

## **The Measure of Success: Some Methodological Considerations for Research on Institutional Performance**

**Paul F. Steinberg**

Central to the aims of the contemporary environmental movement, and the research programs it has inspired, is the hope of having a positive causal impact on environmental quality and on the patterns of social behavior that affect it for good or ill. Such causal aspirations lie at the heart of all environmental institutions, whether these take the form of agencies, laws, or advocacy groups.<sup>1</sup> It is hoped that the Clean Air Act will lead to a decline in sulfur dioxide emissions; that the international whaling regime will cause cetacean populations to rebound; and that stakeholder consortia will increase participation and reduce conflict. As Pressman and Wildavsky (1973, xv) observed a quarter century ago, "Policies imply theories. Whether stated explicitly or not, policies point to a chain of causation between initial conditions and future consequences. If X, then Y."

Accordingly, *while causal theories play a role in all areas of social inquiry, they are vital to research initiatives seeking to provide solutions to environmental problems*. In this paper I provide a collection of observations on methods for evaluating the causes and effects of environmental institutions. Rather than revisit the broader literature on causal inference in social science research (see King, Keohane and Verba, 1994; Przeworski and Tenue, 1970; Mitchell and Bernauer, 1998; George and Bennett, forthcoming), my goal is to highlight methodological dilemmas frequently encountered by researchers hoping to improve the performance of environmental institutions. The paper is divided into two parts. First I argue that research on environmental politics and policy frequently fails to distinguish between the activity and impact its subjects, leading to causal statements that are ambiguous and often misleading. The remainder of the paper focuses on how to craft meaningful measures of institutional performance.

### Activity Analysis versus Impact Analysis

To properly assess the effect of environmental institutions requires that we distinguish between an institution's activities and its impact. Failure to make this simple distinction has led to a profusion of ambiguous causal statements in the published literature. The type of statement that is the target of my critique includes: "The main effect of the EPA program has been to increase public participation." Or, "Greenpeace affects global environmental politics by lobbying governments and by shaping public opinion." In each case, it is unclear whether the phrase is describing activity or causation. Participation may be the main effect of the EPA program in question, but is that effect significant relative to the various forces shaping participation? Greenpeace may spend a lot of time on public opinion, but is public opinion shaped largely by Greenpeace? In each case we have a causal statement that stems entirely from observations on the independent variable. The crux of my argument is that after observing the effects of the institution on a particular outcome, the outcome itself must be studied more systematically to make a determination concerning the relative importance of the institution.

This problem is pervasive in assessments of environmental institutions. The rapidly evolving facts of environmental activity beg explanation, and as researchers we often chase after some interesting new phenomenon - an innovative program, a multinational agreement, an advocacy campaign, a business

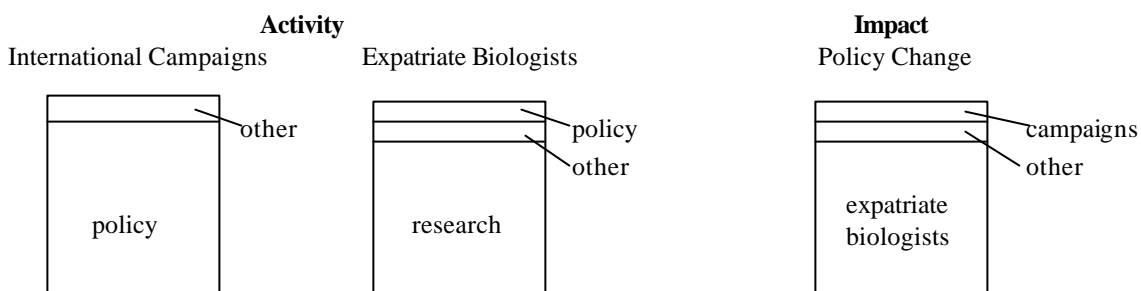
---

<sup>1</sup> Following Young (1994), I define institutions as sets of rules, roles and responsibilities around which expectations converge. Institutions may assume various forms, including markets, laws, organizations, and regimes (constellations of laws and associated organizations specific to an issue area).

