Preventing the Negative Externalities of Development: Aid Compliance, Incomplete Contracts, and State Capacity*

Michael Denly†

June 22, 2025

For the most recent version of this paper, please visit:
https://mikedenly.com/research/aid-compliance-externalities

Abstract

This paper examines the potential negative externalities of foreign aid projects: that is, costs that accrue to people outside the aid transaction between the recipient state and the aid organization overseeing projects (the agent). Both scholars and the media tend to blame agents for aid implementation problems. However, recipients are nowadays responsible for implementation, negative externalities can drastically affect people's livelihoods, and politicians generally want to credit-claim from aid and avoid blame for failures. Accordingly, I argue that both state capacity and agent quality explain the prevention of negative aid externalities. Given that powerful donor countries affect only have limited bandwidth to affect project implementation, I diverge from principal-agent accounts and argue that donor influence is minimal. To the extent that donors have influence on aid implementation, it is via their effects on agents. To test the hypotheses, I compile new project-level datasets on World Bank Task Team Leader quality and recipients' compliance with safeguard policies on involuntary resettlement, indigenous peoples, and the environment. Statistical support for the hypotheses from numerous models and measures suggest that availability and representativeness biases color how scholars and the media approach aid failures. More broadly, future scholarship needs to consider not only principal-agent relationships but also principal-agent-recipient interactions and the role of incomplete contracts between agents and recipients. Otherwise, there is a risk of falling prey to a social engineering fallacy.

* I thank Ivan Besserer, Olgahan Cat, Daniel Chen, Danny Cowser, Mike Findley, Mirko Heinzel, Christopher Kilby, Xiaobo Lü, Laura Meinzen-Dick, Elena McLean, Kyle McNabb, Steve Nelson, Dan Nielson, Xin Nong, Tyler Pratt, Bernhard Reinsberg, Randy Stone, Alex Wais, Rachel Wellhausen, and participants APSA's Annual Meeting, ISA's Annual Meeting, ECPR Joint Sessions, the Institute for Advanced Study in Toulouse, the Workshop on International Financial Institutions (WIFI), the University of Houston, and Texas A&M's Quant Bag seminar for valuable feedback or comments. For funding, I thank the Institute for Advanced Study in Toulouse under grant ANR-17-EURE-0010 (Investissements d'Avenir program), University of Texas at Austin's Innovations for Peace and Development lab (IPD), the Scowcroft Institute of International Affairs at Texas A&M University, and Texas A&M's Lab for Economic Development Research (LEDR). I also thank the following student researchers at IPD and LEDR for assistance with the data collection: Carter Adams, Destiny Alvarez, Alyssa Ashcraft, Roxanne Bankhead, Anita Basavaraju, Sam Bennett, Mary Margaret Burniston, Mariana Caldas, Claudia Canfield, Lauren Childress, Monique Cooper, Ricardo Corona, Ji Na Gil, Colleen Goggin, Hannah Greer, Maria Fernanda Guerrero, Joelean Hall, Tara Jackson, Karan Kanatala, Samiya Javed, Judy Lane, Beomhak Lee, Julianna Leung, Vanessa Lizcano, Amanda Long, Aoife McDonnell, Ruth Mewhinney, Ana Karen Moreno Flores, Nicole Pownall, Briana Rocha, Katelyn Sammon, Mackenzie Sanderson, Kimberly Schuster, Taryn Shanes, Samantha Shoff, Mackenna Shull, Kyle Smith, Ayu Sofyan, Veronica Lulu Stracher, Adityamohan Tantravahi, Ethan Tenison, Skyler Thomas, Jade Tucker, and Jonathan Velasquez. All errors are uniquely my own.

† Assistant Professor, Texas A&M University ✉ mdenly@tamu.edu

In rural Paraguay, an aid project finances the construction of a road. Typically, such roads can help project beneficiaries access markets, schooling, and medical care. However, this road cuts through the land of an indigenous community, dividing its village in half. For its part, the staff from the multilateral development bank (MDB) supervising the project learn that the community does not hold a property title for the land. The three competing claims for the land, allegedly obtained via questionable means, thus endanger the community's rightful claim for compensation from involuntary resettlement. Nearby, another indigenous community with a property title cannot protect themselves from new agricultural settlers, who are a direct result of the road. Reportedly, the settlers threaten the community with violence, burn down its school, and convert precious forests into plots, destroying ecosystem services and people's livelihoods in the process (Tello, 2015, 37-40).

The above example illustrates a larger problem in international development: foreign aid projects administered by MDBs like the World Bank often create severe negative social and environmental externalities. These externalities not only affect people outside the aid transaction between the state and the aid financier but also occur with considerable frequency. For example, each year development projects cause the forced resettlement of about 15 million people (Cernea, 2008; Negi and Ganguly, 2011).[1] To address these problems, all MDBs and wealthier countries giving bilateral aid have social and environmental risk management (safeguard) policies to prevent negative externalities for resettlement, indigenous peoples, and the environment (Greenstein, 2022, 173). By the same token, compliance with these policies is often wanting, resulting in major scandals that can even involve the loss of life. It is thus crucial to know: what drives compliance with safeguard policies? By extension, what prevents or mitigates the negative externalities of development?

In terms of which actor is most to blame for the negative externalities of aid projects, both the media and academic literature tend to put the onus on MDBs. Notably, the International Consortium of Investigative Journalists's (2015) multi-site and -year investigation

[1]For a study on the economic impacts of population resettlement, refer to Bazzi et al. (2016).

excoriated the World Bank in newspapers across the world for "regularly failing to follow its own rules" (e.g., Huffington Post, 2015). For its part, academic literature tends to examine such phenomena using the lens of the principal-agent framework and focus on agency slack (e.g., Hawkins et al., 2006*a*). Although the principal-agent framework is useful for examining governance relationships between powerful donor countries (principals) and MDBs (agents), the framework mostly neglects interactions between MDBs and recipient countries (Gutner, 2005*b*).

Given that agents design and supervise aid projects, but recipients are responsible for implementation following the 2005 Paris Declaration on Aid Effectiveness (OECD, 2008), it is most useful to analyze safeguard policy implementation as a compliance problem. Due to the severe aforementioned consequences of aid externalities, as well as politicians' desires to credit-claim for aid outcomes and avoid citizens blaming them for aid failures, I argue that both high agent quality and state capacity reduce negative aid externalities. For their part, principals are more "reactive bodies" (Kapur, Lewis and Webb, 1997, 10), so I argue that their *average* effect is null—even though principals clearly do intervene strategically (Stone, 2011). By extension, the paper diverges from seminal principal-agent accounts (e.g., Nielson and Tierney, 2003), whose origins actually began from a case study of safeguard failures.

To test the above hypotheses, the paper introduces new data on agents as well as new data on World Bank safeguard policy compliance. For the agent data, I made multiple World Bank transparency requests to receive the necessary input data to reconstruct and update Denizer, Kaufmann and Kraay's (2013) previously confidential measure of Task Team Leader (TTL) quality.[2] For the safeguard data, I individually coded all 2007-2015 World Bank investment lending projects with completed evaluation reports for safeguard policy compliance relating to resettlement, indigenous peoples, and natural habitat destruction, etc. To evaluate the state capacity hypothesis, I use Hanson and Sigman's (2021) state-

[2]The new data differ from those of Limodio (2021) and Heinzel (2022*b*), whose measures do not take into account the TTL at each Implementation Status Report (ISR).

of-the-art Bayesian measure. For principal control, I follow Vreeland (2019) and include measures of World Bank Board membership, United Nations' General Assembly (UNGA) voting alignment with the United States, and temporary membership in the UN Security Council. I complement the above independent variables with numerous country- and project-level control variables.

Given the ordered nature of the safeguard compliance dependent variable, I employ ordered multilevel logit models to estimate the extent to which principals, agents, and recipients matter for safeguard policy compliance. Throughout, I use the bias-corrected Mundlak (1978) specification from Hazlett and Wainstein (2022), but I also consider the robustness of the results to numerous measures and models. All relevant adjustment sets are derived from a Directed Acyclic Graph (DAG).

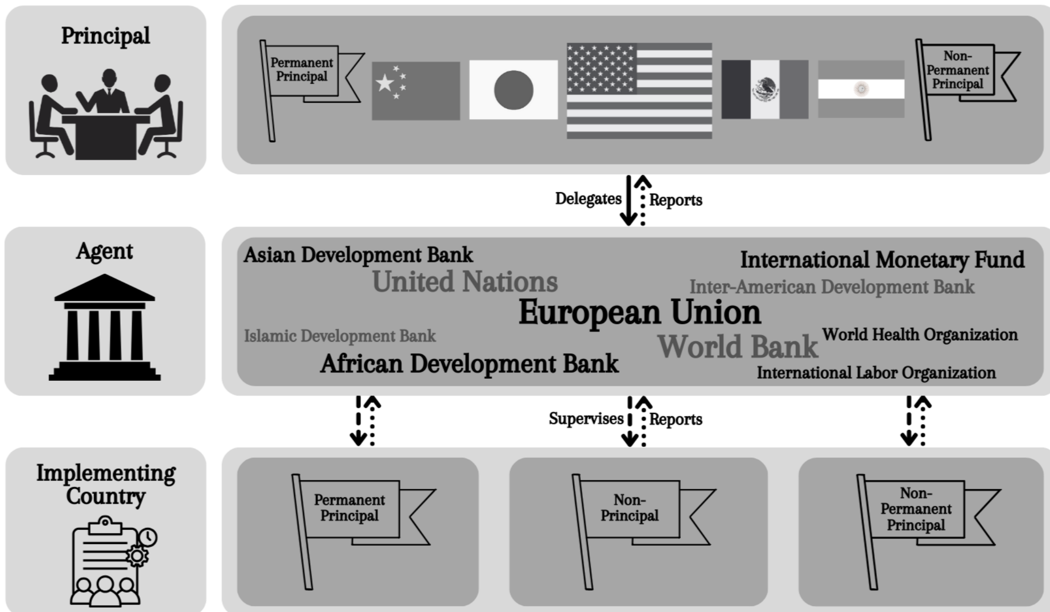
I find that both recipient state capacity and the agent—specifically, TTL quality—both consistently explain very high shares of variation in safeguard compliance outcomes. State capacity's effects primarily manifest when comparing across countries, whereas the TTL's effects are mostly—but not exclusively—discernible within each country. By contrast, none of the measures of principal control consistently explain variation directly. To the extent that powerful donors countries have influence, it is via their ability to moderate the effect of TTL quality.

This paper makes four contributions. First, the paper is one of the few to holistically analyze principals, agents, and recipients together. In doing so, the paper empirically demonstrates that powerful donors exert less influence on salient phenomena than standard principal-agent accounts suggest,[3] reaffirms the importance of the agent,[4] shows that recipient capacity deserves more attention than it receives. Indeed, analyses suggest that the TTL explains only slightly more variance than state capacity, so principal-agent accounts

[3]e.g., Nielson and Tierney (2003), Stone (2011)

[4]e.g., Kilby (2000, 2001), Hawkins and Jacoby (2006), Denizer, Kaufmann and Kraay (2013), Bulman, Kolkma and Kraay (2017), Heinzel and Liese (2021), Heinzel (2022*b*), Ashton et al. (2023)

Figure 1: Principal-Agent and Agent-Implementer Interactions in Multilateral Aid



Sources: Own elaboration and Canva stock images.

that focus on agency slack are very much insufficient.

Second, the paper provides a framework for integrating the principal-agent and agent-implementer relationships. Per Figure 1, scholars can continue to examine delegation and reporting relationships between principals and agents. However, agents also supervise recipients through incomplete contracts,[5] not secondary principal-agent relationships. Consistent with Nielson and Tierney's (2005) clarification on Gutner (2005*b*), MDBs do not grant recipients a (conditional) delegation of authority. Almost the opposite takes place between MDBs and recipients: the latter grant the former authority to operate on its sovereign territory, and recipients can alter the delegation contract at moment's notice. That is not what happens in a true principal-agent relationship, where agents are subordinate and are mostly subject to principal punishment, such as via re-contracting (see Hawkins et al., 2006*a*). In turn, true agents' powers are limited, and they can only overcome principal control via slack, discretion from multiple principals, writing or reinterpreting rules, and buffering (Hawkins and Jacoby,

[5]For an overview of incomplete contracts, see Hart (2017).

2006, 202). Outside of idiosyncratic moments,[6] agents cannot easily re-write the delegation contract once one it is in place.[7]

Fortunately, incomplete contracting can accurately characterize the supervision and reporting relationships between aid agencies and recipients. In particular, an incomplete contracting approach recognizes that recipient countries have residual control by virtue of their ownership status/sovereignty and can "hold-up" the aid agency (e.g., Hart, 2017, 1732-1733). The phenomenon is akin to the obsolescing bargain describing how multinational firms lose leverage after they make investments in recipient countries (Bennon and Fukuyama, 2022). That becomes even more clear when considering agents' disbursement imperative,[8] as well as the facts that agents often supervise powerful members of the UN Security Council or even their own principals (see Figure 1). To be clear, the bottom (implementer) row of Figure 1 has "Non-Principal" in the center to reflect that agents mostly do not supervise principals. Nevertheless, the opposite is still possible. In the case of the World Bank, aid recipients such as China and Brazil mostly enjoy permanent Board member (principal) positions, and countries such as Argentina generally rotate their Board positions within different country groupings (see Vreeland, 2011). Similar dynamics play out at other MDBs, too. Accordingly, the context-specific power of recipient states deserves more attention than the literature currently concedes.

Third, the paper demonstrates that there is heterogeneity in aid compliance problems, as only some are strategic. While some literature focuses on how recipient factors affect aid compliance (e.g., Girod and Tobin, 2016), recipients' revenue options outside of aid, such as natural resources and foreign direct investment, are less relevant for environmental and social

[6] e.g., Chwieroth (2008), Johnson (2014), Denly (2021)

[7] For this reason, Denly (2021) moves away from the principal-agent model to capture bureaucratic autonomy in multilateral aid. Nielson and Tierney (2005, 737) even go so far to as suggest that studies examining relationships between aid agencies and recipient countries via the principal-agent model do not proffer falsifiable hypotheses. I maintain that we can learn from such studies (e.g., Gutner, 2005*a*,*b*; McLean, 2015), even if they miss key details.

[8] MDBs need to disburse to survive and, as a consequence, often do so when corruption or other issues should prevent them from doing so. See, for example, Booth (2011), Buntaine (2016), and Weaver (2008).

risk compliance. The reason why is that environmental and social aid externalities have more immediate political consequences than overall aid non-compliance. For example, politicians all over the world routinely decry the World Bank and International Monetary Fund (IMF) for the conditionality on their loans, and that is often politically popular. Indeed, there is a large literature on strategic noncompliance with international organizations (e.g., König and Mäder, 2014). While some leaders can get away with destroying a rainforest or an indigenous peoples' sacred cultural site, on average that is not a strong political strategy. Even if politicians can deflect some part of the blame on the aid, it is generally bad publicity that they seek to avoid. Additionally, politicians seek to engage in aid credit-claiming and prolong their power (Cruz and Schneider, 2017; Baldwin and Winters, 2023), and environmental and social disasters reduce future aid commitments (Buntaine, 2016), so aid recipients aim to prevent aid externalities if they have the capacity to do so.

Fourth, the new data on safeguard policy compliance help overcome selection biases that impact how academia and the media understand protection of vulnerable people and the environment from the adverse impacts of aid. In particular, the present paper underscores what happens in terms of broader patterns, not just the *selected sample* of failures that reach the newspapers, the World Bank Board, or Inspection Panels.[9] Essentially, the new data and empirical regularities that I show help overcome what Tversky and Kahneman (1974) famously described as availability and representativeness biases: that is, the ability of salient information to shape human thinking in biased ways. In the context of foreign aid, these biases take the form of a social engineering fallacy that devalues local conditions and knowledge (Scott, 1998; Andrews, Pritchett and Woolcock, 2017); overemphasizes the ability of grand plans to solve development challenges (Easterly, 2006; Acemoglu and Robinson, 2012);[10] and undervalues the constraints that agents face in fragile contexts (Honig, 2019).

[9]For more on quasi-judicial bodies like the World Bank Inspection Panel, see Fox (2002) and Zvobgo and Graham (2020). For more on newspaper stories, see International Consortium of Investigative Journalists (2015).

[10]By referencing Acemoglu and Robinson (2012), I am referring to what they call the ignorance hypothesis: that is, that all developing countries need to overcome their problems is Western advice

1. Negative Externalities and Foreign Aid

In the context of economics, "negative externalities are costs that accrue to parties other than [those] that produce them" (Krugman and Obstfeld, 2003, 277). Typically, negative externalities are social in nature, incur transaction costs for monitoring and governance, and outweigh the private benefits or rents that amass to the initiators of the transaction (Coase, 1960; Williamson, 1985). Accordingly, negative externalities are socially inefficient, arise from the lack of institutions such as property rights to correct for market failures, and thus harmfully affect societal provision of public goods (Arrow, 1970; North, 1981). Overcoming negative externalities in economics requires government regulation, a solution to the collective action problem, or an innovative arrangement between the affected parties (Olson, 1965; Ostrom, 1990; Ostrom, Walker and Gardner, 1992).

It is both possible and fruitful to apply what scholars know from negative externalities in economics toward foreign aid. In the context of foreign aid, negative externalities are the costs that accrue to the people outside the aid transaction between the aid financier (the agent) and aid-receiving state (the implementer). Generally, states have different capacities to prevent or mitigate negative externalities to people outside the aid transaction—i.e., project "beneficiaries". They are the very people that aid projects aim to support with, for example, public goods in health (e.g., vaccines), education (e.g., schooling), and infrastructure (e.g., roads, sanitation, flood protection).[11]

What, then, are the potential negative externalities of aid? Rajan and Subramanian (2011) argue that aid causes exchange rate woes and lowers economic competitiveness, but this is clearly only possible under certain circumstances: the aid flows have to be very large as compared a country's national GDP, which is usually not the case (Qian, 2015). Another set of scholars suggest that aid is a fungible non-tax revenue that prolongs the rule

[11] OECD (2017) provides a classification of the economic sectors and activities that aid projects support. By definition, public goods are both non-rival and non-excludable, and a large percentage of aid activities meet these criteria.

of authoritarian leaders, forestalls democratization, and fuels authoritarian reversals in a manner that is worse than oil.[12] However, other more recent studies contest these claims,[13] and aid agencies nowadays take great measures to control corruption in their projects (Rose-Ackerman and Carrington, 2013). Yet another suite of studies argue that aid or aid shocks fuel civil conflict (Nielsen et al., 2011; Crost, Felter and Johnston, 2014; Nunn and Qian, 2014; Dube and Naidu, 2015; Wood and Sullivan, 2015). By the same token, aid did not cause conflict on its own in any of the countries under study (e.g., Colombia, Mali, the Philippines); all of these countries had pre-existing civil conflict or tensions, and aid only added fuel to the fire. Finally, Lee and Platas (2015) examine the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) and find that the targeted program had adverse downstream consequences on neonatal health outcomes.

What, then, are the potential negative externalities that can apply to every aid project and do not require pre-existing country-level development challenges or the study of long and complex causal chains?[14] The negative externalities that accrue to project beneficiaries when the state does not implement the social and environmental risk management policies of the aid financier provide a more complete, direct answer. When states do not adequately implement these policies, the relevant negative externalities that arise might outweigh the benefits of providing aid in the first place. That is why all bilateral donors from OECD countries and all major MDBs have relevant safeguard policies to prevent social and environmental aid externalities (Greenstein, 2022, 173). Failure to adequately implement safeguard protections has resulted in some of the most egregious and embarrassing humiliations in the history of foreign aid and development more broadly (Nielson and Tierney, 2003; Weaver, 2008; Buntaine, 2016). Beatings, forced migration, and large-scale deforestation are just a few examples. What prevents or mitigates such negative aid externalities?

[12] Djankov, Montalvo and Reynal-Querol (e.g., 2008), Bueno de Mesquita and Smith (2009), and Morrison (2012).

[13] e.g., Kono and Montinola (2009), Altincekic and Bearce (2014), Arndt, Jones and Tarp (2015), Bermeo (2016), and Findley et al. (2017).

[14] For more on aid's long and complex causal chains, see Bourguignon and Sundberg (2007) and Denly (2021).

2. Theory

In their role as agents with incomplete contracts, MDBs can take *some* action against negative aid externalities beyond typical supervision activities. Because negative aid externalities hurt bureaucrats' careers and damage agent legitimacy more broadly, safeguard failures prompt agents to decrease countries' future aid funding envelopes (Buntaine, 2016). More immediately, however, agent are subject to the constraint that they do not implement projects (see OECD, 2008). In turn, recipients garner hold-up power consistent with the obsolescing bargain that characterizes foreign direct investment. It stems from MDBs lack sovereignty to fully prevent negative aid externalities on recipients' territories (Nielson and Tierney, 2005) and the costs of canceling projects. The high sunk costs of both money and time due to the MDBs' multi-year project preparation cycles constitute one set of factors. Another is that canceling projects endangers institutional survival, especially given that MDBs' business models depend on disbursing project funds and having loans repaid. In this light, MDBs expend spend less effort on compliance matters that are not easily monitorable (Martens et al., 2002), which makes aid externality prevention highly dependent on recipient actions.

