Credibility and Relevance in Environmental Policy: Measuring Strategies and Performance among Science Assessment Organizations

Ann C. Keller
University of California, Berkeley

ABSTRACT

Organizations that provide scientific information to policy makers face the difficult challenge of maintaining scientific credibility while establishing their political relevance. A growing body of research examines how assessment organizations meet the potentially competing expectations of science and policy communities. However, existing research has failed to produce generalizable findings. This study draws together theoretical approaches in science studies and organization theory to develop a framework that allows for a comparative analysis of multiple cases. The study compares the organizational strategies of the National Research Council, the National Acid Precipitation Assessment Program, and the Intergovernmental Panel on Climate Change. Comparisons among the organizations are made using independent measures of credibility and political relevance. The evidence suggests that organizational strategies do impact assessment effectiveness and that it is possible for organizations to simultaneously achieve scientific credibility and political relevance.

INTRODUCTION

Emerging in tandem with environmentalism in advanced industrialized countries in the 1960s was a novel political form—the science assessment, a document written to provide decision makers and the public with up-to-date scientific information about pressing environmental policy problems. Since the 1972 United Nations Conference on the Human Environment, political elites have increasingly turned to science assessments as a tool that can be used in the development of cooperative agreements of both regional and international scope. As such regimes implicate larger and larger portions of state and even global economies, science assessments take on greater potential significance and, with that, controversy.

All science assessment organizations that address environmental policy issues face the difficult task of producing credible science in a highly charged political environment. When political and economic interests are at stake, scientific credibility frequently rests on the

Address correspondence to the author at annk@berkeley.edu.

perception that an assessment is politically neutral. At the same time, for an assessment to be relevant to decision making, the assessment must speak to policy makers' questions and concerns. This creates a difficult balancing act for organizations producing science assessments in that they must attempt to be both credible and relevant in order to link scientific information to policy decision making.

This study addresses the question of how science assessment organizations confront the difficulties of operating between science and policy communities and what makes for successful outcomes. This research contributes to a growing literature on science assessment organizations by comparing strategies and outcomes associated with three such groups: The National Research Council (NRC), the National Acid Precipitation Assessment Program (NAPAP), and the Intergovernmental Panel on Climate Change (IPCC). This research strongly questions the commonly held view that science assessment organizations can pursue political relevance only at the expense of scientific credibility and vice versa.

Existing Research

Scholarship in science studies has tackled the question of how best to conduct science assessments, in part, through the lenses of boundary work, boundary organizations, and coproduction. Scholars advanced the concept of boundary work to explain how distinctions between what is "scientific" and "nonscientific" are established, contested, and maintained (Gieryn 1983, 1995, 1999; Jasanoff 1990, 2004a, 2004b; Jasanoff and Wynne 1998; Star and Griesemer 1989). This approach rests on the constructivist view that scientific endeavors are fundamentally social and that knowledge production in the sciences relies, in part, on values, norms, and culture.[1] This creates opportunities for actors to debate any particular effort to draw a sharp distinction between the realms of science and nonscience—something that occurs frequently in policy-making settings.

Responding to Gieryn's (1995, 406) claim that any given settlement concerning the science-policy boundary will be temporary, a number of scholars have looked for mechanisms that stabilize the science-policy boundary and, therefore, enable more productive interactions between scientific and policy communities. This has led to two currents in the science studies literature. The first follows Guston's (1999, 2000, 2001) research on "boundary organizations" that act as intermediaries between scientific and political communities by balancing the demands from each. The second current advances the concept of "coproduction" which asserts that natural and social orders reinforce one another such that neither is more fundamental in stabilizing meaning and norms in society (Jasanoff 2004a, 2004b; Jasanoff and Wynne 1998; Miller 2004). This approach breaks with the more traditional view of science and policy that sees each as having the potential to undermine the norms of the other—i.e., science can erode democratic norms by introducing technocracy, whereas politics can introduce bias that erases scientific objectivity. Coproduction sensitizes researchers to the potential for science and politics to interact productively and questions the assumption that the two worlds are naturally in tension.[2] Although this view is

---

1   For examples of work highlighting the social processes that support the creation of scientific knowledge, see Merton (1968), Mahoney (1977), Latour and Woolgar (1979), Collins (1981), and S. Cole, J. Cole, and Simon (1981).
2   For an elegant application of the coproduction framework, see Miller (2004).

critical of the tendency to treat science and politics as fundamentally distinct and, therefore, incompatible, the idea of two communities operating under different norms persists in treatments of the role of science in policy making.

Such perspectives regarding the tension between the expectations of science versus policy communities are reinforced by empirical research that argues for maintaining a separation between science and politics in order to produce successful scientific assessments for use by decision makers. Herrick and Jamieson (1995), in their review of acid rain assessments, argue for lowering expectations about how scientific an assessment can actually be, suggesting that credibility should be knowingly sacrificed for the sake of political relevance. Arguing from the opposite direction, Agrawala (1998a, 617, emphasis in original), in his study of the history of the IPCC, asks how the organization, born out of pure politics, was able to maintain credibility among "high caliber *scientists*" and argues that the answer lies in the fact that the IPCC avoids making policy recommendations and has refused to articulate a scientific research agenda on climate change (Agrawala 1998b, 638–39). The implication here is that distance from politics can preserve an assessment's credibility. Similarly, Hilgartner's (2000) extensive study of the NRC suggests that part of the organization's ability to maintain its credibility stems from its ability to keep the assessment process shielded from public view.

Popular views regarding the role of science in policy making rest on the notion of science as useful in the policy context only when it is objective.[3] This perspective draws from an ideal of science as emerging from the lab with immediate practical relevance which neither requires nor allows for interpretation. The oversimplification that this perspective entails becomes apparent when scientists and policy makers alike find that scientific information injected into a policy debate rarely produces consensus (Mazur 1973, 1981; Nelkin 1979). Moreover, scientists who work in domains where scientists and policy makers frequently interact must make conscious decisions about how involved to become in policy circles given that such involvement can negatively impact their professional reputations and careers (Alpert and Keller 2003; Takacs 1996). Moreover, policy makers often pressure scientists who act as advisors to take policy positions only to discount those same scientists as credible informants in future rounds of policy making (Keller 2001). In these situations, the trade-offs between relevance and credibility may be very real.

Although the tension between relevance and credibility remains a central theme in the current literature on science assessments, a number of studies suggest that organizations are able to strike a workable balance between the demands of their two most prominent audiences (Guston 2000; Miller 2001; Parson 2003). Not all organizations attempting to operate across the science-policy boundary, however, are immediately successful. Agrawala, Broad, and Guston (2001) review the efforts of the International Research Institute for Climate Prediction (IRI) to provide climate predictions to decision makers and find that the political contexts in which such information is used makes it difficult for IRI to act in ways that do not privilege certain stakeholders over others or run afoul of national sovereignty (467–9).

To study how assessment organizations perform, a number of studies begin with successful cases and highlight potential sources of that success. These include frequent

---

3   Such views are frequently portrayed in newspaper treatments of the trials and tribulations of science assessment organizations. See, for example, Duggan (2007), Henderson (2006), and Knickerbocker (2007).

interactions between scientists and policy makers (Farrell and Keating 2006; Selin 2006; VanDeveer 2006), stakeholder involvement in the assessment process (Cash 2000; Kwiatkowski and Ooi 2003; Martello and Iles 2006; Miller 2006; Van der Sluijs 2002), and initial buy-in from end users (Lund 2006). Other studies, however, highlight similar strategies in cases where success has been more elusive (Beck, Asenova, and Dickson 2005; Paterson and Andrews 1995). Such conflicting results make generalizations regarding how to conduct science assessments hard to come by. Further, because most studies treat single cases, generalizability problems persist.

The volume edited by Farrell and Jager (2006) and a subsequent work edited by Mitchell et al. (2006) attempt to address this issue by setting out a framework drawn from insights from the Global Environmental Assessment (GEA) Project (Clark and Dickson 1999) that emphasizes the process of producing assessments over the actual reports themselves and defines assessment success in terms of salience, legitimacy, and credibility. Contributors to the volume edited by Farrell and Jager (2006) focus on elements of organizational design that might explain variations in assessment salience, legitimacy, or credibility. Those contributing to the Mitchell et al. (2006) volume consider either context or assessment design as possible explanatory factors in shaping assessment success. Owing to the diversity of potential explanatory variables considered by each contributor, the collected studies lead to a set of propositions about assessment effectiveness rather than formal conclusions about the impacts of design or context.[4]

This study attempts to move beyond hypothesis generation by creating a framework that allows for systematic evaluation of design features of three assessment organizations and comparison of those features with independent measures of assessment success. Using this framework, this study addresses the question of whether or not scientific credibility and political relevance are necessarily in tension in producing science assessments.

Organization Theory and the Science-Policy Boundary

The literature on science assessment organizations naturally focuses on the boundary between science and politics. Driving this interest is the question of how organizations manage this boundary where definitions of science and policy are often determined by the actors involved and are rarely stable over time or across settings (Jasanoff 1990). Organization theory, not surprisingly, has a long history of studying the dynamics of organizations and their boundaries. Given this rich resource, it is somewhat surprising that the science assessment literature has developed without drawing on insights and tools from organization theory.[5]

Organization theorists, in approaching the concept of an organization's boundaries, argue that every organization must engage in some form of "boundary setting" and

---

4   Notably, these volumes argue that the elements necessary for assessment success—salience, legitimacy, and credibility, according to the GEA framework—are in tension with one another such that increasing one can erode another. For example, Clark, Mitchell, and Cash (2006, 16) write, ". . . these three attributions of salience, credibility, and legitimacy were interconnected . . . in the sense that procedures intended to foster one often undermined another." Similarly, Farrell, Jager, and VanDeveer (2006, 10) argue, "Adding to the challenge, these three determinants [salience, legitimacy, and credibility] are often in tension, because the easiest ways of enhancing any single attribute almost invariably cause declines in another".

5   Though there is little explicit reference to organization theory in the science assessment literature, the concept of "hybrid management" (Miller 2001) contains ideas that are consistent with the notions of linking and buffering.

