table I.2: Summary Statistics of Project Success by Donor (Z-scores)

Direct Effect of IDO Autonomy on Outcomes in Primary Analysis

The models in Table 2 do not incorporate a base term for IDO autonomy as it is collinear to IDO fixed effects. Table I.3 replicates Table 2 incorporating IDO autonomy and dropping IDO fixed effects. Results are substantively identical.

	(1) Project Success (Z-score)	(2) Project Success (Z-score)	(3) Project Success (Z-score)	(4) Project Success (Z-score)
IDO Autonomy	-1.559**	-2.033***	-1.725**	-2.016***
	(0.651)	(0.589)	(0.752)	(0.714)
Environmental Unpredictability	-0.127***	-0.146***	-0.123***	-0.132***
	(0.0297)	(0.0299)	(0.0359)	(0.0380)
Env Unpred*IDO Auton	0.148***	0.174***	0.151***	0.161***
	(0.0462)	(0.0446)	(0.0545)	(0.0562)
Project Size (USD Millions)		0.000562***		0.000353**
		(0.000182)		(0.000141)
Recipient Fixed Effects	N	N	Y	Y
R^2	0.025	0.024	0.077	0.080
Observations	9312	7247	9312	7247

Standard errors in parentheses, clustered by recipient country

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table I.3: Adding Base Term for IDO Autonomy to Table 2. Running regressions without IDO fixed effects but with the "base" autonomy scale leaves the key results on the interaction term substantively unchanged.

Validity of IDO Autonomy Measure

- Principal Components Analysis

The main text explained the construction of the primary IDO autonomy measure, and my decision to use a simple average of the five component measures drawn from the Paris Declaration Monitoring Surveys rather than a principal components approach. A principal components analysis of these five measures – aid predictability, untied aid, use of parallel implementation units, use of country public financial management systems, and use of country procurement systems - yields a first principal component with an eigenvalue of 3.09, thus explaining 62% of the variance in the five measures. This first principal component has quite even loading across the five constituent measures. The second component has eigenvalue of 1.08, just barely above the traditional cutoff of 1. Figure I.1 presents the scree plot.

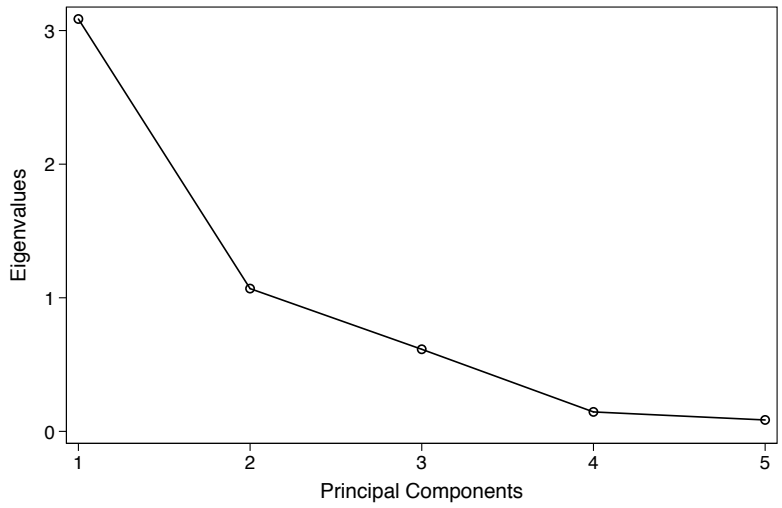

Figure I.1. Scree Plot of Principal Component Eigenvalues from IDO Autonomy Scale Measures

The second principal component, then, is quite marginal to begin with. Table I.4 examines the loading of the variables onto the first three principal components. The loading makes clear that the second component is picking up devolution propensity (with all three of the measures that form part of that subscale positive) where it does not overlap with authorizing environment insecurity (with both the measures that form that subscale negative). Thus combining the two principal components will lead to an overemphasis on devolution propensity relative to authorizing environment.

Variable	Component 1	Component 2	Component 3
Use of PIUs	.2796	.6668	.6767
Aid Predictability	.5254	-.1996	.1177
Use of Country PFM	.5339	.1003	-.2800
Use of Country Procurement	.4980	.1905	-.5030
Untied Aid	.3358	-.6849	.4436

Table I.4: Loading of IDO Autonomy measures onto Principal Components

Using only the first principal component struck me as quite similar, but much less intuitive, then simply averaging the five measures. A cluster analysis (via Stata's clv command) suggests what is implied by both the principal components analysis and intuition, that a single cluster with all five measures – that is, a single scale – is most appropriate here. As such I construct a simple average; the Cronbach's alpha of this scale (.825) suggests to me that this simple averaging is reasonable.