initiative - in an attempt to characterize its impact. Too often these studies take the form of independent variables in search of a dependent variable. An example can be found in Keck and Sikkink's (1998) analysis of transnational activism. Noting that policy reform is one of the goals of international environmental campaigns, these authors observe that successes in this area stems from a "boomerang effect," whereby domestic advocates enlist the help of foreigners to leverage changes in domestic policy. Yet in my own study of foreign influences on environmental policymaking in Costa Rica and Bolivia, the boomerang effect was absent (Steinberg, forthcoming). The reason is simple: High profile international campaigns are relatively unimportant in shaping environmental policymaking in developing countries. The transnational alliances that matter most are of lower profile, higher complexity, and longer duration. Again, policy is important to international campaigns, but the reverse is not true.<sup>2</sup>

My argument is represented schematically in Figure 1. The boxes on the left represent two independent variables, showing the percentage of their activity devoted to one or another outcome. The box on the right represents the dependent variable, showing the relative importance of different independent variables in producing the outcome. To illustrate, we can think of the relative impact of international campaigns, and of expatriate biologists, on conservation policy in developing countries. Campaign organizations expend a great deal of energy trying to affect policies, and these activities will figure prominently in their promotional literature; but because campaigns are rare, focused, short-term events, their impact is small relative to the decades-long process of conservation policy reforms.<sup>3</sup> Foreign scientists, by contrast, expend little of their energy on matters of policy. It is a point of professional pride that they spend most of their time in remote field locations, doing the "real science," and they will frequently focus the interviewer's attention on their dominant activity. But because they are among the few foreigners who stay in one developing country over a period of decades - long enough to acquire domestic political savvy and to see through major policy developments - the small percentage of their time devoted to policy questions has an inordinate amount of influence.

**Figure 1. Activity versus Impact**



<sup>2</sup> To date, research on transnational environmental relations (Princen and Finger, 1994; Wapner, 1996; Lipschutz, 1996) has been driven by this sort of independent variable analysis, focusing on the activities of highly visible multinational environmental NGOs with headquarters in industrialized countries as the empirical basis for broader generalizations about international environmental politics. However, environmentalists in Asia, Africa and Latin America seek out a broad range of foreign allies - from scientific institutions to private philanthropists, religious organizations, universities, overseas development agencies, and a wealth of environmental NGOs large and small - and combine foreign resources with the domestic political expertise needed to see through major policy innovations. If we take our cues from the agendas of multinational environmental NGOs we are left with a small and unrepresentative subset of cross-border environmental activity.

<sup>3</sup> This raises the general question as to why an organization would devote most of its energy to a matter on which it has little influence. This may simply be a function of the limited size, duration, or leverage of the group relative to the outcome (e.g. Ford Motor Company's limited influence on global CO<sub>2</sub> levels) and does not imply inefficiency or irrational behavior. An organization may deem it worthwhile to affect outcomes even slightly, taking full advantage of its limited influence. Moreover, actors do not merely pursue strategies that will maximize their impact; they are constrained by their institutional history and ideology (see Dalton, 1994).

## Operationalizing the Concept of Institutional Success

Any research project aspiring to enhance the performance of environmental institutions is immediately faced with the question of how to operationalize the concept of success, meeting the dual requirements of policy relevance and analytic tractability. What is the axis along which variation in outcomes may be distinguished? Researchers will commonly face a few questions along these lines. (1) What is the appropriate functional form for the dependent variable? (2) Do we measure outputs (laws, policies, budgetary allocations, FTEs) or outcomes? (3) What is the standard by which institutional performance will be measured?