"boundary spanning." Setting boundaries helps delimit the organization in order to specify organizational goals, roles, and tasks to be performed within the organization (Weber 1946). Boundaries are necessary in order to clarify who will define the organization's mission and set out the mechanisms for achieving organizational goals (Barnard 1938). One mechanism of setting boundaries is to define criteria for organizational membership which can, for example, be limited by specific recruitment criteria (Udy 1962; Weber 1946). In addition, a key strategy organizations use for maintaining boundaries, once set, are "buffering" techniques to insulate the central work of the organization from external influences and surprises (Meyer and Rowan 1977; Meyer, Richard Scott, and Deal 1981; Thompson 1967). This helps clarify the distinction between the organization and its environment, define expectations and goals within the organization, and promote stable production.

At the same time, organizational boundaries are porous and organizations depend on their external environment for any number of resources that are crucial to the organization's work. This requires organizations to set up predictable exchanges with actors in their environment in order to secure necessary resources for achieving goals. Organization theorists call creating and maintaining such links to the external boundary spanning. Organizations will coordinate their activities with outside actors when there is a mutual dependence between the organization and external actors (Pfeffer and Salancik 1978). Another form of boundary spanning involves creating structures or processes inside the organization that conform to important complexities in the organization's environment (Buckley 1967; Lawrence and Lorsch 1967). In addition, organizations mimic norms and strategies of other organizations in the same domain, promoting legitimacy with outside actors by demonstrating that organization is similar to other, legitimate actors (DiMaggio and Powell 1983; Meyer and Rowan 1977).

In order to understand how organizations can both set boundaries and span them, a number of organization theorists have highlighted the practice of "loose coupling" in organizations. Loose coupling occurs when organizations establish units or sets of practices within the organization that are relatively independent from other units (Ashby 1968; Buckley 1967). Loose coupling allows organizations to pursue different strategies across sites or levels in the organization (Meyer and Rowan 1977; Meyer, Richard Scott, and Deal 1981; Thompson 1967). Although loose coupling can pose management and coordination challenges for an organization (Lawrence and Lorsch 1967), it can also be a source of dynamism and adaptation in the organization (Cyert and March 1963; Pfeffer and Salancik 1978; Weick 1976).

Organization theory, therefore, predicts that organizations will simultaneously attempt to protect their central work, the so-called "technical core" (Thompson 1967), and establish external links to manage the uncertainties that come with organizational dependence on outside actors and resources. This approach to understanding organizational behavior is clearly relevant to science assessment organizations given their preoccupation with protecting their scientific work from bias while aiming for sufficient ties to decision-making communities to ensure favorable reception for assessment findings. The idea of loose-coupling suggests that common organizational strategies can be used to manage these twin goals.

Drawing on this perspective offered by organization theory, I argue that science assessment organizations pursue buffering strategies to protect their scientific work from bias and politicization. At the same time, they pursue "linking" strategies to maintain ties to their political constituents who might rely on the outputs of science assessments to make

policy decisions.[6] By practicing loose coupling, an organization can simultaneously employ buffering and linking strategies.

Research Hypotheses

This study examines evidence of buffering and linking efforts among science assessment organizations as a means for negotiating the science-policy boundary and sheds light on (a) the extent to which such strategies are used and (b) their relative effectiveness. Several specific hypotheses are considered. First, the existing literature on science assessment organizations poses the question of whether organizational design rather than context is an important factor in shaping assessment effectiveness. Farrell and Jager (2006), in their edited volume, take as a starting assumption that organizational design is an important determinant of assessment success. Mitchell et al., on the other hand, leave open the question of whether design or context is more important in explaining outcomes (Clark, Mitchell, and Cash 2006). Like Farrell and Jager, this study assumes that better organizational design can lead to more robust and well-received assessments. The first hypothesis to be tested, then, is the following:

H[1]   Organizational design has a measurable impact on assessment effectiveness that outweighs context in explaining assessment outcomes.

The second hypothesis derives from the consistency of views in the current assessment literature and in more popular treatments of the science-policy boundary that organizations that manage science and political audiences will have difficulty producing both credibility and relevance, together. Both organization theory—through the mechanism of loose coupling—and the coproduction framework argue that organizations can pursue both simultaneously. The second hypothesis to be tested is

H[2]   Science assessment organizations that employ both buffering and linking strategies can accomplish the twin goals of credibility and relevance without having to sacrifice one in pursuit of the other.

This leads to a third set of considerations. Organization theorists have, by and large, treated buffering and linking (or boundary spanning) as distinct endeavors within the organization. This set of expectations maps nicely on to the conventional way of thinking about the science assessment organization's dilemma—that scientific elements of assessment must be insulated against political bias (buffering) while those same assessments must respond to key concerns of end users, which implies some degree of interaction with and responsiveness to policy audiences (linking). This leads to the following, related hypothesis:

H[3a]   Organizations seeking to protect their scientific credibility will pursue buffering strategies.

H[3b]   Organizations seeking to enhance their political relevance will pursue linking strategies.

---

6   Here, I adopt the term "linking" to emphasize the specific goals of boundary spanning efforts on the part of science assessment organizations.

Although expectations about a lack of credibility-relevance trade-off are shared between organization theory and the coproduction framework, the coproduction approach would not insist that separate activities produce credibility and relevance. Rather, any number of organizational activities might coproduce scientific and political orders. This leads to a fourth hypothesis:

H$_4$ Organizational strategies exist that simultaneously bolster both scientific credibility and political relevance.

Research Design and Case Selection

To address the question of a credibility-relevance trade-off, the study compares the strategies of three organizations that produce scientific assessments of acid rain and climate change: the NAPAP, the IPCC, and the NRC which has produced reports on both topics. These organizations are evaluated through an analysis of the strategies they employ to produce reports that contain credible science and the strategies they use to ensure that potential end users are receptive to the organization's work. Organizational procedures for producing assessments are divided into those designed to protect science from politicization, "buffering strategies," and those intended to link assessment organizations to their political constituencies, or "linking strategies." These strategies are then compared against independent measures of each organization's success with respect to political relevance and maintaining a reputation for producing credible science.

The cases selected here, NAPAP, the NRC, and the IPCC, are chosen for the potential to compare three distinct approaches to scientific assessment in the environmental policy domain. Specifically, each organization exhibits differences in its buffering and linking strategies. In addition, anecdotal evidence collected at the outset of the research suggested that these organizations achieved different levels of success in producing reports that met the needs of their intended audiences. These cases, thus, created the potential for adopting a research design that would highlight systematic differences in independent and dependent variables that might explain observed outcomes.

An ideal research design would compare cases that were similar on all dimensions excepting the independent variables—buffering and linking strategies, here—and the dependent variables—scientific credibility and political relevance, in this case (Przeworski and Teune 1970). Although the organizations considered do address two distinct environmental policy issues, the two issues have many similar properties. Both acid rain and climate change pose threats to the natural and human environment. Both issues are characterized by well-understood mechanisms. These are, respectively, long-range transport and deposition of pollutants and climate forcing as a function of increasing greenhouse gases in the atmosphere. In addition, both issues have significant uncertainties associated with predicting impacts and the effectiveness of proposed mitigation strategies. The buffering capacities of lakes, rivers, and soils make it difficult to predict how quickly ecosystems might succumb to the ill effects of acid rain or how much pollution reduction is required for ecosystem recovery. Manifold uncertainties in the specific responses of the climate to greenhouse forcing—for example, the role of clouds—make it difficult to predict actual regional impacts like drought, severe weather events, and sea level rise. Both issues spent more than 10 years on the formal US policy agenda, and both issues involved transboundary environmental effects which drew the United States into international

negotiations. In addition, both issues challenged the economic positions of powerful interest groups who had incentives to discount claims of severe impacts.

The difference between the two issues, however, must also be noted. Because climate change is global in scope and control of greenhouse gas emissions implies a major shift away from dependence on fossil fuels, one would expect political controversy about climate change to be high. Although acid rain produced its share of political controversy, debate about acid rain was carried out in regional contexts and threatened to impose costs on a considerably smaller fraction of a given country's energy production.[7] Because of this difference, it may be more difficult for science assessments addressing climate change to appear credible given that the issue implicates a larger portion of the US economy and is likely to spur greater political resistance.[8]

This potential contextual difference between acid rain and climate change is joined by several features of the assessment organizations that might impact the analysis. For example, although NAPAP and the NRC both focus their assessments on decision makers in the United States, NAPAP is a government-created institution and received appropriations from Congress for its work. The NRC is quasi-governmental and has a private operating budget, though it relies heavily on government grants and contracts for many of its operations. The IPCC assessments have a much broader audience than NAPAP or the NRC in that they are oriented toward an international audience. This includes, specifically, the governments who participate in IPCC Plenary meetings and, more generally, anyone who is interested in the current state of climate science.[9] Thus, the United States is one target government among many for the IPCC. The IPCC operating budget comes from the IPCC Trust Fund that is supported by World Meteorological Organization (WMO) and United Nations Environment Programme (UNEP) member country contributions. The NRC, therefore, may be more independent of its policy audience than NAPAP and the IPCC in that it is not dependent on them for its budget.[10] In addition, The NRC has been in operating since 1918, providing a track record in performing assessments that predates the specific issues of acid rain and climate change. NAPAP and the IPCC were both created after 1980 in response to concerns about the two issues. As a consequence, they develop their reputations as they go. Another notable contextual difference is that NAPAP carried out the bulk of its work just prior to the advent of the internet and did not have this technological outlet at its disposal in communicating with its stakeholders. The NRC and the IPCC have both made significant use of the internet beginning in the 1990s.

Given that the research design includes cases that are not similar across all potential dimensions of interest, the approach taken here weighs how these contextual factors themselves might impact the study's findings. Where possible, the study will rely on "within-case

---

7   Because of the structure of the Senate, regional conflicts in United States politics can easily produce congressional stalemate. Although the US Congress did eventually pass acid rain legislation, the intensity of the struggle over the issue, which lasted more than a decade, should not be overlooked.

