

THE USE OF SOCIETAL IMPACTS CONSIDERATIONS IN GRANT PROPOSAL PEER REVIEW: A COMPARISON OF FIVE MODELS

J. Britt Holbrook

University of North Texas, Denton, TX, USA

Increasing demands on the part of the public for a demonstrable return on their investment in scientific and technical research have led to the widespread introduction of considerations of societal impacts into the peer review processes at public science and technology funding agencies. This answer to the accountability challenge also introduces a peculiar strain on peer review: expertise in particular areas of scientific and technical research is no guarantee of expertise in addressing the societal impacts of proposed research. Presenting preliminary results of a larger study, this article describes five current models of the peer review of grant proposals and shows that different agencies have very different ways of incorporating societal impacts considerations. The article also elucidates a notion of theoretical adequacy, which will be used to determine whether and how some peer review processes are better than others. The objectives of this article are to lay out the description of the agencies and to offer a preliminary assessment of each model's theoretical adequacy. The objective of our larger study is to determine the best ways to incorporate societal impacts considerations into the peer review of grant proposals, thus helping funding agencies respond to the demand for demonstrable results.

Key words: Peer review; Societal impacts; Accountability; Expertise

INTRODUCTION

Transforming public investments in science and engineering research into social and economic outcomes relies heavily on the process of the peer review of grant proposals. Governments allocate funds for research to various funding agencies, and scientists and engineers receive those funds in the form of grants; but often, who gets what grant for which research is determined to a large degree by the process of peer review. As Chubin and Hackett (4) note, the peer review of grant proposals is the gatekeeper of science funding: its role is to ensure the fair and informed distribution of limited funds for worthy research. Peer review also, therefore, has an unavoidably political dimension (10).

The process of the peer review of grant proposals is theoretically simple: research proposals are reviewed by peers of the research proposer—indi-

viduals who share the same sort of (and usually the same or higher degree of) expertise as the proposer—in order to determine the worth of the proposed research. The process of peer review rests on the theoretical assumption that a research proposer's peers are in the best position to judge the quality of the proposed research. Indeed, for a particularly specialized research proposal, the proposer's peers may be the only ones capable of judging the scientific and technical merit of the proposed research.

Nevertheless, one might still question whether scientific and technical merits alone are, or ought to be, sufficient for making funding decisions—especially when the funds involved are derived from the public treasury (4,10). This is not merely a theoretical musing: there is evidence of increasing demands on the part of the public for a demonstrable return on their investment in science and

engineering research. For example, in the US the 1993 passage of the Government Performance and Results Act (GPRA) during the Clinton Administration and the introduction of the Program Assessment Rating Tool (PART) by the Office of Management and Budget (OMB) under the Bush Administration in 2002 are particularly notable for their impact on federal science and technology (S&T) funding agencies (8). More recently, on October 7, 2009 under the Obama Administration, OMB issued a memo on “Increased Emphasis on Program Evaluation” that “focuses on impact evaluations, or evaluations aimed at determining the causal effects of programs . . . whose expenditures are aimed at improving life outcomes (such as improving health or increasing productivity) for individuals” (24). Similar initiatives have been put forward around the world. The emphasis in each of these initiatives is on the demonstration of results from funding, and so the interest expressed is less about what research is actually funded and more about the societal *outcomes* of that funded research. In other words, each of these initiatives expresses the belief that scientific and technical criteria alone are not sufficient for making funding decisions.

Public S&T funding agencies worldwide have responded to this demand for demonstrable results in an interesting way: the belief that scientific and technical criteria alone are not sufficient for making funding decisions has been combined with our unwavering faith in the soundness of peer review by asking peer reviewers to assess research proposals not only in terms of their intrinsic scientific or intellectual merit, but also in terms of their broader societal impacts. Such societal impact considerations have placed unique burdens on the peer review system, as scientists and engineers are asked to judge the merits of a proposal not only in terms of their own scientific and technical expertise, but also in areas in which they may lack expertise. How has grant proposal peer review adapted to this increased strain? How have different agencies responded differently? Have some agencies responded better than others?