### *Success as a Continuous Variable*

"Success!" The very expression implies a question to which the answer is yes or no, success or failure. That is, it is tempting to operationalize success in terms of a bivariate dependent variable (also referred to as a discrete or qualitative dependent variable), with a clearly defined threshold beyond which we may proclaim that a policy or institution has been a success, and short of which the ruling is failure. Such thresholds may be encoded in law (e.g., a requirement that firms reduce emissions by 30 percent) or may accurately characterize certain decisions, such as whether or not to participate in the Energy Star Program. However, most outcomes are more accurately represented on a graded scale, as continuous variables. We may view attainment of a 30 percent emissions reduction as a case of successful compliance. But if the firm in question actually reduced its emissions by 99 percent, we would not want to lose this information. Even outcomes that at first blush seem to lend themselves well to discrete characterization, such as the outbreak of war or the siting of hazardous waste facilities, in practice contain degrees, and those gradations are relevant to policy. It may be desirable to set a threshold when this captures the essence of the dependent variable or when sample size rules out a more fine-grained analysis. However, *studies of institutional performance should not lower the quality of the data set by forcing a continuous phenomenon into a discrete analysis.*

This has two practical implications. First, while maximum likelihood estimation techniques (such as probit and logit analysis) have gained popularity in social science research in recent years, we should be sure to choose our models to fit the data, rather than vice-versa. When continuous data are available, a simple ordinary least squares analysis (or at least a multivariate MLE analysis) will be more appropriate.

Second, the criteria of necessary and sufficient conditions, used widely in qualitative assessments of causation, will often be inappropriate for studies of environmental institutions because these criteria require a discrete dependent variable. To illustrate, a 20-cylinder twin overhead cam engine is not necessary for the drive from Durham to Washington (an ordinary engine will do); nor is it sufficient, as we need a steering wheel, brakes, and so forth. But it would certainly help! For the necessary and sufficiency criteria to apply, we would have to impose a threshold, specifying "To make the trip in under 3 hours..." Yet in research on the origins and impacts of environmental institutions, the broader range of performance matters, and therefore causal variables will seldom be necessary or sufficient to produce a given outcome, but may nonetheless be extremely important.

### *Outcomes versus Outputs*

While improvements in environmental outcomes are the *raison d'être* of environmental institutions, measuring their impact on environmental quality is often infeasible *and will likely continue to be so.* Where the causes of environmental quality are few; where causal chains between policies, behaviors and outcomes are short; and where environmental quality data sets are available in a useful form and cover a

time span that allows for a statistically distinguishable before-and-after comparison, outcome analysis may be feasible. Rarely do these conditions portend. Environmental quality is typically the result of numerous factors beyond the institutions and behaviors in question. To complicate matters, policies are often introduced in bundles, as different components of a congressional spending bill, bilateral aid program, or county land use plan. In such cases a strong temporal correlation between outcomes and program initiation may provide little indication of its impact.

Given budget constraints, baseline data on environmental quality are not collected unless there is a compelling reason to do so. The starting point for many data sets is the point at which society becomes concerned about a matter - that is, after which the "natural" baseline had been changed. This is true, for example, of climate change measurements and of trends in northern spotted owl population size.<sup>4</sup> To assess the impact of public policy on environmental quality, we need not only a sufficient historical record, but an adequate post-policy evaluation period. Because the success or failure of major policy initiatives generally unfolds over a period of a decade or more (Sabatier, 1991; Baumgartner and Jones, 1993) there is a risk accompanying the study of "hot" topics with little track record, such as the impact of NAFTA on environmental quality (see Mitchell and Bernauer, 1998). Given that such topics are the lifeblood of policy-relevant research, a superior strategy may be to evaluate analogous cases with a longer history, and to use the results to inform current policy. An example is provided by Margaret McKean and Elinor Ostrom, whose research on longstanding common pool resource regimes has been used to inform the current interest in community-level conservation schemes.