Nonetheless, I do retain the first principal component in the data to allow a robustness check; Table I.5 displays the results, which show the same effect as does the measure of autonomy employed in the primary results, e.g. Table 2 (in fact, t-statistics are higher using this principal components approach than with the primary measure).

DV: Project Success	(1) 6 pt scale	(2) Z-score	(3) 6 pt scale	(4) Z-score
Environmental Unpredictability	-0.0587***	-0.0456***	-0.0466***	-0.0393***
	(0.00716)	(0.00626)	(0.0134)	(0.0111)
Env Unpred*IDO Autonomy (Principal Component)	0.0199***	0.0142***	0.0171***	0.0125**
	(0.00517)	(0.00514)	(0.00650)	(0.00612)
IDO Autonomy (Principal Component)		-0.146**		-0.129
		(0.0717)		(0.0833)
Constant	4.341***	0.530***	3.758***	0.510**
	(0.126)	(0.0771)	(0.207)	(0.209)
IDO Fixed Effects	Y	N	Y	N
Recipient Fixed Effects	N	N	Y	Y
R^2	0.098	0.024	0.146	0.076
Observations	9312	9312	9312	9312

Standard errors in parentheses, clustered by recipient country

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table I.5: Results Using First Principal Component Instead of Simple Average for Paris Declaration Monitoring Survey-derived Scale

- Using my Alternate Field Survey Measure

One might be concerned that the Paris Declaration monitoring survey-derived IDO autonomy measure is not actually mapping autonomy. As noted in the main text, I conducted a small survey of aid experts in the field regarding IDO autonomy. A typical role for one of the survey respondents would be a senior position in the aid management unit of a recipient government's ministry of finance. Respondents rated a number of development agencies (including but not limited to those in the sample) on a scale of 1 to 7 in response to the following question:

> To what degree do you believe the in-country field office/bureau of the agencies listed below (presented in random order) are enabled to make decisions with a significant impact on the direction, nature, or quality of development projects? Please only respond for those agencies you have had exposure to either via working with the agencies or discussions with colleagues.

The survey N is 28, with varying coverage for different donors. This is a small but well-informed sample; methodological studies suggest small numbers of high-quality respondents will prove more accurate than significantly larger samples that

lack expertise.[5] The N of 28 is the remaining N after removing surveys which were not substantively responsive or gave indications of nonsensical answers; the two largest reasons for exclusion were (a) rating the Asian Development Bank despite stating that all relevant development-related work experience was in an African country (where the Asian Development Bank does not function) or (b) rating the survey's anchoring vignettes such that the most autonomous text was evaluated as being just as autonomous or less autonomous than the least autonomous text. The survey N is limited by the small number of individuals in any given country who can make expert inter-donor comparisons (this generally excludes employees of development agencies, who can only speak intelligently regarding their own organization).

The correlation between this survey measure and the autonomy scale drawn from the Paris Declaration surveys is .73. Table I.6 substitutes the survey measure for that of the Paris Declaration-derived measure, otherwise paralleling the analysis of Table 2; the results are similar, which should increase confidence in the primary analysis.

	(1)	(2)	(3)	(4)
DV:	6pt scale	Z-score	6pt scale	Z-score
Environmental Unpredictability	-0.102***	-0.0852***	-0.0760**	-0.0724***
	(0.0255)	(0.0228)	(0.0326)	(0.0269)
Env Unpred*IDO Autonomy (Survey)	0.0170***	0.0146**	0.0123*	0.0119*
	(0.00641)	(0.00582)	(0.00713)	(0.00622)
Autonomy (Survey)		-0.142*		-0.132
		(0.0762)		(0.0817)
addlinespace Constant	4.941***	0.892***	4.489***	0.974***
	(0.102)	(0.287)	(0.189)	(0.354)
IDO Fixed Effects	Y	N	Y	N
Recipient Fixed Effects	N	N	Y	Y
R^2	0.094	0.021	0.142	0.072
Observations	8313	8313	8313	8313

Standard errors in parentheses, clustered by recipient country

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table I.6: Robustness to Use of Survey Measure.

Outcome Variance

One might be worried that results are driven by quirks in the variance of outcomes. Table I.7 examines this concern in a simple nonparametric manner, by dividing environmental predictability and autonomy scores at their respective

[5] Leuffen, Shikano, and Walter 2012.

means and then examining the variance in project success z-score by autonomy and environmental predictability quadrant, and finds no substantively large differences. By calculation (see Table I.2), the Z-score outcome measure has a mean near 0 and standard deviation 1 for each IDO. Table I.7 allows us to examine if the variance in this measure differs systematically along the autonomy and environmental predictability axes, thus potentially distorting the interpretation of OLS results. The question, then, is whether any of the quadrants deviate substantially enough from 1 to cause concern. Both low autonomy and high autonomy IDOs do better in contexts of lower environmental unpredictability; the gap between low- and high-SFI contexts is larger for low-propensity to IDOs (approximately .39 SD) than for high-propensity to IDOs (.17 SD).[6]

	Low Autonomy IDOs	High Autonomy IDOs
Low environmental unpredictability	.163 (.863)	.123 (.969)
High environmental unpredictability	-.226 (.998)	-.047 (.983)

Table I.7: Analysis by IDO Autonomy and Environmental Unpredictability Quadrant.