LITERATURE REVIEW

In addition to an ongoing informal discussion within the academic community [e.g., (23)], there

exists some scholarly literature devoted to the peer review process for journals, including most notably philosophical examinations of the epistemological and ethical issues surrounding prepublication peer review (31) and historical accounts of the origins of prepublication peer review (15,26,32). However, much of the literature devoted to peer review is concerned with the peer review of grant proposals. Kostoff (14) provides the best account of the use of peer review at the level of research programs, as well as a very extensive list of references covering most of the literature on peer review up to 1997.

Despite occasional calls for its abandonment from critics of the process of peer review of grant proposals [e.g., (12,28,29)], the most common question surrounding grant proposal peer review is not *whether* peer review is the best model for allocating funding, but rather how and *how well* peer review of grant proposals works to allocate funds. For example, as part of their study of peer review in the National Science Foundation, Cole, Rubin, and Cole (5) divide work on proposal peer review into three categories: a) general studies of the effectiveness of peer review; b) studies of factors affecting the granting of awards; and c) studies of outcomes of the review process. Chubin and Hackett (4) also divide studies of the grant proposal review process into three groups, although their system of division is simultaneously more helpful and more provocative: a) agency-sponsored, or “insider,” studies, which benefit from greater access to information, but suffer from greater political pressure than; b) agency-funded studies, which allow more freedom of inquiry, but restrict access to “inside” information; and c) independent studies, which allow maximum latitude and minimal (if any) funding for inquiry. Chubin and Hackett go on to examine several empirical studies of grant proposal peer review in terms of the values of effectiveness, efficiency, responsiveness, and fairness.

Other scholars have targeted their research on one or more of these values. For instance, Wenerås and Wold (34) focused on the value of fairness in their influential study on nepotism and sexism in the peer review of proposals for postdoctoral fellowships from the Swedish Medical Research Council (MRC). Their work sparked a rush of responses, including similar studies regarding sexism at other funding agencies, voluntary self-studies by

agencies, studies of sexism in peer review for journal publication (33), and even a recent attempt to replicate the original study (30), which found that, while bias against women was no longer evident at the MRC, nepotism was still a serious issue.

Horrobin (12) suggests that grant proposal peer review is neither effective—“Forget the reviewers. Just flip a coin,” he writes, citing Rothwell and Martyn (27)—nor responsive—“The peer review process may be stifling innovation” [see also (13)]. Others have focused on the values of effectiveness, responsiveness, and fairness in the context of the review of proposals for interdisciplinary research (16,19,20). Interest in the value of efficiency is quite prevalent among funding agencies themselves, which tend to study the efficiency of their own systems of proposal review on an ongoing basis. For instance, at the US National Science Foundation (NSF), the merit review system undergoes continual examination to inform writing an Annual Report on Merit Review to the National Science Board, and each program’s handling of the merit review process is scrutinized every 3 years by a Committee of Visitors.

Another interesting feature of the existing literature on grant proposal peer review is that there seems to be a growing interest in the process of peer review, as well as its *outcomes*. Lamont and Mallard (17), arguing against those who suggest that peer review on the basis of intrinsic scientific merit is “automatically fair,” suggest that “biases in evaluation against certain categories of people can enter *the process of evaluation* through the favoring of epistemological styles and other cognitive categories” (emphasis added). Lamont and Mallard base their conclusion on a comparison of peer review models from the US, the UK, and France. With a similar nod toward the process of peer review, Langfeldt (19) writes: “There are different models of peer review for different kinds of policy instruments. Different countries have different traditions of peer review, and so have different disciplines. However, the implications of various review processes are often incompletely understood and often not considered when designing review processes” [see also (18)].

THE STUDY

This article presents the preliminary results of an on-going 3-year study, the Comparative Assess-

ment of Peer Review (CAPR), which is based at the University of North Texas and funded by NSF’s Science of Science and Innovation Policy (SciSIP) program. CAPR aims to develop a taxonomy of five current models of peer review of grant proposals that incorporate societal impacts considerations across three US federal agencies: the National Science Foundation (NSF), the National Institutes of Health (NIH), and the National Oceanic and Atmospheric Administration (NOAA) and related efforts in two non-US contexts: the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Dutch Technology Foundation (STW).

