
Theory, Data, and Hypothesis Testing: World Bank Environmental Reform Redux

Daniel L. Nielson and Michael J. Tierney

We thank Tamar Gutner for her thoughtful comments on our article, “Delegation to International Organizations: Agency Theory and World Bank Environmental Reform” (*International Organization*, Spring 2003). While championing principal-agent (P-A) theory as the most promising approach to the study of international organizations (IOs) in general, she questions the novelty of our P-A model and criticizes both our methods and data. We are grateful for the opportunity to clarify our contributions and offer concrete solutions to the research questions that she raises.

Gutner’s dissent contributes to the debates over World Bank reform and environmental lending, but her empirical critique is also a study in contrasts. She overstates our emphasis on the year 1994 in the history of World Bank environmental reform but understates the actual impact of the 1994 reforms. She rightly calls attention to the need for research on project implementation, but she exaggerates the disconnect between commitments and disbursements. She reasonably asks for more systematic and more nuanced coding of development projects, but she fails to provide specific coding rules to accurately capture projects’ likely environmental impact. We do attempt some of this work here and show that more systematic and nuanced coding rules do not weaken—rather, they strengthen—the substantive conclusions of our original article.

As with her empirical critique, Gutner’s limited discussion of conceptual and theoretical issues is troubling. Ours was not a “straightforward application of the basic insights of the P-A model”¹ to the case of the World Bank, but an extension and refinement of P-A theory that generated new observable implications. We argued that IO settings highlight issues that most conventional agency models ignore: longer delegation chains, multiple principals, and, in particular, collective principals.

We thank Brad Parks and Sue Peterson for comments on an early version of this article. For research assistance, we are grateful to Daniel Magleby, Dan Maliniak, and Jess Sloan.

1. Gutner 2005b, 773.

Moreover, Gutner relies on implicit and flawed definitions of core concepts such as principal, agent, and delegation. If such imprecise definitions of these key concepts become widely accepted, we fear that cumulative research within the P-A tradition will suffer. Hence we address these conceptual and theoretical issues first.

Concepts, Theory, and Hypotheses

While she applauds the current interest in agency theory for the study of IOs, Gutner argues that analysts should study other points along the chain of delegation and should recognize “that IOs themselves may also be usefully analyzed as principals when they enter into contracts with recipient countries.”² We agree that agency theory can be usefully employed to analyze multiple links in any delegation chain. We also agree that private contractors and nongovernmental organizations (NGOs) are frequently hired by IOs to implement policy decisions and that IOs thus delegate real authority to them. However, we find it problematic (but increasingly common) for researchers to similarly conceive of recipient governments as agents of IOs.³ This approach stretches the meaning of principal and agent beyond utility.

We see a grave conceptual error in employing agency theory wherever bargaining occurs or wherever one actor can affect another’s incentives. Recall that delegation requires “a conditional grant of authority from a *principal* to an *agent* in which the latter is empowered to act on behalf of the former.”⁴ The Ugandan government has the authority to implement environmental policies on its territory because it is a sovereign state, not because the World Bank or any other IO has authorized Uganda to do so.⁵

More permissive definitions extend agency theory too far beyond its conceptual roots, which focused on employment contracts, the delegation of authority, and bargaining under hierarchy.⁶ It is similarly inconsistent with the seminal P-A appli-

2. *Ibid.*, 780.

3. See Gilbert, Powell, and Vines 2000; Brett 2003; and Gutner 2005a.

4. For an extended discussion of these conceptual issues see Hawkins et al. Forthcoming 2006.

5. A more empirically accurate and theoretically consistent way to conceive of the relationship between the World Bank and a borrowing government is suggested by standard cooperation theory. See Oye 1984. For example, Uganda may agree to adjust its environmental policy in exchange for a loan or a grant that comes from a group of other governments—a group that has delegated authority to an MDB to allocate foreign aid. See Milner 2004. If Uganda fails to comply with the conditions of the agreement, we do not conclude that Uganda is a shirking agent. Unlike a Bank staffer, an NGO, or a private contractor, the Ugandan government cannot be fired by the Bank for failing to comply with its conditions. The most the Bank can do is cut off access to future finance, which has nothing to do with the government’s authority to continue making and implementing policy as it sees fit.

6. See Coase 1937; Alchian and Demsetz 1972; and Fama 1980. Williamson 1975 focuses on identifying the actor (or group of actors) that maintain the “rights of residual control” in a bargaining relationship helps to provide agency theory with its analytic rigor. See also Lake 1996.

cations in political science and international relations.⁷ Contracts negotiated in lateral bargaining settings do have some similarities to P-A relationships (for example, hidden action and hidden information, which can be addressed through screening and selection and in contract provisions). But these contracts lack other critical attributes of P-A relationships in the political realm, especially Madison's dilemma, where delegated authority can be used against the principal. Since no authority is delegated in horizontal bargaining situations, P-A theory will lose traction here. Moreover, many of the "principal's" tools of monitoring and administrative checks will require the consent of any "agent" that is a sovereign government. Conversely, in a hierarchical P-A relationship, the authority of the principal to intrusively monitor and unilaterally impose administrative checks is inherent in the relationship. This is not true for relations between IO donors and sovereign recipients. Hence we should be careful about applying P-A models outside of hierarchical settings. If we expand the definition in the way that Gutner suggests, it becomes difficult to deduce falsifiable hypotheses or distinguish these P-A hypotheses from alternative explanations. Similarly, if principals are misidentified in empirical work, analysts may conclude that delegation succeeds or fails in situations where in fact no delegation takes place at all.