Features of the modeling (e.g. Overfitting concerns, 2nd-level N distortions, etc.)

One might also worry, particularly given the small number of IDOs in this multilevel model, results are driven by features of the modeling. To address this concern, I first calculated the simple mean of project success (expressed as a z-score) for each IDO for projects above and below the mean of environmental unpredictability, yielding eighteen observations (two per IDO). I then calculated the gap between each IDO's high unpredictability and low unpredictability project mean success (thus leaving one observation per IDO). I then used this gap as the dependent variable in a regression with only a single explanatory variable, IDO autonomy. This result is Table I.8 below. There remains a clear relationship between IDO autonomy and the impact of environmental unpredictability on performance, significant at the 90% confidence interval (and just short of

[6] Given the large N, the analysis can of course confirm that that these variances are not equal; the question is whether they are substantively different enough to potentially bias results. I would argue the answer to this is in the negative.

significance at the 95% level, with a t-statistic of 2.32 but only nine observations). The R-squared is also .43, suggesting IDO autonomy explains a great deal of the difference in this gap in comparing IDOs. [7]

	(1) Gap Between High and Low Unpredictability Success by IDO
Autonomy (PD Scale)	1.319^{+}
	(0.569)
Constant	-1.177*
	(0.375)
R^2	0.434
Observations	9

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < .01$

Table I.8: Difference in Difference (in Z-scores) of Gap Between High and Low Unpredictability Project Success (split at mean of Environmental Unpredictability)

Table I.9 below further examines the underlying relationship between IDO autonomy and project success at the IDO-by-IDO level, summarizing the relationship between environmental unpredictability and overall project success for each donor in isolation; that is, using only data from one donor at a time and implementing nine different regressions.[8] In each case, the model is of the form

$$\text{Project Success}_{i,j} = \beta_1 * \text{Environmental Unpredictability (State Fragility Index)}_j + \varepsilon_i.$$

[7] The coefficient is positive even though the gap between high unpredictability and low unpredictability projects is smaller as IDO autonomy rises because the dependent variable (the gap itself) is always negative; every IDO has less success in high unpredictability environments than low unpredictability environments. For more autonomous IDOs this is a smaller negative number than for less autonomous IDOs.

[8] This is intuitively similar to a rank-based regression.

IDO	Autonomy Score	Correlation between Env Unpred & Success (Z-score) for this donor with only this donor's data in regression
EC	.559	-0.0249**
		(0.0103)
Global Fund	.594	-0.0471***
		(0.0112)
World Bank	.608	-0.0365***
		(0.0043)
GiZ	.666	-0.0525***
		(0.0175)
KfW	.666	-0.0331***
		(0.0101)
JICA	.667	-0.0221
		(0.0133)
Asian DB	.669	-0.0671***
		(0.0217)
IFAD	.681	-0.0183
		(0.0362)
DFID	.799	-0.0019
		(0.0084)

Standard errors in parentheses, clustered by recipient country
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table I.9: IDO-by-IDO Regressions. This table allows a direct examination of the 2nd level N that drives results. IDOs with lower levels of autonomy see a greater negative correlation between environmental unpredictability and project success.

As expected, greater environmental predictability has a more negative and statistically significant relationship with overall project success for less autonomous donors. This confirms—using an approach that does not rely on the parameterization of the interaction term—that higher levels of autonomy mitigate the inverse relationship between the environmental unpredictability measure (the State Fragility Index) and overall project success. A figure presenting the data underlying table I.9 is presented as Figure I.2 below. [9]

[9] Credit to Chris Kilby, who as a discussant at NEUDC 2014 first generated this graph (that is, the graph is generated by me, but is inspired by a similar graph generated by Kilby).