The rationale for choosing these particular agencies for CAPR is twofold:

- First, this selection of agencies provides a broad spectrum of approaches to science and engineering research extending from basic to applied, and including both mission-oriented and generic (or what NSERC refers to as “free”) research funding agencies.
- Second, each of these agencies incorporates considerations of societal impacts into its grant proposal peer review process, and each does so in a different manner.

The ultimate objective of CAPR is to compare how different agencies have incorporated societal impacts considerations into their peer review processes, with the aim of determining whether some agencies have done so in a way that might help other agencies perform better. This article presents the CAPR team’s preliminary results, including a description of the five agencies’ models and an account of theoretical adequacy, which is our preliminary evaluative frame.

Every peer review process rests on implicit theoretical assumptions, answers to questions such as:

- What criteria are used in a given peer review process, and how are they balanced?
- What goal or goals do they serve?
- Who qualifies as a peer?
- What range of interests should be represented in peer review?

Alternative answers to such questions will result in different models of peer review. Moreover, alternative answers to such questions will instantiate different values. Examining the values that different models of peer review instantiate and comparing those instantiated values to the expressed goals of particular funding agencies will make it easier for funding agencies to identify flaws in the structure of, and if necessary to restructure, their existing peer review processes.

The CAPR team is taking a mixed-methods approach to the study. Much of the research involves a collection and review of administrative documents from each of the funding agencies under study (1). This review of the administrative documents related to peer review has enabled the creation of a survey (2). The survey is currently being used to generate a convenience sampling of stakeholders in the peer review process at the agencies being studied. The results of the survey will then be used to generate questions for semistructured interviews of targeted stakeholders. The CAPR team is also holding a series of workshops [see (3)] to engage stakeholders in the coproduction of useful knowledge.

Despite the fact that the methodologies employed for the study are well established, CAPR's approach to research on the process of peer review of grant proposals is different from anything in the existing literature in two important ways. First, our study focuses on the process of peer review for grant proposals in terms of the adequacy of theory underlying various models of the proposal review process. This notion of underlying theoretical adequacy (or lack thereof) can perhaps best be captured with reference to the old saw, "It sounds good in theory, but it doesn't work in practice." With reference to a practical process, such as the peer review of grant proposals, to say that it is "good in theory, not in practice" amounts to a sort of rational failure. If a particular model of peer review "doesn't work in practice," this is evidence of a mistake somewhere in the theoretical underpinnings or structure of that model. Second, CAPR focuses especially on the use of various models of peer review to assess grant proposals beyond intrinsic merit and specifically in terms of societal impact (SI).

Assessing Theoretical Adequacy

Although it may be legitimate for some purposes to judge the effectiveness of a decision-making process in terms of its outcomes (what one might call a "proof is in the pudding" approach), CAPR's preliminary research focuses on the theoretical adequacy (TA) of different models of peer review. Applied to a practical means–end process, TA refers to the way the process is structured so as to realize, in principle, certain values (see Fig. 1 for an example). Peer review is itself such a practical means–end process, and peer review is structured in a variety of ways, depending on the end one wishes to achieve.

For instance, one may "blind" a proposer to the identity of her reviewers in order, in principle, to allow those reviewers to speak freely, without fear of negative repercussions. While such a single-blind approach is theoretically adequate to protect the reviewers from fear of repercussion, and therefore to promote a certain level of honesty in the reviews, it is not theoretically adequate to prevent reviewers from exercising whatever biases they might have in favor of or against a particular proposer whose identity (gender, race, age, etc.) they know: hence, the widespread use of double-blind peer review, at least in the context of peer review of publications. Single-blind peer review is far more common among public S&T funding agencies, because the identity of the proposer is deemed to be essential to making a good decision about funding.