8   A number of studies show that politically controversial topics generate controversy about scientific research methods and findings much more so than topics about which there is general agreement. See, for example, Collingridge and Reeve (1986), Ezrahi (1980), Nelkin (1979), and Mazur (1973, 1981).

9   Although the IPCC may think of its audience as including the general public, the reports are clearly written with policy makers in mind as demonstrated by the "Summary for Policy Makers" included in each of its assessments.

10   The NRC does not receive direct appropriations from Congress for its work. It relies on a combination of public and private resources including federal agency grants, grants from foundations, and its own endowment. The NRC also relies heavily on the donated time of the scientists who participate on report committees.

comparisons" to gain some leverage over the contextual factors that are not well controlled by the case selection (Przeworski and Teune 1970). For example, the inclusion of the NRC in the study offers partial control over any "issue-effect" stemming from the comparison of acid rain and climate change. The NRC, because it produces reports for both acid rain and climate change and uses the same organizational strategies in doing so, provides some insight into whether one of the issues poses greater credibility or relevance challenges than the other. Another opportunity for within-case comparison comes from NAPAP owing to the fact that it changed organizational strategy after producing its first assessment. By making use of within-case comparisons and by noting and accounting for contextual differences across the cases, the study will highlight potential competing explanations.

METHODOLOGY

In order to capture the strategies that each organization uses in producing reports and measure their respective levels of success, this study relies on several measures, focusing on the structures and procedures that each organization employs and cataloging these as either buffering or linking strategies. In some cases, a single organizational strategy will satisfy the requirements of both buffering and linking. To compare the use of buffering and linking strategies across the three organizations, the study uses a measure that attempts to capture the level at which a particular strategy is used by the assessment organization. The scale is ordinal, where "not used" = 0, "low-level use" = 1, and "high level use" = 2. Judgments about whether the strategy is used or not used are fairly straightforward. The distinction between high- and low-level use, however, is more subjective. The codings are relative such that receiving a "high-level use" score suggests only that one organization uses a given strategy more intensively than one or both of the other organizations in the study.

The identified buffering strategies are combined into a single buffering and single linking score for each organization. As a consequence, the study does not test the effect of any particular organizational strategy on assessment effectiveness. Rather, the approach calls attention the combination of strategies an organization uses to manage its two audiences and relies on theories about the different goals associated with buffering and linking to support the logic of combining strategies into their respective buffering and linking categories.[11]

One important point to note regarding the treatment of each organization's strategies is that this study is based on publicly available information about the respective organizations. As a consequence, this analysis might not capture all the relevant practices each organization uses to produce science assessments.[12] However, it makes use of the same data that are available to stakeholders in making their own judgments about assessment credibility and relevance. Earlier work on environmental assessments argues strongly that process rather than product contributes to assessment effectiveness (Clark and Dickson 1999; Farrell and Jager 2006; Miller et al. 1997; Mitchell et al. 2006). This suggests that hidden practices, though they might impact the final product, play less of a role than those processes that can be assessed by stakeholders.

---

11   The strategy of "data reduction," appropriate when treating a small number of cases, combines variables of interest into a smaller set of meaningful variables to be tested in the study (Collier 1993, 112).

12   See Hilgartner (2000) for an example where formal organizational procedures are, at best, an incomplete account of organizational activities.

In order to measure political relevance and assessment credibility, the study makes use of three independent measures. Political relevance is defined as the organization's ability to capture the attention of decision makers and the public. This is measured in two ways: (1) by the number of congressional hearings that devote attention to the work of a science assessment organization and (2) by the total number of newspaper stories written about each organization in major, English-language newspapers.[13]

In order to measure an assessment's credibility, defined as an organization's capacity for maintaining a reputation as an authoritative source for scientific information about acid rain or climate change, the study turns to the content of print media treatments of science assessment organizations.[14] Major newspapers tend to cover activities of assessment organizations—new reports, changes in personnel, etc.—in addition to related congressional activity and public debates. Newspapers, therefore, can track emerging scientific information, can offer a window on congressional attention and temperament in response to assessment activities, and can include information about debates that emerge around science assessments.

In addition to providing a window on events surrounding science assessments as they occur, research suggests that the mass media play a role in shaping public opinion. Although the correspondence between exposure to media stories and comprehension of those stories can be fairly loose,[15] research demonstrates that the quantity and quality of stories in the media affects public perceptions about which issues are most important.[16] This gives rise to the possibility that the quantity and type of coverage the media gives to science assessments can impact public views about the importance and quality of those assessments.

Studies of the role of the media in shaping public opinion about salient policy issues struggle with how to model the media. Is the media the primary actor in shaping public agendas? Do the media merely respond to elite cues about which issues to cover? Or, is the media trying to cover issues they think will sell papers, that is, anticipating public interest? For the purposes of this study, we can assume that all three models describe some aspect of the media's role. First, the media have an independent effect on the attention its readership pays to science assessments. Second, the media transmit information about political elites' interest in science assessments about acid rain and climate change. Finally, editors make strategic decisions about what stories to cover and what prominence to give those stories based on their judgment about what sells papers. If all three models describe some aspect of the role of the media, then newspaper stories act as an important conduit for opinions about science assessment organizations. If science assessment organizations achieve credibility or are viewed as controversial, we would expect the media to pick up on and convey these views in their coverage of science assessment organizations. Through this process, the media might play an independent role in shaping perceptions about assessment credibility.

13   For the NRC, only those newspaper stories that focus on the organization's work on acid rain or climate change are counted. The limitations of using English-language newspapers to estimate political relevance are discussed in the analysis.

14   The approach used in this study is similar to that used by Weart (1988) and by Baumgartner and Jones (1993).

15   For research on the differences between exposure to a media message and actual reception of that message, see, for example, Neuman (1976) and Neuman, Just, and Crigler (1992).

16   For research on media effects on public opinion, see Iyengar and Kinder (1987), Zaller (1999), and Brody (1991).

Credibility, then, is measured through content analysis of newspaper articles that mention science assessments of NAPAP, the IPCC, or the NRC.[17] The credibility measure divides newspaper stories into those that treat the science assessment organization as a reliable or authoritative source of scientific information and those that suggest the science assessment organization, its members, reports, or procedures are the subject of controversy. This approach attempts to code the overall framing of the assessment organization in the newspaper article and asks, on balance, whether the article portrays the assessment organization as a reliable source for scientific information.[18] An assessment's "credibility" score is the percentage of stories that are coded as "authoritative."[19]

Science Assessment Strategies for Three Cases

This section provides a brief overview of the organizational strategies employed by NAPAP, the IPCC, and the NRC in producing science assessments on acid rain and climate change.[20] The NAPAP was created by legislation that was signed into law by President Carter in 1980. NAPAP, on paper, looked like its governance was divided among scientists and agency heads. In practice, scientists dominated the production of its reports, though agency heads did have some influence over the distribution of research dollars allocated under NAPAP. To ensure the quality of its reports, NAPAP engaged in two levels of peer review: (1) including in its reports only scientific findings that had been published in peer-reviewed journals and (2) sending out its report chapters for peer review. An exception to its peer review process led to heavy criticism when NAPAP's director penned the executive summary of its 1987 Interim Report without submitting it for review.[21] Following the controversy, NAPAP reorganized with a new director, made a stronger effort to keep decision makers and the public informed of its plans for its final report, and created opportunities for decision makers to comment on circulated drafts. In spite of these efforts to make stronger links to its policy-making audience, its final report was published without an executive summary and was criticized as being irrelevant to the policy-making process (Herrick and Jamieson 1995; Roberts 1991; Rubin, Lave, and Morgan 1991).

The IPCC, created in 1988 by the United Nations Environment Programme and the World Meteorological Organization, adopted a range of practices in preparing its reports that were intended to ensure its ability to provide credible science to decision makers. First, its major reports are published in three separate volumes that cover, respectively, science,

---

17 Print stories were found using the LexisNexis database of English language newspapers. Searches were conducted among "Major Newspapers" in the "General News" category of the database for all available dates. For a more complete description of the methods used to conduct the content analysis, contact the author.

18 For a description of the coding scheme used for this analysis, contact the author.

19 Because controversy is "newsworthy," this approach might overestimate the level of controversy that truly exists around science assessment organizations. However, it is unlikely that this effect of "newsworthiness" would apply more strongly to one science assessment organization than another. The credibility scores, therefore, should be understood as relative measures among the organizations rather than an objective measure of the amount of controversy that arises around each one.

20 Because of restrictions on article length, this section is substantially redacted. For access to the original paper, contact the author. For single case-study treatments of NAPAP, see Herrick and Jamieson (1995) and Herrick (2000). For an incredibly detailed look at the NRC, see Hilgartner (2000). For detailed treatments of the IPCC, see Agrawala (1998a, 1998b), Edwards and Schneider (2001), Miller (2004, 2006), and Siebenhuner (2006).

21 The details of the director's role in producing the executive summary are discussed in a congressional hearing held after the Interim Assessment (1987) fell under heavy criticism from the scientific community (HCSST 1988).

impacts, and policy options.[22] These volumes are produced by working groups that are located in separate countries. This approach allows the IPCC to produce a summary of the science of climate change that stands alone—a buffering strategy, although it allows the IPCC to weigh in on expected impacts and policy options—a linking strategy.[23]

The IPCC's review process goes considerably beyond that of standard scientific peer review. Notably, the IPCC takes pains to gather a geographically diverse group of scientists into the production of its reports. It does this to counter arguments that its findings serve a given country or region to the detriment of others.[24] Like NAPAP, the IPCC draws from the published scientific literature. It then subjects its reports to peer review but also invites comments on its reports during plenary sessions held by each working group that are attended by representatives of member states and of NGOs participating in the UN Framework Convention on Climate Change (FCCC). In addition, the IPCC invites its end users to participate in plenary sessions to approve, line-by-line, its summaries for policy makers. Decision makers, through the UN FCCC, can request that the IPCC produce reports on topics of specific concern to the Convention. In this way, decision makers have some influence over the IPCC's report agenda.