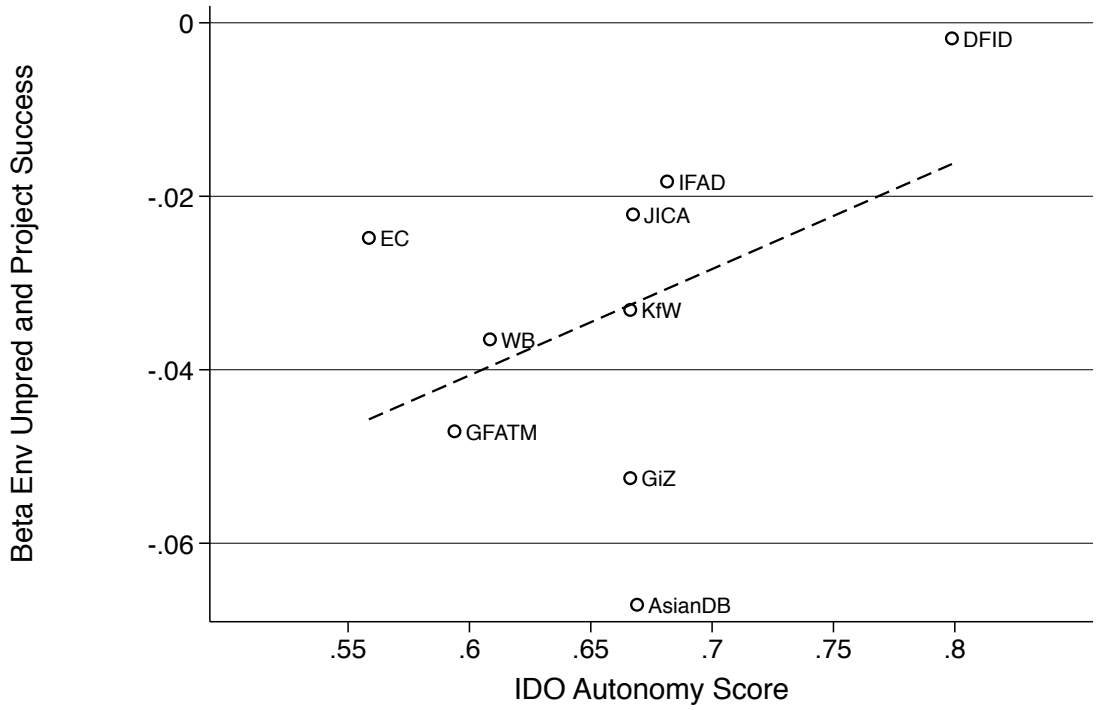

Figure I.2: Graph of IDO-by-IDO slopes.

While Figure 2 in the main text chooses the highest and lowest observed values of IDO Autonomy in demonstrating effects, Figure I.3 shows that the 25th and 75th percentile observations of IDO Autonomy are still differentiable from one another, drawing from the same model as figure 2 in the main text.

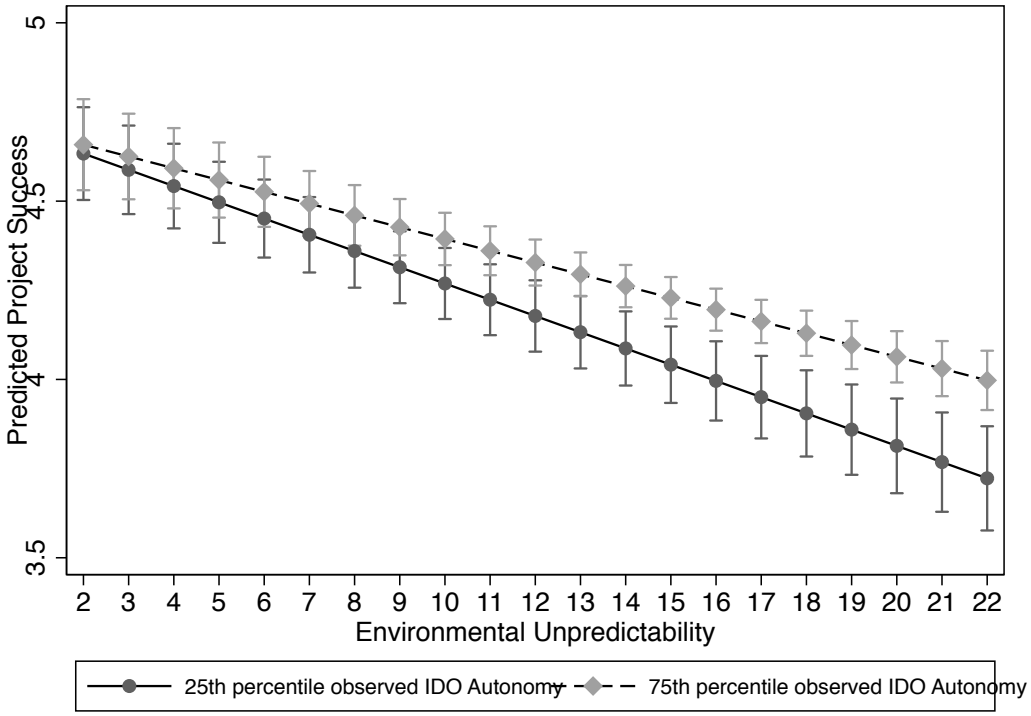

Figure I.3: Differentiating the 25th and 75th percentile of IDO Autonomy from one another

Clustering of Standard Errors

The primary analysis in this paper has clustered standard errors at the recipient country level to adjust for the possibility that project success may be correlated within a given recipient country. The results presented in the primary analysis are robust to alternative clustering strategies; that is, to clustering standard errors at the IDO level or, where practicable, to double clustering at both the IDO and the recipient country level.