All processes of the peer review of grant proposals aim to fund the best possible research in the fairest possible way; but different funding agencies have different mandates for the specific types of research they are to fund and different processes by which they arrive at their funding decisions. Thus, what counts as the best research will vary from agency to agency, depending on the agency's specific mandate. By comparing different models of the peer review of grant proposals, we aim to show not merely that different models exist, but also that some models are better than others in terms of theoretical adequacy.

Theoretical adequacy describes the appropriateness of means to ends, and whether a particular process is theoretically adequate will depend on whether the means utilized in a process are in prin-

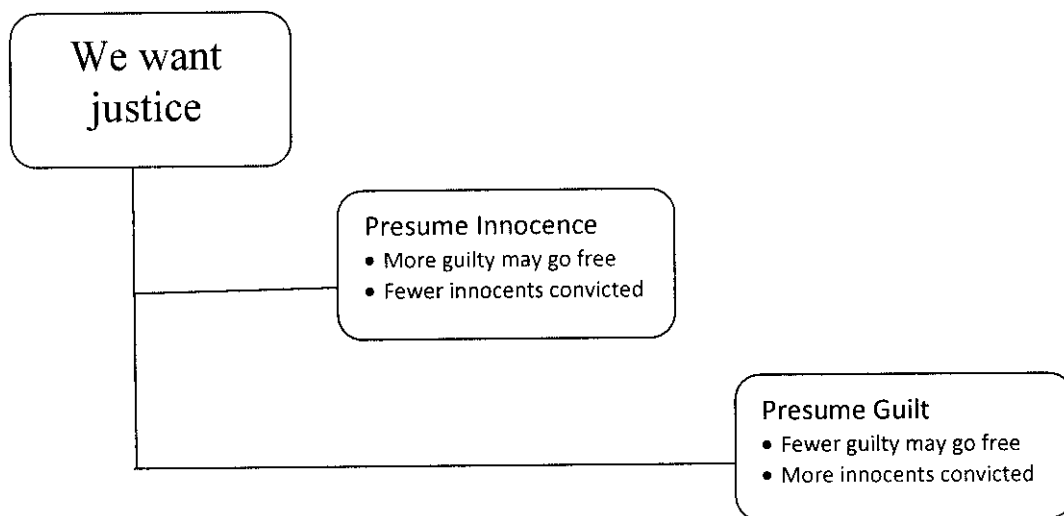


Figure 1. Theoretical adequacy. The US criminal justice system is set up in such a way that someone accused of a crime is automatically presumed innocent until proven guilty. The justice system is structured in this way because we in the US have decided that, in order to render the system consistent with our valuation of justice, it would be better to set things up so that more guilty people are acquitted for crimes they did commit than innocent people are convicted for crimes they did not commit. Were our values different—say, if we were to hold that it is better to convict more actually innocent people to make sure that as few actually guilty people as possible were acquitted—then our current legal system would be theoretically inadequate: we ought, instead, to presume guilt. In order to determine theoretical adequacy, we must first determine the intended end, then evaluate the means to that end.

ciple well-suited to achieving the desired end. In cases in which the means are less consistent with the desired ends, the process will be deemed less theoretically adequate; and when the means are more consistent with the desired ends, the process will be deemed more theoretically adequate. The degree of theoretical adequacy thus varies according to context.

Assessing the TA of Specific Models of Peer Review

A comparative assessment of different models of the peer review of grant proposals will allow funding agencies to see where they fall in terms of the theoretical adequacy of their model of peer review relative to the models of other agencies, as well as opening up possibilities for restructuring their existing models of peer review to attain a greater degree of theoretical adequacy (i.e., seeing other possible models of peer review that are more theoretically adequate may encourage or guide restructuring). CAPR pays particular attention to the theoretical adequacy of the ways in which these

different agencies incorporate societal impacts considerations. *Nota bene*, however, this article presents merely a description of different models of peer review and a preliminary assessment of TA.