For their part, politicians in recipient countries value aid, so they prefer to avoid negative aid externalities that imperil potential future funding streams. Ostensibly, there are exceptions. *On average*, though, the destruction of a rainforest or an indigenous people's cultural heritage site presents potential political risks. A safer strategy is for politicians to spend their time engendering compliance and claiming credit for aid-financed public goods that can deliver votes (e.g., Cruz and Schneider, 2017; Baldwin and Winters, 2023). The challenge is that it can only happen if states have capable and independent bureaucracies across their territories, yielding:

> Hypothesis 1 (state capacity): Aid recipients with higher state capacity are more likely to prevent or mitigate negative social and environmental externalities via

their higher potential to implement.[15]

As compliance scholars will discern, Hypothesis 1 is related to both the management school and scholars emphasizing the role of domestic constituencies. The hypothesis aligns with the management school due its focus on capacity constraints (e.g., Chayes and Chayes, 1993). Some scholars divide the management school into resource constraints and neoinstitutionalist-oriented autonomy constraints (e.g., Börzel et al., 2010), but it is difficult to distinguish between the two subtypes in foreign aid, as both traits are *mostly* lacking in recipient countries (e.g., Fukuyama, 2004). Additionally, the management school not only ignores incomplete contracts but also the distribution of costs and benefits, which is where scholars emphasizing domestic constituencies offer insights (e.g., McLean, 2015). Notably, MDB procurement contracts often stay within the same country (McLean, 2017; Heinzel, 2022*a*), so recipients have an incentive to avoid negative aid externalities and potentially curry favor with domestic suppliers, who can be politically powerful.

It is also worth underscoring why the predictions of the enforcement school (e.g., Keohane and Nye, 1977) or constructivist school (e.g., Finnemore and Sikkink, 1998) are not in conflict with the state capacity hypothesis. For its part, the constructivist school is more about changing longer-term norms, consistent with the critical juncture that made safeguard violations salient. By contrast, the enforcement school is more relevant, though not in conflict with Hypothesis 1. The enforcement school emphasizes the utility of better monitoring (e.g., Downs, Rocke and Barsoom, 1996; Tallberg, 2002). From an incomplete contracts perspective, better monitoring is still relevant and unequivocally helps with compliance. Following numerous studies that empirically emphasizing the crucial role of agents (e.g., Denizer, Kaufmann and Kraay, 2013; Bulman, Kolkma and Kraay, 2017; Heinzel and Liese, 2021; Heinzel, 2022*b*; Ashton et al., 2023), I thus offer:

[15] The most used definitions of state capacity stress implementation (e.g., Soifer and vom Hau, 2008; Besley and Persson, 2010; Centeno, Kohli and Yashar, 2017). All of these scholars draw from the Mann's (1984) definition of infrastructural power.

Hypothesis 2 (Agent quality): Higher-quality agents prevent or mitigate negative aid externalities more frequently.

Finally, it is necessary to specify the role of principal. Although Nielson and Tierney's (2003) original contribution on principal-agent theory in MDBs focuses on a case study of safeguard failure, the Boards overseeing MDBs are more "reactive bod[ies]: ratifier[s], occasionally naysayer[s]" (Kapur, Lewis and Webb, 1997, 10). Additionally, principals maintain only small staffs at MDBs (Buntaine, 2016, 64), making principal monitoring more sporadic than consistent for implementation issues. In line with Stone's (2011) findings that principals only selectively intervene on high-salience issues, I proffer the final hypothesis:

Hypothesis 3 (Limited principal hypothesis): Principals' influence on the prevalence of negative aid externalities in the normal course of operations is, on average, limited, even though principals clearly intervene in the most salient cases.

3. Research Design

3.1. Dataset and Dependent Variable

To test the above hypotheses, I coded a new dataset of states' compliance with social and environmental risk management—i.e., safeguard policies—in World Bank projects. These policies are the product not just of World Bank staff but numerous consultations not just with recipient governments but also NGOs, interest groups, and local communities negatively affected by previous aid projects (Greenstein, 2022). The World Bank was the first MDB to adopt safeguard measures in the late 1980s following severe negative externalities for failed projects in Brazil and India. In Brazil, these externalities included large-scale deforestation and the spread of tuberculosis and malaria to local, indigenous populations. In India, forced displacement led to a "long-march" of protests, ultimately resulting in 140 arrests and beatings of affected populations by local authorities (Weaver, 2008, 22-23). Since these low moments in the history of the World Bank, the institution rebounded and has served as a leader in the development of safeguard policies, yielding significant policy emulation across the different MDBs (Buntaine, 2016; Greenstein, 2022).

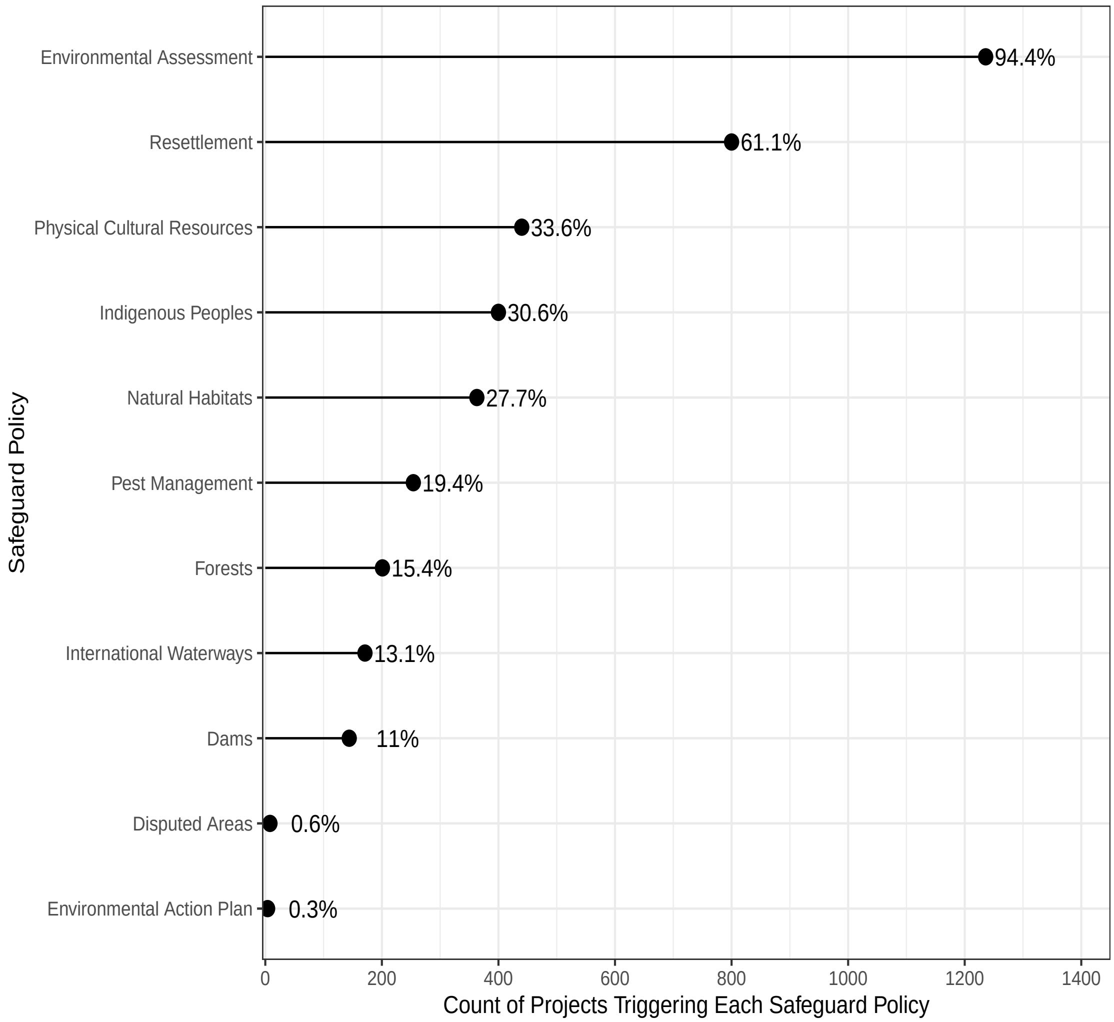
The dataset in the present paper only covers World Bank investment projects, which serve as the unit of analysis for this study. I exclude structural adjustment/policy projects because they do not have safeguards policies. For their part, Program for Results (PforR) projects do not have the same uniform usage of social and environmental risk management measures due to the flexible nature of the PforR lending instrument,[16] so I also exclude PforR loans.

With respect to the time period of study, the dataset covers 1,309 investment projects from 2007-2015. I chose 2007 as the starting year for two main reasons. First, although aid recipients implemented their own World Bank projects prior to the 2005 Paris Declaration on Aid Effectiveness, starting the analysis after 2005 made recipient ownership even more salient. Second, the World Bank finished converting the social and environmental risk management measures in Figure 2 from Operational Manual Statements to official Policies in 2006 (Independent Evaluation Group, 2010, 7). Starting with 2007 thereby prevents any potential problems with staff treating Operational Manual Statements different than Policies. Similarly, starting with 2007 reduces biases arising from implementers using the Operational Manual Statement and Policy distinction to reduce their safeguard policy compliance burden. Essentially, by 2007 safeguard policy inclusion happened a matter of course, as opposed to being subject to bargaining outcomes that may produce endogeneity. Given that it usually take 4-8 years to implement projects, another 6-12 months for relevant evaluation documents to be ready, and it is only possible to fully evaluate safeguard compliance on completed projects, the current ending year of the dataset is 2015. I also stopped in 2015 because the World Bank introduced a revised safeguards policy framework in 2016

[16] See Winters (2010) and Winters and Kulkarni (2014) for more on the different types of World Bank lending instruments as well as when the institution decides to use one over the other.

(Greenstein, 2022).

Figure 2: Safeguard Policies Triggered in the Sample



Source: Own coding.

Figure 2 provides a numerical breakdown of the safeguard policies triggered in the sample of 1,309 projects. The most frequently policies triggered include those regarding required environmental assessment (94%) and resettlement (61%). Projects trigger policies regarding physical cultural resources (34%), indigenous peoples (31%), natural habitats (28%), and pest management (19%) with relatively high frequency as well. With relatively less frequency, projects sometimes trigger policies regarding forests (15%), international waterways (13%), and dams (11%). Projects in the sample almost never trigger policies on disputed areas or required environmental action plans.

To measure the dependent variable, project-level compliance with World Bank safeguard policies, I coded the available evaluation documents for each project. Notably, I focused on Project Performance Assessment Reports (PPARs), Implementation Completion and Results Report Reviews (ICRRs), and project documents examined by the Independent Evaluation Group (IEG). Although IEG produces an overall borrower compliance score for each project that it evaluates, its scope is much broader than merely safeguards,[17] which is why the manual, safeguard-specific coding was necessary. Implementation Completion Reports (ICRs) are generally written by consultants hired by each project's respective TTL, making them at least somewhat independent. However, the IEG PPARs and ICRRs provide another level of insulation against the potential downplaying safeguard issues in projects. In particular, the PPAR is very extensive, often entailing visits to the implementing country. By the same token, IEG does not evaluate all investment projects and exhibit some biases in the projects that it chooses to evaluate (Kilby and Michaelowa, 2019), so I always examine both the ICR and IEG PPAR/ICRR. When other relevant documents are available, such as Project Papers detailing the safeguard performance of the first project in a supplemental financing project, I examine those documents as well.

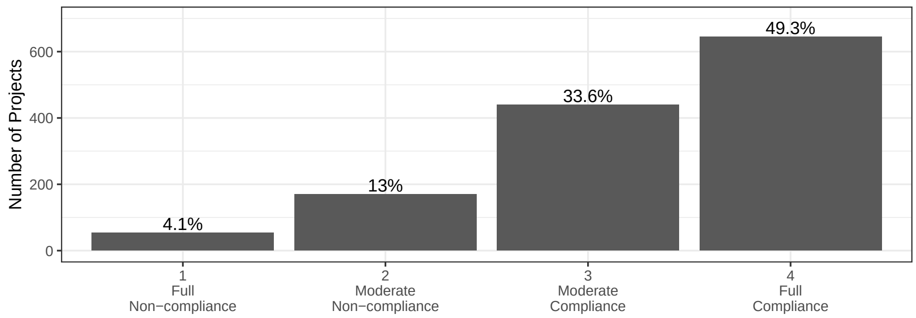
Following Buntaine (2016, 92-93),[18] the safeguard compliance scores in the present study range from 1 (low compliance) to 4 (full compliance) corresponding with the policies in Figure 2. The difference between the present study's coding and that of Buntaine (2016) only pertains to the years under study. On that score, Buntaine (2016) examines 1990-2009, which is mostly prior to the World Bank's conversion of the safeguards Operational Directives to official Policies.

[17] As Girod and Tobin (2016, 220) explain, citing Smets, Knack and Molenaers (2013), overall borrower compliance captures "the extent to which the borrower complied with covenants and agreements. The following criteria are taken into account: government ownership and commitment to achieving objectives, adequacy of stakeholder involvement, timely resolution of implementation issues, adequacy of M&E [Monitoring and Evaluation] arrangements and relationship with donors/partners."

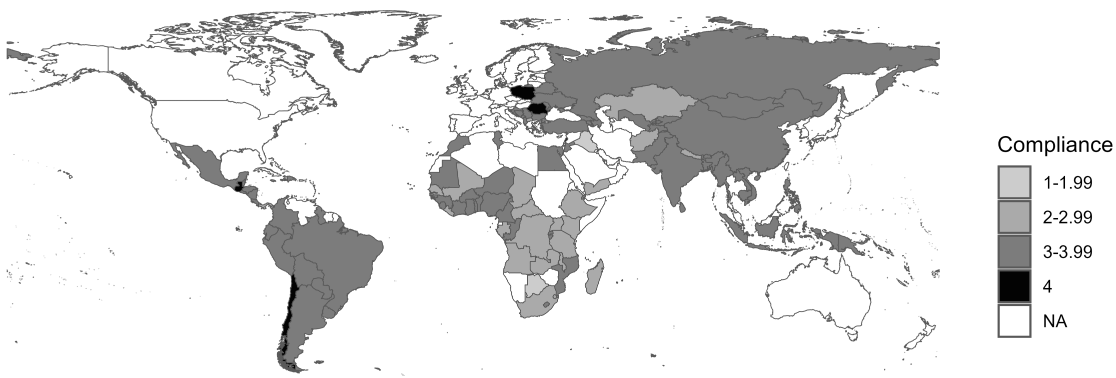
[18] Buntaine (2016) uses safeguard compliance as an independent variable, not a dependent variable, to explain commitment patterns and how safeguard failures impact bureaucrat careers.

Figure 3: Summary Statistics of World Bank Safeguard Policy Compliance (2007-2015)

(a) Distribution of Compliance Levels



(b) Map of Average Compliance Levels by Country



Source: Own coding.

To ensure quality in the coding of safeguard compliance, all projects underwent at least two rounds of coding. One team member performed an initial review, then a more experience team member or I performed a double-check of the first team member's work. Additionally, I performed random triple-checks of some projects. When I did so, I never changed any of the final compliance scores. Appendix G provides example projects and accompanying coding details for eight projects, encompassing two projects for each of the four compliance outcomes.

Figure 3a provides summary statistics for projects that coded to date, and Figure 3b provides a map of average country-level compliance scores. As Figure 3a indicates, around 83% projects generally comply with the safeguard policies, as indicated by the share of projects with a compliance score of 3 (moderate compliance) or 4 (full compliance). Circa

17% of projects exhibit moderate-to-full non-compliance, as suggested by the shares of compliance scores with only 1 (full non-compliance) or 2 (moderate non-compliance). Table A.1 provides the average results by country in numerical format.

3.2. Independent Variables

3.2.1. Primary Independent Variable

I employ Hanson and Sigman's (2021) measure for the primary independent variable, state capacity. Hanson and Sigman's (2021) state capacity variable stands out relative to competing frequentist measures.[19] Notably, Hanson and Sigman's (2021) measure uses a Bayesian measurement model to combine multiple indicators and overcome missing data challenges in a way that competing frequentist measures cannot (see Fariss, Kenwick and Reuning, 2020). To capture the multiple, latent dimensions of state capacity, Hanson and Sigman (2021, 1502) include indicators on administrative efficiency, bureaucratic quality, budgetary quality, census frequency, fiscal capacity, information capacity, law and order, military expenditures, police expenditures, public administration, and taxation. The variable's large scope is not only useful from a measurement perspective, but it also reduces the risk of endogeneity from previous World Bank projects contributing to large shares of current state capacity.[20] In the present dataset, Hanson and Sigman's (2021) state capacity variable ranges from -1.667 to 1.55.

3.2.2. Agent (Project-Level) Variables

Following Denizer, Kaufmann and Kraay (2013) and Bulman, Kolkma and Kraay (2017), outcomes for safeguards, or any other project component, depend on project-specific features, especially the quality of the Task Team Leader (TTL). To capture the quality of the TTL, I made a transparency request to the World Bank to obtain full data on the name of

[19]See, for example, Hendrix (2010), Lee and Zhang (2017), and O'Reilly and Murphy (2022).

[20]Such a risk is minimal because projects are generally targeted at particular sectors, so state capacity only increases over long periods of time with many projects from various donors.

the TTL at each mandatory, 6-month Implementation Status Report (ISR) for each project. After three rounds of back-and-forth, the World Bank sent me the data. I then combined these data with IEG data on project outcomes to calculate a TTL quality score. Following the detailed description in Appendix E, I (i) code the name of the TTL at each ISR; (ii) compute each TTL's daily average IEG outcome score for his/her completed projects, excluding the current and future ones, leaving a TTL-day dataset of 40 million observations; (iii) merge those daily scores back into each ISR for each project; and (iv) average across all ISRs for each respective project to obtain a project-specific TTL-quality score.

Other potentially important project-level confounding variables include the safeguard risk category determined prior to project approval and commitment amounts. The safeguard risk category, which has three principal levels,[21] captures *ex-ante* expectations about compliance difficulties. Given that the variable only correlates at 0.08 with both TTL quality and overall IEG project outcomes, the variable helps show that potential endogeneity from TTL assignment into more difficult safeguard projects is not a concern here. For their part, projects with higher commitment amounts might garner more attention. I thus take into accounts commitments, deflate the variable to constant 2015 US dollars, and take its natural log to decrease the risk of overdispersion affecting the results. Although supervision costs may proxy for supervision effort designed to reduce safeguard compliance issues, I exclude them because they are clearly post-treatment to any safeguard issues that arise—or what Angrist and Pischke (2008) call a "bad control". Additionally, the direction of post-treatment bias is hard to predict from a statistical perspective (Montgomery, Nyhan and Torres, 2018, 760).

[21] As Table 1 demonstrates, the variable has four categories. However, the fourth category "FI" refers to financial intermediary loans. I recode those to missing given that risk levels vary across financial intermediaries, and the "FI" risk level does not necessarily imply higher or lower pre-implementation risk, consistent with the other categories.

Table 1: World Bank Safeguard Risk Categories and their Corresponding Sample Statistics

Category	Details	Sample Frequency	Mean Compliance Score
A	Projects qualify when they are likely to have significant "*sensitive, diverse, or unprecedented*" adverse impacts. Required assessment for Category A projects must compare the impacts with alternatives, including foregoing the project, and recommending prevention, mitigation or compensation measures. The recipient must also prepare a full Environmental Impact Assessment (EIA) or an equivalent regional/sectoral EA document.	216	3.08
B	Projects qualify when potential adverse impacts are *less severe, more site-specific, largely reversible*, and easier to mitigate than Category A. The required EA is narrower in scope but still analyzes impacts and specifies mitigation and performance-improvement measures.	959	3.29
C	Projects qualify when they are expected to have *limited or no* adverse impacts. Beyond the initial assessment, no further EA is necessary.	83	3.63
FI	Projects qualify the World Bank routes funds via through a *financial intermediary* (e.g., bank, micro-finance institution) for sub-projects that could generate adverse environmental impacts.	51	3.51

Sources: World Bank (2013), World Bank (2022), and own coding.

3.2.3. Principal (Donor) Variables

Especially given that Nielson and Tierney's (2003) seminal article on the relevance of principal-agent theory to the MDBs focuses on a case study of a safeguard policy failure, it is essential to control for donor variables. Following Vreeland's (2019) review, I use three variable to capture powerful donor countries principals' interests. To capture formal influence these donor countries, I follow Kaja and Werker (2010) and add an indicator of whether the aid-receiving state was a member of the World Bank Board. Next, I follow Dreher, Sturm and Vreeland (2009) and include an indicator variable of whether the country was a temporary member of the UN Security Council (UNSC). Additionally, I include Bailey, Strezhnev and Voeten's (2017) measure of the aid-receiving state's ideal point distance from the United States, the World Bank's most powerful shareholder. As Bailey, Strezhnev and

Voeten (2017) explain, their ideal point measure captures dynamic state preferences through UN General Assembly (UNGA) voting and correlates at 0.92 with the separate ideal point measures using only votes deemed "important" votes by the US State Department. For all of these donor variables, I follow Kilby and Michaelowa (2019) and merge based on the evaluation year,[22] whereas I merge on the project approval year or fiscal year, as appropriate, for the other covariates—i.e., given that they pertain more to implementation.