The NRC is the "operating arm" of the National Academies of Science and Engineering and was created in 1918 for the purpose of providing policy advice to the government on matters of science. The NRC responds to external requests for reports but ultimately controls its own report agenda.[25] Although the organization engages in traditional peer review practices, it stands out in that it convenes a new committee for each study it produces. This allows the organization to attempt to: (1) access the most qualified experts on a given topic, (2) create balance on the report committee with respect to major scientific or political cleavages, and (3) maintain a reputation for the NRC that is distinct from any given committee, especially as a committee disbands once its report is completed. Of the three organizations in the study, the NRC has been in existence the longest and, thus, has the most experience in producing scientific reports for policy makers.

FINDINGS

Buffering and Linking Scores

Looking at the strategies across cases yields some interesting patterns (table 1). Four out of nine of the buffering strategies are used by all three organizations and all three receive similar scores across these four strategies. The one exception is that the NRC has more

---

22   For additional information on the IPCC structure and report process, see the organization's Web site: http://www.ipcc.ch/about/index.htm. Also, note that the procedures described here have evolved over time. For analyses of events that have led to reforms in the IPCC's approach, see Edwards and Schneider (2001), who discuss the IPCC response to controversy erupting in 1996, and Miller (2004), who treats reforms put in place in 1991.

23   To manage the complex procedures that govern the IPCC, the organization created Technical Support Units (TSUs) to aid each working group. There is very little written about the role of the TSUs; however, Working Group II does include a brief overview of its TSU and states, "Their prime function is to ensure that the work of WGII proceeds according to schedule and in line with the Principles and Procedures governing the work of the IPCC." Miller (2004, 60–1) argues that the TSUs provide a buffer between the experts producing IPCC reports and the stakeholders who try to influence those reports.

24   For a discussion of the shortcomings in the ability of the IPCC's to support participation of scientists from developing countries, see Agrawala (1998b, 632).

25   See http://sites.nationalacademies.org/nrc/PoliciesandProcedures/index.htm for the NRC's own description of its procedures. For a critical analysis of its approach, see Hilgartner (2000).

Table 1
Buffering Strategies

	Pre-1987 NAPAP	Post-1987 NAPAP	NRC	IPCC
Scientists write reports[a]	2	2	2	2
Reliance on published literature[a]	2	2	2	2
Peer review by science experts[a]	2	2	2	2
Scientists control research agenda[a]	1	1	2	1
Peer review of summary for policy makers	0	0	2	2
Intraorganizational buffering	0	0	2	2
Distinct publications for treating science and policy	0	2	0	2
Closed meetings	0	0	2	0
Formal buffering unit with/in organization	0	0	0	2
Total buffering score	**7**	**9**	**14**	**15**

[a]These strategies are found in all three organizations. $0 = $ strategy absent; $1 = $ low-level or symbolic use of strategy; $2 = $ high-level use of strategy.

control over its agenda than the other two in that it ultimately decides whether or not it will produce a report on a given topic. NAPAP and IPCC scientists do exercise influence over the content of reports, but each is also responsive to policy makers in producing reports. Recall that NAPAP reserved a role for agency heads to disburse NAPAP research dollars and that the IPCC responds to research requests from the UN FCCC. Most notable in the buffering scores is the fact that the NRC and the IPCC look fairly similar and they outscore NAPAP, even after NAPAP's 1988 changes. NAPAP's lower score reflects the fact that, when it had a summary for policy makers, it did not subject that summary to peer review. It also reflects the fact that NAPAP did not have any intraorganizational mechanisms for buffering like the NRC's temporary committee structure or the IPCC's three working groups. Both of these strategies potentially deflect controversy by allowing it to adhere to one part of the organization without necessarily casting the entire organization under suspicion.

Turning to linking strategies, these exhibit more variation across the three organizations (table 2). There are nine linking strategies highlighted by this research. Only two of these are shared across organizations. Also notable is the larger difference in linking scores as compared to buffering scores. This spread reflects the fact that the IPCC substantially outscores both the NRC and NAPAP in terms of linking strategies. The IPCC stands out in that it allows decision makers to participate in processes that are reserved for scientists in the other assessment organizations. In particular, the IPCC invites decision makers to nominate scientists to work on its reports, requires their review and acceptance of IPCC reports, and it involves decision makers in a line-by-line approval of its summary for policy makers.

One interesting note of comparison is that the three organizations all vary with respect to transparency. The IPCC process is more open to its end users than that of NAPAP and the NRC. NAPAP did not seem to give transparency much thought until after the 1987 controversy. In response to the fallout, NAPAP publicized its plan for producing its *Integrated Assessment* (1991) and opened a draft report to decision makers and the public for comments. The NRC, on the other hand, although it publicizes its procedures, actively guards its closed-committee approach and only allows scientific review of draft reports.

Combining the buffering and linking scores show that the IPCC has the most elaborated set of strategies and these are evenly divided between buffering and linking. The

Table 2
Linking Strategies

	NAPAP pre-1988	NAPAP Post-1988	NRC	IPCC
Decision-maker input to assessment agenda[a]	1	1	2	2
Transparency of procedure[a]	1	1	1.5 (increase over time)	2
Summary for policy makers	2	0	2	2
Decision-maker review of reports	0	2	0	2
Decision-maker nomination of expert participants	0	0	0	2
Decision-maker participation in summaries for policy makers	0	0	0	2
Decision-maker review of procedure	0	2	0	0
Public review of reports	0	2	0	0
Representative science	0	0	2 (policy balance)	2 (geographic)
Total linking score	**4**	**8**	**7.5**	**14**

[a]These strategies are found in all three organizations. 0 = strategy absent; 1 = low-level or symbolic use of strategy; 2 = high-level use of strategy.

NRC, on the other hand, relies much more on buffering. NAPAP increased both its buffering and linking after the 1987 controversy but still falls behind the other two organizations in terms of its overall use of buffering and linking strategies.

Relevance Indicators

Reviewing the indicators of political and public relevance, the number of congressional hearings in which each organization is mentioned does not demonstrate an obvious pattern (table 3). NAPAP, the IPCC, and NRC's climate change work all achieve similar scores, whereas NRC's acid rain efforts receive less attention. However, all three organizations do appear numerous times on the congressional agenda. Also, congressional attention to the latecomer assessment organizations, because of the time spans involved, is more concentrated (figure 1).

Turning to media attention, one notes that there are far more stories about assessments of climate change than there are of assessments of acid rain (see table 4). Also of interest is the NRC's relative position with respect to media attention for the two issues. Specifically,

Table 3
Number of Hearings that Mention the Science Assessment Organization and Its Work

Organization	Number of Congressional Hearings	Time Frame	Ratio of Hearings by Year
NAPAP	20	8 years (1983–1990)	2.5
NRC acid rain	12	15 years (1976–1990)	0.8
IPCC	20	17 years (1989–2005)	1.2
NRC climate change	19	26 years (1980–2005)	0.7

Figure 1
Congressional Hearings

Congressional Hearings
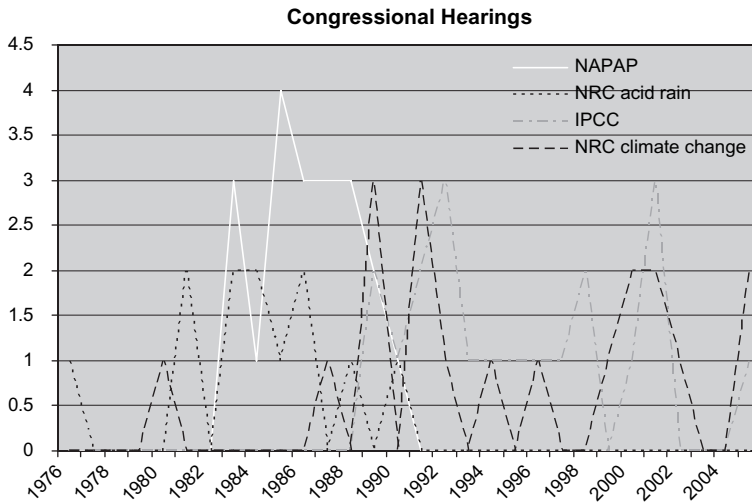
more print media stories mention the NRC's work on acid rain than NAPAP's, whereas the NRC receives far fewer stories on climate change than the IPCC.

An important factor in comparing the media scores, however, is the fact that the IPCC's audience is an international one, whereas NAPAP reports were written for American policy makers. The NRC's reports do attract international attention, though it lacks an organizational structure that guarantees an international audience. To account for this, the study uses the percentage of North American newspaper stories from the IPCC and the NRC climate change samples to estimate the total number of North American stories about each organization in the database (table 4). Although the magnitude of the difference changes when one compares only stories from North American newspapers, the relative positions are the same with the IPCC garnering 40% more stories than its nearest competitor, the NRC for its work on climate change, and outscoring NAPAP by more than a factor of 10 (figure 2).

Credibility Scores

The credibility scores of the different organizations yield considerable variation and provide some basis for comparing the assessment organizations (table 5). NAPAP achieves the lowest credibility score in the group (59%), whereas the NRC achieves the highest scores both for its acid rain and climate change reports (84% and 82%, respectively). The IPCC's credibility score falls slightly behind that of the NRC's with respect to climate change (77% as compared to the NRC's 82%). Another interesting finding is the NAPAP's credibility score, when broken down into the time periods before and after its 1988 reforms, changes dramatically going from a low of 29% to the much higher 71%. The results reported in table 6 show that credibility scores are robust with respect to regional variation and foreign

**Table 4**
Number of Print Media Stories

Organization	Number of Print Media Stories	Number of Stories Coded	Percentage of Coded Stories from North American Papers	Number of Stories in North American Papers	Number of Front-page Stories (# in North American Papers)	Percentage Front-page Stories (% in North American Papers)
NAPAP	70	70	95.7	67	4 (4)	5.7 (5.7)
NRC acid rain	105	105	96.2	101	9 (9)	8.6 (8.6)
IPCC	1956[a]	97	49.5	968[b]	6 (2)	6.2 (2.1)
NRC climate	759[a]	98	79.6	594[b]	8 (6)	8.2 (6.1)

[a]These numbers are estimates based on the true number of stories in the random samples.
[b]These numbers are estimates based on the percentage of stories in the sample that come from North American newspapers.