As noted in the main text, it is also possible that project success is correlated within IDOs. While the small number of clusters when clustering at the IDO level may negatively affect the coverage properties of clustered standard errors (one of the motivations for clustering at the recipient level in the primary analysis), Table I.10 shows that results with standard errors clustered on the IDO are strongly consistent with the main text. The most conservative clustering strategy would be to double-cluster at both the IDO and recipient level. However, doing so precludes inclusion of fixed effects; that is, the limited degrees of freedom (given the 2nd-level N of nine) makes the inclusion of either donor or recipient fixed effects and double-clustering

simultaneously impossible.[10] Table I.11 implements double-clustering in the only case where it is viable to do so, paralleling model 2 in Table I.11. Using the Z-score of project success as the dependent variable (and thus no IDO fixed effect), Table I.11 suggests that the primary results are also robust to simultaneous clustering at the recipient and IDO level, to the extent that estimating such a model is possible.

	(1)	(2)	(3)	(4)
DV:	6pt scale	Z-score	6pt scale	Z-score
Environmental Unpredictability	-0.170***	-0.127***	-0.149***	-0.123***
	(0.0272)	(0.0235)	(0.0310)	(0.0192)
Env Unpred*IDO Autonomy	0.205***	0.148***	0.187***	0.151***
	(0.0398)	(0.0372)	(0.0440)	(0.0277)
IDO Autonomy		-1.559**		-1.725***
		(0.671)		(0.354)
Constant	4.423***	1.383**	3.807***	1.564***
	(0.0367)	(0.418)	(0.381)	(0.283)
IDO Fixed Effects	Y	N	Y	N
Recipient Fixed Effects	N	N	Y	Y
R^2	0.099	0.025	0.147	0.077
Observations	9312	9312	9312	9312

Standard errors in parentheses, clustered by IDO

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table I.10: Main Results with Standard Errors Clustered by IDO

	Project Success (Z-score)
Environmental Unpredictability	-0.127***
	(0.0245)
IDO Autonomy	-1.559**
	(0.644)
Env Unpred*IDO Autonomy	0.148***
	(0.0388)
Constant	1.383***
	(0.386)
R^2	0.025
Observations	9312

Standard errors in parentheses, double clustered by IDO and recipient country via cgmreg

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table I.11: Main Results with Double Clustering, outcome as z-score, base term for autonomy, and no fixed effects for recipient country, IDO, or sector

Additional Robustness Tests

[10] Estimation is via Cameron, Gelbach, and Miller 2006 and their *cgmreg* routine. I mean to say that *cgmreg* cannot estimate standard errors – the routine fails – when fixed effects are included.

In addition to the robustness checks discussed here, the results above are robust to:

- Using ordered logit models on six point project outcome scales (rather than OLS)
- Using z-scores as outcomes (rather than the six-point scale where employed)
- Compressing success and failure to a binary outcome and employing logit models
- Restricting SFI to common support; that is, only the range of SFI realized in all donors' data (2-22, rather than 0-25 in the main analysis)
- Dropping the latter two waves of the Paris Declaration survey in generating the IDO autonomy measure (to allay concerns that donors responded to measurement by changing their practices)
- Dropping either subscale of the state fragility index (legitimacy or effectiveness)
- Using any of the four domains of the state fragility index (security, political, economic, or social)

IDO Autonomy Measure in Full

Table I.12 provides a full list of all organizations for which IDO autonomy scores were generated and those scores, for full transparency regarding the measure.

IDO	Autonomy Score	Rank
Ireland	0.878579795	1
UK (DFID)	0.798823953	2
Norway	0.796352506	3
Netherlands	0.773272038	4
Sweden	0.719851851	5
IMF	0.714166641	6
Finland	0.689640522	7
IFAD	0.681465507	8
Denmark	0.678942561	9
Canada	0.677956104	10
AsianDB	0.669080436	11
Japan (JICA)	0.667425275	12
Germany (GiZ/KfW)	0.666281819	13
France	0.628306508	14
WB	0.608462632	15
Switzerland	0.605228841	16
GFATM	0.593850553	17

New Zealand	0.593333304	18
EC	0.558577597	19
Austria	0.535915732	20
Spain	0.533711374	21
Belgium	0.501379311	22
Luxembourg	0.492137939	23
African DB	0.488045961	24
Australia	0.480275869	25
Portugal	0.476367801	26
Italy	0.459770113	27
IADB	0.392873555	28
Korea	0.375316083	29
United States	0.36240229	30
GAVI Alliance	0.330833346	31
Turkey	0.285268188	32
United Nations	0.234992817	33

Table I.12: Full List of IDO Autonomy Scores. Note that the Paris Declaration Monitoring Surveys are, for bilateral donors, at the country level; thus KfW and GiZ share Germany's score. For the other bilateral donors in the sample I have added the IDO name to the country where appropriate.