NSF's Merit Review Model and the Broader Impacts Criterion

NSF is the US federal agency that supports basic research across all fields of science and engineering (with the exception of medical research, which is supported by NIH); it boasts an annual budget of about \$6 billion—over \$9 billion in FY09 because of the American Recovery and Reinvestment Act (ARRA) of 2009—and is the funding source for approximately 20% of all federally supported basic research conducted by America's colleges and universities. Almost all (usually over 95%) of the funding proposals submitted to NSF (now routinely over 40,000 annually) undergo the process of merit review, either via mail-only, panel-only, or combination (mail + panel) methods. Every proposal submitted that does undergo the NSF process of merit

review is judged according to two generic merit review criteria.

Intellectual Merit: What is the intellectual merit of the proposed activity? How important is the proposed activity to advancing knowledge and understanding within its own field or across different fields? How well qualified is the proposer (individual or team) to conduct the project? (If appropriate, the reviewer will comment on the quality of prior work.) To what extent does the proposed activity suggest and explore creative and original concepts? How well conceived and organized is the proposed activity? Is there sufficient access to resources?

Broader Impacts: What are the broader impacts of the proposed activity? How well does the activity advance discovery and understanding while promoting teaching, training, and learning? How well does the proposed activity broaden the participation of underrepresented groups (e.g., gender, ethnicity, disability, geographic, etc.)? To what extent will it enhance the infrastructure for research and education, such as facilities, instrumentation, networks and partnerships? Will the results be disseminated broadly to enhance scientific and technological understanding? What may be the benefits of the proposed activity to society?

NSF has explicitly chosen not to assign a particular weight to either criterion, and reviewers are asked to comment only on aspects of the proposal they feel qualified to judge. In theory, it seems to make sense to ask reviewers not to comment outside their areas of competency. In practice, however, there exists a marked tendency on the part of both proposers and reviewers to emphasize the intellectual merits of a proposal to the exclusion of its broader impacts (5–10). That the merit review process has two criteria in theory, but often uses only one in practice indicates a rational breakdown in NSF's model of peer review.

TA Assessment of NSF. NSF's mission is "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense." Thus, societal impacts are built in to the mission of NSF. NSF's two merit review criteria are meant to link the progress of science with beneficial societal outcomes. Yet, if proposers and reviewers emphasize only the intellectual merit of proposals, beneficial societal outcomes will not be considered in making funding decisions. NSF

ought to ensure that both merit review criteria are used in evaluating grant proposals.

Should NSF consider assigning a specific weight to the criteria? Do experts in a particular area of science or engineering possess the relevant expertise to articulate and assess the broader impacts of proposed research? Should NSF abandon the Broader Impacts Criterion? Alternatively, should NSF use reviewers who are experts in SI research to assist in assessing proposals according to the Broader Impacts Criterion? What about the possibility of including non-experts as reviewers (e.g., not just scientists and engineers from disciplines other than that represented in the proposal, not just experts in SI research, but also members of the lay public)? How would such possibilities affect the NSF model of merit review?

NIH's Two-Tiered Model and Employment of Lay Reviewers

NIH, which is composed of 27 Institutes and Centers (IC), is the primary federal agency charged with conducting and supporting medical research in the US. Its mission is to foster "science in pursuit of fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to extend healthy life and reduce the burdens of illness and disability." In pursuing its mission, NIH invests roughly \$30 billion annually (over \$36 billion in FY09 due to ARRA) to support medical research, of which around 80% is awarded in the form of competitive grants.

NIH uses a two-tiered model for peer review of grant proposals, which first employs scientific experts in what are known as Scientific Review Groups (SRGs; sometimes also IRGs for Initial Review Groups) to judge the scientific and technical merits of the proposal, followed by review by an IC Advisory Council to determine the relevance of the proposed research to specific IC priorities and public health needs. These IC Advisory Councils are composed of both scientific experts and "public representatives" (usually in a ratio of two scientists to one layperson) in order to "ensure that the NIH receives advice from a cross-section of the US population in the process of its deliberation and decisions." NIH is explicit about the fact that, in tier one, scientific and technical review committees are asked to make judgments on the basis of scientific

and technical merit *only* (i.e., they are required *not* to make judgments on the basis of policy). Advisory Councils then meet to “review the summary statements for scientific merit, and, against a broad background of considerations that include the total pattern of biomedical and behavioral research that is being supported by the Institute or Center, the adequacy and appropriateness of peer review as performed by the IRGs, the need for research to be initiated in new areas, the relevance of the proposed research to the Institute or Center’s mission and programmatic priorities, the potential public health or categorical disease impact, and other matters” (21).