Figure 2
Newspaper Stories
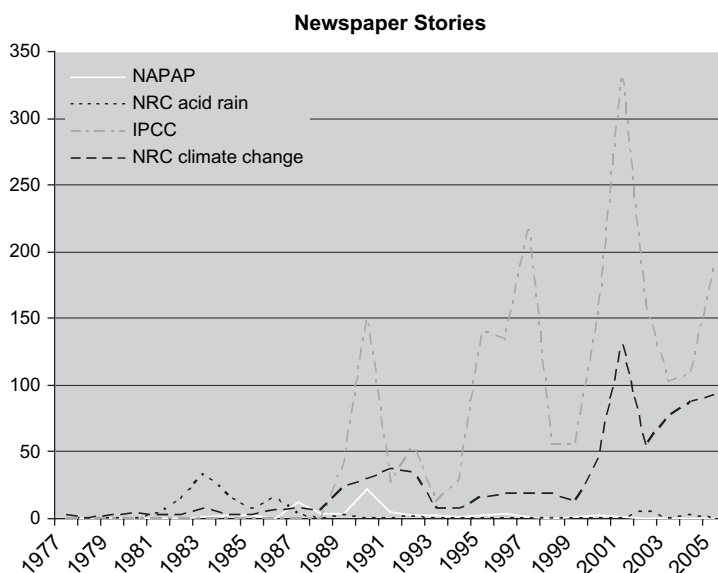
publication, even for the IPCC which gets just over half of its coverage from newspapers outside Canada and the United States.

Considering Measurement Validity

Thus far in studies of assessment organizations, most scholars have left implicit the indicators they use for the concepts of interest in their study.[26] This study represents a departure

Table 5
Credibility Scores[a]

	Number of Stories Found in LexisNexis Search	Number of True Stories	Number of Stories Coded	Percent Credible
NAPAP (all years)	70	70	70	58.6
NAPAP (1983–1988)	21	21	21	28.6
NAPAP (1989–2001)	49	49	49	71.4
NRC acid rain	105	105	105	83.8
IPCC	2445	1956[b]	97	77.3
NRC climate	1167	759[b]	98	81.6

[a]Contact the author for descriptions of the data-gathering and coding methods underlying these figures.
[b]These numbers are estimates based on the true number of stories in the random samples.

26   VanDeveer's (2006) study of LRTAP is an exception where VanDeveer uses attendance at LRTAP meetings as an indicator of country participation. VanDeveer (2006, 45–8) also explores qualitative aspects of participation, such as delegation size and preparation, which are not well captured by counting attendance.

Table 6
Credibility Scores Highlighting North American Newspapers

Organization	Number of Print Media Stories	Number of Stories Coded	Percentage of Coded Stories from North American Papers	Total Stories Credibility Score	Credibility Score in North American Newspapers
NAPAP	70	70	95.7	58.6	56.7
NRC acid rain	105	105	96.2	83.8	84.2
IPCC	1956[a]	97	49.5	77.3	77.1
NRC climate	759[a]	98	79.6	81.6	80.8

[a]These numbers are estimates based on the true number of stories in the random samples.

from the existing literature in that it devotes considerable attention to creating a standard set of measures that can be applied across cases. Having presented the data derived from these measures, one can assess how successful several of these proxies are in elucidating the concepts of interest.

The buffering and linking scores are fairly direct measures of organizational structures and procedures. Moreover, the derivation of the scores is both explicit and transparent. Although it is possible to overlook an important strategy or weigh too heavily an unimportant one, there are several reasons to feel confident that the final buffering and linking scores do capture important features of each organization. First, finding a strategy in play in one organization disciplines the researcher to look for evidence of a similar strategy in the other two. Thus, any errors in characterization are likely to apply evenly across the cases. Further, the weights added to the buffering and linking scores are used to capture relative differences observed among the organizations rather than attempting to assess some absolute commitment to a given strategy. These two features of the buffering and linking scores increase the confidence that one has in the relative positioning of the assessment organizations.

The concept of political relevance is treated by measuring how much attention the assessment organizations are given in two settings: Congress and the media. Although the stakeholder groups for each assessment organization extend beyond these two institutions, they are undoubtedly important audiences for assessment organization. The measures used here are fairly direct and provide simple ways to gauge the relative attention that each assessment organization gets in two distinct venues. Although there may be biases in the way that the media covers science assessment organizations, for example, a tendency to highlight controversies, or in the kinds of assessment events that members of Congress respond to, such biases should apply to all assessment organizations equally, making the relative scores among the organizations more informative than the raw score given to each.

The measure for credibility comes from content analysis of the newspaper stories about science assessment organizations.[27] This measure gains content validity from the fact that the media do respond to controversies surrounding the work of science

---

27   For similar approaches, see Weart (1988) and Baumgartner and Jones (1993).

Figure 3
NAPAP Newpaper Stories—Total of 70 Stories in Database
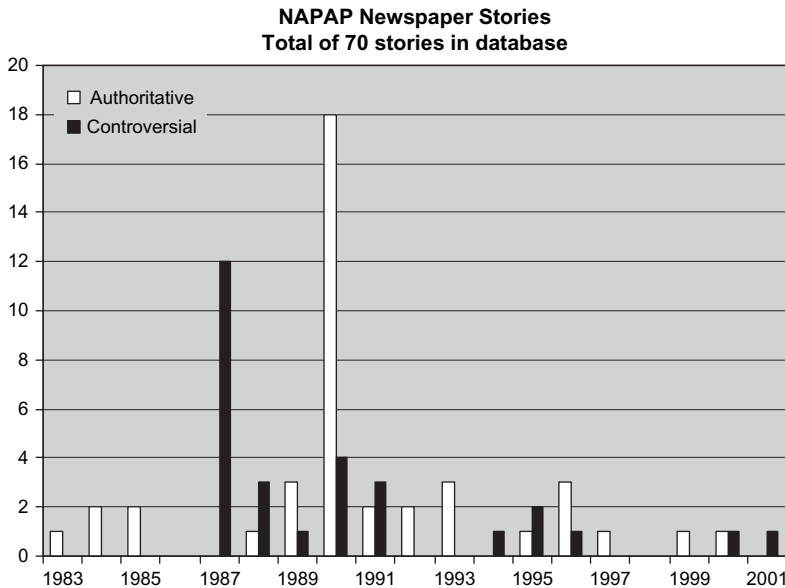
assessments. For instance, the content analysis picks up the controversy surrounding the *NAPAP (1987) Interim Assessment* in that every story that covers NAPAP's work in 1987 makes some mention of the controversy (figure 3). Similarly, the year that the IPCC received the highest proportion of controversial stories was 1996 (see figure 5), the year its initially controversial Second Assessment Report was released (Houghton et al. 1996). That media attention captures these important, known instances when science assessment organizations ran afoul of one or more of their stakeholders raises confidence in the measure. Another feature of these data that support the content validity of this indicator is the fact that controversies do not appear to dominate the coverage that science assessment organizations receive. This demonstrates that print media outlets are not only interested in science assessment organizations when they are generating controversy. These organizations appear to be newsworthy in their own right in that a significant proportion of the media attention devoted to each organization casts that organization as authoritative (figures 3– 6).

The credibility measure gains additional validity from the relative positions among the cases, in particular, the superior scores that the NRC receives. The NRC had an established reputation when it began writing reports on acid rain and climate change. Further, the NRC was not testing out a novel organizational structure or procedures when it produced reports on these two issues. Rather, it was using strategies that it had employed for years and to which the public was already accustomed. The NRC undertook no significant reforms of its procedures although it was producing reports on acid rain and climate change. NAPAP and the IPCC, on the other hand, had to adjust their approaches as they produced reports on their respective topics in response to negative reactions from a portion of their

Figure 4
NRC Acid Rain Newspaper Stories—Total of 150 Stories in Database

NRC Acid Rain Newspaper Stories
**Total of 105 stories in database**
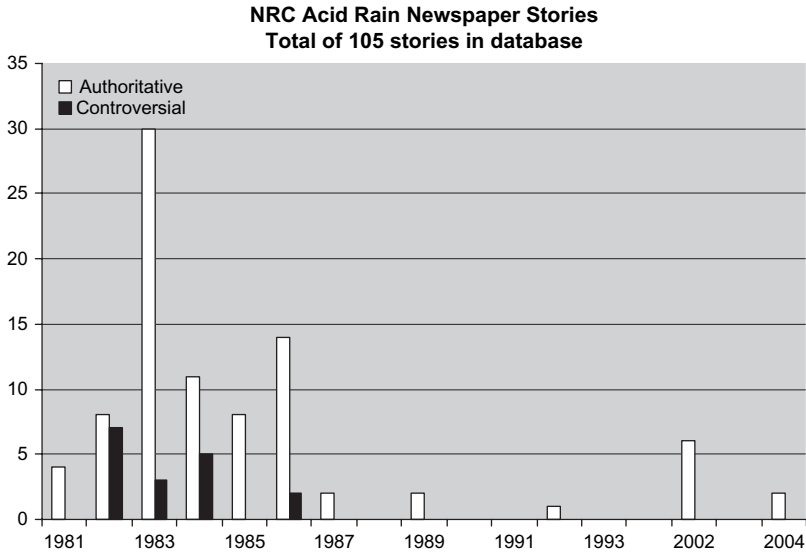
stakeholders—something that reflects the potential difficulties faced by emerging organizations that conduct scientific assessments. That the credibility score reproduces this expected relative position between the NRC on the one hand and NAPAP and the IPCC on the other adds face validity to the measure.