Case Study Interviewees

Interviewee Data and Numbering Schemes

The following table indicates all individuals that provided information (mostly by interview, but in a handful of instances by email correspondence) that informs the broader qualitative data in *Navigation by Judgment* (Oxford University Press, 2018), not only those interviewees upon whose comments this article draws. A few notes on these tables – first, the number in the left-hand column does not correspond to the number in the interview citations (e.g. interview 63, 6/25/13). The cited numbers are randomized to maintain the promised confidentiality to interviewees. The dates of the interviews are omitted from the tables below, as including dates would make it much easier to infer the identity of a given speaker. For the same reason, individuals who contributed in more than one domain – e.g. speaking to South African interventions in both municipal governance and health – are given a new randomly generated number for use in each section, as to do otherwise would make it exceedingly easy to identify these speakers. As such, though 147 interviews are listed below, there is a degree of overlap; it is small, however. There are more than 135 unique interviews on which these cases collectively draw.

SOUTH AFRICA

Table I.13: South Africa Municipal Governance Capacity Building Interviews					
#	Surname	First Name	Position	USAID	DFID
1	Anonymous	Anonymous	COGTA Senior Official		
2	Bester	Angela	Former DFID staff, then Deloitte; also former DG, Public Service Commission		
3	Brooks	Frikkie	Head of KZN provincial planning department		
4	Chipkin	Ivor	Executive Director, PARI		
5	Chrystal	Blake	Supervisory Program Officer, USAID SA		
6	Dei	Colleen	Former USAID SA Mission Chief		
7	Fortuin	Joe	Director of Aid, COGTA		
8	Francis	Virginia	USAID Health Team, former RTI SA staff		
9	Glasser	Matt	Former USAID advisor in SA on municipal financing		
10	Hackner	Allan	USAID SA Financial Sector Manager (former COR on Municipal)		
11	Harding	Joel	DFID Governance Advisor		
12	Heymans	Chris	Former CMTP chief architect, now WB		
13	Hofmeyr	Beatie	Head of Education and Training Unit, LGSP implementing sub-contractor		
14	Horn	Steve	former ISLGS CoP		
15	Kolker	Joel	Former USAID staff, municipal program		
16	Konig	Ferdie	CMTP ISF in Phalaborwa, Mpumalanga		
17	Layte	Michelle	former RTI LGSP CoP (Vaz's successor)		
18	Madurai	David	Chief Director, Norms, Standards, Policy and Research, COGTA; former Chief Director, Delopment Planning & Local Economic Development		
19	Mangokwena	Andries	Advisor in Thulamela under CMTP		
20	Mathivha	Makonde	Municipal Manager, Thulamela, Limpopo		
21	Matomela	Bongani	Former Deputy Project Director, LGSP		
22	Naidoo	Subethri	Former Governance Advisor, DFID; former Local Government sector manager, USAID; former Deloitte program manager on CMTP		
23	Olver	Chippy	Former Deputy Director General DPLG		
24	Powell	Derek	Former Deputy DG, DPLG		
25	Rambulana	Wilson	former LGSP Revenue Enhancement Advisor (trainer)		

26	Sadan	Mastoera	Programme Manager, PSPPD, Office of the SA Presidency	[redacted]		
27	Savage	David	Former WB staff, now SA Treasury head of Cities Support program	[redacted]		
28	Snook	Steve	former USAID Democracy and Governance deputy team leader	[redacted]		
29	Tazewell	Littleton	Deputy Mission Director, USAID South Africa Regional Program	[redacted]		
30	Thomas	Richard	Former DFID South Africa Governance Advisor on CMTP		[redacted]	
31	Timm	Jeremy	Former CMTP now Treasury muni gov support		[redacted]	
32	Toli	Robin	Chief Director, International Development Coordination, SA Treasury	[redacted]		
33	TV	Pillay	Head of Municipal Finance, SA Treasury	[redacted]		
34	Vaz	Peter	former RTI LGSP CoP	[redacted]		
35	Yako	Pam	Former municipal manager, Amathole District; former DG, Environmental Affairs, Water Affairs	[redacted]		