Increasingly, groups of governments agree to delegate their sovereign authority to IOs, including the World Bank. But this context also suggests the limitations of standard P-A models. Governments delegate authority jointly, as a collective body. They also draft elaborate sets of rules and procedures that govern their collective decision making. However, extant P-A models lack conceptual tools adequate to deal with collective principals, which are ubiquitous in politics.⁸ In our article, we suggested steps by which P-A theory might be extended to model collective delegation.

From this theoretical extension, we derive key hypotheses: convergence of preferences among members of a collective principal (or between multiple principals) will increase the probability and amount of institutional reform; and institutional reform will decrease the amount of agency slack. Earlier P-A applications were prone to "false negatives," where analysts attributed agency slack to failures in delegation instead of acknowledging a prior problem of coordination among members of a collective principal. Our 2003 study launched a research program examining the simultaneous problems of collective action and delegation that all collective principals face. This is not a "straightforward" P-A application, but an extension of existing theory that specifies both P-A dynamics and intraprincipal politics that influence delegation outcomes.⁹

7. See Moe 1984; Kiewiet and McCubbins 1991; and Pollack 2003.

8. Lyne 2005.

9. In subsequent work we extend and test the observable implications of both multiple principal and collective principal models. See Lyne 2005; Lyne, Nielson, and Tierney 2004; and Nielson and Tierney 2004.

Similar misconceptions may have led Gutner to conflate distinct independent and dependent variables that were critical to our hypotheses. As our core causal mechanisms we argued that (1) credible threats by the principal to recontract and (2) convergence of preferences within the collective principal should induce variance in our two dependent variables: institutional reform and behavioral change. Perhaps confusion arose because our model is dynamic: we argued that institutional reforms will also provoke behavioral change. Gutner worries that we assume that institutional reform results smoothly in behavioral change and in fact that we appear to believe the two are the same.

This critique is puzzling, because we kept institutional reform and behavioral change distinct in our discussion of hypotheses, and we addressed them in different sections of the article as distinct observable implications of our model. The case study examined institutional reforms that occurred (or failed to occur) at the Bank, and the statistics on loan approvals served as our proxy for behavioral change (or stasis). We did not assume that lending behavior followed smoothly from institutional change—in fact, we reported instances where it did not—but instead discussed the evidence as it supported or failed to support our hypotheses.

Methods and Empirics

In summing up her empirical critique, Gutner states: “At the end of Nielson and Tierney’s article, one is no closer to understanding whether significant environmental reform has occurred at the Bank, and if so, why.”¹⁰ We disagree. The dispute boils down to a fundamental difference over how much one can infer about World Bank behavior from the changes one can observe in Bank institutions and lending commitments. Of course, this dispute hinges on the meaning of “significant.” The loan approval patterns reported in our 2003 article demonstrated substantively and statistically significant changes among lending categories at the Bank.¹¹ Most importantly, the data suggested that the Bank has shifted its lending away from categories that Bank critics and Western governments claim cause the most environmental harm and toward categories whose environmental impact is negligible, nonexistent, or even positive.

Gutner is almost certainly correct at the margins, where implementation issues and other problems of hidden information manifest. But the broad patterns of Bank lending commitments, which overwhelmingly drive disbursements, are unmistakable to any unbiased observer. Given the information disclosure reforms imposed in 1994, it is not credible to suggest that the Bank could successfully masquerade, say, oil exploration loans as public health projects. Yet this is where Gutner’s logic leads, suggesting that lending commitments provide no meaningful information

10. Gutner 2005b, 778.

11. Nielson and Tierney 2003, 267–71.

about Bank behavior. The changes in types of projects funded are not marginal, they are dramatic—whether we measure the proportion of projects¹² or proportion of dollars¹³—and both substantively and statistically significant enough to infer that the Bank’s portfolio was less damaging to the environment by the mid to late 1990s than it was in the mid-1980s. We explore these patterns in greater detail below.

Commitments versus Disbursements

We argued in our 2003 article that loan commitments provide important information about Bank behavior. At a minimum, Executive Board approval of loans is a necessary condition for the disbursement of any funds. However, if the gap between disbursements and commitments is large, then the value of commitment data as a proxy for actual flows is reduced. In actual fact, commitments and disbursements are highly correlated. For example, from our project data set, we randomly sampled 250 International Development Association (IDA) projects between 1980 and 2000 and gathered data on disbursements from the IDA’s “Statement of Development Credits.”¹⁴ Of the 250 projects in the sample, fifty-eight had not yet closed by March 2003. Only two of the sampled projects had been canceled completely.