I do not wish to overstate the difficulty of outcome analysis. In some cases, these constraints will be absent or overcome, with the use of large-N studies (which may control for the influence of confounding variables), post-facto construction of longitudinal data sets (such as ice cores to measure trends in CO<sub>2</sub>), or simulations that can distinguish the important uncertainties from the trivial ones. But speaking to the area I know best - biodiversity conservation - *there is a tendency on the part of natural and social scientists alike to underestimate the data requirements of collecting the most rudimentary ecological data*, such as the location and abundance of various species in a given locale.<sup>5</sup> Suffice it to say that in studies aspiring to evaluate the origins and impacts of environmental institutions, environmental outcome analysis it is not something to be tacked on to an already ambitious research project.

Given the difficulty of measuring the impact of institutions on environmental outcomes, environmental policy research typically measures success in terms of outputs - funds committed to wetlands conservation, factories inspected per month, or decisions by firms to join voluntary environmental programs. I strongly suspect that a study comparing the first and second drafts of environmental policy grant proposals would find the strategic retreat from outcome analysis to output analysis to be a recurrent theme.

---

<sup>4</sup> The reader may protest that the problem would not have been noticed absent at least a rough impression of longitudinal trends in environmental quality. However, often what we notice are increases in activities (such as CO<sub>2</sub> emissions) that *may* cause a change in environmental quality. Moreover, even if we notice an unmistakable change in environmental quality, we need to understand longer-term trends in order to distinguish natural variance from changes in the trend line.

<sup>5</sup> President Clinton's plan to reduce timber conflicts in the Pacific Northwest ("Option 9") envisioned adaptive management units in the national forests, but establishing baseline data on species abundance in these units would require so much labor and expertise that the adaptive management component is moribund. Similarly, southern California's Natural Communities Conservation Planning process, touted by Secretary of Interior Bruce Babbitt as a model for the nation, seriously underestimated the difficulty of collecting information on the location of high-quality species habitat. The result is a conservation plan with little scientific credibility. Costa Rica's famous "all-taxa biodiversity inventory" of the Guanacaste region was deemed overly ambitious and ultimately abandoned.

One of the major pitfalls in using outputs as a measure of performance, particularly in cross-national comparisons, is that a given policy instrument can be used for widely divergent purposes (what we might call functional divergence),<sup>6</sup> while diverse policies may produce similar outcomes (functional convergence).<sup>7</sup> In this context, *the construct validity of output measures is paramount*. Simply put, we must be sure that our output variables are measuring what we claim. Indices that use principal components analysis or prior theory to incorporate several measures of the phenomenon in question can help. We should also be aware that the strategic retreat from outcome analysis to output analysis leads us to focus on compliance rather than effectiveness; in cases where the policy itself is flawed, high levels of compliance may be accompanied by poor environmental outcomes (see Chayes and Chayes, 1995; Mitchell, 1994). We will often need to settle for output analysis, but we should be untiring in expounding for our audience the potential shortcomings of our measures of success.

### *The Social Construction of Success*

What level of institutional performance constitutes a successful outcome? Whether studying outputs or outcomes, we are left with the question of how high to place the bar. This problem cannot be sidestepped through rank ordering of performance. If the institutions in question are all performing below an acceptable standard (as established by the lawmaker, the citizen, or the analyst), the causal mechanism accounting for the difference between the best and worst performer may be different from the mechanism needed to elevate performance to an acceptable level. In modeling terms, this requires projecting beyond the range of the data. In practical terms, a rank order analysis absent an explicit standard may be nibbling at the edges of the problem rather than addressing the tough questions.