Table I.14: South Africa Health Interviews						
#	Surname	First Name	Position	USAID	DFID	CDC
1	Agenbag	Rentia	Government and Civil Society Support Manager, SANAC	[redacted]		
2	Anonymous	Anonymous	CDC & USAID PEPFAR Implementer	[redacted]		
3	Anonymous	Anonymous	Senior DC-based PEPFAR official	[redacted]		
4	Anonymous	Anonymous	Senior CDC Official in Another Southern African Country	[redacted]		
5	Anonymous	Anonymous	USAID and CDC PEPFAR Implementer	[redacted]		
6	Barker	Pierre	Senior VP, Institute for Health Care Improvement		[redacted]	
7	Barron	Peter	Public health specialist & advisor to DDG Pillay	[redacted]		
8	Coovadia	Jerry	Director, MaTCH	[redacted]		

9	Coovadia	Ashraf	Head of pediatric HIV,Rahima Moosa Mother and Child Hospital, Johannesburg			
10	Dei	Colleen	Former USAID SA Mission Chief			
11	Desmond	Chris	Chief research specialist, Human Sciences Research Council			
12	Fryatt	Bob	Former DFID Health Advisor, SA			
13	Giddy	Janet	Former HIV program coordinator, McCord Hospital, Durban			
14	Goga	Ameena	Senior Specialist Scientist, MRC			
15	Gorna	Robin	Former Senior Regional Health and AIDS adviser for DFID Southern Africa			
16	Grant	Ken	HLSP Programme Director, SARRAH			
17	Harding	Joel	DFID Governance Advisor			
18	Holst	Helga	CEO, McCord Hospital, Durban			
19	Kok	Michelle	Advisor to Precious Robinson, NDOH			
20	Kumar	Smita	USAID PMTCT Lead			
21	Lesole	Lerato	PMTCT Specialist, CDC SA; previous NDOH			
22	Mahasela	Lusanda	Deputy, Research & M&E, Johns Hopkins Health and Education in South Africa			
23	Mazibuko	Ntombi	RTC PMTCT Project Manager; former EGPAF & NDOH; former NDOH			
24	Ngubane	Gugu	former HLSP A-Plan Project Manager and Technical Adviser on PMTCT			
25	Nkulu	Hilary	former DFID SA Programme Manager			
26	Pattinson	Robert	Director, MRC Maternal and Infant Health Care Strategies Unit, University of Pretoria			
27	Pillay	Yogan	NDOH Deputy Director General			

28	Robinson	Precious	NDOH Deputy Director in charge of PMTCT			
29	Sanne	Ian	CEO, Right To Care			
30	Schneider	Helen	Director, School of Public Health, University of the Western Cape; former SANAC, MRC			
31	Slingers	Nevilene	Donor Coordination Manager, SANAC			
32	Taback	Rayna	Senior Public Health Advisor, CDC South Africa			
33	Tazewell	Littleton	Deputy Mission Director, USAID South Africa Regional Program			
34	Toledo	Carlos	Chief, HIV Prevention Branch, CDC South Africa			
35	Toli	Robin	Chief Director, International Development Coordination, SA Treasury			
36	Venter	Francois	Deputy Executive Director, Wits Reproductive Health Institute			
37	Vranken	Peter	CDC Senior Technical Advisor, PEPFAR			
38	Wilson	John	HLSP Programme Manager - MSP, RRHF, SARRAH			

LIBERIA

Table I.15: Liberia Health Sector Interviews					
#	Surname	First Name	Position	USAID	DFID
1	Anonymous	Anonymous	Former Liberia NACP Advisor		
2	Anonymous	Anonymous	MoHSW senior personnel		
3	Anonymous	Anonymous	Senior official, USAID Liberia		
4	Anonymous	Anonymous	USAID Liberia international staff		
5	Augustin	Randolph	Lead Health Officer, USAID		
6	Benson	Angela	FARA Coordinator, MoHSW		
7	Bility	Kalipha	Former Program Coordinator, NACP; in 2013 Deputy Minister, Ministry of Agriculture		
8	Bruce	Lwopu	Head of Blood Safety, MoHSW; former deputy head, NACP		