Of the 192 closed projects, the mean disbursement rate was 91.6 percent (including the canceled projects).¹⁵ We would expect a similar disbursement rate for the International Bank for Reconstruction and Development (IBRD). The sample for all categories suggests that commitments are an excellent predictor of disbursements. In fact the R-squared is 0.95 for a simple regression, suggesting that commitments account for as much as 95 percent of the variance in disbursements for the sample.¹⁶ Thus Gutner’s concern about the gap between commitments and disbursements appears overblown.

Implementation

Gutner’s critique is better placed when she notes that we do not provide data demonstrating that the reformed World Bank implements projects in ways that are more environmentally sound than before. We also agree that evidence on implementation would allow additional tests of our P-A model and should be collected and

12. *Ibid.*, 268.

13. *Ibid.*, 267.

14. World Bank 2003.

15. The disbursement rate for total dollars committed for projects in the sample was nearly identical: 91.7 percent. Interestingly, the disbursement rate for environmental projects was slightly higher than the rate for dirty projects, 93 percent to 91 percent, suggesting that Gutner’s analytic point may be less relevant for environment-friendly projects than for others. However, this difference is not significant statistically.

16. To us, this seems good enough for intergovernment work.

analyzed systematically. Of course, this is true of numerous other dependent variables as well. P-A models have a large number of observable implications and could be fruitfully explored by scholars of IOs. Here we agree with Gutner that many problems of implementation may not result from agency slack, but instead from problems in contract enforcement between the Bank and recipients. However, because we have argued that recipient countries are not Bank agents, we would not see failures to explain or predict these problems as flaws in agency theory but rather in the misapplication of P-A models to issues beyond their reach.

Gutner's evidence that the Bank has not significantly altered its behavior on the environment does not consist of new data on implementation, data on disbursements, a recoding of our data, or any other form of systematically collected evidence.¹⁷ Instead, she dismisses our empirical work—both the case study and the quantitative analysis—with reference to just two internal Bank reports. These Operations Evaluation Department (OED) reports are notorious, however, for offering contradictory assessments and are routinely quoted by both the Bank and its critics in their battle for rhetorical advantage. Gutner quotes a 2002 report as if it were the final nail in the coffin for our article: “The modest extent of mainstreaming the environment into the Bank’s overall program is disturbing.”¹⁸ However, additional reading of this report reveals “support” for the opposite position: “Among the Bank’s environmental projects and programs are many successful examples both of direct environmental project lending and of projects that have mainstreamed the environment into other operations, some of which have served as models for other projects.”¹⁹ Our point is not that there are more quotes in this particular document that favor our interpretation of reform and behavioral change at the Bank, but that such “cherry-picked” quotes are not comparable to evidence that is systematically collected and analyzed using the standards of social science. Whether the methods are qualitative or quantitative, valid inferences require no less.

Institutional Reforms

Gutner's strongest critiques stem from her expertise as a specialist in the area of World Bank environment policy. She claims that our emphasis on the year 1994 is misplaced: “This period of time was not a benchmark in terms of Bank reform.”²⁰ Here, we note that she overstates our emphasis on 1994 and understates its actual importance. We devoted considerable energies (and five pages) to discussing the

17. Interestingly, Gutner's previous analysis of World Bank loans to Central and Eastern Europe, which relies in part on internal Bank assessments, suggests that implementation problems were less acute for environmental projects than for projects in other sectors. See, especially, Gutner 2002, 162, 236–37.

18. World Bank 2002.

19. *Ibid.*, 17.

20. Gutner 2005b, 778.

earlier period of reform beginning in the mid-1980s. Clearly, important institutional reforms occurred then, and the figures and analytical statistics below suggest that the institutional reforms were accompanied by significant changes in lending patterns (particularly in dirty and neutral projects).

However, the critics of the Bank remained unsatisfied with these changes, which they labeled “green-washing.” In the early 1990s, U.S. Rep. Barney Frank (D-Mass.) threatened to withhold \$3 billion in replenishment from the IDA unless the Bank undertook specific institutional reforms. He later partially followed through on the threat by withholding \$1 billion. Frank held hearings and began making the threat in 1992. Thus, if our model is correct, meaningful institutional reforms “that stuck” should have begun after the Frank hearings in late 1992, persisted through 1994, and continued until Congress finally appropriated the withheld \$1 billion several years later. The year 1994 is important only in the sense that, given the average two-year project cycle, it is the first year we would expect to see changes to the loan portfolio resulting from Frank’s threats.