3.2.4. State-Level Control Variables

Given that many resettlement and social safeguards issues often arise from lack of property rights (Tello, 2015), I also control for the Varieties of Democracy (V-Dem) project's measure property rights protection (see Lindberg et al., 2014). Because democracies are generally better at environmental protection than autocracies (e.g., Bernauer and Koubi, 2009), I further include V-Dem's overall democracy measure, which does not have the same measurement challenges as the commonly-used Polity measure (see Vreeland, 2008).

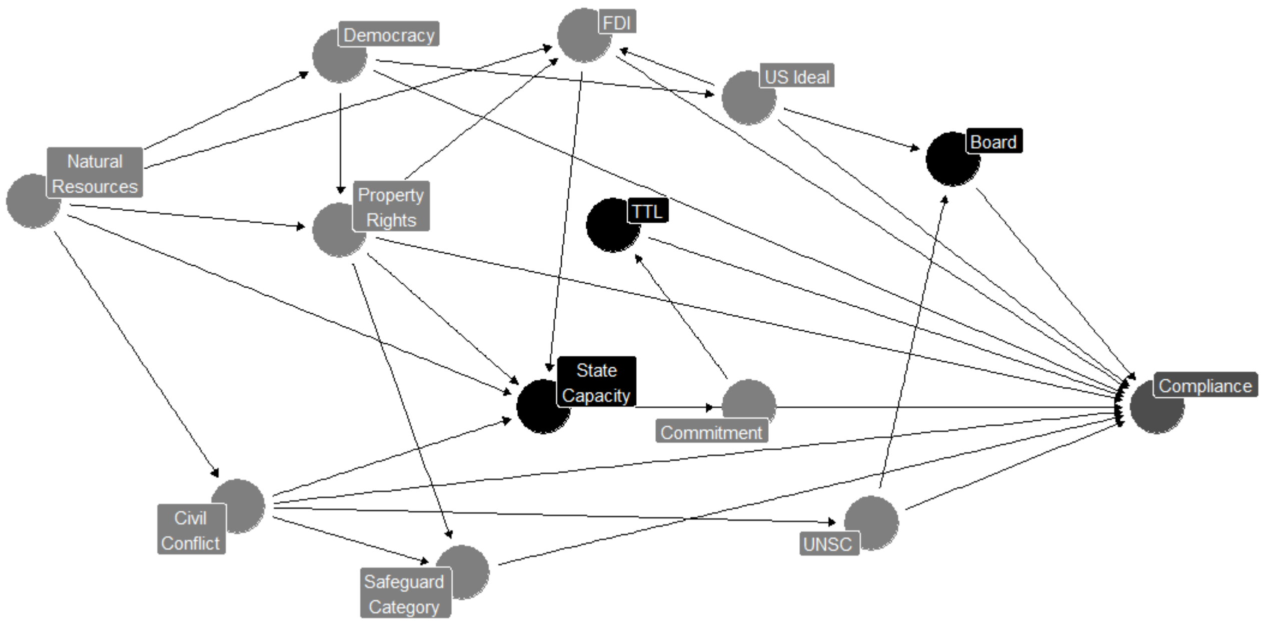
The final set of state-level control variables come from the World Bank's (2017) World Development Indicators (WDI). In line with Girod and Tobin (2016), I control for natural resource rents as a share of GDP and foreign direct investment as a share of GDP. Although safeguards are only one component of Girod and Tobin's (2016) measure of borrower compliance, it is feasible that these variables are relevant for safeguard compliance as well.

3.2.5. Identification and Determining the Final Covariates

The previous sections indicate that there are many potential control variables, but some of them are irrelevant. Notably, some variables may be mediators that can change the estimand from a total effect to a direct effect; others may be colliders that produce spurious results; and still others may be descendants that must be excluded. To obviate such possibilities and estimate the total effect, I present a Directed Acyclic Graph in Figure 4.

[22] The main results do not change when I merge by the approval year instead of the evaluation year.

Figure 4: Covariate Selection DAG



Note: See Appendix F for path descriptions. DAG drawn with the `ggdag` R package (Barrett, 2025).

Appendix F provides a verbal description for all paths.

Table 2 details the relevant model specifications that I derive from the DAG and pairwise correlations between all variables in Figure A.1. Under selection on observables,[23] all "base" specifications following the DAG satisfy Pearl's (2009) backdoor criterion. That is theoretically sufficient for unbiasedness, but controlling for additional variables yields efficiency gains in terms of reducing variance. I thus also consider Perković et al.'s (2018) canonical, "full" adjustment sets—that is, a richer set of controls. Finally, I supplement those full adjustment sets with ones that exclude potentially collinear variables in the "full lc" specifications, according to Figure A.1 and Table 2.

The key reason underpinning the necessity of using multiple adjustment sets—and not estimating effects from the same model—is that commitments mediate the effect of state capacity on compliance. Per Pearl's (2009) rules of do-calculus, controlling for either

[23]Selection on observables means that there are no unaccounted for control variables. Of course, it is *possible* that the DAG is missing a variable, but the DAG does formalize all assumptions transparently.

Table 2: Adjustment Sets and Models Derived from the DAG

Model	Treatment	Set Type	Variables Included	Notes
base1	State capacity	Minimal	Civil conflict, democracy, FDI, property rights, US ideal	Dropped commitment (mediator) and TTL (descendant)
base2	State capacity	Minimal	Civil conflict, FDI, natural resources, property rights	Dropped commitment (mediator) and TTL (descendant)
full	State capacity	Canonical	Board, civil conflict, democracy, FDI, natural resources, property rights, safeguard category, US ideal, UNSC	
full lc	State capacity	Canonical	Board, civil conflict, democracy, FDI, natural resources, safeguard category, US ideal, UNSC	Removing property rights due to potential collinearity
base	TTL quality	Minimal	Commitment	
full	TTL quality	Canonical	Board, civil conflict, commitment, democracy, FDI, natural resources, property rights, safeguard category, state capacity, UNSC, US ideal	TTL is upstream of state capacity, so the latter is OK to include here
full lc	TTL quality	Canonical	Board, civil conflict, commitment, FDI, natural resources, property rights, safeguard category, state capacity, UNSC	Removing US ideal and democracy due to potential collinearity
base	Board	Minimal	UNSC and US ideal	
full	Board	Canonical	Board, civil conflict, commitment, democracy, FDI, natural resources, property rights, safeguard category, state capacity, UNSC, US ideal	
full lc	Board	Canonical	Board, civil conflict, commitment, democracy, FDI, natural resources, safeguard category, state capacity, UNSC, US ideal	Removing property rights due to potential collinearity

Note: See Figure 4 for the DAG. Adjustment sets created with `dagitty` (Textor et al., 2016). State capacity has two minimal adjustment sets that satisfy the backdoor criterion, but the others only have one.

commitments or its descendant, TTL quality, would change the estimand from the targeted total effect to a direct effect for the state capacity treatment. By contrast, the TTL and Board treatments do not suffer from the same drawbacks, so it is useful to control for state capacity in models where the Board and the TTL are the treatments. Table A.2 presents the summary statistics for all covariates included in the final adjustment sets that I detail in Table 2.

3.3. Empirical Strategy

The primary empirical strategy involves a bias-corrected, Mundlak (1978) multilevel ordered logit model with cluster-robust standard errors, as advanced by Hazlett and Wainstein (2022):[24]

$$\Pr\left(y^*_{compliance\,(i,j)}\right) = \Lambda\left(\alpha_{country\,j[i]} + \underbrace{\beta_w\left(X_{ij} - \bar{X}_j\right)}_{\text{within-country effect}} + \underbrace{\beta_b\,\bar{X}_j}_{\text{between-country effect}} \right) \tag{1}$$

where subscript i indexes projects and j their implementing country; X_{ij} corresponds to the project-level values of the time-varying independent variables; $\bar{X}_j$ is the country-level mean for each independent variable; and $\alpha_{country\,(j[i])}$ is a random intercept that captures the unobserved, time-invariant country-level factors j for project i.[25] Because the dependent variable, $y^*_{compliance(i,j)}$, has four ordered categories, it is possible to classify it as follows, where τ_i are the cutpoints for each imposed category:

$$y_{compliance\,(i,j)} = \begin{cases} 1 & \text{if } y^*_{ij} \leq \tau_1 \\ 2 & \text{if } \tau_1 < y^*_{ij} \leq \tau_2 \\ 3 & \text{if } \tau_2 < y^*_{ij} \leq \tau_3 \\ 4 & \text{if } \tau_3 < y^*_{ij}, \end{cases} \tag{2}$$

The bias-corrected Mundlak (1978) specification is useful for at least five reasons. First, as Appendix C demonstrates, state capacity changes little within countries. Accordingly, a fixed effects approach that only estimates within-unit changes is inappropriate: although it is still consistent, it discards almost all relevant information and produces imprecise estimates (Bell and Jones, 2015, 139). Second, the Mundlak (1978) specification enables the model to capture the effects of time-variant variables and overcome omitted variable bias in the

[24] Wooldridge (2010) and Schunck (2013), as well as other econometricians, often refer to the Mundlak (1978) approach as the correlated random effects model.

[25] I focus the main results on those entailing a single intercept per Kropko and Kubinec (2020), who show that two-way fixed effects and their multilevel equivalents are very difficult to interpret. Nevertheless, I consider the robustness of the results to two random intercepts in Table 4.

same way as the fixed effects model (Hazlett and Wainstein, 2022).[26] Third, the Mundlak (1978) specification enables for comparison of between/across-country effects (β_b) and within-country (β_w) ones. Fourth, the use of cluster-robust standard errors enables the model to overcome typical biases in the multilevel model and recover the same standard errors as the fixed effects model (Hazlett and Wainstein, 2022). Fifth, I allow for minimal deviation from the fixed effects coefficients and standard errors by incorporating partial pooling. In turn, the estimates regularize the coefficients toward their respective global means and overcome small-sample biases when countries receive few projects over the sample time period (see Gelman and Hill, 2007, 252). Table 4 summarizes the numerous robustness tests considered in the paper.

4. Results

4.1. State Capacity

Table 3 presents the main results in terms of average marginal effects.[27] In the base models corresponding to the minimal adjustment sets to satisfy the backdoor criterion, a one-unit increase in state capacity between countries raises the probability that a project has full, category-4 safeguard compliance by 15-16 percentage points. In the models with full covariates corresponding to the canonical adjustment sets, that number is 13 percentage points. For the the lower-compliance categories of 3, 2, and 1, a one-unit increase in state capacity between countries reduces the probability of compliance by 2-7 percentage points in the full models, respectively. Given that Hanson and Sigman's (2021) state capacity variable ranges from circa -1.67 to 1.55 in this dataset, a one-unit increase in state capacity corresponds to a 32 percentage point increase in the variable.[28] These magnitudes are very high. Although the between-country estimates are highly statistically significant throughout

[26]The fixed effects model is unbiased under conditional independence (Hazlett and Wainstein, 2022, 56).
[27]See Arel-Bundock, Greifer and Heiss (2024) for more on the different types of marginal effects.
[28]$1/(1.67 + 1.55) = 0.32$

Table 3: Compliance with World Bank Safeguard Policies 2007-2015
(Mundlak Ordered Logit Models with Country Random Intercepts)

	(1) State Capacity (base1)	(2) State Capacity (base2)	(3) State Capacity (full)	(4) State Capacity (full lc)	(5) TTL Quality (base)	(6) TTL Quality (full)	(7) TTL Quality (full lc)	(8) Board (base)	(9) Board (full)	(10) Board (full lc)
Panel A: Between-Country Effects										
Compliance = 1	-0.0242***	-0.0280***	-0.0282***	-0.0270***	-0.0488***	-0.0144	-0.0249*	-0.0061	-0.0172	-0.0143
	(0.0056)	(0.0063)	(0.0071)	(0.0064)	(0.0155)	(0.0108)	(0.0128)	(0.0168)	(0.0114)	(0.0116)
Compliance = 2	-0.0617***	-0.0710***	-0.0723***	-0.0693***	-0.1157***	-0.0370	-0.0624*	-0.0145	-0.0442	-0.0368
	(0.0132)	(0.0135)	(0.0176)	(0.0160)	(0.0309)	(0.0277)	(0.0322)	(0.0402)	(0.0277)	(0.0284)
Compliance = 3	-0.0617***	-0.0665***	-0.0614***	-0.0641***	-0.0974***	-0.0300	-0.0490*	-0.0131	-0.0358	-0.0295
	(0.0153)	(0.0134)	(0.0148)	(0.0148)	(0.0254)	(0.0219)	(0.0262)	(0.0371)	(0.0226)	(0.0228)
Compliance = 4	0.1476***	0.1655***	0.1619***	0.1604***	0.2619***	0.0814	0.1362*	0.0337	0.0972	0.0806
	(0.0313)	(0.0294)	(0.0365)	(0.0346)	(0.0665)	(0.0600)	(0.0697)	(0.0940)	(0.0609)	(0.0623)
Panel B: Within-Country Effects										
Compliance = 1	-0.0074	-0.0042	-0.0019	-0.0112	-0.0225***	-0.0190***	-0.0196***	0.0121	0.0095	0.0083
	(0.0228)	(0.0241)	(0.0236)	(0.0229)	(0.0062)	(0.0054)	(0.0056)	(0.0079)	(0.0114)	(0.0107)
Compliance = 2	-0.0189	-0.0105	-0.0048	-0.0288	-0.0535***	-0.0487***	-0.0491***	0.0289	0.0245	0.0214
	(0.0579)	(0.0612)	(0.0605)	(0.0588)	(0.0114)	(0.0117)	(0.0118)	(0.0189)	(0.0285)	(0.0269)
Compliance = 3	-0.0189	-0.0099	-0.0041	-0.0266	-0.0450***	-0.0396***	-0.0386***	0.0260	0.0199	0.0172
	(0.0578)	(0.0571)	(0.0513)	(0.0543)	(0.0094)	(0.0098)	(0.0097)	(0.0179)	(0.0235)	(0.0219)
Compliance = 4	0.0453	0.0246	0.0107	0.0666	0.1210***	0.1073***	0.1074***	-0.0670	-0.0540	-0.0470
	(0.1385)	(0.1424)	(0.1354)	(0.1359)	(0.0240)	(0.0249)	(0.0245)	(0.0441)	(0.0631)	(0.0594)
Observations	1268	1162	1112	1217	1148	1055	1063	1190	1055	1055

Marginal effects with country-clustered robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: “base” refers to the minimal adjustment set to satisfy Pearl’s (2009) backdoor criterion;
“full” refers to the variables in Perković et al.’s (2018) canonical adjustment set;
“full lc” refers to the canonical adjustment set, minus any variables excluded due to potential collinearity (see Table 2)

($p < .001$), the within-country estimates do not reach statistical significance. These results are consistent with the slow-moving nature of state capacity (see Appendix C). The robustness tests summarized in Table 4 indicate very similar results regardless of the specification. Overall, state capacity between countries is a very strong predictor of safeguard compliance.

4.2. TTL Quality

Per Table 3, the TTL variable is a highly statistically significant, positive predictor of category-4 compliance outcomes, as well as a statistically significant, negative predictor of category 1-3 compliance outcomes within countries. In terms of average marginal effects, a one-unit increase in the TTL category raises the probability of receiving a category-4, full compliance score by 11-12 percentage points within countries. For the lower-compliance categories (1-3), a one-unit increase in TTL quality decreases the probabilities of receiving these lower compliance outcomes by 5-7 percentage points. The results are also similar across the numerous robustness tests detailed in Table 4, suggesting that TTL quality is a robust predictor of safeguard compliance within countries.

Effects of TTL quality between countries has a less consistent, though still overall, positive effect on compliance. In all base models that minimally satisfy Pearl's (2009) backdoor criterion under selection on observables, all estimates in Tables 3 and 4 are statistically significant at the 1% level. By contrast, TTL quality between countries often loses statistical significance after augmenting the controls to correspond with Perković et al.'s (2018) canonical adjustment set (see Table 2). Given that models with many controls are subject to potential collinearity or positivity challenges, I remove some variables according to Table 2. TTL quality between countries often regains statistical significance in these models with a smaller number of covariates, though usually only at the 10% level. The broader takeaway is that TTL quality between countries likely matters most of the time but not always. Practically, the result suggests that a lot—though not all—TTL knowledge is country-specific (Heinzel, 2022*b*), so re-assigning better TTLs to other countries may not always yield better

Table 4: Robustness Tests for the Main Effects

Model	Table	State Capacity (between)	TTL Quality (between)	TTL Quality (within)	Board (between)	Board (within)	Notes
Traditional ordered logit without Mundlak structure	B.1	0.1472*** 0.1638*** 0.1763*** 0.1817***	0.2219*** 0.0653 0.1240*		-0.0184 0.0284 0.0245		Models lack a Mundlak structure here, so between column reflects overall coefficients.
Linear model with country random intercept	B.2	0.1635*** 0.1718*** 0.1735*** 0.1744***	0.2219*** 0.0653 0.1240*	0.0909*** 0.0770*** 0.0864***	0.0223 0.0864 0.0651	-0.0303 -0.0127 -0.0057	
Logit model with country random intercept (compliance categories 1–3 coded to 0; category 4 coded to 1)	B.3	0.7345*** 0.7700*** 0.7990*** 0.7957***	0.9925*** 0.3151 0.5844*	0.4141*** 0.3628*** 0.3694***	0.0833 0.4413 0.3233	-0.1292 -0.0607 -0.0258	The coding follows the distribution of compliance (see Figure 3a).
Cross-sectional Mundlak ordered logit model	B.4	0.1598*** 0.1694*** 0.1579*** 0.1543***	0.2970*** 0.0881 0.1907***	0.1209*** 0.1078*** 0.1082***	0.1138 0.1187** 0.1059*	-0.0844 -0.0541 -0.0465	
Ordered logit with country and year random intercepts	B.5	0.1480*** 0.1663*** 0.1626*** 0.1607***	0.2631*** 0.0808 0.1359**	0.1176*** 0.1048*** 0.1056***	0.0333 0.0990 0.0812	-0.0693 -0.0546 -0.0478	Year nested to maintain clustered robust SEs.
Subsetting to only projects with an IEG evaluation	B.6	0.1445*** 0.1616*** 0.1530*** 0.1488***	0.2619*** 0.0814 0.1362*	0.1210*** 0.1073*** 0.1074***	0.0402 0.0972 0.0806	-0.0353 -0.0540 -0.0470	Kilby and Michaelowa (2019) suggest potential favoritism in which projects IEG evaluates—hence the reason for the model.
Subsetting to only projects triggering indigenous peoples, cultural resources, and resettlement safeguards	B.7	0.1511*** 0.1666*** 0.1467*** 0.1474***	0.2661*** 0.2546***	0.1322*** 0.1199***	0.0695 0.1348**	-0.1435** -0.1094	
Switching state capacity measure to Government Effectiveness from the Worldwide Governance Indicators	B.8	0.1120*** 0.1320*** 0.1353*** 0.1410***	0.2619*** 0.1263*** 0.1074***	0.1210*** 0.1075*** 0.1301***	0.0337 0.1304** 0.1128*	-0.0670 -0.0494 -0.1427	
Estimating the direct effects of the UNSC and US ideal point distance variables with appropriate adjustment sets	B.9						No US Ideal point estimate reaches statistical significance. Only 1/24 UNSC estimates reaches significance at the 10% level.

Note: see Appendix B for full tables. All effects, except for those corresponding the linear model, are average marginal effects; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. State capacity (within) coefficient excluded because it is insignificant throughout, consistent with Appendix C and Table 3. Coefficients for ordered logit models correspond to full, category-4 compliance outcomes.

compliance outcomes.

4.3. Principal (Donor) Variables

The Board variable neither consistently predicts the noncompliance outcomes nor the full, category-4 compliance outcome between or within countries. In the main results captured by Table 3, the Board variable never reaches statistical significance. In one base model and some of the full model robustness checks corresponding to Perković et al.'s (2018) canonical adjustment sets (see Table 2), the Board variable does reach statistical significance. However, the coefficients are difficult to interpret. First, the fact that the effect only appears after introducing many controls suggests that Board is not very strong by itself. Second, these robustness are not preferred specifications (see Section 3.3). Third, the coefficients are negative for the within-country effects, whereas those for between countries are positive. *A priori*, a true effect in either direction would be interesting. For example, it is possible to surmise that Board membership could spur increased compliance via the spotlight that it brings. Alternatively, decreased compliance from Board membership could suggest that agents have difficulty supervising principals, consistent with Figure 1. The challenge is that having both effects take place at once, varying within and between countries, is rather implausible and difficult to unpack—particularly given the lack of consistent evidence.

Donors' abilities to influence compliance outside of their formal Board positions is similarly weak. Both the UNSC variable and the US ideal point distance, which reflect informal influence, show similarly poor abilities to predict compliance. For example, none of the US ideal point estimates reach statistical significance, and only 1/24 UNSC estimates reach statistical significance at the 10% level (see Table B.9).