9	Curran	Desmond	DFID Representative in Liberia 2007-2009		
10	Dahn	Eunice	Chief Medical Officer, MoHSW		
11	Davis	Natty B.	Chairman and CEO, NiC; former Minister without Portfolio and National Coordinator, LRDC		
12	Dolopeh	Dr.Eugene	Former Program Manager, NACP		
13	Duncan	Julie	Commissioner, NAC; former Assistant Minister for Preventive Services, MoHSW		
14	Dworku	Tanu	Former USAID Health Officer, Former NACP Coordinator		
15	Dzokoto	Agnes	Senior Technical Officer, AWARE (responsible for Liberia)		
16	Flomo	Matthew	Deputy Minister for Administration, MoHSW		
17	Freeman	Josephine	Former PMTCT Coordinator, NACP		
18	Gabelle	Chris	Former lead Liberia Governance Advisor, DFID		
19	Gaddis	Beth	Health Officer, USAID		
20	Gwenigale	Walter	Minister of Health, MoHSW		
21	Hughes	Jacob	Head of Liberia Health PF Management firm; Former PwC		
22	Hymowitz	Dan	Advisor to the Monserrado County Ebola Response, African Governance Initiative		
23	Jones	Janyaj	M&E Deputy, NACP		
24	Karzon	Toagee	Controller, MoHSW		
25	Lippevald	Theo	RBHS/JSI Deputy CoP		
26	Logan	David	Global Fund Coordinator, MoHSW; former deputy coordinator, NACP		
27	Macaulay	Rose	RBHS/JSI CoP		
28	Manuel	Marcus	Former DFID Deputy Director for West Africa		
29	Mapleh	Louise	PBF Coordinator		
30	Martin	Bill	Former Senior Adviser to the Minister, MoHSW; now PF Manager		
31	McDermott	Chris	Former health lead, USAID		
32	Nartey	Alex	Former lead of PwC team to MoHSW		
33	Niyuhire	Floride	RBHS Advisor on PBF to MoHSW		
34	Nyoweh	Moses	STI Officer, NACP		

35	Sanvee	Dr.Lilly	Head implementer, Catholic Hospital, AWARE		
36	Scheening	Sarah	Senior Policy and Implementation Advisor, USAID Global Health Bureau		
37	Sieh	Sonpon	Program Coordinator (head), NACP; former M&E on HIV, NACP		
38	Sirleaf	Momolu	Head of Aid Coordination, MoHSW		
39	Subah	Pewu	Head of Project Implementation Unit, MoHSW		
40	Tamattey	Felix	Senior Partner leading PwC Engagement, MoHSW		

Table I.16: Liberia Capacity Building Interviews

#	Surname	First Name	Position	USAID	DFID
1	Anonymous	Anonymous	Senior CSA Official		
2	Anonymous	Anonymous	Senior DC-based USAID Official		
2	Allen	William	Former Director General, Liberia Civil Service		
3	Atuanya	Jenkins	Former Deputy Director General, CSA; now assistant minister, ministry of Lands Mines & Energy		
4	Baki	Shadi	Head of Biometrics, CSA		
5	Belleh	Willie	Partner Subah Belleh Associates; local partner for CISCAB		
6	Callender	Elizabeth	Deputy Head, OTI Liberia		
7	Cooper	Vicky	Former WB consultant on Civil Service Pay Reform; current Chief of Party, GEMS		
8	Cooper	Lloyd	Grants Manager, BRDG		
9	Curran	Desmond	DFID Representative in Liberia 2007-2009		
10	Davis	Natty B.	Chairman and CEO, NiC; former Minister without Portfolio and National Coordinator, LRDC		
11	Drosaye	Alfred	CSA PAO		
12	Fahnbulleh	Louise	former OTI staff, Liberia		
13	Fn'Piere	Pat	Consultant, BRDG; OTI Field Advisor		
14	Gabelle	Chris	Former DFID Governance Advisor in Liberia		

15	Gattorn	John	Former Africa Program Manager, OTI		
16	Glentworth	Garth	OBE; former senior Governance Advisor, DFID		
17	Hare	Sam	Former Deputy Minister, Ministry of Youth and Sports		
18	Hunter	Rosslyn	M&E team, BRDG		
19	Johnson	Mimi	HR team, BRDG		
20	Kialain	David	Former principal deputy, GRC		
21	Lauer	Barb	Former CoP for BRDG, DAI		
22	Liberty	T. Edward	Director General, LISGIS		
23	Logan	James	Former Deputy Minister, Ministry of Agriculture		
24	Mayshak	Nellie	Former head, ASI CISCAB team 2007-2009		
25	Muhula	Raymond	Public Sector Specialist, World Bank		
26	Neymah	Oblayon	Former Reform Directorate CSA; current head of LIPA		
27	O'Neill	Dominic	Head of DFID Sierra Leone 2008-2011		
28	Panton	Richard	Deputy Director General, Training, LIPA		
29	Patel	Jalpa	Former coordinator, ASI CISCAB project, 2009-2010		
30	Sigrist	Ken	Former head, ASI CISCAB team 2009-2010		
31	Tarpeh	Dominic	Former CISCAB consultant, now with GRC		
32	Thompson	James	Subah Belleh staff; former member of CISCAB core team		
33	Wilson	Peter	Program Development Officer, BRDG		
34	Wilson	Mark	Grants Manager, BRDG		