Key institutional reforms in fact took place when we predicted. However, Gutner understates the importance of 1993–94 as a reform period. Several important institutional changes did occur then, including (1) the initial operations of the Inspection Panel in 1994; (2) the implementation of sunshine policies in 1994 that provided public access to Bank documents; (3) changes to the loan approval process in 1994, involving board members much earlier; and (4), the assignment in 1994 of an environment specialist to every environmentally sensitive project. Each of these reforms conforms closely to the mechanisms of principal control detailed in our 2003 article, including, respectively, “fire-alarm” monitoring, information requirements, “police-patrol” oversight, and administrative checks.²¹

Gutner cites Fox to argue that the Inspection Panel has not been an effective check on Bank staff, in large measure because the panel serves at the discretion of the Executive Board and a voting majority on the board can reject its findings.²² However, this is additional evidence for our argument. The Executive Board is the collective principal, and thus it is authorized to filter resulting information and decide the fate of projects. Gutner is correct that this institutional design might inhibit the ability of environmental NGOs to realize all their preferences, but, as we argued in our article, it also is likely that Board majorities rarely share the extreme environmental preferences of NGO activists initiating these complaints. Instead, the Board uses the Panel to gather information from interested parties so that decisions conform as closely as possible to Board members’ actual preferences. Environmentalist scholars and many NGOs would likely prefer a Panel with greater autonomy from the Board (as we do), but our P-A model does not predict this.

21. See Kiewiet and McCubbins 1991.

22. See Fox 2002.

Gutner writes that “the link between reform and a reduction in the gap between Bank intention and activities must be explained and not assumed.”²³ That is definitely right. But institutional change is one of our key dependent variables. It is entirely reasonable to conclude that the evidence supports our hypotheses—that recontracting threats and preference convergence lead to institutional reform—if we in fact observe institutional reform following from such threats and convergence. This is what we reported in our 2003 article.

Coding

There is room for improvement in the methods that we employed in our 2003 research. We coded all loans in a simple fashion, as Gutner correctly recognizes. For example, in coding stand-alone environmental loans, we categorized projects as “primarily environmental” if they were aimed at pollution abatement, including improvements in air, water, and solid waste disposal or at natural resource preservation, including protection of biodiversity, forests, wetlands, and other habitats.

We based coding for projects from 1980 to 1994 solely on the project descriptions found in the World Bank Annual Reports (1980–2000). After 1994, we were able to examine Staff Appraisal Reports (SARs) and Project Appraisal Documents (PADs) to sort out ambiguities, which were rare. Our coding was independent of the World Bank’s own categorization, two researchers coded each project independently, and few disagreements between coders occurred. It is possible, of course, that before 1994 Bank project descriptions grossly distorted the nature of the actual projects. But given the public scrutiny of the Bank since the mid-1980s, it requires extreme cynicism to conclude that such distortions were both significant and widespread. After 1994, when project documents became publicly available, we also included in our data the “mainstreamed” environmental components of traditional sector loans. So, at least after 1994, we explicitly accounted for those traditional projects with “significant environmental components” that Gutner claims we fail to capture.²⁴

That said, we agree with Gutner that our coding—although it improved on what had been available before—was too simple. Accordingly, we have undertaken the Project-Level Aid Database (PLAID) effort, which is now funded by National Science Foundation grant SES-0454384. This database recodes all World Bank loans, in addition to nearly all other bilateral and multilateral development agency projects, for a variety of characteristics, including the degree to which each project causes or ameliorates damage to the natural environment.

Our coding rules for this variable (ENV_IMPACT) are thorough and specific. Our codebook currently runs to more than 6,000 words and is too long to reproduce

23. Gutner 2005b, 779.

24. *Ibid.*, 776. See Nielson and Tierney 2003, 269, fn. 111.

here; however, interested readers can access it at on the Internet.²⁵ To ensure the validity of our data, we precoded projects for six months and consulted numerous environmental specialists in five different disciplines while constructing and revising these coding rules. In addition, the structure of the PLAID database permits scholars who are unsatisfied with our coding rules to revise our substantive assumptions and rerun numbers according to their own criteria.

In short, each project in the database is categorized on a five-point ordinal scale that ranges from most damaging to most beneficial for the natural environment. The values that each project can take are dirty strictly defined (DSD), dirty broadly defined (DBD), neutral (N), environmental broadly defined (EBD), and environmental strictly defined (ESD). DSD projects have a direct negative impact on the environment and are associated with sixteen specific types of projects detailed in the codebook. DSD projects may strip the environment of irreplaceable natural resources, as in the case of extractive industries (such as mining or logging). Projects that severely pollute or degrade the environment also fall under DSD; examples include heavy industry, such as fertilizer, tire, and brick factories, and most energy sector loans (obvious exceptions include renewable energy projects and some efficiency-enhancing projects). By comparison, DBD projects have a moderate but negative impact on the environment, with thirty-nine types of projects specified in the codebook. Overwhelmingly, these finance agricultural projects.²⁶ Other projects that appear frequently in this category include electrical transmission, fisheries, pharmaceuticals, and textiles.