"A policy problem," Charles Anderson (1978) reminds us, "is a political condition that does not meet some standard." The emergence of a new policy problem may therefore result from a perceived worsening of conditions or by raising the standards by which long-standing conditions are judged. The same can be said of the search for environmental solutions; our assessments of institutional performance are derived partly from the actions and attributes of the institutions themselves, and partly from the normative orientations and empirical expectations the analyst brings to bear. Accordingly, *a research program aspiring to produce cumulative, inter-subjectively verifiable knowledge on institutional performance requires careful front-end reflection - and tail-end exposition - regarding our measures of success*. The height of the bar marks the extent and nature of the problem, defines the anomalous cases, and shapes our recommendations for reform.

The relevance of this point for cumulative research programs can be seen in the historical evolution of the policy sciences. In the spirit of optimism that characterized policy research in the 1960s, the failure of President Johnson's War on Poverty was egregious. If we can land a man on the moon, these researchers asked, why are we so unable to alleviate poverty in the inner cities? (Nelson, 1977) The policy implementation studies spawned by these failures highlighted the preponderance of veto points in the American political system, leading to Pressman and Wildavsky's (1973) famous title, "Implementation: How Great Expectations in Washington Are Dashed in Oakland; Or, Why It's Amazing that Federal Programs Work at All..." With this more sober view, in later years the instances of success were the

---

<sup>6</sup> Land redistribution programs gained prominence in Latin America because they were seen as essential to changing at least three distinct outcomes: curbing inflation, halting the spread of communism, and helping the poor - roughly in that order. If rural poverty reduction is the outcome of interest, land redistribution programs (the outputs) would be a potentially misleading proxy.

<sup>7</sup> For example, Japan has low levels of state spending on social welfare relative to other industrial democracies, but is nonetheless a comparatively egalitarian society; it would be a mistake to infer the social outcome in Japan from its levels of social spending (the output). (Castles, 1989)

anomalous cases begging explanation, a task engaged by the recent wave of research on "new public management" (Barzeley, 1992; Bardach, 1998). The point I wish to emphasize is that the institutional accomplishments studied by new public management researchers would not even have come into focus absent a decades-long critical engagement of what constitutes exceptional institutional performance.

Clarity regarding the standard of success is particularly pertinent to cross-disciplinary work, because our expectations concerning institutional performance are shaped by the prevailing views of our respective fields. In their work on international environmental institutions, Haas, Keohane and Levy (1993) construe success as the ability to have some positive impact on environmental quality - a standard that would strike some as unduly timid given the scale of environmental problems and the institutional resources at hand. But it makes sense from the perspective of international relations, in which a large body of research in the realist tradition suggests that these institutions should have no impact independent of the power configurations underlying them. In contrast, programs having "some impact" on increasing women's participation in development is unacceptable to feminist scholars such as Mayoux (1995), who argue that the persistence of an overall pattern of nonparticipation is a mark of failure.<sup>8</sup> The standard wisdom shapes the standard of performance. A workers' uprising is unsurprising to Truman, astonishing to Olson, and long overdue to Marx.

It is perfectly appropriate for researchers to derive a measure of success that is anchored to the difficulty of the task at hand. In California, where a two-thirds majority is required to raise tax revenues for local land preservation, the measure of success (and corresponding designations of leaders and laggards) will differ from a similar analysis in a state with more forgiving voting rules (see Press, 1998). In my research, I conclude that Costa Rica and Bolivia have achieved impressive gains *relative* to the immense challenges of institution building in poor countries. Any attempt to derive general propositions across research traditions regarding determinants of successful environmental institutions should be aware that the explanatory variables may differ from one field to the next in part because we are employing different scales with respect to the dependent variable.<sup>9</sup>

### *Willingness versus Ability*

"If citizen participation programs are only symbolic, it is because agency administrators want them that way," or so argues one of the country's foremost experts on participatory institutions (Berry, 1981). This raises the general point that when measuring institutional performance, it is crucial to distinguish between externally imposed standards (what a researcher or social constituency deems to be a successful institutional outcome) and internal ones; is this in fact a goal shared by the institution in question? Indeed, is it a goal shared by the society in question?