More broadly, the inconsistent donor interest results take on special meaning given that Nielson and Tierney's (2003) seminal article on principal-agent theory's relevance to MDBs uses safeguard failure for a confirming case study. Following Abadie (2020), statis-

Table 5: Model Performance Metrics (Country Random Intercept Models)

Specification	AIC	BIC	Conditional R^2	Marginal R^2	RMSE	σ residuals
SC Traditional	2757.75	2783.50	0.11	0.04	2.96	2.96
TTL Traditional	2522.89	2548.13	0.12	0.03	2.95	2.95
SC Mundlak	2759.14	2790.04	0.11	0.05	2.96	2.96
TTL Mundlak	2519.94	2550.22	0.13	0.04	2.95	2.96

Note: "SC" corresponds to state capacity; "Traditional" corresponds to the regular multilevel ordered model without the Mundlak structure. "Mundlak" refers to the models with both between and within treatments. The SC models exclude the TTL variable, and the TTL models exclude the state capacity variable. All models are bivariate, with compliance as the dependent variable. R^2 calculations follow the latent scale according to Nakagawa and Schielzeth (2013) and Lüdecke et al. (2021).

tical nonsignificance is often more informative than significance itself. I do not interpret further control variables due to the Table 2 fallacy and the difficulty associated with their interpretation (see Westreich and Greenland, 2013). As the DAG clarifies (see Figure 4), each treatment requires its own unique set of controls.

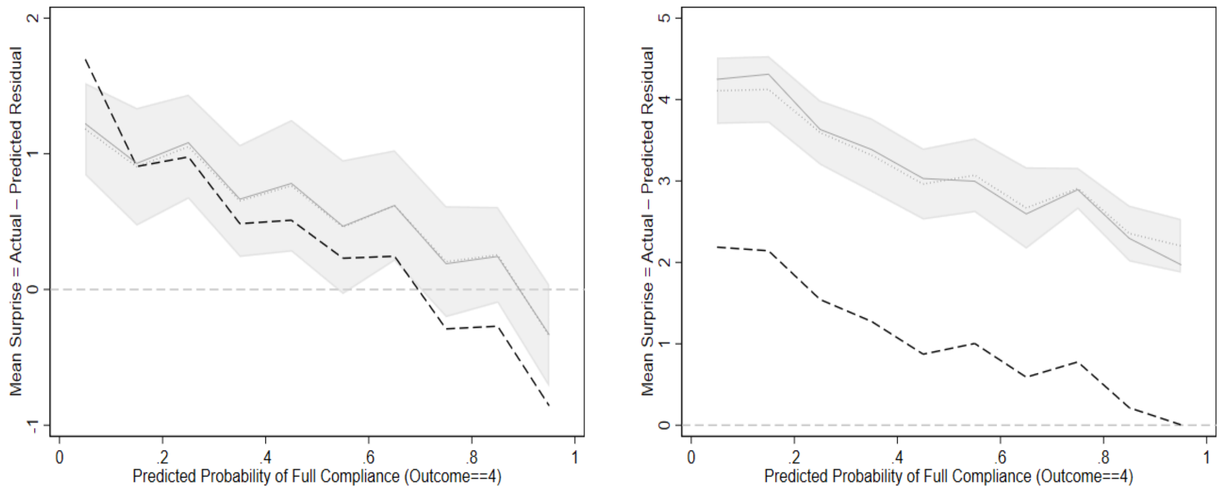
4.4. Variance and Residual Decomposition

To answer the question of whether the TTL variable or state capacity explains more variance, I re-run bivariate multilevel ordered logit models, emplying both traditional and Mundlak specifications. Then, I assess the models across numerous metrics in Table 5. As the results suggest, the models with the TTL variable fit slightly better according to the Akike Information Criterion (AIC), Bayesian Information (BIC), and root mean squared error (RMSE). Furthermore, the TTL models generally have slightly higher conditional and marginal R^2 values, suggesting that TTL explains more variance. Nevertheless, these differences are small throughout, indicating that a model without either state capacity or TTL quality runs a very high risk of omitted variable bias and, in turn, endogeneity.

To further probe model fit, I conduct residual decomposition ("surprise") analysis along the lines of Card and Dahl (2011). The latter show that, after controlling for Las Vegas point spreads, U.S. football game outcomes are essentially as-if random—the average surprise is

Figure 5: Residual Decomposition by Safeguard Risk Category

(a) State Capacity (Overall Safeguard Risk) (b) TTL (Overall Safeguard Risk)



Note: Diagnostic plots adapted from Card and Dahl (2011). Plots with mean surprise closer to zero indicate better regression fit. Dashed black lines = addition of between coefficients; solid gray lines = addition of within coefficients; dotted gray lines = baseline estimates without treatments. 95% confidence intervals correspond to baseline estimates, which reflect the adjustment sets from Table 2. All plot contain 10 equally-spaced bins to measure preducted probabilities on the x-axis. Figure D.2 provides estimates by safeguard category.

zero. In World Bank safeguards, the closest analogue to the spread is the pre-implementation risk category, which closely tracks actual compliance outcomes (see Table 1). By controlling for the category as well as the full adjustment set in Table 2, I reduce *ex-ante* predictability relative to each category's starting point. Assuming that the model is correctly specified, omitted variables are the source of any remaining surprises that deviate from a residual mean of 0 on the y-axis in Figure 5. Its x-axis, which reflects deciles bins, capture the differing predicted probabilities of attaining full, category-4 compliance. The modeling strategy remains an Mundlak ordered logit model with a country random intercept and clustered robust standard errors by country. Both the between coefficient in dashed black and within coefficients in solid, dark gray are relative to the baseline without either state capacity or the TTL in dotted, light gray. The 95% confidence interval bands refer to the baseline.

Overall, the between-country coefficients in the black-dashed lines reduce residual surprise the most. The solid-gray within-country coefficients for both state capacity and the TTL stay similar to dotted-gray baseline estimates without them. Although only the TTL between-country estimates fall outside the baseline confidence interval, the state capacity model residual surprises are closer to zero and better fitting, so both sets of estimates remain informative. To the extent that it is possible make inferences from the small-sample plots by category in Figure D.2,[29] state capacity reduces residual surprises the most in high-risk, category A projects, but less so for other categories. By contrast, TTLs do not reduce residual surprises much in high-risk, category-A projects but do so in moderate-risk, category B situations. Neither the TTL nor state capacity have much impact in low-risk, category-C projects, which is unsurprising given the lower difficulty of the task.

4.5. Moderation Analyses

There are two types of potential moderation that are potentially relevant for discerning the universality of the results. The first is whether TTL quality can moderate state capacity, such that one can substitute for the other. Unfortunately, as the DAG suggests, TTL is a descendant of the commitment mediator (see Figure 4), so such an analysis is empirically fraught.[30] Second, although I do not find that the principal has any immediate effects on safeguard compliance outcomes, it is possible that the principal could moderate the effect of state capacity or TTL quality on safeguard compliance. These analyses are empirically straightforward.

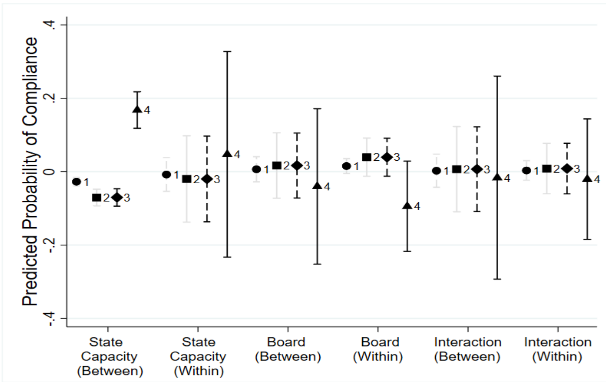
While the Board's ability to moderate state capacity is low (see Figure 6a), Board membership has some marginal impact on TTLs' ability to engineer compliance across countries. I characterize the impact as marginal because the relevant p-values hover around 0.1, and the effect is not present within countries. Figure 6b demonstrates that full compliance (category 4) tends to decrease the effect of the TTL with Board membership between countries,

[29] Sample sizes by category follow Table 1.

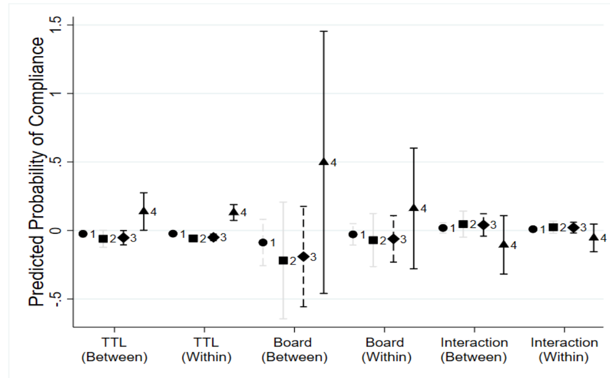
[30] The analysis is technically possible, but performing it changes the estimand to a direct effect, which is not the quantity of interest. This analysis focuses on the total effect.

Figure 6: Moderation Analyses (Country Intercept Models with Full Controls)

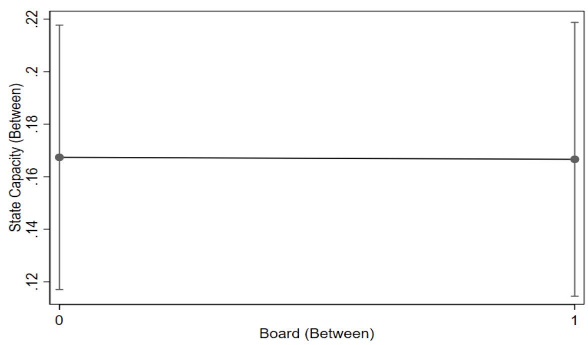
(a) State Capacity × Board (Coefficients)



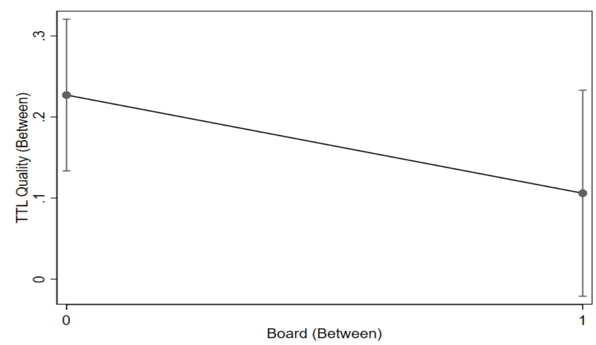
(b) TTL Quality × Board (Coefficients)



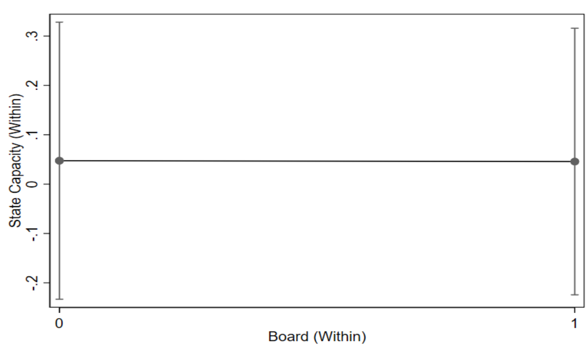
(c) State Capacity × Board (Between Plot)



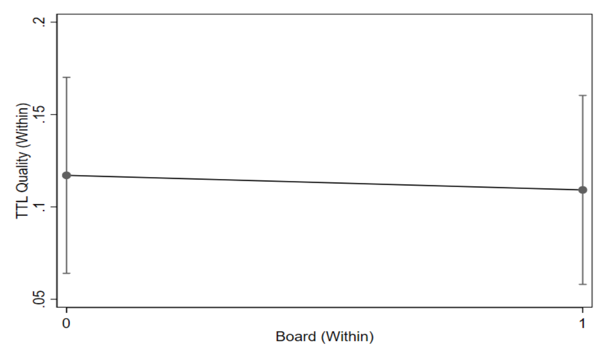
(d) TTL Quality × Board (Between Plot)



(e) State Capacity × Board (Within Plot)



(f) TTL Quality × Board (Within Plot)



Note: All coefficients correspond to marginal effects from Mundlak multilevel ordered logit models with country random intercepts, controls from the canonical adjustment sets, and country-clustered robust standard errors. Interaction plots correspond to the full compliance category (category 4). The DAG-derived adjustment set for the state capacity interactions includes: civil conflict, democracy, FDI, natural resources, property rights, safeguard category, UNSC, and US ideal point distance. The DAG-derived adjustment set for the TTL interactions adds the commitment.

whereas the opposite pattern takes place with less than full compliance (categories 1-3). Per Figure 6d, the size of the relevant slope change is substantial, too. By contrast, the slope changes in the other interaction plots in Figure 6 are minimal.

I also assess the extent to which such findings travel beyond the formal influence of the Board to reflect informal influence. Because the DAG indicates that the US ideal point is an ancestor of state capacity (see Figure 4), I forgo analysis with the US ideal point and focus on moderation with the UNSC variable. Figure D.1 present the results for both the state capacity and TTL quality variables. Neither the between nor within effects achieve statistical significance, and the relevant sizes of the slope changes are minimal.

Overall, consistent with Figure 1, there is some evidence that MDB staff have some difficulty supervising powerful countries via formal channels, not informal ones. In line with Nielson and Tierney (2005), evidence suggests that staff-implementer interactions are not principal-agent relationships. Otherwise, staff's ability to engineer safeguard compliance outcomes would not change based the power of the recipient country in the respective institution.

5. Implications for Theory, Policy, and External Validity

It is worth repeating that the above results correspond to the World Bank, which is the aid agency with the most developed environmental and social safeguard policies to prevent negative aid externalities. Over time, other MDBs and bilateral aid agencies have emulated the World Bank's policies (Greenstein, 2022), but that emulation and implementation after emulation takes time. It is thus clear that the above results correspond to what case study scholars call an "extreme case" (see Gerring, 2017). What that means here is that the World Bank is, *on average*, more likely than other aid agencies to have the ability to design projects and deploy high-quality agents to prevent or mitigate negative aid externalities. For this reason, state capacity is likely to be even more essential for determining safeguard outcomes at other MDBs and bilateral aid agencies with less safeguard policy experience.

More broadly, the importance of state capacity in driving results calls into question the focus of previous media exposés on safeguards as well as the previous academic literature. In line with the 2005 Paris Declaration on Aid Effectiveness, the World Bank generally does not implement its own projects, so it also does not implement its own safeguard policies. Furthermore, the cases that reach the World Bank Board and quasi-judicial accountability bodies like the World Bank Inspection Panel are a *selected sample* extreme safeguard policy failures.[31] For this reason, the focus of media exposés—including that of the highly-respected International Consortium of Investigative Journalists (2015)—and the academic literature on inspection panels and their findings does not provide a full picture safeguard compliance. More specifically, previous media exposés and literature are subject to what Tversky and Kahneman (1974) famously called availability and representative biases. Only by examining the full range of safeguard policy outcomes can scholars and lay audiences understand that compliance, not significant policy failure, is the norm.

Another implication of the findings is that the dominant framework for theorizing about foreign aid, the principal-agent model, likely needs augmentation to better capture dynamics with recipient countries. To be clear, the seminal work from Pollack (1997), Nielson and Tierney (2003), and Hawkins et al. (2006*b*), among others, remains highly relevant and useful for capturing dynamics between principals and agents. However, the literature's focus on principals and agents has come at the neglect of recipient state capacity issues, and the statistical results in the present paper suggest that neglect is costly. To remedy this issue, I propose supplementing the principal-agent model with a second level focusing on the incomplete contract between the agent and aid recipient country implementer (see Figure 1). Doing so will enable the academic literature to make better sense of the true nature of foreign

[31] For more on inspection panels, see Fox (2002).

aid, which is subject to hold-up problems and other agent-implementer power dynamics that prevent agents from *engineering* outcomes in the way that the principal-agent model focusing on agency slack suggests. After all, all foreign aid—regardless of whether it is bilateral or multilateral in nature—has not two but at least three main actors: principals, agents, and recipients/implementers. In cases where procurement is relevant for describing outcomes (e.g., Malik and Stone, 2018), there are even more actors to consider, too.

6. Conclusion

Even if an aid project accomplishes all of its objectives, its potential negative externalities—such as destruction of habitats, involuntary resettlement, and the loss of indigenous cultural property—can outweigh the benefits of undertaking an aid project in the first place. The present paper is thus not about reaching aid effectiveness targets but making sure that those targets are not achieved at all costs. In contrast to the previous literature on aid externalities, which focuses on externalities deriving from long causal chains, the present paper examines the direct social and environmental negative externalities of aid. It does so by examining aid recipients' project-level compliance with World Bank social and environmental safeguard policies (see Figure 2).

Consistent with past literature, I find that agency matters, but also I find that state capacity is a primary predictor of safeguard policy compliance—and, by extension, the prevention or mitigation of negative aid externalities. These results are consistent with my broader theory. It stresses that agent-implementer interactions are subject to incomplete contracts, hold-up problems, and political incentives that differ from typical aid compliance scenarios.[32] In the case of safeguard policy compliance, noncompliance in the form of, say, destruction of a rainforest or indigenous people's land rarely constitutes strong political strategy. Although politicians may wish to punish some outgroups on the margins, their

[32]See Girod and Tobin (2016) for more typical aid compliance scenarios.

larger incentive is to credit claim from aid and prolong their power (Cruz and Schneider, 2017; Baldwin and Winters, 2023). Along those lines, recipient states wish to avoid the penalty in terms of less future aid commitments that Buntaine (2016) documents. That is why aid recipients, regardless of their political regime, generally attempt to avoid negative aid externality outcomes if their state has the capacity to do so.

Finally, the present study responds to what Falleti (2021) called the "invisibility" of indigenous issues to political science and provides some cross-country, quantitative data to better understand these topics and related environmental ones.[33] I say "better understand" because the literature and media has to date focused on anthropological case studies and a very *selected sample* of projects that reach accountability bodies like the Inspection Panels (e.g., Fox, 2002; Randeria, 2003; Randeria and Grunder, 2011; International Consortium of Investigative Journalists, 2015; Tello, 2015; Zvobgo and Graham, 2020). To further overcome availability and representativeness biases, future research can use the new data advanced in this article to investigate when state capacity and agency break down. By doing so, scholars and policymakers will be able to better protect indigenous communities and the environment from the negative externalities of foreign aid.

[33]Notably, Falleti (2021) finds that that the top-3 journals—*American Journal of Political Science*, *American Political Science Review*, and *Journal of Politics*—only published 11 articles with the word "indigenous" in the title or abstract from 1990-2020.

References

Abadie, Alberto. 2020. "Statistical Nonsignificance in Empirical Economics." *American Economic Review: Insights* 2(2):193–208.

Acemoglu, Daron and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York: Crown Business.

Altincekic, Ceren and David H. Bearce. 2014. "Why There Should Be No Political Foreign Aid Curse." *World Development* 64(October 2012):18–32.

Andersen, Jørgen Juel and Michael L. Ross. 2014. "The Big Oil Change: A Closer Look at the Haber-Menaldo Analysis." *Comparative Political Studies* 47(7):993–1021.

Andrews, Matt, Lant Pritchett and Michael Woolcock. 2017. *Building State Capability: Evidence, Analysis, Action*. New Haven, Connecticut: Oxford University Press.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, New Jersey: Princeton University Press.

Arel-Bundock, Vincent, Noah Greifer and Andrew Heiss. 2024. "How to Interpret Statistical Models Using marginaleffects for R and Python." *Journal of Statistical Software* 111(9):1–32.

Arndt, Channing, Sam Jones and Finn Tarp. 2015. "Assessing Foreign Aid's Long-run Contribution to Growth and Development." *World Development* 69:6–18.

Arrow, Kenneth J. 1970. Political and Economic Evaluation of Social Effects and Externalities. In *The Analysis of Public Output*, ed. Julius Margiolis. NBER chapter 1, pp. 1–23.

Ashton, Louise, Jed Friedman, Diana Goldemberg, Mustafa Zakir Hussain, Thomas Kenyon, Akib Khan and Mo Zhou. 2023. "A Puzzle with Missing Pieces: Explaining the Effectiveness of World Bank Development Projects." *World Bank Research Observer* 38:115–146.

Bailey, Michael A., Anton Strezhnev and Erik Voeten. 2017. "Estimating Dynamic State Preferences from United Nations Voting Data." *Journal of Conflict Resolution* 61(2):430–456.

Baldwin, Kate and Matthew S. Winters. 2023. "Foreign Aid and Political Support: How Politicians' Aid Oversight Capacity and Voter Information Condition Credit-Giving." *World Politics* 75(1):1–42.

Barrett, Malcolm. 2025. "ggdag: Analyze and Create Elegant Directed Acyclic Graphs. R package.".
URL: *https://github.com/r-causal/ggdag*

Bazzi, Samuel, Arya Gaduh, Alexander D. Rothenberg and Maisy Wong. 2016. "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia." *American Economic Review* 106(9):2658–2698.

Bell, Andrew and Kelvyn Jones. 2015. "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data." *Political Science Research and Methods* 3(1):133–153.