Neutral (N) projects have no immediate impact on the environment. Thirty-one specific types of projects are listed in the codebook as typically neutral. Examples include anticorruption projects, export promotion, education, telecommunications, and disaster aid. Although we can imagine how export promotion could hurt the environment (depending on what is being exported) or how disaster relief could prevent environmental degradation, the environmental effects are typically indirect and not easily predictable.

EBD projects generally have fewer obvious environmental benefits and tend to be more preventative than ESD projects. We specify seventeen distinct types of EBD projects, including nuclear safety and erosion-control projects, which do not immediately improve the environment. Instead, they insure against environmental disasters that nuclear power plants and erosion could cause. By contrast, ESD projects have an immediate positive impact on the environment, for example, projects designed to reduce air pollution, chlorofluorocarbon (CFC) production, or

25. See (<http://mjtier.people.wm.edu/intlpolitics/aid/>). Accessed 10 March 2005.

26. Although we recognize that different crops can have dramatically different environmental impacts; for instance, drought tolerant crops such as millet do less damage than cotton, which typically requires large amounts of water and fertilizer. Future refinements of this coding scheme rely on surveys of environmental scientists which will help us to place different crops (and industries) into specific categories.

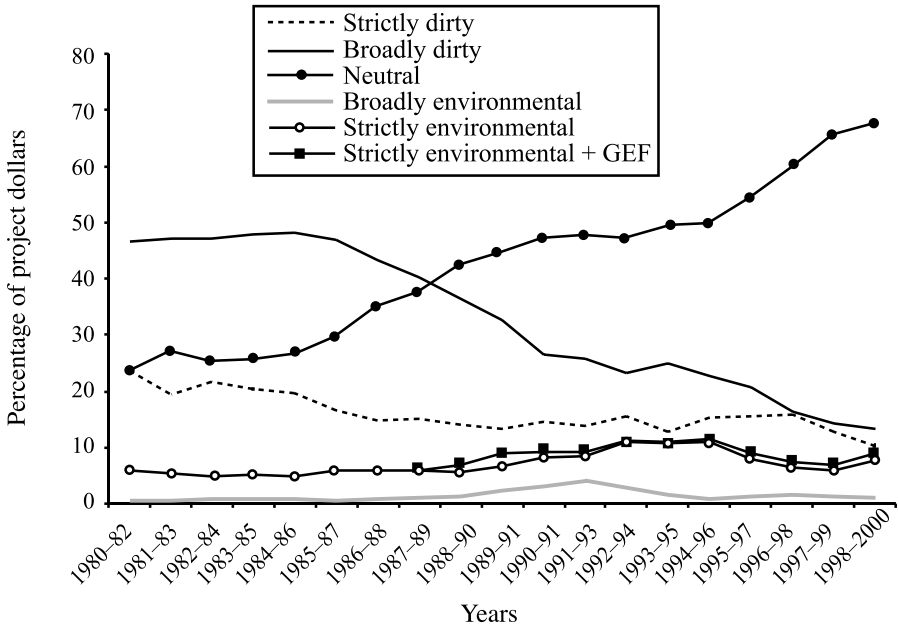
carbon dioxide emissions. The codebook specifies twenty-eight specific ESD project types.

Like the data in our 2003 research, eventually every project in the PLAID database will be independently coded by two researchers. When discrepancies emerge, a senior member of the research team makes a final decision after gathering more detailed information on the expected environmental impact of the project. PLAID categories thus should overcome some of Gutner's harshest criticisms of our methods. Using a beta version of PLAID, we have replicated all of the descriptive statistics from the 2003 article and report them below. In addition, we employ an ordered probit regression model to gain additional leverage on two of the hypotheses specified in our original article.

Descriptive Statistics

While her critique of our coding methodology for our 2003 article is on the mark, it applies equally to other scholars in the field, including Gutner herself.²⁷ Gutner correctly argues that the coding criteria in our article could be more explicit, nuanced, and rigorous. Hence, in Figures 1 and 2 below we use the new PLAID database to reassess the patterns of project approval at the World Bank from 1980