Figure 5
IPCC Newspaper Stories—97 Stories Sampled from an Estimated 1956 Stories in Database

IPCC Newspaper Stories
97 stories sampled from and estimated 1956 stories in database
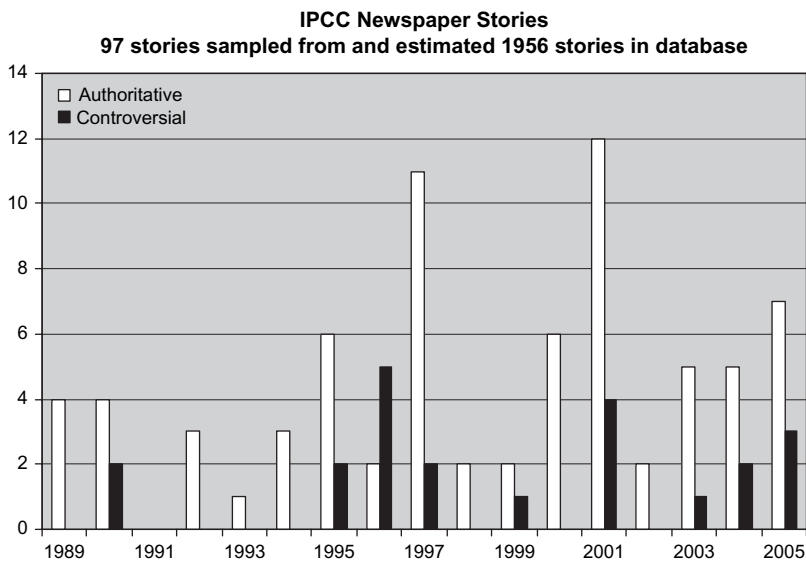
Figure 6
NRC Climate Change Newspaper Stories—98 Stories Sampled from an estimated 746 Stories in Database

NRC Climate Change Newspaper Stories
98 stories sampled from an estimated 746 stories in database
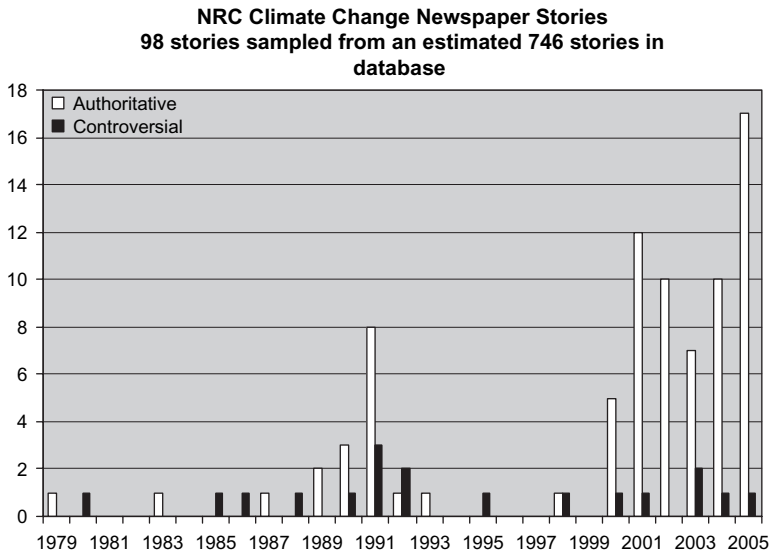
Overall, there is reason to believe that each of the measures, though proxies, capture meaningful aspects of the concepts of interest. With increased confidence in the measures, one can use the study findings to evaluate the stated hypotheses.

ANALYSIS

The research design employed here offers useful leverage over the question of context versus organizational design (H1) by providing a way to test for an issue-effect.

In this case, if organizational capacity is trumped by issue, the study would support the idea that context is more important than design in shaping assessment effectiveness. At first blush, there is evidence that an issue-effect may be at play. For instance, print media stories on the two climate change assessment organizations greatly outnumber those for acid rain. This might suggest that the political relevance scores are tracking general interest in an issue rather than specific interest in these organizations' work.

To test the idea that the issue of climate change, by itself, grabs media attention, the study included an additional search for stories in the LexisNexis database on the Subsidiary Body for Scientific and Technological Advice (SBSTA) created by the UN FCCC and the United States Global Change Research Program (USGCRP).[28] SBSTA works closely with decision makers involved in climate change negotiations to provide science and technical advice, whereas the USGCRP organizes federally funded research on climate change in the

---

28   For a comparison of the IPCC and SBSTA, see Miller (2006). For a detailed history of the USGCRP, see Pielke (2000a, 2000b).

United States and produces synthesis reports on the latest science (e.g., USGCRP 2000). These two organizations offer an interesting point of comparison in that SBSTA has an international scope similar to that of the IPCC, whereas the USGCRP, like the NRC, is primarily focused on a US audience. The results of the search yielded 8 stories about SBSTA and 21 about the USGCRP. This comparison casts doubt on the conclusion that an issue area, by itself, will generate media attention for science assessment organizations.[29]

When looking at congressional attention to the two issues, there is little evidence supporting the idea that there is an issue-effect driving congressional activity. The raw scores for congressional hearings that mention NAPAP and the IPCC are the same. When one considers the time frames over which hearings are held, Congress pays more concentrated attention to NAPAP. However, given that Congress created NAPAP, this concentrated attention is not surprising. Comparing the attention given to the NRC on acid rain and climate change, where the NRC case provides a control for organizational factors, it does seem like Congress pays more attention to the NRC's work on climate change. However, once one considers the time frames involved, congressional attention to the NRC's work on climate change is only slightly more intensive than that for its work on acid rain.

One obvious place to look for an issue-effect is in the comparison of the NRC's credibility scores for the two issues. Since the NRC uses the same organizational strategies in both cases, if issue were driving perceptions about credibility, one should see that across the two NRC cases. In fact, the data show no significant difference in credibility between the two issues. The NRC within-case comparison, by itself, cannot completely dismantle arguments for an issue-effect. For example, it could be that both issues are sufficiently controversial that we do not see much issue-effect in the credibility scores. Or it could be that the NRC's credibility, which does not rest solely on its work on these two issues, is strong enough to wash out any issue-effect. However, taking the NRC results along with the data regarding media and congressional attention, a strong case is made for accepting the idea that organizational factors are at play in shaping assessment outcomes.

With confidence that organizational design has a role in shaping outcomes, one can evaluate the second hypothesis (H2) in light of the research findings. The key research question for the study—do science assessment organizations face a credibility/relevance tradeoff—is addressed by evidence from several sources. First, the strong credibility scores that the IPCC receives in light of its greater use of linking strategies when compared to the other two organizations suggest that ties to user audiences do not necessarily undermine credibility. The comparison between the IPCC and the NRC further underscores this finding. The fact that the IPCC achieves almost the same level of credibility as the NRC for its work on climate change is somewhat surprising given that the IPCC, unlike the NRC, cannot borrow from an existing base of credibility in reporting on the science of climate change. Coupled with the IPCC's intensive use of linking strategies, it is difficult to accept the claim that creating ties with policy end users will result in a significant loss of perceived scientific credibility.

In addition, NAPAP, which uses the fewest linking strategies, receives the lowest credibility score of the three organizations. The within-case treatment of NAPAP helps establish this point. NAPAP's credibility score is lowest before it institutes changes in

---

29   For an analysis of the strengths of SBSTA as a science assessment organization, see Miller (2006).

1988 (table 5). The 1988 reforms involve more substantial changes on the linking side of the equation than it does on the buffering side in that NAPAP adds a set of science-only publications along with three new linking strategies that increased transparency and input from the public and decision makers. During this second phase, NAPAP's credibility score approaches that of the IPCC (table 5). One should not conclude that NAPAP's addition of several linking strategies produced its higher credibility score; the documented changes establish only a correlation. At the same time, the within-case comparison provides evidence that the addition of linking strategies does not necessarily erode an organization's perceived credibility. Taking these findings together, there is substantial evidence to support the hypothesis that science assessment organizations can pursue both credibility and relevance.

The next set of hypotheses attempt to untangle the question of whether or not particular organizational strategies are linked to credibility versus relevance. Although the research design cannot speak to the issue of the impact of specific strategies, some evidence exists to reinforce the notion that buffering supports credibility and linking improves assessment relevance. However, close treatment of the IPCC case complicates this picture.

The NRC cases do lend support to the idea that buffering and credibility are strongly correlated (H3a). Certainly, the NRC relies more heavily on buffering strategies than linking ones and achieves the highest credibility scores of the three organizations. Although some of its score is likely to stem from its past performance, its credibility score does encourage careful consideration of its approach; the NRC tackles two sticky policy issues and is treated very well in the media for its work on these two issues. The results for NAPAP do not tell such a straightforward story about buffering and credibility. NAPAP does better in the period (post-1988) where its buffering activities are balanced by a similar number of linking strategies. When NAPAP relied mostly on buffering strategies, it received the lowest credibility score among the three organizations. Interestingly, the NAPAP case points to the quality or the coverage of buffering activities across the organization. The controversy that arose surrounding the 1987 Interim Assessment stemmed from the fact that the summary for policy makers was not peer reviewed. So, in this case, the existing buffering strategies were not helpful in overcoming criticism stemming from the organization's omission of what appears to have been a crucial buffering strategy. This suggests that it is not merely the number of buffering strategies employed but the quality and coverage of those strategies that contribute to an assessments reputation for credibility.

Moving to political visibility, there appears to be a correlation between linking scores and the visibility indicators (H3b). The IPCC, with the highest linking score receives significantly more media attention than the other two organizations. The comparison between the IPCC and two other climate science assessment organizations—SBTSA and the UCGCRP—suggests that the IPCC's visibility is not merely a product of its subject matter and that media is responding, at least in part, to specific assessment organizations rather than simply the issue of climate change. Also, again turning to the within-case comparison for NAPAP, NAPAP's media coverage achieves its peak in 1990, the year that Congress passed an acid rain control program.[30] This period follows NAPAP reforms that

---

30   Also in 1990, NAPAP began circulating the draft of its Integrated Assessment, which would be published the following year (NAPAP 1991).

substantially increased its linking strategies. Notably, the highpoint for media attention to NAPAP is not 1987, the period when the organization was subject to the greatest controversy.