Too often this very simple analytic distinction between willingness and ability goes unheeded in studies of institutional success. Ostrom's (1990) requisites of local collective action will not produce environmentally sustainable management schemes absent a collective commitment to sustainability - a

---

<sup>8</sup> Similarly, when asked which countries have been most successful in protecting biological diversity, conservation biologists will typically reply "none," because the pace of extinction everywhere far exceeds that envisioned by the normative underpinnings of the field.

<sup>9</sup> Another implication of the subjective nature of institutional assessment is that we can expect biases in the self-assessments offered by environmental actors. Environmentalists with high standards for environmental quality will not only judge agency performance more harshly; it is my experience that their view of their own impact is more pessimistic than that of outside observers. In a similar vein, private firms and government agencies with the greatest commitment to the environment will generally provide more complete information concerning the severity of the environmental problems under their purview. In such cases we may wish to supplement our queries with a subjective ranking from the organization in question of its performance relative to that of its peers.

variable that figures nowhere in the analysis. Similarly, an increase in social capital (see Putnam, 1993) may not be correlated with improvements in the provision of public goods because networks of trust and reciprocity are also the social technology underpinning political cronyism and terrorist cells. Putnam was able to control for social preferences in his analysis of institutional performance in Italy by examining a wide range of policy outcomes. Studies hoping to use social capital as a predictor of institutional success in the environmental arena must take into account the fact that some communities will apply their ample capacity to efficiently pave over bird habitats.

To flesh out the distinction between willingness and ability in assessments of institutional performance, Figure 2 categorizes institutional failures according to whether society, and the institution's leaders, believe that the institution in question should pursue environmental ends. I use government agencies for illustrative purposes. In each scenario, the outcome is assumed to be detrimental to environmental quality. "Inefficient" institutions are those for which social preferences and the agency's core mission both stress environmental protection, but were unable to prevent the observed outcome. In such cases an agency may be suffering from corruption, principal-agent problems, an external regulatory environment that prohibits it from carrying out its mission (red tape), or poorly designed policies. The U.S. Superfund program would fall into this category (see Mazmanian and Morell, 1992). With "Weak" institutions, the agency's core mission stresses environmental goals, but it does not enjoy enough public support to enable it to carry out its mission. The United Nations Environment Programme is an example of a weak institution whose meager resources have marginalized its role in international environmental affairs. In some cases a weak agency may simply be entrepreneurial, in the process of proving its worth and advancing a cause that the society will later come to embrace. In other instances this may be a case of environmental clientelism, where the agency is pursuing an agenda plainly at odds with popular will, and as a result is not able to serve its environmental "clients" well.<sup>10</sup>

**Figure 2. Anatomy of Institutional "Failure"**

		Public Environmental Will <sup>†</sup>	
		high	low
Agency Environmental Will <sup>‡</sup>	high	INEFFICIENT principal-agent problems corruption, red tape, poor program design (Superfund)	WEAK environmental clientelism, entrepreneurship (UNEP)
	low	CONFLICT-RIDDEN outdated, "whipsawed" (US Forest Service)	UNWILLING faithfully shunning environmental goals (Orange County, CA Planning Commission)

<sup>†</sup> Public preference for this particular agency to pursue environmental goals

<sup>‡</sup> Importance accorded environmental goals in the agency's core mission

<sup>10</sup> The terms "entrepreneurial" and "client" agencies are from Wilson, 1989.