27. In her 2002 book on MDB environmental performance in Central and Eastern Europe, Gutner looks at a much smaller sample of projects but employs a coding scheme that is similar to our 2003 article. As with our article, Gutner concludes that the Bank's criteria for classifying projects lacks precision and undercounts environmental lending. See Gutner 2002, 234, fn. 46. She attempts to identify "projects with primary environmental goals or significant environmental components" (147–48). Ultimately, Gutner codes each Bank project from 1990–99 for the Central and Eastern European region by focusing on "project objectives" (154) that are inferred from project descriptions in official Bank documents. This seems an entirely reasonable approach, and it is qualitatively similar to the method we used in our article, except that we ensured reliability by having every project coded twice. If Gutner's method were substantially different from ours, we might observe significant differences between our list of environmental projects and hers. In fact, Gutner identifies twenty environmental projects (148–49). Of the twenty projects, we coded nineteen exactly as Gutner did (95 percent agreement). The only exception was a District Heating Rehabilitation Project in Estonia from 1994. We did not think that it had a significant environmental component because only \$187,000 out of \$38.4 million were designated for the environment in project documents. However, Gutner's description of such projects in chap. 5 convinces us that it probably should be characterized as environmental because of the positive externalities associated with improved energy efficiency. Even so, in our 2003 article, all of the environmental spending on this project that comes from the IBRD was captured in our measure of project dollars, where we included mainstreamed environmental components (fig. 5, 267). However, we did not capture this as an environmental project in our data on project counts (fig. 6, 268). At the suggestion of a discussant, we also dropped all forestry loans that were coded as environmental in the project count data. However, we included all the mainstreamed environmental dollars in loans from all sectors, including forestry. Hence, all three of the forestry loans that Gutner (accurately) classifies as environmental do not appear in fig. 6 for our 2003 article. Note that all our classification rules were conservative, which made it more difficult to support our hypotheses. If we had used Gutner's coding scheme, our results would almost certainly have offered even stronger evidence in support of our hypotheses. This comparison suggests that our coding system used in 2003 produces empirical patterns quite similar to those produced by at least one leading specialist of multilateral environmental aid and thus should increase the reader's confidence in both the validity and the reliability of our measures.



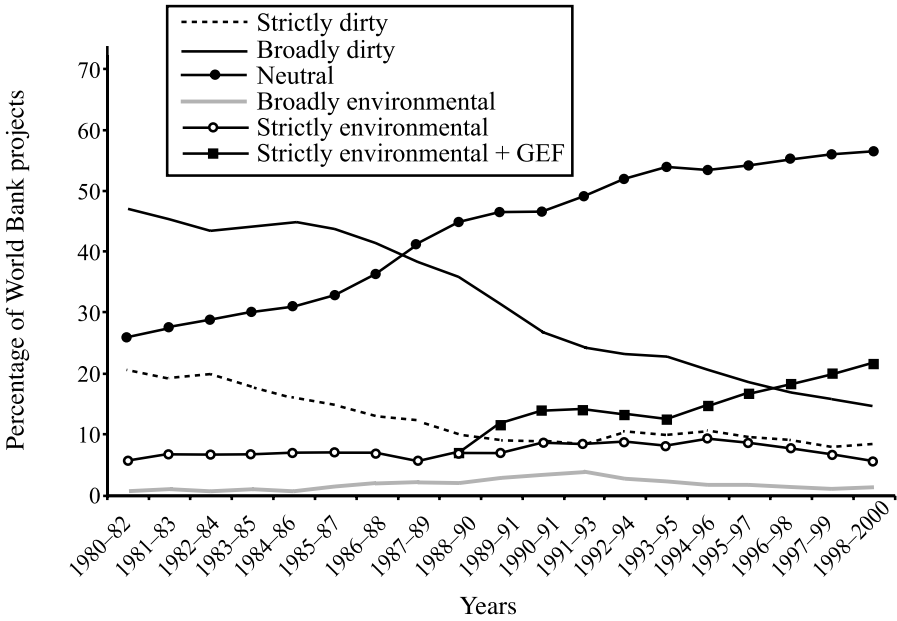
Note: GEF = Global Environment Facility.

FIGURE 1. Percent of World Bank project dollars by category (three-year rolling average)

to 2000, the same period we analyzed in our 2003 article. Reflecting our previous analysis, Figures 1 and 2 show the percent of total project dollars and total project counts, respectively, dedicated to each of the project categories described above.

We also include in the figures a separate line (marked by closed squares) that shows the percent of total dollars and proportion of total projects that Global Environment Facility (GEF) projects add. We agree with Gutner that GEF grants are different than IBRD/IDA loans, which is why we differentiated them from IBRD/IDA projects in our 2003 article and here as well. But since the World Bank staff is managing all GEF projects included here, these data do reflect the fact that additional staff resources must be dedicated to environmental project design, implementation, and oversight for Bank-implemented GEF projects.²⁸ Because the actions of the agent (Bank staff) are an important observable implication of many P-A

28. Gutner is persuasive when she argues that staff activities on GEF projects cannot be fully explained by recontracting on the part of the Bank board. After all, the GEF has its own board with different voting shares and different voting rules. Interestingly, these two distinct collective principals share many of the same agents (Bank staffers who wear two hats). This observation suggests a potentially interesting exploration of the multiple principal model within an IO setting.



Note: GEF = Global Environment Facility.

FIGURE 2. Percent of World Bank projects (count) approved by category (three-year rolling average)

models, Bank-managed GEF projects remain germane, so we include them in the figures. However, prompted by Gutner’s legitimate concern that the GEF projects may be driving our results, we omit all GEF projects from the statistical tests reported below.