The congressional attention the IPCC receives matches the other two when looking at the raw scores. When considering the ratio of hearings to time frame on the congressional agenda, the IPCC outscores the NRC, its competitor in terms of credibility scores (table 3). That it trails NAPAP on the ratio measure is almost certainly a function of the fact that Congress created NAPAP and was responsible for its oversight, whereas the IPCC is independent of the US Congress. Although this study cannot rule out other factors that might lead to high visibility for a science assessment organizations, the data here support the argument that attention to linking strategies is consistent with organizational visibility.

Efforts to categorize IPCC strategies into either buffering or linking categories provide initial evidence that the categories of buffering and linking are not always specifically linked to credibility and relevance, respectively (H4). Certainly, the Working Group structure of the IPCC buffers the science of the organization from potential political battles fought over concrete proposals for action, whereas it creates an opportunity for the IPCC to weigh in on the likely outcomes associated with various response strategies. Another strategy that clearly serves both buffering and linking goals is the involvement of government representatives in review of IPCC reports. On the linking side, this involvement increases the knowledge that participating governments have of IPCC output and conclusions. At the same time, when representatives from governments who have opposing views on climate policy can agree on the validity of a single IPCC report, this bolsters the scientific credibility of that report by increasing its perceived universality. Acceptance of IPCC reports among a disparate group of political actors increases confidence that the science contained in the reports is not exceedingly biased. Although the NRC does incorporate the notion of representative science into its panel composition when an issue raises particular controversy, this strategy is not nearly as extensive as allowing end users to review and have input in the assessment process. Here, the IPCC looks substantially different than the other two organizations and provides a convincing example of coproduction.

The comparison among the cases is also important in that it demonstrates the expectation laid out in the organization theory literature that organizations will mirror complexity in their environments by establishing structures and procedures in the organization that take into account that complexity.[31] In comparing NAPAP with the NRC and the IPCC, one is struck by the inventive ways that the latter two organizations have developed to survive in their complex environments. The finding that higher buffering and/or linking scores correlate with higher credibility scores may be an indication that the science assessment organizations that are more successful understand the complexity of their respective environments and have established organizational strategies to manage that complexity. Additional support for this conclusion comes from the fact that NAPAP, though it started out with few linking strategies, quickly added these after the 1987 controversy. These changes point to the fact that NAPAP realized its lack of attention to its political constituency and, ultimately, increased its responsiveness to the democratic norms in operation there.

---

31   See Miller (2001, 487) for a related argument.

Although the study contains several important findings, it also has several limitations. Because the research draws on correlations among the cases, the inferences drawn are preliminary and no causal relationships can be inferred. Future work could address this shortcoming by treating a larger set of cases. Future research might also devise additional measures to those used here to capture the concepts of interest.

Another limitation of this approach is the fact that it does not differentiate among the strategies studied. Although the effort to measure buffering and linking strategies across organizations is important in terms of trying to establish consistent measures as a basis for comparison, it may be that one design feature is more important than the overall buffering and linking scores in bolstering an assessment organization's credibility. For example, there are two interesting design elements that are shared by the NRC and the IPCC. The first is the use of loose coupling as a mechanism for producing assessment reports. The NRC creates one committee for each report that disbands once the report is complete, whereas the IPCC produces three distinct reports, each of which is authored by a distinct group of actors. A second notable feature of both the NRC and the IPCC's approaches is that of representative science. Rather than relying solely on mechanisms for peer review developed by science communities, both the NRC and the IPCC take a cue from democratic politics and incorporate representation into their efforts to substantiate the science contained in their respective reports. This study cannot isolate these strategies as the crucial ones that lead to their higher credibility scores, but the lack of such strategies in the NAPAP case makes these strategies an interesting target for future research.

CONCLUSION

This study calls into question the commonly held view that science assessment organizations can pursue political relevance only at the expense of scientific credibility and vice versa. Instead, this study shows that organizations can pursue buffering and linking strategies simultaneously to support the organization's twin goals. The IPCC case demonstrates that this is possible even when linking strategies are a significant portion of the organization's larger approach.

The study raises several interesting questions about the analytical categories of buffering and linking. Certainly, all three organizations use both buffering and linking strategies with, for the most part, buffering strategies used to protect scientific credibility and linking strategies adopted to increase connections to political end user. At the same time, the study casts doubt on the assumption that buffering is necessarily associated with credibility and linking with relevance. In fact, some strategies are hard to characterize as obviously fitting into one of the two categories. This finding supports the coproduction framework that views scientific and political order as mutually supportive where acceptance from one's scientific audience should be correlated with one's political standing. Although buffering and linking do show up in each organization in expected patterns, there are also strategies that defy this categorization.

Evidence in the study also strongly supports the conclusion that organizational strategies have an impact on measured outcomes and overwhelms any issue-effect. Although it is possible to think of contexts that would challenge even the most successful science assessment organizations, this study shows that organizational performance does vary according to organizational strategy and that this variation surfaces even when different

issue areas are considered. This suggests that attention to organizational design is a worthwhile pursuit in strengthening the performance of science assessment organizations.

From a methodological standpoint, this study takes an important step in the assessment domain by creating specific measures of the concepts of interest and developing a research design that allows for a reasonable test of the expected relationships across the cases. This approach is clearly not a substitute for the single case study work that has dominated the assessment literature thus far. Rather, it provides a complement to that work and encourages future work that applies a common analytical framework across multiple cases. Tackling a larger number of cases could help identify the extent to which specific strategies enhance science assessment organizations' effectiveness. Two strategies identified here, loose coupling and representative science, are particularly good candidates for future research. Studies that treat a larger number of cases may be able to isolate the role that these or other strategies play in helping assessment organizations manage conflicting demands in their environments.

FUNDING

REFERENCES

Agrawala, Shardul. 1998a. Context and early origins of the Intergovernmental Panel on Climate Change. *Climatic Change* 39:605–20.
———. 1998b. Structural and process history of the Intergovernmental Panel on Climate Change. *Climatic Change* 39:621–42.
Agrawala, Shardul, Kenneth Broad, and David H. Guston. 2001. Integrating climate forecasts and societal decision making: Challenges to an emergent boundary organization. *Science, Technology & Human Values* 26:454–77.
Alpert, Peter, and Ann Keller. 2003. The ecology-policy interface: How close is too close? *Frontiers in Ecology and the Environment* 1:1.
Ashby, W. Ross. 1968. Principles of the self-organizing system. In *Modern systems research for the behavioral scientist*, ed. Walter Buckley, 108–18. Chicago, IL: Aldine.
Barnard, Chester I. 1938. *The functions of the executive*. Cambridge, MA: Harvard Univ. Press.
Baumgartner, Frank R., and Bryan D. Jones. 1993. *Agendas and instability in American politics*. Chicago, IL: Univ. of Chicago Press.
Beck, Matthias, Darinka Asenova, and Gordon Dickson. 2005. Public administration, science, and risk assessment: A case study of the U.K. bovine spongiform encephalopathy crisis. *Public Administration Review* 65:396–408.
Brody, Richard A. 1991. Assessing the President: The media, elite opinion, and public support. Stanford, CA: Stanford Univ. Press.
Buckley, Walter. 1967. *Sociology and modern systems theory*. Englewood Cliffs, NJ: Prentice Hall.
Cash, David W. 2000. Viewpoint: Distributed assessment systems: An emerging paradigm of research, assessment and decision-making for environmental change. *Global Environmental Change* 10: 241–44.
Clark, William, and Nancy Dickson. 1999. The Global Environmental Assessment Project: Learning from efforts to link science and policy in an interdependent world. *Acclimations* 8:6–7.
Clark, William C., Ronald B. Mitchell, and David W. Cash. 2006. Evaluating the influence of Global Environmental Assessments. In *Global environmental assessments: Information and influence*, eds. Ronald B. Mitchell, William C. Clark, David W. Cash, and Nancy M. Dickson, 1–28. Cambridge, MA: MIT Press.

Cole, Stephen, Jonathan Cole, and Gary A. Simon. 1981. Chance and consensus in peer review. *Science* 214:881–6.

Collier, David. 1993. The comparative method. In *Political science: State of the discipline II*, ed. Ada Finifter, 105–19. Washington, DC: American Political Science Association.

Collingridge, David, and Colin Reeve. 1986. *Science speaks to power: The role of experts in policy making*. New York: St. Martin's Press.

Collins, H. M. 1981. Son of seven sexes: The social destruction of a physical phenomenon. *Social Studies of Science* 11:33–62.

Cyert, Richard M., and James G. March. 1963. *A behavioral theory of the firm*. Englewood Cliffs, NJ: Prentice Hall.

DiMaggio, Paul J., and Walter W. Powell. 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review* 48:147–60.

Duggan, Joe. 2007. Group stresses wise water use in corn ethanol boom. *Lincoln Journal Star,* October 11: B1.

Edwards, Paul N., and Stephen H. Schneider. 2001. Self-governance and peer review in science-for-policy: The case of the IPCC Second Assessment Report. In *Changing the atmosphere: Expert knowledge and environmental governance*, eds. Clark A. Miller and Paul N. Edwards, 219–46. Cambridge: MIT Press.

Ezrahi, Yaron. 1980. Utopian and pragmatic rationalism: The political context of scientific advice. *Minerva* 18:111–31.

Farrell, Alexander E., Jill Jager, and Stacy D. VanDeveer. 2006. Overview: Understanding design choices. In *Assessments of regional and global environmental risks: Designing processes for the effective use of science in decision making*, eds. Alexander E. Farrell and Jill Jager, 1–24. Washington, DC: Resources for the Future.

Farrell, Alexander E., and Jill Jager, eds. 2006. Assessments of regional and global environmental risks: designing processes for the effective use of science in decision Making. Washington, DC: Resources for the Future.

Farrell, Alexander E., and Terry J. Keating. 2006. Dissent and trust in multilateral assessments comparing LRTAP and OTAG. In *Assessments of regional and global environmental risks: Designing processes for the effective use of science in decision making*, eds. Alexander E. Farrell and Jill Jager, 64–83. Washington, DC: Resources for the Future.