Bennon, Michael and Francis Fukuyama. 2022. "The Obsolescing Bargain Crosses the Belt and Road Initiative: Renegotiations on BRI Projects." *Oxford Review of Economic Policy* 38(2):278–301.

Bermeo, Sarah Blodgett. 2016. "Aid Is Not Oil: Donor Preferences, Heterogeneous Aid, and the Aid-Democratization Relationship." *International Organization* 70(1):1–32.

Bernauer, Thomas and Vally Koubi. 2009. "Effects of Political Institutions on Air Quality." *Ecological Economics* 68(5):1355–1365.

Besley, Timothy and Torsten Persson. 2010. "State Capacity, Conflict, and Development." *Econometrica* 78(1):1–34.

Booth, David. 2011. "Aid, Institutions and Governance: What Have We Learned?" *Development Policy Review* 29:s5–s26.

Börzel, Tanja A., Tobias Hofmann, Diana Panke and Carina Sprungk. 2010. "Obstinate and Inefficient: Why Member States Do Not Comply with European Law." *Comparative Political Studies* 43(11):1363–1390.

Bourguignon, François and Mark Sundberg. 2007. "Aid Effectiveness: Opening the Black Box." *American Economic Review: Papers and Proceedings* 97(2):316–321.

Bueno de Mesquita, Bruce and Alastair Smith. 2009. "A Political Economy of Aid." *International Organization* 63(2):309–340.

Bulman, David, Walter Kolkma and Aart Kraay. 2017. "Good Countries or Good Projects? Macro and Micro Correlates of World Bank and Asian Development Bank Project Performance." *Review of International Organizations* 12:335–363.

Buntaine, Mark T. 2016. *Giving Aid Effectively: The Politics of Environmental Performance and Selectivity at Multilateral Development Banks.* Oxford: Oxford University Press.

Card, David and Gordon B. Dahl. 2011. "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior." *Quarterly Journal of Economics* 126(1):103–143.

Centeno, Miguel Angel, Atul Kohli and Deborah J. Yashar. 2017. Unpacking States in the Developing World: Capacity, Performance, and Politics. In *States in the Developing World*, ed. Miguel Angel Centeno, Atul Kohli and Deborah Yashar. Cambridge: Cambridge University Press chapter 1, pp. 1–31.

Cernea, Michael M. 2008. Compensation and Investment in Resettlement: Theory, Practice, Pitfalls, and Needed Policy Reform. In *Can Compensation Prevent Impoverishment? Reforming Resettlement Through Investments and Benefit-Sharing*, ed. Michael M. Cernea and Hari Mohan Mathur. Oxford: Oxford University Press chapter 1, pp. 15–82.

Chayes, Abram and Antonia Handler Chayes. 1993. "On Compliance." *International Organization* 47(2):175–205.

Chwieroth, Jeffrey M. 2008. "Normative Change from Within: The International Monetary Fund's Approach to Capital Account Liberalization." *International Studies Quarterly* 52(1):129–158.

Coase, Ronald H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3(1):1–44.

Crost, Benjamin, Joseph H. Felter and Patrick B. Johnston. 2014. "Aid Under Fire: Development Projects and Civil Conflict." *American Economic Review* 104(6):1833–1856.

Cruz, Cesi and Christina J. Schneider. 2017. "Foreign Aid and Undeserved Credit Claiming." *American Journal of Political Science* 61(2):396–408.

D'Arcy, Michelle, Marina Nistotskaya and Ola Olsson. 2024. "Cadasters and Economic Growth: A Long-Run Cross-Country Panel." *Journal of Political Economy* 132(11):3785–3826.

Denizer, Cevdet, Daniel Kaufmann and Aart Kraay. 2013. "Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance." *Journal of Development Economics* 105(2013):288–302.

Denly, Michael. 2021. Institutional Autonomy and Donor Strategic Interest in Multilateral Foreign Aid: Rules vs. Influence. In *Political Economy of International Organizations*.

Denly, Michael, Michael G. Findley, Joelean Hall, Andrew Stravers and James Igoe Walsh. 2022. "Do Natural Resources Really Cause Civil Conflict? Evidence from the New Global Resources Dataset." *Journal of Conflict Resolution* 66(3):387–412.

Djankov, Simeon, Jose G. Montalvo and Marta Reynal-Querol. 2008. "The Curse of Aid." *Journal of Economic Growth* 13(3):169–194.

Downs, George W., David M. Rocke and Peter N Barsoom. 1996. "Is the Good News about Compliance Good News about Cooperation?" *International Organization* 50(3):379–406.

Dreher, Axel, Jan-Egbert Sturm and James Raymond Vreeland. 2009. "Development Aid and International Politics: Does Membership on the UN Security Council Influence World Bank Decisions?" *Journal of Development Economics* 88(1):1–18.

Dreher, Axel, Matthew Gould, Matthew D. Rablen and James Raymond Vreeland. 2014. "The Determinants of Election to the United Nations Security Council." *Public Choice* 158(1-2):51–83.

Dube, Oeindrila and Suresh Naidu. 2015. "Bases, Bullets, and Ballots: The Effect of U. S. Military Aid on Political Conflict in Colombia." *Journal of Politics* 77(1):249–267.

Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. New York: Cambridge University Press.

Easterly, William. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. New York: Penguin Books.

Falleti, Tulia G. 2021. "Invisible to Political Science: Indigenous Politics in a World in Flux." *Journal of Politics* 83(1):e5–e12.

Fariss, Christopher J., Michael R. Kenwick and Kevin Reuning. 2020. Measurement Models. In *SAGE Handbook of Research Methods in Political Science and International Relations*, ed. Luigi Curini and Robert J. Franzese. London: Sage Publications chapter 20, pp. 353–370.

Findley, Michael G., Adam S. Harris, Helen V. Milner and Daniel L. Nielson. 2017. "Who Controls Foreign Aid? Elite Versus Public Perceptions of Donor Influence in Aid-Dependent Uganda." *International Organization* 71(4):633–663.

Finnemore, Martha and Kathryn Sikkink. 1998. "International Norm Dynamics and Political Change." *International Organization* 52(4):887–917.

Fox, Jonathan A. 2002. The World Bank Inspection Panel and the Limits of Accountability. In *Reinventing the World Bank*, ed. Jeffrey A Winters. Ithaca and London: Cornell University Press pp. 131–163.

Fukuyama, Francis. 2004. *State-Building: Governance and World Order in the 21st Century*. Ithaca, New York: Cornell University Press.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Modelling*. Cambridge: Cambridge University Press.

Gerring, John. 2017. *Case Study Research: Principles and Practices*. Second ed. Cambridge: Cambridge University Press.

Girod, Desha M. and Jennifer L. Tobin. 2016. "Take the Money and Run: The Determinants of Compliance with Aid Agreements." *International Organization* 70(1):209–239.

Greenstein, Gus. 2022. "The Influence of Alternative Development Finance on the World Bank's Safeguards Regime." *Global Environmental Politics* 22(3):171–193.

Gutner, Tamar. 2005*a*. "Explaining the Gaps between Mandate and Performance: Agency Theory and World Bank Environmental Reform." *Global Environmental Politics* 5(2):10–37.

Gutner, Tamar. 2005*b*. "World Bank Environmental Reform: Revisiting Lessons from Agency Theory." *International Organization* 5(3):773–783.

Hanson, Jonathan K. and Rachel Sigman. 2021. "Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research." *Journal of Politics* 83(4):1495–1510.

Hart, Oliver. 2017. "Incomplete Contracts and Control." *American Economic Review* 107(7):1731–1752.

Hawkins, Darren, David A. Lake, Daniel L. Nielson and Michael J. Tierney. 2006*a*. Delegation under Anarchy: States, International Organizations, and Principal-Agent Theory. In *Delegation and Agency in International Organizations*. New York: Cambridge University Press chapter 1, pp. 3–38.

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

Hawkins, Darren and Wade Jacoby. 2006. How Agents Matter. In *Delegation and Agency in International Organizations*. Cambridge: Cambridge University Press chapter 7, pp. 199–228.

Hazlett, Chad and Leonard Wainstein. 2022. "Understanding, Choosing, and Unifying Multilevel and Fixed Effect Approaches." *Political Analysis* 30(1):46–65.

Heinzel, Mirko. 2022*a*. "Divided Loyalties? The Role of National IO Staff in Aid-Funded Procurement." *Governance* 35:1183–1203.

Heinzel, Mirko. 2022*b*. "International Bureaucrats and Organizational Performance. Country-Specific Knowledge and Sectoral Knowledge in World Bank Projects." *International Studies Quarterly* 66(sqac013):1–13.

Heinzel, Mirko and Andrea Liese. 2021. "Managing Performance and Winning Trust: How World Bank Staff Shapes Recipient Performance." *Review of International Organizations* 16:625–653.

Hendrix, Cullen S. 2010. "Measuring State Capacity: Theoretical and Empirical Implications for the Study of Civil Conflict." *Journal of Peace Research* 47(3):273–285.

Honig, Dan. 2019. "When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation." *International Organization* 73(1):171–201.

Huffington Post. 2015. "World Bank Safeguards.".
URL: *http://www.huffingtonpost.com/news/world-bank-safeguards/*

Independent Evaluation Group. 2010. Safeguards and Sustainability Policies in a Changing World: An Independent Evaluation of the World Bank Group Experience. Technical report Washington, DC: .

International Consortium of Investigative Journalists. 2015. "Evicted and Abandoned: The World Bank's Broken Promise to the Poor.".
URL: *http://www.icij.org/project/world-bank*

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

Kaja, Ashwin and Eric Werker. 2010. "Corporate Governance at the World Bank and the Dilemma of Global Governance." *World Bank Economic Review* 24(2):171–198.

Kapur, Devesh, John P Lewis and Richard Webb. 1997. *The World Bank: Its First Half Century (Volume 1: History)*. Vol. 1 Washington, DC: Brookings University Press.

Keohane, Robert and Joseph Nye. 1977. *Power and Interdependence: World Politics in Transition*. Boston: Little, Brown, and Company.

Kilby, Christopher. 2000. "Supervision and Performance: The Case of World Bank Projects." *Journal of Development Economics* 62(1):233–259.

Kilby, Christopher. 2001. "World Bank-Borrower Relations and Project Supervision." *Canadian Journal of Development Studies* 22(1):191–217.

Kilby, Christopher and Katharina Michaelowa. 2019. What Influences World Bank Project Evaluations? In *Lessons on Foreign Aid and Economic Development: Micro and Macro Perspectives*, ed. Nabamita Dutta and Claudia R. Williamson. Cham, Switzerland: Palgrave Macmillan chapter 6, pp. 109–150.

König, Thomas and Lars Mäder. 2014. "The Strategic Nature of Compliance: An Empirical Evaluation of Law Implementation in the Central Monitoring System of the European Union." *American Journal of Political Science* 58(1):246–263.

Kono, Daniel Y. and Gabriella R. Montinola. 2009. "Does Foreign Aid Support Autocrats, Democrats, or Both?" *Journal of Politics* 71(2):704–718.

Krasner, Stephen D. 2020. "Learning to Live with Despots: The Limits of Democracy Promotion." *Foreign Affairs* 5(1):43–54.

Kropko, Jonathan and Robert Kubinec. 2020. "Interpretation and Identification of Within-Unit and Cross-Sectional Variation in Panel Data Models." *Plos One* 15(4):1–22.

Krugman, Paul and Maurice Obstfeld. 2003. *International Economics: Theory and Policy*. Seventh ed. Boston: Pearson Education.

Lee, Melissa M. and Melina R. Platas. 2015. "Aid Externalities: Evidence from PEPFAR in Africa." *World Development* 67:281–294.

Lee, Melissa M. and Nan Zhang. 2017. "Legibility and the Informational Foundations of State Capacity." *Journal of Politics* 79(1):118–132.

Limodio, Nicola. 2021. "Bureaucrat Allocation in the Public Sector: Evidence from the World Bank." *Economic Journal* 131(January):3012–3040.

Lindberg, Staffan I., Michael Coppedge, John Gerring and Jan Teorell. 2014. "V-Dem: A New Way to Measure Democracy." *Journal of Democracy* 25(3):159–169.

Lüdecke, Daniel, Mattan S. Ben-Shachar, Indrajeet Patil and Dominique Makowski. 2021. "performance: An R Package for Assessment, Comparison and Testing of Statistical Models Statement of Need." *Journal of Open Source Software* 6:1–8.

Malik, Rabia and Randall W. Stone. 2018. "Corporate Influence in World Bank Lending." *Journal of Politics* 80(1):103–118.

Mann, Michael. 1984. "The Autonomous Power of the State: Its Origins, Mechanisms and Results." *Archives européennes de sociologie* 25(1984):185–213.

Martens, Bertin, Uwe Mummert, Peter Murrell and Paul Seabright. 2002. *The Institutional Economics of Foreign Aid*. Cambridge: Cambridge University Press.

Masi, Tania, Antonio Savoia and Kunal Sen. 2024. "Is there a Fiscal Resource Curse? Resource Rents, Fiscal Capacity and Political Institutions in Developing Economies." *World Development* 177(January).

McLean, Elena V. 2015. "Multilateral Aid and Domestic Economic Interests." *International Organization* 69(01):97–130.

McLean, Elena V. 2017. "The Politics of Contract Allocation in the World Bank." *Review of International Organizations* 12(2):255–279.

Menaldo, Victor. 2016. *The Institutions Curse: Natural Resources, Politics, and Development*. Cambridge: Cambridge University Press.

Montgomery, Jacob M., Brendan Nyhan and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62(3):760–775.

Morrison, Kevin M. 2012. "What Can We Learn about the Resource Curse from Foreign Aid?" *World Bank Research Observer* 27(1):52–73.

Morrison, Kevin M. 2013. "Membership No Longer Has Its Privileges: The Declining Informal Influence of Board Members on IDA Lending." *Review of International Organizations* 8(2):291–312.

Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica* 46(1):69–85.

Nakagawa, Shinichi and Holger Schielzeth. 2013. "A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-Effects Models." *Methods in Ecology and Evolution* 2:133–142.

Negi, Nalin Singh and Sujata Ganguly. 2011. "Development Projects vs. Internally Displaced Populations in India: A Literature Based Appraisal.".

Nielsen, Richard A., Michael G. Findley, Zachary S. Davis, Tara Candland and Daniel L. Nielson. 2011. "Foreign Aid Shocks as a Cause of Violent Armed Conflict." *American Journal of Political Science* 55(2):219–232.

Nielson, Daniel L. and Michael J. Tierney. 2003. "Delegation to International Organizations: Agency Theory and World Bank Environmental Reform." *International Organization* 57(02):241–276.

Nielson, Daniel L. and Michael J. Tierney. 2005. "Theory, Data, and Hypothesis Testing: World Bank Environmental Reform Redux." *International Organization* 59(3):785–800.

North, Douglass C. 1981. *Structure and Change in Economic History*. New York: W.W. Norton and Company.

Nunn, Nathan and Nancy Qian. 2014. "US Food Aid and Civil Conflict." *American Economic Review* 104(6):1630–1666.

OECD. 2008. "Paris Declaration of Aid Effectiveness and Accra Agenda for Action.".

OECD. 2017. "Purpose Codes: Sector Classification.".
URL: *http://www.oecd.org/dac/stats/purposecodessectorclassification.htm*

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, Massachusetts: Harvard University Press.

O'Reilly, Colin and Ryan M. Murphy. 2022. "An Index Measuring State Capacity: 1789-2018." *Economica* 89(355):713–745.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Ostrom, Elinor, James Walker and Roy Gardner. 1992. "Covenants With and Without a Sword: Self-Governance is Possible." *American Political Science Review* 86(2):404–417.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Perković, Emilija, Johannes Textor, Markus Kalisch and Marloes H. Maathuis. 2018. "Complete Graphical Characterization and Construction of Adjustment Sets in Markov Equivalence Classes of Ancestral Graphs." *Journal of Machine Learning Research* 18:1–62.

Pollack, Mark A. 1997. "Delegation, Agency, and Agenda Setting in the European Community." *International Organization* 51(1):99–134.

Qian, Nancy. 2015. "Making Progress on Foreign Aid." *Annual Review of Economics* 7(1):277–308.

Rajan, Raghuram G. and Arvind Subramanian. 2011. "Aid, Dutch Disease, and Manufacturing Growth." *Journal of Development Economics* 94(1):106–118.

Randeria, Shalini. 2003. "Glocalization of Law: Environmental Justice, World Bank, NGOs, and the Cunning State in India." *Current Sociology* 51(3/4):305–328.

Randeria, Shalini and Ciara Grunder. 2011. The (Un)Making of Policy in the Shadow of the World Bank: Infrastructure Development, Urban Resettlement and the Cunning State in India. In *Policy Worlds: Anthropology and the Analysis of Contemporary Power*, ed. Cris Shore, Susan Wright and Davide Però. New York: Berghahn Books chapter 10, pp. 187–204.

Rose-Ackerman, Susan and Paul D. Carrington. 2013. *Anti-Corruption Policy: Can International Actors Play a Constructive Role?* Durham, N.C.: Carolina Academic Press.

Ross, Michael L. 2015. "What Have We Learned about the Resource Curse?" *Annual Review of Political Science* 18(1):239–259.

Schunck, Richard. 2013. "Within and Between Estimates in Random-Effects Models: Advantages and Drawbacks of Correlated Random Effects and Hybrid Models." *Stata Journal* 10(3):288–308.

Scott, James C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, Connecticut: Yale University Press.

Smets, Lodewijk, Stephen Knack and Nadia Molenaers. 2013. "Political Ideology, Quality at Entry and the Success of Economic Reform Programs." *Review of International Organizations* 8(4):447–476.

Soifer, Hillel and Matthias vom Hau. 2008. "Unpacking the Strength of the State: The Utility of State Infrastructural Power." *Studies in Comparative International Development* 43(3-4):219–230.

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

Tallberg, Jonas. 2002. "Paths to Compliance: Enforcement, Management, and the European Union." *International Organization* 56(3):609–643.

Tello, Rodolfo. 2015. *Social Safeguards: Avoiding the Unintended Impacts of Development*. Arlington, VA: Amakella Publishing.

Textor, Johannes, Benito van der Zander, Mark S. Gilthorpe, Maciej Liśkiewicz and George TH Ellison. 2016. "Robust Causal Inference Using Directed Acyclic Graphs: The R Package 'dagitty'." *International Journal of Epidemiology* 45(6):1887–1894.

Tversky, Amos and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185(4157):1124–1131.

Vreeland, James Raymond. 2008. "The Effect of Political Regime on Civil War: Unpacking Anocracy." *Journal of Conflict Resolution* 52(3):401–425.

Vreeland, James Raymond. 2011. "Foreign Aid and Global Governance: Buying Bretton Woods - the Swiss-bloc Case." *Review of International Organizations* 6(3):369–391.

Vreeland, James Raymond. 2019. "Corrupting International Organizations." *Annual Review of Political Science* 23:205–224.

Vreeland, James Raymond and Axel Dreher. 2014. *The Political Economy of the United Nations Security Council: Money and Influence*. New York: Cambridge University Press.

Weaver, Catherine E. 2008. *Hypocrisy Trap: The World Bank and the Poverty of Reform*. Princeton, New Jersey: Princeton University Press.

Weber, Max. 1978. *Economy and Society: An Outline of Interpretive Sociology*. Berkeley, California: University of California Press.

Westreich, Daniel and Sander Greenland. 2013. "The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients." *American Journal of Epidemiology* 177(4):292–298.

Williamson, Oliver. 1985. *The Economic Institutions of Capitalism*. New York: Simon and Schuster.

Winters, Matthew S. 2010. "Choosing to Target: What Types of Countries Get Different Types of World Bank Projects." *World Politics* 62(3):422–458.

Winters, Matthew S. and Shyam Kulkarni. 2014. The World Bank in the Post-Structural Adjustment Era. In *Handbook of Global Economic Governance: Players, Power and Paradigms*, ed. Manuela Moschella and Catherine Weaver. New York: Routledge chapter 17, pp. 251–266.

Wood, Reed M. and Christopher M. Sullivan. 2015. "Doing Harm by Doing Good? The Negative Externalities of Humanitarian Aid Provision during Civil Conflict." *Journal of Politics* 77(3):736–748.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross-Section and Panel Data*. 2nd editio ed. Cambridge, Massachusetts: MIT Press.

World Bank. 2013. "Operational Manual: OP 4.01 - Environmental Assessment.".

World Bank. 2017. "World Development Indicators.".
URL: *http://data.worldbank.org/data-catalog/world-development-indicators*

World Bank. 2022. "Environmental Screenning: Project Categories A, B, C, and FI.".

Zvobgo, Kelebogile and Benjamin A.T. Graham. 2020. "The World Bank as an Enforcer of Human Rights." *Journal of Human Rights* 19(4):425–448.