Both figures display qualitatively similar patterns to those found in our 2003 article. The proportion of dollars committed to environmental projects at the Bank increases beginning in the late 1980s, peaks in the early 1990s, and decreases through the mid-1990s until the last period. The same is generally true of the number of environmental projects, except that the inclusion of GEF projects displays a sharp increase in the proportion of environmental projects as a share of the total number. The general increase in the proportion of environmental projects and project dollars through the early 1990s is broadly consistent with our argument, though the downturns in the late-1990s are not. This is precisely what we reported in 2003.

The similarities in these data compared to the findings in the 2003 article also manifest in the patterns for other types of projects. Here, we encounter the strongest evidence for our hypotheses. In general, the number of “dirty” projects defined broadly remains stable until the late 1980s, decreases sharply until it flattens in

the early 1990s, and then again decreases sharply beginning in the middle 1990s through 2000. A similar—though much less pronounced—downward trend can be seen for dirty projects strictly defined.

The trend for neutral projects—which have no directly discernable impact on the environment—are inversely related to the dirty projects and are even more pronounced for project dollars. These data strongly suggest that the World Bank has shifted projects away from activities that threaten the environment. While the Bank received substantial pressure to pursue environmental projects, the demands from member states to “stop doing harm” were indisputably greater.²⁹ Figures 1 and 2 suggest that Bank staff members changed lending behavior accordingly. The trends for dirty and neutral projects thus provide very strong evidence for our argument.

Analytical Statistics

Regression analysis lends greater confidence that the patterns we see in the figures are not the result of hidden variables. As described above, we coded each loan for a single ordinal dependent variable from least to most environmental (strictly dirty < broadly dirty < neutral < broadly environmental < strictly environmental). Responding to Gutner’s reasonable critique regarding the different nature of GEF projects, we excluded them from this sample. Using ordered probit regression analysis, we tested the effects of the periods after 1986 and 1993, along with a large set of control variables seeking to hold constant the effects of demand for environmental loans,³⁰ on the likelihood that loans would appear in any of the categories ranging from strictly dirty to strictly environmental. Table 1 displays our results.

Along with the control variables reported in Table 1, we coded dummy variables for each year (compared to the baseline of 1980) to hold constant the effects of year-to-year fluctuations. The key independent variables for post-1986 and post-1993 were positively and significantly related to more environmental loans at the 0.05 and 0.001 levels, respectively. These results support our argument that the two periods should have seen shifts in loan commitments toward environmental projects and away from dirty projects. They also suggest that the shift after 1993 toward environmental lending was more significant than after 1986. Controlling for yearly variation, our two key independent variables performed better than nearly all of the control variables, save GDP PER CAPITA and FORESTATION RATE. Richer countries were significantly (at the 0.001 level) more likely to receive more environmental loans, as were countries with high deforestation rates. The other control variables were not significantly related to loan type.

29. See Wade 1997; and Rich 1994.

30. We lagged the control variables by two years in an attempt to account for the average length of the project cycle during the period under study.

TABLE 1. Ordered probit regression results on project type

<i>Independent variables</i>	<i>Coefficient</i>	<i>Standard error</i>	<i>z value</i>	<i>p value</i>
AFTER 1986	0.281	0.115	2.44	0.015
AFTER 1993	0.542	0.100	5.42	0.000
GDP (1995) IN \$100 BILLION	-0.017	0.012	-1.44	0.149
GDP GROWTH	-0.002	0.003	-0.64	0.522
GDP PER CAPITA	0.040	0.011	3.51	0.000
SECONDARY SCHOOL ENROLLMENT	0.001	0.001	0.97	0.334
SULFUR DIOXIDE EMISSIONS	0.008	0.016	0.49	0.625
THREATENED MAMMALS	0.019	0.185	0.10	0.917
COAL USE PER LAND AREA	-0.388	0.467	-0.83	0.406
ORGANIC WATER POLLUTION	0.018	0.037	0.49	0.625
FORESTATION RATE	-0.575	0.133	-4.34	0.000
POTABLE WATER	0.000	0.001	0.07	0.944
<i>Number of observations</i>	5,130			
<i>Log likelihood</i>	-6271			
<i>Chi²</i>	370			

Note: Dummy variables for each year (1981–2000) are omitted from table but included in regression.

We can predict probabilities for each loan type given specified changes in the key independent variables. Holding all control variables at their means or modes, the probability of a strictly dirty loan shifted from a baseline before 1987 of 0.21 to 0.13 after 1986 and to 0.05 after 1993. Concomitantly, the probability of a strictly environmental loan went from a baseline of 0.03 before 1987 to 0.06 after 1987 and to 0.15 after 1993. The other categories saw corresponding changes across the two periods. These results move us beyond the descriptive statistics above and the simple analytical statistics presented in our 2003 article, and they reinforce our conclusion that Bank-financed projects became significantly more environmentally friendly after 1987 and especially after 1993.