Gieryn, Thomas F. 1983. Boundary work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review* 48:781–95.

———. 1995. Boundaries of science. In *Handbook of science and technology studies*, eds. Sheila Jasanoff, Gerald E. Markle, James C. Peterson, and Trevor J. Pinch, 393–443. Thousand Oaks, CA: Sage.

———. 1999. *Cultural boundaries of science*. Chicago, IL: Univ. of Chicago Press.

Guston, David. 1999. Stabilizing the boundary between U.S. politics and science: The role of the office of technology transfer as a boundary organization. *Social Studies of Science* 29:87–111.

———. 2000. Between politics and science: Assuring the integrity and productivity of research. Cambridge, UK: Cambridge Univ. Press.

———. 2001. Boundary organizations in environmental policy and science: An introduction. *Science, Technology & Human Values* 26:399–408.

Henderson, Tom. 2006. Hey, sailor, care to go upstairs for a little science. *Lewiston Morning Tribune* July 31: 8A.

Herrick, Charles. 2000. Predictive modeling of acid rain: Obstacles to generating useful information. In *Prediction: Science, decision making, and the future of nature*, eds. Daniel Sarewitz, Roger A. Pielke, and Radford Byerly, 251–68. Washington, DC.

Herrick, Charles, and Dale Jamieson. 1995. The social construction of acid rain: Some implications for science/policy assessment. *Global Environmental Change* 5:105–12.

Hilgartner, Stephen. 2000. *Science on stage: Expert advice as public drama*. Stanford, CA: Stanford Univ. Press.

Houghton, J. T., et al., eds. 1996. *Climate change 1995: The science of climate change*. Cambridge, UK: Cambridge Univ. Press.

House Committee on Science, Space and Technology (HCSST). 1988. *National Acid Precipitation Assessment Program: Hearing before the Subcommittee on Natural Resources, Agriculture Research and Environment of the House Committee on Science, Space and Technology.* 100th Congress, 2nd Session. Washington, DC: U.S. Government Printing Office.

Iyengar, Shanto, and Donald R. Kinder. 1987. *News that matters*. Chicago, IL: Chicago Univ. Press.

Jasanoff, Sheila. 1990. *The fifth branch: Science advisors as policy makers*. Cambridge, MA: Harvard Univ. Press.

———. 2004a. The idiom of co-production. In *States of knowledge: The co-production of science and social order*, 1–12. New York: Routledge.

———. 2004b. Ordering knowledge, ordering society. In *States of knowledge: The co-production of science and social order*, 13–35. New York: Routledge.

Jasanoff, Sheila, and Brian Wynne. 1998. Science and decisionmaking. In *Human choice and climate change: Volume I: The Societal Framework*, eds. Steve Rayner and Elizabeth L. Malone, 1–87. Columbus, OH: Batelle Press.

Keller, Ann. 2001. Legislating for acid rain: Substance, symbol, and science in policy decision making. *Annual Meeting of the American Political Science Association,* San Francisco, CA, August 30–September 2.

Knickerbocker, Brad. 2007. Climate summary fuels worry. *Christian science monitor,* November 15: 4.

Kwiatkowski, Roy E., and Maria Ooi. 2003. Integrated environmental impact assessment: A Canadian example. *Bulletin of the World Health Organization* 81:434–38.

Latour, Bruno, and Steve Woolgar. 1979. *Laboratory life: The social construction of scientific facts*. Beverly Hills, CA: Sage.

Lawrence, Paul R., and Jay W. Lorsch. 1967. *Organization and environment: Managing differentiation and integration*. Boston, MA: Graduate School of Business Administration, Harvard.

Lund, David C. 2006. Limits to assessment: An example from regional abrupt climate change assessment in the United States. In *Assessments of regional and global environmental risks: Designing processes for the effective use of science in decision making*, eds. Alexander E. Farrell and Jill Jager, 138–65. Washington, DC: Resources for the Future.

Mahoney, Michael J. 1977. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research* 1:161–75.

Martello, Marybeth Long, and Alastair Iles. 2006. Making climate change meaningful. Framing, methods, and process in coastal zone and agriculture assessments. In *Assessments of regional and global environmental risks: Designing processes for the effective use of science in decision making*, eds. Alexander E. Farrell and Jill Jager, 101–18. Washington, DC: Resources for the Future.

Mazur, Allan. 1973. Disputes between experts. *Minerva* 8:243–62.

———. 1981. *The dynamics of technical controversy*. Washington, DC: Communications Press.

Merton, Robert K. 1968. The Matthew effect in science. *Science* 159:56–63.

Meyer, John W., and Brian Rowan. 1977. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology* 83:340–63.

Meyer, John W., W. Richard Scott, and Terrence E. Deal. 1981. Institutional and technical sources of organizational structure: Explaining the structure of education organizations. In *Organization and the human services: Cross disciplinary reflections*, ed. Herman D. Stein, 151–79. Philadelphia, PA: Temple Univ. Press.

Miller, Clark A. 2001. Hybrid management: Boundary organizations, science policy, and environmental governance in the climate regime. *Science, Technology & Human Values* 26:478–500.

———. 2004. Climate science and global political order. In *States of knowledge: The co-production of science and social order*, ed. Shelia Jasanoff, 46–66. New York: Routledge.

———. 2006. The design and management of international scientific assessments: Lessons from the climate regime. In *Assessments of regional and global environmental risks: Designing processes for the effective use of science in decision making*, eds. Alexander E. Farrell and Jill Jager, 187–205. Washington, DC: Resources for the Future.

Miller, Clark, et al. 1997. *Assessment as a communications process. The global environmental assessment team, a critical evaluation of global environmental assessments: The climate experience.* Washington, DC: IGES/CARE.

Mitchell, Ronald B., William C. Clark, David W. Cash, and Nancy M. Dickson, eds. 2006. *Global environmental assessments: Information and influence.* Cambridge, MA: MIT Press.

NAPAP (National Acid Precipitation Assessment Program [U.S.]). 1991. *The U.S. national acid precipitation assessment program: 1990 integrated assessment report.* Washington, DC: NAPAP Office of the Director.

———. 1987. *NAPAP interim assessment: The causes and effects of acidic deposition, Volume I, executive summary.* Washington, DC: US Government Printing Office.

Nelkin, Dorothy, ed. 1979. *Controversy: Politics of technical decisions.* Beverly Hills, MA: Sage.

Neuman, W. Russel. 1976. Patterns of recall among television news viewers. *Public Opinion Quarterly* 40:115–23.

Neuman, W. Russel, Marion Just, and Ann Crigler. 1992. *Common knowledge.* Chicago, IL: Univ. of Chicago Press.

Parson, Edward A. 2003. *Protecting the ozone layer: Science and strategy.* New York: Oxford Univ. Press.

Paterson, Christopher J., and Richard N. L. Andrews. 1995. Procedural and substantive fairness in risk decisions: Comparative risk assessment procedures. *Policy Studies Journal* 23:85–95.

Pfeffer, Jeffrey, and Gerald R. Salancik. 1978. *The external control of organizations.* New York: Harper and Row.

Pielke, Roger A. 2000a. Policy history of the US global change research program: Part I. Administrative development. *Global Environmental Change* 10:9–25.

———. 2000b. Policy history of the US global change research program: Part II. Legislative process. *Global Environmental Change* 10:133–44.

Przeworski, Adam, and Henry Teune. 1970. *The logic of comparative social inquiry.* Malabar: Krieger.

Roberts, Leslie. 1991. Learning from an acid rain program. *Science* 251:1302–5.

Rubin, Edward S., Lester B. Lave, and Granger Morgan. 1991. Keeping climate research relevant. *Issues in Science and Technology* 8:47–55.

Selin, Noelle Eclely. 2006. Applying assessment lessons to new challenges: From sulfur to POPs. In *Assessments of regional and global environmental risks: Designing processes for the effective use of science in decision making*, eds. Alexander E. Farrell and Jill Jager, 84–100. Washington, DC: Resources for the Future.

Siebenhüner, Bernd. 2006. Can assessments learn, and if so, how? A study of the IPCC. In *Assessments of regional and global environmental risks: Designing processes for the effective use of science in decision making*, eds. Alexander E Farrell and Jill Jager, 166–86. Washington, DC: Resources for the Future.

Star, Susan Leigh, and James R. Griesemer. 1989. Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's museum of vertebrate zoology. *Social Studies of Science* 19:387–420.

Takacs, David. 1996. *The idea of biodiversity: Philosophies of paradise.* Baltimore, MD: Johns Hopkins Univ. Press.

Thompson, J. D. 1967. *Organizations in action.* New York: McGraw Hill.

Udy, Stanley H., Jr. 1962. Administrative rationality, social setting, and organizational development. *American Journal of Sociology* 68:299–308.

USGCRP. 2000. *Climate change impacts on the United States: The potential consequences of climate variability and change.* Washington, D.C.: U.S. Global Change Research Program.

van der Sluijs, Joroen P. 2002. A way out of the credibility crisis of models used in integrated environmental assessment. *Futures* 34:133–46.

VanDeveer, Stacy D. 2006. European politics with a scientific face: Framing, asymmetrical participation, and capacity in LRTAP. In *Assessments of regional and global environmental risks: Designing processes for the effective use of science in decision making*, eds. Alexander E. Farrell and Jill Jager, 25–63. Washington, DC: Resources for the Future.

Weart, Spencer. 1988. *Nuclear fear: A history of images*. Cambridge, MA: Harvard Univ. Press.

Weber, Max. 1946. In *From Max Weber: Essays in sociology*, eds. Hans H. Gerth and C. Wright Mills. New York: Oxford Univ. Press.

Weick, Karl E. 1976. Educational organizations as loosely coupled systems. *Behavioral Science* 19: 357–67.

Zaller, John. 1999. The myth of massive media impact revived: New support for a discredited idea. In *Political persuasion and attitude change*, eds. Diana C Mutz, Paul Sniderman, and Richard A. Brody, 17–78. Ann Arbor: The Univ. of Michigan Press.