Appendix A Descriptive Statistics **App-2**

Appendix B Additional Regression Tables **App-6**

Appendix C Model Choice Diagnostics **App-16**

Appendix D Additional Figures **App-17**

Appendix E TTL Data Coding and Cleaning Details **App-19**

Appendix F DAG Paths **App-20**

Appendix G Safeguards Coding Examples **21**

Appendix A Descriptive Statistics

Table A.1: Compliance Scores by Country (Coded Projects)

Country	N	Mean	Std. Dev.	Minimum	Maximum
Afghanistan	21	2.952381	.8646497	1	4
Albania	9	3.888889	.3333333	3	4
Angola	5	3	1.224745	1	4
Antigua and Barb	1	1		1	1
Argentina	26	3.307692	.6793662	2	4
Armenia	13	3.538462	.5188745	3	4
Azerbaijan	12	3.583333	.7929615	2	4
Bangladesh	30	3.5	.5723515	2	4
Belarus	7	3.857143	.3779645	3	4
Belize	2	3.5	.7071068	3	4
Benin	11	3.363636	.504525	3	4
Bhutan	5	3.6	.5477226	3	4
Bolivia	12	3.583333	.6685579	2	4
Bosnia and Herzegovina	11	3.545455	.6875517	2	4
Botswana	2	2	0	2	2
Brazil	50	3.28	.7835034	1	4
Bulgaria	4	3.75	.5	3	4
Burkina Faso	14	3.214286	.6992932	2	4
Burundi	10	3.2	.9189366	2	4
Cape Verde	2	4	0	4	4
Cambodia	6	3.833333	.4082483	3	4
Cameroon	15	3.066667	.8837151	2	4
Central African Republic	6	2.333333	.8164966	1	3
Chad	8	2.75	1.164965	1	4
Chile	1	4		4	4
China	110	3.645455	.6145986	1	4
Colombia	11	3.636364	.6741999	2	4
Comoros	3	3.666667	.5773503	3	4
Congo, Democratic of	15	2.733333	.8837151	1	4
Congo, Republic of	7	3.571429	.5345225	3	4
Costa Rica	2	3.5	.7071068	3	4
Cote d'Ivoire	11	3.272727	1.00905	1	4
Croatia	10	3.6	.5163978	3	4
Djibouti	8	3.25	.46291	3	4
Dominican Republic	7	3.857143	.3779645	3	4
Ecuador	5	3.2	.83666	2	4
Egypt	15	3.133333	.9904304	1	4
El Salvador	4	3.5	.5773503	3	4
Ethiopia	27	2.555556	1.050031	1	4

Continued on next page

Table A.1: Compliance Scores by Country (Coded Projects) – *continued*

Country	**N**	**Mean**	**Std. Dev.**	**Minimum**	**Maximum**
Gabon	4	2.5	1.290994	1	4
Gambia, The	4	2.75	1.258306	1	4
Georgia	11	3.272727	.4670994	3	4
Ghana	17	3.117647	.9926198	1	4
Guatemala	3	4	0	4	4
Guinea	9	3.444444	.5270463	3	4
Guinea-Bissau	5	3.2	1.30384	1	4
Guyana	2	3.5	.7071068	3	4
Haiti	17	2.647059	.8617697	1	4
Honduras	12	3.5	.6741999	2	4
India	82	3.195122	.8949657	1	4
Indonesia	23	3.130435	.8688732	1	4
Iraq	3	2	1	1	3
Jamaica	6	3.833333	.4082483	3	4
Jordan	5	3.8	.4472136	3	4
Kazakhstan	6	3	1.264911	1	4
Kenya	19	2.842105	1.118688	1	4
Kiribati	3	3.333333	.5773503	3	4
Kosovo	4	3.75	.5	3	4
Kyrgyz Republic	13	3.615385	.6504436	2	4
Lao PDR	16	3.75	.5773503	2	4
Lebanon	2	4	0	4	4
Lesotho	6	3.5	.83666	2	4
Liberia	11	2.818182	1.250454	1	4
Macedonia, FYR	6	2.833333	.7527727	2	4
Madagascar	7	3	.8164966	2	4
Malawi	11	2.818182	.8738629	2	4
Maldives	2	2	1.414214	1	3
Mali	12	2.916667	.7929615	1	4
Marshall Islands	1	4		4	4
Mauritania	2	3.5	.7071068	3	4
Mauritius	1	4		4	4
Mexico	16	3.625	.7187953	2	4
Micronesia	2	4	0	4	4
Moldova	8	3.5	.7559289	2	4
Mongolia	10	3.7	.6749486	2	4
Montenegro	7	3.714286	.48795	3	4
Morocco	8	3.375	.7440238	2	4
Mozambique	19	3.105263	.875261	2	4
Myanmar	6	3.166667	.7527727	2	4
Nepal	25	2.84	.6244998	2	4
Nicaragua	12	3.583333	.6685579	2	4

Continued on next page

Table A.1: Compliance Scores by Country (Coded Projects) – *continued*

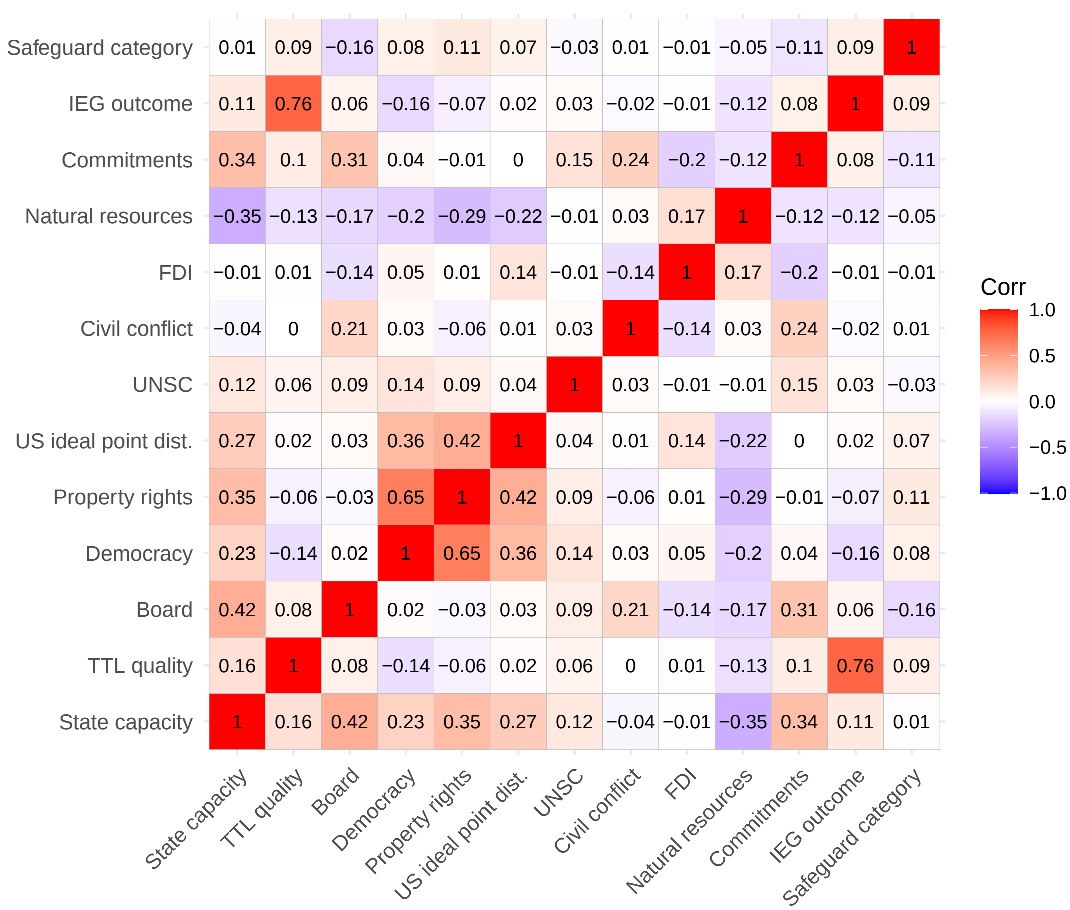
Country	**N**	**Mean**	**Std. Dev.**	**Minimum**	**Maximum**
Niger	13	3.076923	.7595545	2	4
Nigeria	25	3.16	.8981462	1	4
Pakistan	21	3.285714	.9561829	1	4
Panama	6	2.5	1.048809	1	4
Papua New Guinea	9	3.222222	.6666667	2	4
Paraguay	4	3.25	.9574271	2	4
Peru	16	3.1875	.9105859	1	4
Philippines	9	3.555556	.8819171	2	4
Poland	1	4		4	4
Romania	1	4		4	4
Russia	4	3.75	.5	3	4
Rwanda	9	3.777778	.6666667	2	4
Samoa	7	3.285714	1.112697	1	4
Sao Tome and Principe	2	3.5	.7071068	3	4
Senegal	14	3	.877058	1	4
Serbia	9	3.222222	1.092906	1	4
Sierra Leone	5	3.2	.83666	2	4
Solomon Islands	4	3.75	.5	3	4
South Africa	1	3		3	3
South Sudan	2	3	0	3	3
Sri Lanka	16	3.25	.8563488	1	4
St. Lucia	2	3.5	.7071068	3	4
St. Vincent and	1	4		4	4
Swaziland	2	3	0	3	3
Tajikistan	9	3.444444	1.013794	1	4
Tanzania	20	2.95	.6863327	2	4
Timor-Leste	6	3.5	.83666	2	4
Togo	6	3.166667	.7527727	2	4
Tonga	5	3.4	.5477226	3	4
Tunisia	7	3.142857	.8997354	2	4
Turkey	12	3.583333	.7929615	2	4
Tuvalu	1	4		4	4
Uganda	20	2.55	.9445132	1	4
Ukraine	6	3.666667	.5163978	3	4
Uruguay	5	3.6	.5477226	3	4
Uzbekistan	11	3.181818	.8738629	1	4
Vanuatu	2	3	1.414214	2	4
Vietnam	48	3.625	.5309566	2	4
Yemen	4	2.25	1.258306	1	4
Zambia	8	2.875	1.356203	1	4

Source: Own coding.

Table A.2: Summary Statistics

	Mean	Std. Dev.	Min	Max
Compliance score	3.280367	.8423543	1	4
State capacity	.2514543	.5888828	-1.667	1.55
TTL	4.081379	.5858504	1.687143	5.968615
Board	.2154316	.4112786	0	1
Safeguard category	2.006944	.7492274	1	6
GDP per capita (log)	7.593406	1.008442	5.424327	9.629243
Democracy	.4676114	.2240747	.087	.929
Property rights	.6218248	.1977725	.122	.923
US ideal point dist.	-2.998228	.49765	-4.263185	-.835578
Temp. UNSC	.0756303	.2645066	0	1
Civil war	.236822	.4252946	0	1
FDI	4.493641	6.979562	-4.84583	103.3374
Natural resources	7.707047	8.618377	0	56.9313
Commitments (log)	17.99747	1.209989	13.73863	22.12996

Figure A.1: Correlation Matrix



Appendix B Additional Regression Tables

Table B.1: Compliance with World Bank Safeguard Policies 2007-2015
(Non-Mundlak Ordered Logit Models with Country Random Intercepts)

	(1) State Capacity (base1)	(2) State Capacity (base2)	(3) State Capacity (full)	(4) State Capacity (full lc)	(5) TTL Quality (base)	(6) TTL Quality (full)	(7) TTL Quality (full lc)	(8) Board (base)	(9) Board (full)	(10) Board (full lc)
Compliance=1	-0.0240***	-0.0276***	-0.0307***	-0.0317***	-0.0262***	-0.0184***	-0.0201***	0.0034	-0.0051	-0.0044
	(0.0054)	(0.0064)	(0.0065)	(0.0065)	(0.0066)	(0.0050)	(0.0051)	(0.0060)	(0.0065)	(0.0065)
Compliance=2	-0.0614***	-0.0705***	-0.0792***	-0.0816***	-0.0601***	-0.0469***	-0.0500***	0.0081	-0.0131	-0.0113
	(0.0127)	(0.0135)	(0.0154)	(0.0146)	(0.0114)	(0.0110)	(0.0109)	(0.0141)	(0.0165)	(0.0164)
Compliance=3	-0.0617***	-0.0657***	-0.0664***	-0.0684***	-0.0458***	-0.0368***	-0.0381***	0.0069	-0.0102	-0.0088
	(0.0144)	(0.0129)	(0.0141)	(0.0134)	(0.0090)	(0.0089)	(0.0092)	(0.0120)	(0.0128)	(0.0128)
Compliance=4	0.1472***	0.1638***	0.1763***	0.1817***	0.1320***	0.1021***	0.1082***	-0.0184	0.0284	0.0245
	(0.0297)	(0.0291)	(0.0317)	(0.0296)	(0.0229)	(0.0226)	(0.0224)	(0.0320)	(0.0357)	(0.0356)
Observations	1268	1162	1112	1112	1148	1055	1063	1190	1055	1055

Marginal effects with country-clustered robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: "base" refers to the minimal adjustment set to satisfy Pearl's (2009) backdoor criterion;
"full" refers to the variables in Perković et al.'s (2018) canonical adjustment set;
"full lc" refers to the canonical adjustment set, minus any variables excluded due to potential collinearity (see Table 2)

Table B.2: Compliance with World Bank Safeguard Policies 2007-2015
(Linear Model with Country Random Intercepts)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	State Capacity (base1)	State Capacity (base2)	State Capacity (full)	State Capacity (full lc)	TTL Quality (base)	TTL Quality (full)	TTL Quality (full lc)	Board (base)	Board (full)	Board (full lc)
State capacity (between)	0.1635***	0.1718***	0.1735***	0.1744***		0.1826***	0.1791***		0.1826***	0.2012***
	(0.0317)	(0.0292)	(0.0348)	(0.0338)		(0.0364)	(0.0377)		(0.0364)	(0.0358)
State capacity (within)	0.0140	-0.0131	-0.0230	0.0319		0.0217	0.0335		0.0217	-0.0407
	(0.1573)	(0.1757)	(0.1777)	(0.1595)		(0.1799)	(0.1714)		(0.1799)	(0.1742)
Civil war (between)	-0.0357	-0.0357	-0.0553	-0.0559		-0.0400	-0.0361		-0.0400	-0.0360
	(0.0615)	(0.0591)	(0.0553)	(0.0571)		(0.0553)	(0.0635)		(0.0553)	(0.0590)
Civil war (within)	-0.1083**	-0.1047*	-0.1062**	-0.1208**		-0.1169**	-0.1171***		-0.1169**	-0.1180***
	(0.0530)	(0.0553)	(0.0540)	(0.0512)		(0.0461)	(0.0453)		(0.0461)	(0.0444)
FDI (between)	0.0059**	0.0071***	0.0084***	0.0080***		0.0083***	0.0053**		0.0083***	0.0076***
	(0.0025)	(0.0026)	(0.0027)	(0.0026)		(0.0026)	(0.0022)		(0.0026)	(0.0025)
FDI (within)	0.0058*	0.0053	0.0058*	0.0059*		0.0054	0.0056*		0.0054	0.0056
	(0.0035)	(0.0037)	(0.0034)	(0.0033)		(0.0035)	(0.0033)		(0.0035)	(0.0035)
Natural resources (between)	0.0001		-0.0007	-0.0006		-0.0004	0.0011		-0.0004	-0.0009
	(0.0020)		(0.0020)	(0.0019)		(0.0023)	(0.0022)		(0.0023)	(0.0023)
Natural resources (within)	-0.0050		0.0002	-0.0017		0.0035	0.0032		0.0035	0.0055
	(0.0055)		(0.0065)	(0.0054)		(0.0063)	(0.0061)		(0.0063)	(0.0064)
Property rights (between)	-0.0169	0.1736	0.2104*	0.1734		0.1955*	-0.0288		0.1955*	
	(0.1182)	(0.1199)	(0.1166)	(0.1120)		(0.1156)	(0.1153)		(0.1156)	
Property rights (within)	-0.7274**	-0.4724	-0.6209*	-0.8584***		-0.7140**	-0.7287**		-0.7140**	
	(0.3333)	(0.3791)	(0.3618)	(0.2860)		(0.3637)	(0.3500)		(0.3637)	
Democracy (between)		-0.2741**	-0.4537***	-0.4442***		-0.4123***			-0.4123***	-0.3120***
		(0.1075)	(0.1090)	(0.1053)		(0.1063)			(0.1063)	(0.0891)
Democracy (within)		-0.5064*	-0.4804*	-0.2768		-0.5304*			-0.5304*	-0.5686**
		(0.2691)	(0.2707)	(0.2473)		(0.2818)			(0.2818)	(0.2799)
US ideal pt. dist. (between)		-0.0141	-0.0191			-0.0281		0.0699*	-0.0281	-0.0118
		(0.0510)	(0.0429)			(0.0402)		(0.0412)	(0.0402)	(0.0387)
US ideal pt. dist. (within)		0.0448	0.0518			0.0953		0.0340	0.0953	0.0879
		(0.0789)	(0.0850)			(0.0758)		(0.0731)	(0.0758)	(0.0733)
Safeguard category (between)			0.2870***	0.2932***		0.2476***	0.1768*		0.2476***	0.2496***
			(0.0710)	(0.0714)		(0.0823)	(0.0973)		(0.0823)	(0.0836)
Safeguard category (within)			0.1460***	0.1470***		0.1280***	0.1322***		0.1280***	0.1223***
			(0.0352)	(0.0335)		(0.0390)	(0.0399)		(0.0390)	(0.0392)
Board (between)			0.0551	0.0810		0.0864	0.0289	0.0223	0.0864	0.0651
			(0.0658)	(0.0589)		(0.0620)	(0.0965)	(0.0965)	(0.0620)	(0.0665)
Board (within)			-0.0391	-0.0918*		-0.0127	-0.0147	-0.0303	-0.0127	-0.0057
			(0.0481)	(0.0470)		(0.0556)	(0.0554)	(0.0498)	(0.0556)	(0.0529)
Temp. UNSC (between)			0.2458	0.1249		0.3039*	0.2263	0.3002*	0.3039*	0.3017*
			(0.1721)	(0.1542)		(0.1681)	(0.1881)	(0.1648)	(0.1681)	(0.1721)
Temp. UNSC (within)			-0.0079	-0.0129		-0.0179	-0.0128	-0.0007	-0.0179	-0.0235
			(0.0538)	(0.0568)		(0.0595)	(0.0618)	(0.0526)	(0.0595)	(0.0569)
TTL (between)					0.2219***	0.0653	0.1240*		0.0653	0.0534
					(0.0641)	(0.0638)	(0.0722)		(0.0638)	(0.0656)
TTL (within)					0.0909***	0.0770***	0.0776***		0.0770***	0.0776***
					(0.0232)	(0.0250)	(0.0248)		(0.0250)	(0.0248)
Log commitment (between)					-0.0235	-0.0406*	-0.0476*		-0.0406*	-0.0456*
					(0.0184)	(0.0238)	(0.0250)		(0.0238)	(0.0249)
Log commitment (within)					-0.0886***	-0.0701***	-0.0716***		-0.0701***	-0.0694***
					(0.0187)	(0.0168)	(0.0167)		(0.0168)	(0.0167)
Constant	0.4449***	0.4059**	-0.1253	-0.0541	-0.0016	0.3490	0.4151	0.6706***	0.3490	0.6124
	(0.0907)	(0.2039)	(0.2193)	(0.1509)	(0.4274)	(0.5495)	(0.6070)	(0.1209)	(0.5495)	(0.5617)
Country intercept	-2.1570***	-2.2874***	-2.6476***	-2.6181***	-2.0208***	-3.0585**	-2.3849***	-1.9203***	-3.0585**	-2.7831***
	(0.1440)	(0.2454)	(0.4803)	(0.4038)	(0.1635)	(1.2363)	(0.2028)	(0.1186)	(1.2363)	(0.6098)
Observations	1268	1162	1112	1217	1148	1055	1063	1190	1055	1055

Note: Country-clustered robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: "base" refers to the minimal adjustment set to satisfy Pearl's (2009) backdoor criterion;
"full" refers to the variables in Perković et al.'s (2018) canonical adjustment set;
"full lc" refers to the canonical adjustment set, minus any variables excluded due to potential collinearity (see Table 2)

Table B.3: Compliance with World Bank Safeguard Policies 2007-2015
(Logit Models with Country Random Intercepts)