"Conflict-Ridden" institutions show a mismatch between social preferences for environmental quality and the agency's core mission. The concept of *core* mission is crucial, because we must not only ask whether the institution shares the goal, but the importance it places on this goal relative to its central purpose. The distinction between inefficient and conflict-ridden institutions is crucial. The observation that the U.S. Forest Service has "failed" to protect biological diversity says less about organization efficacy and more about the mismatch between the agency's enduring mission and changing social preferences. In such cases the agency in question may give lip service to environmental protection, creating environmental advisory boards and revising certain provisions of its organic legislation, while in practice continuing to pursue its core mission.<sup>11</sup>

Finally, the "Unwilling" institutions are those for which the agency and the relevant citizenry agree that it should not pursue environmental ends. The Orange County Planning Commission in southern California provides a prototype for such an outcome. In these scenarios, the outcome is a failure only from the perspective of an outside observer whose interests differ from those of the polity in question. Such an observer may wish to engage in consciousness-raising (fostering the rise of new values), to provide information bearing on pre-existing preferences, or to make a claim based on externalities. Although unwilling institutions are often characterized as poor performers, the concept of performance refers to progress toward an agreed-upon end. Where there is disagreement on the goals, this is not merely a different outcome on the variable of institutional performance, but a different equation altogether.

#### Future Directions: The Relation between Policy Failure and Policy Success

Confronted with underperforming institutions, how can analysts contribute to the search for solutions? Let me conclude with an observation concerning the relation between the success and failure of environmental institutions. There will always be more causes of failure than causes of success. This is so because a policy success is, by definition, faithful adherence to a complex set of causal events needed to conclude a treaty, to manage a park, or to clean up a waterway. Any significant deviation from this particular chain of events (i.e., all other imaginable outcomes) constitutes a failure. As a result, failure has random events on its side. Three practical implications follow. First, many and diverse variables are sufficient to cause institutional failure, while this is not true of success. Therefore once we identify the causes of failure, we must remember that just as the cessation of cigarette smoking does not cause immortality, removing the cause of failure may not be sufficient for success, for some other problem may arise in its place. Success does not equal one minus failure. The second implication is that the controlled-case comparison method may be inappropriate for studies of institutional success, because this method - vastly overrated in my opinion - can do little more than demonstrate that a variable is insufficient to produce a particular outcome. In light of the above comments, this requires controlling for straw man hypotheses, like the idea that development of a modern industrial economy is sufficient to produce a strong environmental regulatory agency.

Finally, the observation that there are more causes of institutional failure than success may account for the greater role that human agency plays in public administration research - a field highly attuned to sources of success - compared to other areas of social science. What variable could possibly guard against so

---

<sup>11</sup> We must further distinguish between the stated and actual goals of the institution (Steinberg, 1998). The EPA is as committed to following the Administrative Procedures Act as it is to implementing the Clean Air Act, but only the latter makes it onto the agency's press releases. Society wants EPA to follow both; Americans want both clean air and clean procurement systems. We may wish to evaluate EPA performance on clean air only, but we should bear in mind that "failure" to reach an environmental goal may be the result of overzealous pursuit of another legitimate goal. We may need to give up some of the latter in order to achieve more of the former (see Wilson, 1989).



many potential causes of failure as to produce a successful outcome? Such a variable would need to be constantly on the lookout for pitfalls and looming dangers. Policy entrepreneurs, particularly "fixers" (Mazmanian and Sabatier, 1989), would seem to fit the bill. One of the greatest contributions of studies of policy failure (Ascher, 1999; Peluso, 1992) will be to elucidate the many and varied challenges such entrepreneurs will face.