Conclusion

Many of Gutner's empirical and methodological critiques help to advance our thinking on IO reform and behavior. She suggests a number of areas for empirical research that are both substantively interesting and amenable to analysis within a P-A framework. While there are currently a large number of studies on these topics employing nonrandom case selection, small-*n* samples, unsystematic coding rules, and research methods that defy replication, we believe our work on institutional reform, financial commitments, and disbursements avoids these pitfalls and contributes to the growing literature on P-A theory in international relations and

political science. We are hopeful that Gutner and other scholars will launch complementary studies on aid implementation and effectiveness. Unsurprisingly, we agree with Gutner that “P-A models, properly applied, can greatly advance rationalist approaches to IOs”³¹ and, more importantly, advance our knowledge of IO behavior and change in world politics.

References

- Alchian, Armen A., and Harold Demsetz. 1972. Production, Information Costs, and Economic Organization. *American Economic Review* 62 (5):775–95.
- Brett, E.A. 2003. Participation and Accountability in Development Management. *Journal of Development Studies* 40 (2):1–29.
- Coase, Ronald. 1937. The Nature of the Firm. *Economica* 5 (17):386–405.
- Fama, Eugene. 1980. Agency Problems and the Theory of the Firm. *Journal of Political Economy* 88 (2):288–307.
- Fox, Jonathan A. 2002. The World Bank Inspection Panel and the Limits of Accountability. In *Reinventing the World Bank*, edited by Jonathan Pincus and Jeffrey Winters, 131–63. Ithaca, N.Y.: Cornell University Press.
- Gilbert, Christopher, Andrew Powell, and David Vines. 2000. The World Bank’s Structure: The Bank as an Institution. In Gilbert and Vines, *The World Bank: Structure and Policies*. Cambridge: Cambridge University Press.
- Gutner, Tamar. 2002. *Banking on the Environment: Multilateral Development Banks and Their Environmental Performance in Central and Eastern Europe*. Cambridge, Mass.: MIT Press.
- . 2005a. Explaining the Gaps Between Mandate and Performance: Agency Theory and World Bank Environmental Reform. *Global Environmental Politics* 5 (2):10–37.
- . 2005b. World Bank Environmental Reform: Revisiting Lessons from Agency Theory. *International Organization* 59 (3): 773–83.
- Hawkins, Darren, David Lake, Daniel Nielson, and Michael Tierney. Forthcoming 2006. States, International Organizations, and Principal-Agent Theory. In *Delegation Under Anarchy*, edited by Darren Hawkins, David Lake, Daniel Nielson, and Michael Tierney. Cambridge: Cambridge University Press.
- Kiewiet, D. Roderick, and Matthew D. McCubbins. 1991. *The Logic of Delegation: Congressional Parties and the Appropriations Process*. Chicago: University of Chicago Press.
- Lake, David. 1996 Anarchy, Hierarchy, and the Variety of International Relations. *International Organization* 50 (1):1–33.
- Lyne, Mona. 2005. *The Voter’s Dilemma: Clientelism, Collective Goods and Democratic Accountability*, unpublished manuscript, University of South Carolina. Columbia, SC.
- Lyne, Mona, Daniel L. Nielson, and Michael J. Tierney. 2004. A Problem of Principals: Common Agency and Social Lending at the Multilateral Development Banks. Paper presented at the 100th Annual Meeting of the American Political Science Association, September, Chicago.
- Milner, Helen. 2004. Why Delegate the Allocation of Foreign Aid to Multilateral Organizations? Principal-Agent Problems and Multilateralism. Paper presented at the Conference on Delegation to International Organizations, September, La Jolla, Calif.
- Moe, Terry M. 1984. The New Economics of Organization. *American Journal of Political Science* 28 (3):739–77.

31. Gutner 2005b, 782.

- Nielson, Daniel L., and Michael J. Tierney. 2003. Delegation to International Organizations: Agency Theory and World Bank Environmental Reform. *International Organization* 57 (2):241–76.
- . 2004. Principals and Interests: Common Agency and Multilateral Development Bank Lending. Paper prepared for the 100th Annual Meeting of the American Political Science Association, September, Chicago.
- Oye, Kenneth. 1984. *Cooperation under Anarchy*. Princeton, N.J.: Princeton University Press.
- Pollack, Mark A. 2003. *The Engines of European Integration: Delegation, Agency, and Agenda Setting in the EU*. New York: Oxford University Press.
- Rich, Bruce. 1994. *Mortgaging the Earth: The World Bank, Environmental Impoverishment, and the Crisis of Development*. Boston: Beacon Press.
- Wade, Robert. 1997. Greening the Bank: The Struggle over the Environment, 1970–1995. In *The World Bank: Its First Half Century*, Vol. 2, edited by Devesh Kapur, John P. Lewis, and Richard Webb, 611–734. Washington, D.C.: Brookings Institution Press.
- Williamson, Oliver. 1975. *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization*. New York: Free Press.
- World Bank. 2002. *Promoting Environmental Sustainability in Development: An Evaluation of the World Bank's Performance*. Washington, D.C.: World Bank.
- . 2003. *Special Purpose Financial Statements and Internal Control Report of the International Development Association*. Washington, D.C.: World Bank.