	(1) State Capacity (base1)	(2) State Capacity (base2)	(3) State Capacity (full)	(4) State Capacity (full lc)	(5) TTL Quality (base)	(6) TTL Quality (full)	(7) TTL Quality (full lc)	(8) Board (base)	(9) Board (full)	(10) Board (full lc)
State capacity (between)	0.7345***	0.7700***	0.7990***	0.7957***		0.8572***	0.8403***		0.8572***	0.9447***
	(0.1517)	(0.1389)	(0.1649)	(0.1608)		(0.1965)	(0.1881)		(0.1965)	(0.1873)
State capacity (within)	0.0424	-0.0851	-0.1684	0.0827		0.0923	0.1344		0.0923	-0.1855
	(0.7362)	(0.8363)	(0.8685)	(0.7688)		(0.8825)	(0.8372)		(0.8825)	(0.8568)
Civil war (between)	-0.1616	-0.1575	-0.2545	-0.2638		-0.1865	-0.1800		-0.1865	-0.1663
	(0.2769)	(0.2618)	(0.2467)	(0.2577)		(0.2533)	(0.3042)		(0.2533)	(0.2731)
Civil war (within)	-0.4885**	-0.4752*	-0.4791**	-0.5492**		-0.5541***	-0.5506***		-0.5541***	-0.5580***
	(0.2346)	(0.2482)	(0.2364)	(0.2273)		(0.2071)	(0.2047)		(0.2071)	(0.1997)
FDI (between)	0.0276**	0.0324***	0.0387***	0.0370***		0.0397***	0.0268**		0.0397***	0.0351***
	(0.0120)	(0.0118)	(0.0125)	(0.0122)		(0.0133)	(0.0122)		(0.0133)	(0.0134)
FDI (within)	0.0276	0.0243	0.0271	0.0284		0.0266	0.0285		0.0266	0.0265
	(0.0199)	(0.0192)	(0.0186)	(0.0190)		(0.0209)	(0.0217)		(0.0209)	(0.0197)
Natural resources (between)	0.0007		-0.0031	-0.0034		-0.0024	0.0052		-0.0024	-0.0042
	(0.0091)		(0.0093)	(0.0092)		(0.0112)	(0.0109)		(0.0112)	(0.0115)
Natural resources (within)	-0.0234		-0.0003	-0.0084		0.0170	0.0151		0.0170	0.0254
	(0.0256)		(0.0302)	(0.0256)		(0.0305)	(0.0300)		(0.0305)	(0.0307)
Property rights (between)	-0.0690	0.7770	0.9847*	0.8044		0.9879*	-0.1012		0.9879*	
	(0.5282)	(0.5393)	(0.5398)	(0.5231)		(0.5584)	(0.5464)		(0.5584)	
Property rights (within)	-3.2773**	-2.1220	-2.9149*	-3.9979***		-3.3308**	-3.4150**		-3.3308**	
	(1.5357)	(1.7073)	(1.6681)	(1.3450)		(1.6776)	(1.6618)		(1.6776)	
Democracy (between)		-1.1934**	-2.0677***	-2.0235***		-1.9717**			-1.9717**	-1.4177***
		(0.4974)	(0.5895)	(0.5390)		(0.8941)			(0.8941)	(0.4748)
Democracy (within)		-2.2774*	-2.1454*	-1.2171		-2.3617*			-2.3617*	-2.6201**
		(1.1925)	(1.1789)	(1.0926)		(1.2216)			(1.2216)	(1.2159)
US ideal pt. dist. (between)		-0.0707	-0.0894			-0.1357		0.3183*	-0.1357	-0.0574
		(0.2275)	(0.1881)			(0.1964)		(0.1854)	(0.1964)	(0.1755)
US ideal pt. dist. (within)		0.1952	0.2163			0.4270		0.1497	0.4270	0.4002
		(0.3606)	(0.3927)			(0.3567)		(0.3234)	(0.3567)	(0.3464)
Safeguard category (between)			1.3221***	1.3678***		1.1734***	0.8649*		1.1734***	1.1844***
			(0.3431)	(0.3508)		(0.3854)	(0.4715)		(0.3854)	(0.4011)
Safeguard category (within)			0.6680***	0.6751***		0.5965***	0.6246***		0.5965***	0.5748***
			(0.1646)	(0.1581)		(0.1892)	(0.1930)		(0.1892)	(0.1856)
Board (between)			0.2663	0.3895		0.4413	0.1606	0.0833	0.4413	0.3233
			(0.3216)	(0.2758)		(0.4458)	(0.4423)	(0.4237)	(0.4458)	(0.3295)
Board (within)			-0.1669	-0.4063*		-0.0607	-0.0720	-0.1292	-0.0607	-0.0258
			(0.2087)	(0.2120)		(0.2472)	(0.2524)	(0.2142)	(0.2472)	(0.2359)
Temp. UNSC (between)			1.0752	0.5552		1.3562	1.0217	1.4088*	1.3562	1.3779
			(0.7912)	(0.6773)		(0.9475)	(0.9383)	(0.7929)	(0.9475)	(0.8640)
Temp. UNSC (within)			-0.0315	-0.0539		-0.0684	-0.0483	-0.0020	-0.0684	-0.0938
			(0.2373)	(0.2496)		(0.2706)	(0.2820)	(0.2270)	(0.2706)	(0.2564)
TTL (between)					0.9925***	0.3151	0.5844*		0.3151	0.2532
					(0.2998)	(0.3141)	(0.3444)		(0.3141)	(0.3126)
TTL (within)					0.4148***	0.3628***	0.3694***		0.3628***	0.3682***
					(0.1094)	(0.1216)	(0.1200)		(0.1216)	(0.1178)
Log commitment (between)					-0.1064	-0.1922	-0.2178*		-0.1922	-0.2167*
					(0.0843)	(0.1250)	(0.1223)		(0.1250)	(0.1244)
Log commitment (within)					-0.3981***	-0.3231***	-0.3332***		-0.3231***	-0.3195***
					(0.0853)	(0.0779)	(0.0771)		(0.0779)	(0.0766)
Constant	-0.2620	-0.4660	-2.9052***	-2.5980***	-2.2253	-0.7799	-0.6136	0.7797	-0.7799	0.5130
	(0.4042)	(0.9238)	(1.0200)	(0.7268)	(1.9573)	(2.6820)	(2.8777)	(0.5416)	(2.6820)	(2.7098)
Country intercept	0.2593***	0.1958*	0.0864	0.0953	0.3577***	0.0224	0.1766**	0.4177***	0.0224	0.0731
	(0.0814)	(0.1101)	(0.1116)	(0.0931)	(0.1349)	(0.2208)	(0.0819)	(0.1130)	(0.2208)	(0.1238)
Observations	1268	1162	1112	1217	1148	1055	1063	1190	1055	1055

Note: marginal effects with country-clustered robust standard errors.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: Compliance with World Bank Safeguard Policies 2007-2015
(Cross-Sectional Mundlak Ordered Logit Models)

	(1) State Capacity (base1)	(2) State Capacity (base2)	(3) State Capacity (full)	(4) State Capacity (full lc)	(5) TTL Quality (base)	(6) TTL Quality (full)	(7) TTL Quality (full lc)	(8) Board (base)	(9) Board (full)	(10) Board (full lc)
Panel A: Between-Country Effects										
Compliance = 1	-0.0258***	-0.0285***	-0.0273***	-0.0259***	-0.0520***	-0.0154	-0.0337***	-0.0189	-0.0208*	-0.0185*
	(0.0061)	(0.0064)	(0.0071)	(0.0064)	(0.0163)	(0.0102)	(0.0120)	(0.0175)	(0.0110)	(0.0111)
Compliance = 2	-0.0669***	-0.0735***	-0.0707***	-0.0669***	-0.1289***	-0.0400	-0.0863***	-0.0483	-0.0539**	-0.0482*
	(0.0145)	(0.0130)	(0.0182)	(0.0169)	(0.0327)	(0.0255)	(0.0287)	(0.0439)	(0.0246)	(0.0252)
Compliance = 3	-0.0671***	-0.0675***	-0.0599***	-0.0615***	-0.1160***	-0.0326	-0.0707***	-0.0466	-0.0440**	-0.0391**
	(0.0166)	(0.0112)	(0.0152)	(0.0155)	(0.0316)	(0.0199)	(0.0234)	(0.0445)	(0.0194)	(0.0194)
Compliance = 4	0.1598***	0.1694***	0.1579***	0.1543***	0.2970***	0.0881	0.1907***	0.1138	0.1187**	0.1059*
	(0.0340)	(0.0265)	(0.0378)	(0.0365)	(0.0738)	(0.0549)	(0.0612)	(0.1051)	(0.0539)	(0.0549)
Panel B: Within-Country Effects										
Compliance = 1	-0.0092	-0.0048	-0.0026	-0.0120	-0.0212***	-0.0189***	-0.0191***	0.0140	0.0095	0.0081
	(0.0226)	(0.0241)	(0.0235)	(0.0226)	(0.0063)	(0.0053)	(0.0054)	(0.0088)	(0.0113)	(0.0106)
Compliance = 2	-0.0238	-0.0124	-0.0067	-0.0310	-0.0525***	-0.0490***	-0.0490***	0.0358	0.0246	0.0212
	(0.0583)	(0.0623)	(0.0608)	(0.0587)	(0.0122)	(0.0118)	(0.0117)	(0.0221)	(0.0286)	(0.0271)
Compliance = 3	-0.0239	-0.0114	-0.0057	-0.0285	-0.0472***	-0.0399***	-0.0401***	0.0346	0.0200	0.0172
	(0.0584)	(0.0570)	(0.0515)	(0.0538)	(0.0093)	(0.0099)	(0.0098)	(0.0220)	(0.0237)	(0.0223)
Compliance = 4	0.0570	0.0286	0.0151	0.0716	0.1209***	0.1078***	0.1082***	-0.0844	-0.0541	-0.0465
	(0.1392)	(0.1433)	(0.1358)	(0.1350)	(0.0247)	(0.0251)	(0.0247)	(0.0520)	(0.0634)	(0.0598)
Observations	1268	1162	1112	1217	1148	1055	1063	1190	1055	1055

Marginal effects with country-clustered robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: "base" refers to the minimal adjustment set to satisfy Pearl's (2009) backdoor criterion;
"full" refers to the variables in Perković et al.'s (2018) canonical adjustment set;
"full lc" refers to the canonical adjustment set, minus any variables excluded due to potential collinearity (see Table 2)

Table B.5: Compliance with World Bank Safeguard Policies 2007-2015
(Mundlak Ordered Logit Models with Country and Year Random Intercepts)

	(1) State Capacity (base1)	(2) State Capacity (base2)	(3) State Capacity (full)	(4) State Capacity (full lc)	(5) TTL Quality (base)	(6) TTL Quality (full)	(7) TTL Quality (full lc)	(8) Board (base)	(9) Board (full)	(10) Board (full lc)
Panel A: Between-Country Effects										
Compliance = 1	-0.0248***	-0.0287***	-0.0290***	-0.0275***	-0.0501***	-0.0147	-0.0253*	-0.0061	-0.0180	-0.0148
	(0.0057)	(0.0066)	(0.0073)	(0.0066)	(0.0157)	(0.0110)	(0.0130)	(0.0172)	(0.0117)	(0.0118)
Compliance = 2	-0.0621***	-0.0717***	-0.0729***	-0.0696***	-0.1166***	-0.0369	-0.0623*	-0.0144	-0.0451	-0.0371
	(0.0134)	(0.0137)	(0.0177)	(0.0161)	(0.0309)	(0.0275)	(0.0320)	(0.0403)	(0.0277)	(0.0282)
Compliance = 3	-0.0612***	-0.0659***	-0.0608***	-0.0636***	-0.0964***	-0.0293	-0.0483*	-0.0128	-0.0359	-0.0292
	(0.0149)	(0.0130)	(0.0144)	(0.0145)	(0.0248)	(0.0212)	(0.0255)	(0.0364)	(0.0221)	(0.0221)
Compliance = 4	0.1480***	0.1663***	0.1626***	0.1607***	0.2631***	0.0808	0.1359**	0.0333	0.0990	0.0812
	(0.0313)	(0.0294)	(0.0364)	(0.0346)	(0.0661)	(0.0591)	(0.0689)	(0.0938)	(0.0607)	(0.0615)
Panel B: Within-Country Effects										
Compliance = 1	-0.0069	-0.0020	0.0001	-0.0110	-0.0224***	-0.0190***	-0.0197***	0.0128	0.0099	0.0087
	(0.0228)	(0.0238)	(0.0237)	(0.0230)	(0.0062)	(0.0054)	(0.0056)	(0.0084)	(0.0116)	(0.0111)
Compliance = 2	-0.0173	-0.0050	0.0001	-0.0278	-0.0522***	-0.0478***	-0.0484***	0.0300	0.0249	0.0219
	(0.0570)	(0.0594)	(0.0595)	(0.0582)	(0.0116)	(0.0119)	(0.0119)	(0.0195)	(0.0285)	(0.0272)
Compliance = 3	-0.0170	-0.0046	0.0001	-0.0253	-0.0431***	-0.0380***	-0.0375***	0.0265	0.0198	0.0172
	(0.0561)	(0.0544)	(0.0496)	(0.0531)	(0.0094)	(0.0099)	(0.0097)	(0.0178)	(0.0230)	(0.0216)
Compliance = 4	0.0412	0.0116	-0.0003	0.0641	0.1176***	0.1048***	0.1056***	-0.0693	-0.0546	-0.0478
	(0.1359)	(0.1376)	(0.1329)	(0.1342)	(0.0243)	(0.0252)	(0.0247)	(0.0451)	(0.0629)	(0.0597)
Observations	1268	1162	1112	1217	1148	1055	1063	1190	1055	1055

Marginal effects with country-clustered robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; year nested within country.

Note: “base” refers to the minimal adjustment set to satisfy Pearl’s (2009) backdoor criterion;

Note: “full” refers to the variables in Perković et al.’s (2018) canonical adjustment set;

Note: “full lc” refers to the canonical adjustment set, minus any variables excluded due to potential collinearity (see Table 2)

Table B.6: Compliance with World Bank Safeguard Policies 2007-2015
(Mundlak Ordered Logit Models with Country and Year Random Intercepts [Only IEG-Evaluated Projects])

	(1) State Capacity (base1)	(2) State Capacity (base2)	(3) State Capacity (full)	(4) State Capacity (full lc)	(5) TTL Quality (base)	(6) TTL Quality (full)	(7) TTL Quality (full lc)	(8) Board (base)	(9) Board (full)	(10) Board (full lc)
Panel A: Between-Country Effects										
Compliance = 1	-0.0261***	-0.0289***	-0.0280***	-0.0275***	-0.0488***	-0.0144	-0.0249*	-0.0076	-0.0172	-0.0143
	(0.0061)	(0.0065)	(0.0068)	(0.0068)	(0.0155)	(0.0108)	(0.0128)	(0.0176)	(0.0114)	(0.0116)
Compliance = 2	-0.0627***	-0.0701***	-0.0691***	-0.0670***	-0.1157***	-0.0370	-0.0624*	-0.0175	-0.0442	-0.0368
	(0.0139)	(0.0132)	(0.0168)	(0.0166)	(0.0309)	(0.0277)	(0.0322)	(0.0406)	(0.0277)	(0.0284)
Compliance = 3	-0.0556***	-0.0626***	-0.0558***	-0.0543***	-0.0974***	-0.0300	-0.0490*	-0.0151	-0.0358	-0.0295
	(0.0151)	(0.0139)	(0.0146)	(0.0144)	(0.0254)	(0.0219)	(0.0262)	(0.0359)	(0.0226)	(0.0228)
Compliance = 4	0.1445***	0.1616***	0.1530***	0.1488***	0.2619***	0.0814	0.1362*	0.0402	0.0972	0.0806
	(0.0322)	(0.0295)	(0.0353)	(0.0350)	(0.0665)	(0.0600)	(0.0697)	(0.0940)	(0.0609)	(0.0623)
Panel B: Within-Country Effects										
Compliance = 1	-0.0119	-0.0077	-0.0032	-0.0080	-0.0225***	-0.0190***	-0.0196***	0.0067	0.0095	0.0083
	(0.0290)	(0.0288)	(0.0279)	(0.0277)	(0.0062)	(0.0054)	(0.0056)	(0.0088)	(0.0114)	(0.0107)
Compliance = 2	-0.0287	-0.0187	-0.0079	-0.0194	-0.0535***	-0.0487***	-0.0491***	0.0154	0.0245	0.0214
	(0.0694)	(0.0697)	(0.0689)	(0.0675)	(0.0114)	(0.0117)	(0.0118)	(0.0201)	(0.0285)	(0.0269)
Compliance = 3	-0.0254	-0.0167	-0.0064	-0.0157	-0.0450***	-0.0396***	-0.0386***	0.0132	0.0199	0.0172
	(0.0612)	(0.0618)	(0.0555)	(0.0545)	(0.0094)	(0.0098)	(0.0097)	(0.0177)	(0.0235)	(0.0219)
Compliance = 4	0.0660	0.0430	0.0175	0.0430	0.1210***	0.1073***	0.1074***	-0.0353	-0.0540	-0.0470
	(0.1595)	(0.1603)	(0.1523)	(0.1497)	(0.0240)	(0.0249)	(0.0245)	(0.0464)	(0.0631)	(0.0594)
Observations	1116	1107	1057	1066	1148	1055	1063	1134	1055	1055

Marginal effects with country-clustered robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; year nested within country.
Note: "base" refers to the minimal adjustment set to satisfy Pearl's (2009) backdoor criterion;
Note: "full" refers to the variables in Perković et al.'s (2018) canonical adjustment set;
Note: "full lc" refers to the canonical adjustment set, minus any variables excluded due to potential collinearity (see Table 2)

Table B.7: Compliance with World Bank Safeguard Policies 2007-2015
(Mundlak Ordered Logit with Country Random Intercepts [Only Projects Triggering Social & Resettlement Safeguards])

	(1) State Capacity (base1)	(2) State Capacity (base2)	(3) State Capacity (full)	(4) State Capacity (full lc)	(5) TTL Quality (base)	(6) TTL Quality (full lc)	(7) Board (base)	(8) Board (full lc)
Compliance = 1	-0.0258***	-0.0291***	-0.0261***	-0.0251***	-0.0484***	-0.0458***	-0.0123	-0.0242*
	(0.0065)	(0.0072)	(0.0084)	(0.0077)	(0.0163)	(0.0159)	(0.0175)	(0.0129)
Compliance = 2	-0.0709***	-0.0788***	-0.0708***	-0.0695***	-0.1342***	-0.1274***	-0.0334	-0.0667*
	(0.0160)	(0.0163)	(0.0215)	(0.0198)	(0.0375)	(0.0381)	(0.0475)	(0.0345)
Compliance = 3	-0.0544***	-0.0587***	-0.0498***	-0.0528***	-0.0835***	-0.0815***	-0.0238	-0.0439**
	(0.0166)	(0.0153)	(0.0154)	(0.0157)	(0.0259)	(0.0273)	(0.0359)	(0.0224)
Compliance = 4	0.1511***	0.1666***	0.1467***	0.1474***	0.2661***	0.2546***	0.0695	0.1348**
	(0.0353)	(0.0340)	(0.0425)	(0.0405)	(0.0718)	(0.0749)	(0.1005)	(0.0680)
Panel B: Within-Country Effects								
Compliance = 1	-0.0306	-0.0475	-0.0443*	-0.0386	-0.0241***	-0.0215***	0.0254**	0.0196
	(0.0280)	(0.0298)	(0.0268)	(0.0270)	(0.0078)	(0.0071)	(0.0114)	(0.0143)
Compliance = 2	-0.0840	-0.1287*	-0.1201*	-0.1069	-0.0667***	-0.0600***	0.0690**	0.0541
	(0.0725)	(0.0734)	(0.0674)	(0.0693)	(0.0153)	(0.0162)	(0.0296)	(0.0375)
Compliance = 3	-0.0645	-0.0959*	-0.0844*	-0.0812	-0.0415***	-0.0384***	0.0491**	0.0356
	(0.0565)	(0.0539)	(0.0465)	(0.0531)	(0.0098)	(0.0111)	(0.0247)	(0.0253)
Compliance = 4	0.1791	0.2721*	0.2488*	0.2268	0.1322***	0.1199***	-0.1435**	-0.1094
	(0.1556)	(0.1541)	(0.1382)	(0.1476)	(0.0284)	(0.0310)	(0.0631)	(0.0761)
Observations	976	888	863	950	865	820	907	816

Note: marginal effects with country-clustered robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: "base" refers to the minimal adjustment set to satisfy Pearl's (2009) backdoor criterion;

Note: "full" specifications capturing Perković et al.'s (2018) canonical adjustment set excluded due to inability to calculate marginal effects.