## References

- Anderson, Charles W., The Logic of Public Problems: Evaluation in Comparative Policy Research, in Douglas E. Ashford (ed.), *Comparing Public Policies: New Concepts and Methods*, Sage Publications, Beverly Hills, 1978.
- Ascher, William, *Why Governments Waste Natural Resources: Policy Failures in Developing Countries*, Johns Hopkins University Press, Baltimore, MD, 1999.
- Bardach, Eugene. *Getting Agencies to Work Together: The Practice and Theory of Managerial Craftsmanship*, Brookings Institution Press, Washington, DC, 1998.
- Barzelay, Michael with the collaboration of Babak J. Armajani, *Breaking Through Bureaucracy: A New Vision for Managing in Government*, University of California Press, Berkeley, 1992.
- Baumgartner, Frank and Bryan D. Jones, *Agendas and Instability in American Politics*, University of Chicago Press, Chicago, 1993.
- Bennett, Andrew, *Causal Inference in Case Studies: From Mill's Methods to Causal Mechanisms*, Presented at the American Political Science Association Conference, Atlanta, 1999.
- Berry, Jeffrey M. (1981) Beyond Citizen Participation: Effective Advocacy Before Administrative Agencies, *The Journal of Applied Behavioral Science* 17(4):463-477.
- Castles, Frances G., Introduction: Puzzles of Political Economy, in Castles (ed.), *The Comparative History of Public Policy*, Oxford University Press, New York, 1989.
- Chayes, Abram and Antonia Handler Chayes, *The New Sovereignty: Compliance with International Regulatory Agreements*, Harvard University Press, Cambridge, MA, 1995.
- Dalton, Russel J., *The Green Rainbow: Environmental Groups in Western Europe*, Yale University Press, New Haven, 1994.
- George, Alexander and Andrew Bennett, *Case Studies and Theory Development*, forthcoming.
- Haas, Peter M., Robert O. Keohane, and Marc A. Levy (eds.), *Institutions for the Earth: Sources of Effective International Environmental Protection*, The MIT Press, Cambridge, MA, 1993.
- Hirschman, Albert O., *Essays in Trespassing: Economics to Politics and Beyond*, Cambridge University Press, Cambridge, 1981.

- Margaret E. Keck and Kathryn Sikkink, *Activists Beyond Borders: Advocacy Networks in International Politics*, Cornell University Press, Ithaca, 1998.
- Mayoux, Linda (1995) Beyond Naivety: Women, Gender Inequality and Participatory Development, *Development and Change* 26:235-258.
- Mazmanian, Daniel and David Morell, *Beyond Superfailure: America's Toxics Policy for the 1990s*, Westview Press, Boulder, CO, 1992.
- Mazmanian, Daniel A., and Paul A. Sabatier, *Implementation and Public Policy*, University Press of America, New York, 1989.
- Mitchell, Ronald B., *Intentional Oil Pollution at Sea: Environmental Policy and Treaty Compliance*, The MIT Press, Cambridge, MA, 1994.
- Mitchell, Ronald and Thomas Bernauer (1998) Empirical Research on International Environmental Policy: Designing Qualitative Case Studies, *Journal of Environment & Development* 7(1):4-31.
- Nelson, Richard, *The Moon and the Ghetto*, Norton, New York, 1977.
- Ostrom, Elinor, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, New York, 1990.
- Peluso, Nancy Lee, *Rich Forests, Poor People: Resource Control and Resistance in Java*, University of California Press, Berkeley, 1992.
- Press, Daniel (1998) Local Environmental Policy Capacity: A Framework for Research, *Natural Resources Journal* 38:29-52.
- Przeworski, Adam and Henry Tenue, *The Logic of Comparative Social Inquiry*, John Wiley & Sons, New York, 1970.
- Putnam, Robert D., with Robert Leonardi and Raffaella Nanetti, *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton University Press, Princeton, NJ, 1993.
- Sabatier, Paul (1991) Toward Better Theories of the Policy Process, *PS: Political Science & Politics*, June:147-156.
- Steinberg, Paul F. (1998) Consensus by Design, Policy by Default: Implementing the Convention on Biological Diversity, *Society and Natural Resources* 11(4):375-385.
- Steinberg, Paul F., *Environmental Leadership in Developing Countries: Transnational Relations and Biodiversity Policy in Costa Rica and Bolivia*, forthcoming.
- Wilson, James Q., *Bureaucracy: What Government Agencies Do and Why They Do It*, BasicBooks/Harper Collins, 1989.
- Young, Oran R., *International Governance: Protecting the Environment in a Stateless Society*, Cornell University Press, Ithaca, 1994.