Note: "full lc" refers to the canonical adjustment set, minus any variables excluded due to potential collinearity (see Table 2)

Table B.8: Compliance with World Bank Safeguard Policies 2007-2015
(Mundlak Ordered Logit Models with Country and Year Random Intercepts [With Government Effectiveness])

	(1) Govt. Eff. (base1)	(2) Govt. Eff. (base2)	(3) Govt. Eff. (full)	(4) Govt. Eff. (full lc)	(5) TTL Quality (base)	(6) TTL Quality (full)	(7) TTL Quality (full lc)	(8) Board (base)	(9) Board (full)	(10) Board (full lc)
Panel A: Between-Country Effects										
Compliance = 1	-0.0187***	-0.0228***	-0.0239***	-0.0239***	-0.0488***	-0.0226**	-0.0242*	-0.0061	-0.0233**	-0.0203*
	(0.0067)	(0.0071)	(0.0077)	(0.0070)	(0.0155)	(0.0112)	(0.0126)	(0.0168)	(0.0112)	(0.0117)
Compliance = 2	-0.0473***	-0.0575***	-0.0609***	-0.0611***	-0.1157***	-0.0578**	-0.0606**	-0.0145	-0.0597**	-0.0519*
	(0.0180)	(0.0181)	(0.0208)	(0.0189)	(0.0309)	(0.0277)	(0.0305)	(0.0402)	(0.0257)	(0.0271)
Compliance = 3	-0.0460**	-0.0516***	-0.0506***	-0.0560***	-0.0974***	-0.0459**	-0.0453**	-0.0131	-0.0474**	-0.0406*
	(0.0185)	(0.0169)	(0.0166)	(0.0167)	(0.0254)	(0.0212)	(0.0225)	(0.0371)	(0.0209)	(0.0212)
Compliance = 4	0.1120***	0.1320***	0.1353***	0.1410***	0.2619***	0.1263**	0.1301**	0.0337	0.1304**	0.1128*
	(0.0419)	(0.0402)	(0.0432)	(0.0407)	(0.0665)	(0.0588)	(0.0642)	(0.0940)	(0.0566)	(0.0590)
Observations	1278	1168	1118	1227	1148	1061	1061	1190	1061	1061
Panel B: Within-Country Effects										
Compliance = 1	-0.0010	-0.0057	-0.0047	-0.0033	-0.0225***	-0.0192***	-0.0200***	0.0121	0.0088	0.0077
	(0.0191)	(0.0211)	(0.0228)	(0.0203)	(0.0062)	(0.0054)	(0.0056)	(0.0079)	(0.0115)	(0.0109)
Compliance = 2	-0.0025	-0.0144	-0.0120	-0.0084	-0.0535***	-0.0492***	-0.0500***	0.0289	0.0226	0.0196
	(0.0483)	(0.0537)	(0.0588)	(0.0524)	(0.0114)	(0.0118)	(0.0119)	(0.0189)	(0.0288)	(0.0272)
Compliance = 3	-0.0024	-0.0129	-0.0099	-0.0077	-0.0450***	-0.0390***	-0.0374***	0.0260	0.0180	0.0154
	(0.0469)	(0.0481)	(0.0487)	(0.0479)	(0.0094)	(0.0094)	(0.0093)	(0.0179)	(0.0230)	(0.0213)
Compliance = 4	0.0059	0.0330	0.0266	0.0193	0.1210***	0.1075***	0.1074***	-0.0670	-0.0494	-0.0427
	(0.1143)	(0.1227)	(0.1303)	(0.1206)	(0.0240)	(0.0246)	(0.0243)	(0.0441)	(0.0631)	(0.0592)
Observations	1278	1168	1118	1227	1148	1061	1061	1190	1061	1061

Marginal effects with country-clustered robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; year nested within country.
Note: "base" refers to the minimal adjustment set to satisfy Pearl's (2009) backdoor criterion;
Note: "full" refers to the variables in Perković et al.'s (2018) canonical adjustment set;
Note: "full lc" refers to the canonical adjustment set, minus any variables excluded due to potential collinearity (see Table 2)

Table B.9: Compliance with World Bank Safeguard Policies 2007-2015
(Mundlak Ordered Logit Models with Country Random Intercepts)

	(1) US Ideal (base)	(2) US Ideal (base)	(3) US Ideal (base)	(4) Temp UNSC (base)	(5) Temp UNSC (full)	(6) Temp UNSC (full lc)
Panel A: Between-Country Effects						
Compliance = 1	-0.0129	-0.0043	-0.0093	-0.0268	-0.0499	-0.0484
	(0.0087)	(0.0071)	(0.0073)	(0.0280)	(0.0324)	(0.0334)
Compliance = 2	-0.0317	-0.0106	-0.0227	-0.0652	-0.1265	-0.1228
	(0.0206)	(0.0174)	(0.0175)	(0.0677)	(0.0777)	(0.0809)
Compliance = 3	-0.0269	-0.0080	-0.0171	-0.0628	-0.0993*	-0.0958
	(0.0173)	(0.0131)	(0.0131)	(0.0650)	(0.0602)	(0.0622)
Compliance = 4	0.0714	0.0229	0.0491	0.1549	0.2757	0.2670
	(0.0460)	(0.0375)	(0.0375)	(0.1601)	(0.1679)	(0.1741)
Panel B: Within-Country Effects						
Compliance = 1	-0.0139	-0.0185	-0.0168	0.0079	0.0066	0.0080
	(0.0122)	(0.0146)	(0.0137)	(0.0095)	(0.0109)	(0.0104)
Compliance = 2	-0.0342	-0.0452	-0.0410	0.0191	0.0168	0.0202
	(0.0296)	(0.0355)	(0.0332)	(0.0228)	(0.0268)	(0.0255)
Compliance = 3	-0.0290	-0.0340	-0.0309	0.0184	0.0132	0.0157
	(0.0248)	(0.0267)	(0.0250)	(0.0218)	(0.0212)	(0.0200)
Compliance = 4	0.0771	0.0977	0.0887	-0.0454	-0.0366	-0.0439
	(0.0662)	(0.0760)	(0.0713)	(0.0539)	(0.0587)	(0.0558)
Observations	1168	1118	1118	1309	1055	1055

Note: marginal effects with country-clustered robust standard errors.

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

US ideal base adjustment set: democracy

US ideal full adjustment set: civil conflict, democracy, natural resources, property rights, safeguard category, UNSC

US ideal full lc adjustment set: civil conflict, democracy, natural resources, safeguard category, UNSC

UNSC base adjustment set: civil conflict

UNSC full adjustment set: civil conflict, commitment, democracy, FDI, natural resources, property rights, safeguard category, state capacity, US ideal

UNSC full lc adjustment set: civil conflict, commitment, democracy, FDI, natural resources, safeguard category, state capacity, US ideal

Appendix C Model Choice Diagnostics

Table C.1: Panel-level Summary of State Capacity Variable

	Mean	Std. Dev.	Min	Max	Observations
state capacity (overall)	0.251	0.589	-1.667	1.550	$N = 1{,}274$
between	—	0.651	-1.400	1.550	$n = 107$
within	—	0.085	-0.156	0.694	$\bar{T} = 11.91$

Note: the above reflects `xtsum` in Stata

Table C.2: Panel Diagnostics and Random–Intercept Results

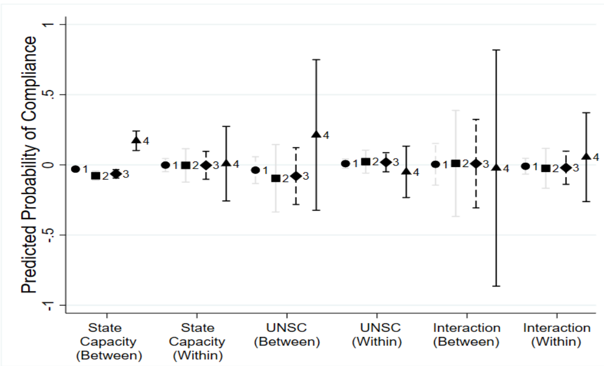
	Estimate	Interpretation
Intraclass correlation, $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$	0.98	98 % of variance at country level
LR test vs. pooled OLS, $\chi^2(1)$	$4\,161.4^{***}$	Country dummies matter

Note: Linear regression model with no covariates, a country random intercept, and state capacity as the dependent variable. $N = 1{,}274$ observations, $n = 107$ countries. Variance components from the random-intercept model: $\sigma_u^2 = 0.417$ (country intercept), $\sigma_e^2 = 0.008$ (residual). *** $p < 0.001$.

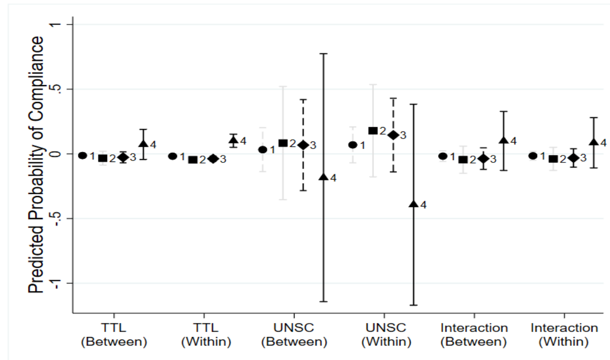
Appendix D Additional Figures

Figure D.1: UNSC Moderation Analyses

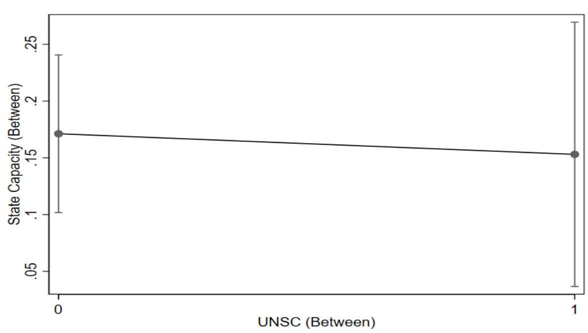
(a) State Capacity × UNSC (Coefficients)



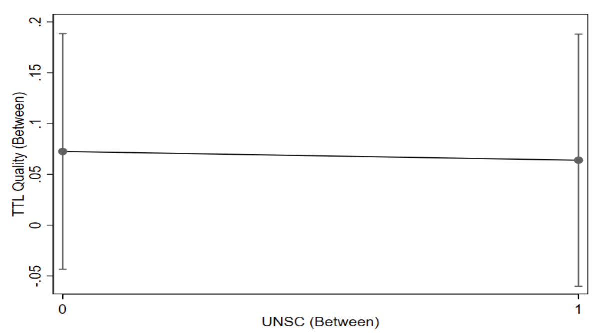
(b) TTL Quality × UNSC (Coefficients)



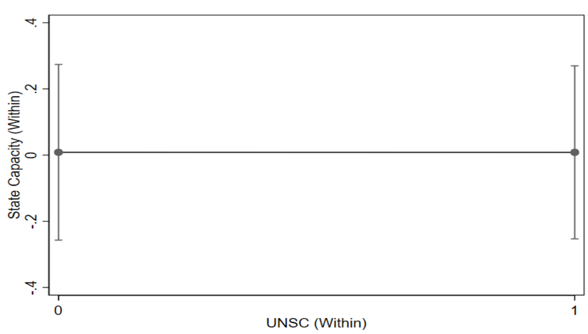
(c) State Capacity × UNSC (Between Plot)



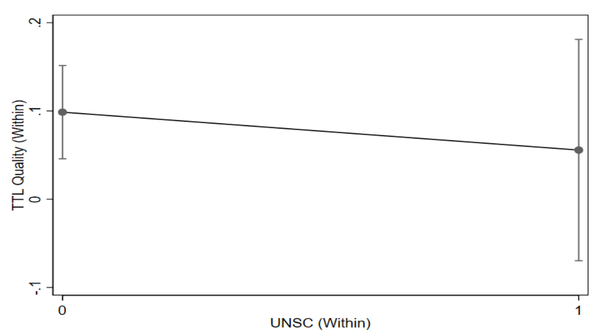
(d) TTL Quality × UNSC (Between Plot)



(e) State Capacity × UNSC (Within Plot)



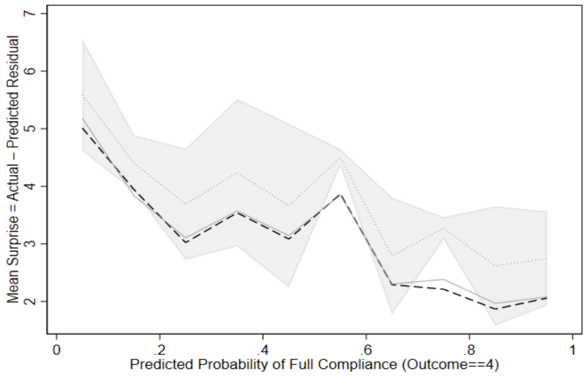
(f) TTL Quality × UNSC (Within Plot)



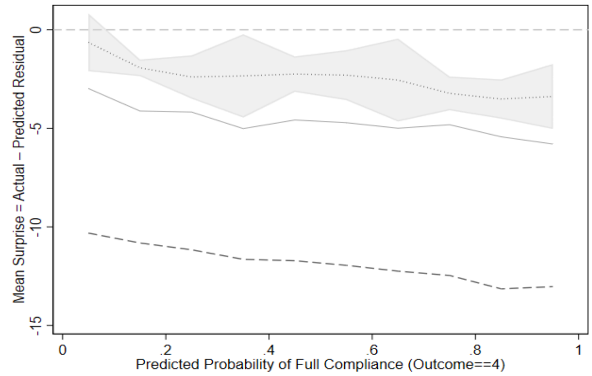
Note: All coefficients correspond to marginal effects from multilevel ordered logit models with country random intercepts, full controls, and country-clustered robust standard errors. Plots are marginal effects plots for the between coefficients for full compliance (category 4). The DAG-derived adjustment set for the state capacity interactions includes: civil conflict, democracy, FDI, natural resources, property rights, safeguard category, US ideal point distance. The DAG-derived adjustment set for the TTL interactions adds commitments.

Figure D.2: Residual Decomposition by Safeguard Risk

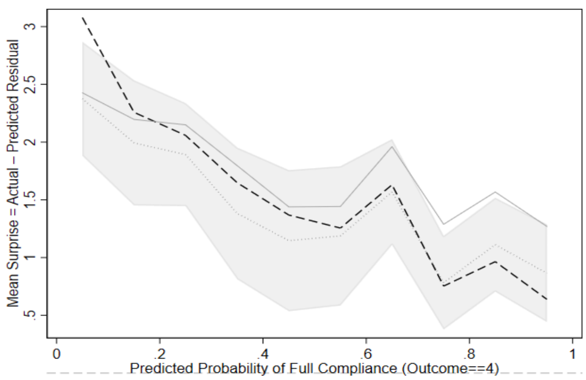
(a) State Capacity (Safeguard Risk Category A)



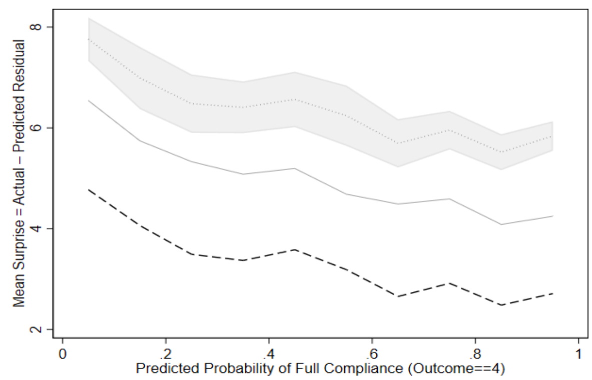
(b) TTL (Safeguard Risk Category A)



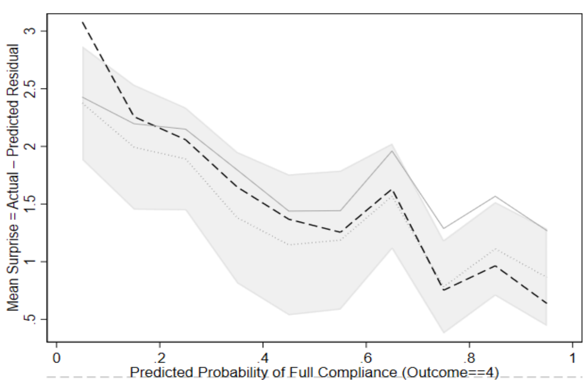
(c) State Capacity (Safeguard Risk Category B)



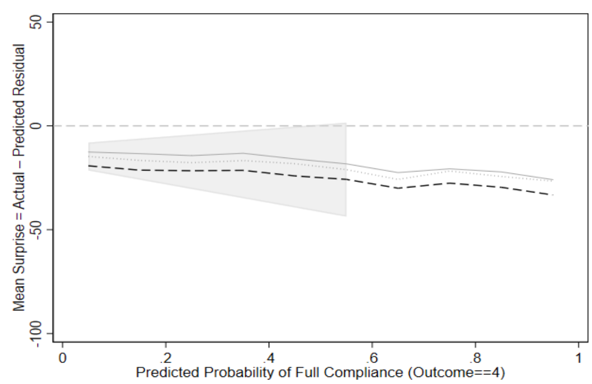
(d) TTL Quality (Safeguard Risk Category B)



(e) State Capacity (Safeguard Risk Category C)



(f) TTL Quality (Safeguard Risk Category C)



Note:Note: Plots inspired by Card and Dahl (2011). Plots with mean surprise closer to zero indicate better regression fit. Dashed black lines = addition of between coefficients; solid gray lines = addition of within coefficients; dotted gray lines = baseline estimates without treatments. 95% confidence intervals correspond to baseline estimates, which reflect the adjustment sets from Table 2. All plot contain 10 equally-spaced bins to measure preducted probabilities on the x-axis. See Figure 5 for the overall models.

Appendix E TTL Data Coding and Cleaning Details

To calculate the average TTL rating by project, it was necessary to use two sets of data: (i) the TTL name at each (mostly) bi-annual Implementation Status Report (ISR) for each project; and (ii) the Independent Evaluation Group (IEG) outcome rating of each project. The IEG outcome data are easily accessible online, whereas the TTL name at each ISR is only available from the transparency request that I made. Using both sets of data, I calculated the rolling weighted average TTL outcome rating for each day that the TTL is in the World Bank system. The daily rating is necessary because projects frequently take more than five years to implement, and TTLs often change jobs. Indeed, the World Bank uses a 3-5-7 staff rotation timeline for its staff.

After creating the daily dataset of circa 40 million observations, I merge the rolling weighted average TTL outcome score for each IEG-rated project up to each particular day with the ISR dataset. That merge allows us to pinpoint the average TTL rating at each ISR date for every project. Given that the average TTL outcome rating for each date represents a weighted rolling average, it only considers IEG ratings for closed projects up to each particular day. By extension, the rating does not improperly take into account IEG ratings for projects that close at a later time in the dataset.

Finally, it is necessary to clarify what the “weighted” part refers to in the rolling weighted average. On that score, the final average TTL rating score is weighted because, after completing the merge with the ISR dataset, the rating takes into account the number of ISRs that each TTL completed. To make this more concrete, take, for example, a project that is under implementation for five years. Given that ISRs take place approximately every six months, let us assume that the project has a total of 10 ISRs. If a TTL named Jim completed 4 of the ISRs and another TTL named Valerie completed 6 of the ISRs, the final TTL outcome rating for the project will reflect 4 day-specific ISR ratings for Jim as well as 6 day-specific ratings for Valerie. By extension, 40% of the weighting for the project-specific TTL rating will capture Jim’s average IEG outcome for all of his previous projects, whereas 60% of the rating will reflect Valerie’s average IEG outcome for all of her previous projects.

Appendix F DAG Paths

In this section, I describe the decisions capturing underpinning each path in the DAG (see Figure 4):

- **State capacity**:
  - *Property rights* affect state capacity because they give states legibility to collect taxes and exert control over a territory (Scott, 1998; Lee and Zhang, 2017; D'Arcy, Nistotskaya and Olsson, 2024).
  - *Natural resources* affect state capacity via fiscal substitution (Ross, 2015; Masi, Savoia and Sen, 2024).
  - *FDI* affects state capacity not only because it contributes to revenue and gross domestic product (GDP) by definition,[34] but also because investors often make demands of states to safeguard their investments, which require states to exert greater control over their territory.
  - *Civil conflict* affects state capacity per Weber (1978), who famously suggested that strong states have a monopoly of violence across their territories.
- **Commitment**
  - *State capacity* affects commitment levels because the World Bank allocates its money dedicated to lower-income countries using a performance-based allocation system (Morrison, 2013).
- **TTL quality**
  - *Commitments* affect TTL quality, because the World Bank will not assign poor or inexperienced TTLs to projects with high commitment values.
- **Safeguard risk category**
  - *Civil conflict* affects the safeguard risk category because it makes projects more difficult to access and, in turn, supervise.
  - *Property rights* affect the safeguard risk category because they directly affect the World Bank and country's ability to settle disputes.
- **Board**
  - *US ideal point distance* affects the Board because, by convention, the World Bank president is always American, and US values affect how the Board is run
  - *UNSC* membership affects the Board because countries use this position of power to affect international institutions in many ways (Vreeland and Dreher, 2014)
- **Democracy**

[34] The equation is: $GDP = Consumption + Investment + GovernmentExpenditures + Exports - Imports$

  - *Natural resources* affect democracy because they allow rulers to consolidate power, particularly after the nationalization of many countries' oil industries (Andersen and Ross, 2014)

- **Property rights**
  - *Democracy* affects property rights because it foster the rule of law, by which all people receive fair treatment under the law, regardless of status, background, etc. With the rule of law, it is possible to enforce property rights.
  - *Natural resources* affect property rights because natural resource discovery often happens in remote territory and necessitates cadasters and, in turn, property rights.

- **FDI**
  - *Property rights* affect FDI because clear property rights allow investors to protect themselves against expropriation.
  - *Natural resources* affect FDI, because they are very lucrative and often spur investment (Menaldo, 2016).
  - *US ideal point distance* fosters FDI because countries that share US values generally are more favorable for operating and protecting business investments.

- **Civil conflict**
  - *Natural resources* foment civil conflict, at least in Africa (Denly et al., 2022)

- **US ideal point distance**
  - *Democracy* affects the US ideal point distance because, until the presidencies of Donald Trump, whose time periods do not coincide with the present study, the US engaged in significant democracy promotion around the world (Krasner, 2020).

- **UNSC**
  - Civil conflict affects temporary UN Security Council memberships because, as Dreher et al. (2014) show, there is a norm against countries receiving UNSC membership if they are undergoing civil conflict.

Appendix G Safeguards Coding Examples

Coming soon.