TA Assessment of NIH. The mission of NIH is “science in pursuit of fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to extend healthy life and reduce the burdens of illness and disability.” NIH thus aims not only to pursue knowledge for the sake of knowledge, but also the application of that knowledge to enhance health. For this reason, NIH pursues what they term “translational research,” which is aimed at the application of basic research (conducted at “the bench”) to clinical problems (the patient’s “bedside”). Dividing the peer review of grant proposals into a “scientific” tier and an “advisory” tier, however, implies that considerations of application can in fact be separated from questions of basic research. Yet this assumption has been widely criticized in the science policy community [e.g., (25)].

Does using lay reviewers as part of the review process help or hinder its overall operation? What sorts of difficulties are encountered by scientists asked to judge a proposal only according to its scientific merits (as in tier 1)? What about difficulties encountered by scientists who participated as members of the IC Advisory Councils (in tier 2)? Does NIH’s use of a two-tiered model make translation from bench to bedside more difficult? Does NIH’s use of a two-tiered model obviate some of the difficulties encountered by NSF?

The NOAA Merit Review Model: Establishing Order of Rank

NOAA is a US federal agency focused generally on the condition of the oceans and atmosphere. In

particular, NOAA supplies environmental information (most familiarly through the forecasts and warnings of the National Weather Service), acts as the steward of US coastal and marine environments, and sponsors applied research on the environment. NOAA’s mission is “To understand and predict changes in the Earth’s environment and conserve and manage coastal and marine resources to meet our nation’s economic, social and environmental needs.” NOAA awards roughly \$1 billion per year in competitive grants.

For competitive grants for research projects, NOAA uses a merit review model to establish an order of rank for proposals. NOAA merit review is conducted by mail reviewers and/or peer panel reviewers, and requires a minimum of three reviewers per proposal. Each reviewer *individually* evaluates and ranks proposals (i.e., NOAA requires that “No consensus advice shall be provided by either merit review group if there are any non-Federal members”). Reviewers are asked to rate project proposals according to the following criteria:

1. **Importance and/or relevance and applicability of proposed project to the program goals:** This ascertains whether there is intrinsic value in the proposed work and/or relevance to NOAA, federal, regional, state, or local activities.
2. **Technical/scientific merit:** This assesses whether the approach is technically sound and/or innovative, if the methods are appropriate, and whether there are clear project goals and objectives.
3. **Overall qualifications of applicants:** This ascertains whether the applicant possesses the necessary education, experience, training, facilities, and administrative resources to accomplish the project.
4. **Project costs:** The budget is evaluated to determine if it is realistic and commensurate with the project needs and time frame.
5. **Outreach and education:** NOAA assesses whether this project provides a focused and effective education and outreach strategy regarding NOAA’s mission to protect the Nation’s natural resources.

The merit reviewer's ratings are then used to produce a rank order of the proposals: each reviewer's individual ratings and rankings are relayed to a NOAA Program Officer, who reviews them as a group and makes recommendations to a Selecting Official. Both the Program Officer and the Selecting Official may alter the merit reviewers' rankings of proposal in light of additional considerations, including: availability of funding, geographic or institutional balance/distribution of funds, whether this project duplicates other projects funded or considered for funding by NOAA or other federal agencies, program priorities and policy factors, applicant's prior award performance, and partnerships and/or participation of targeted groups.

TA Assessment of NOAA. Like both NSF and NIH, NOAA is concerned to advance scientific understanding for the benefit of society, although in the case of NOAA the scientific understanding pursued is always related to the environment. This understanding of the environment is supposed to allow NOAA to "conserve and manage" the environment. Interestingly, NOAA peer reviewers are not asked to rate or rank proposals according to their societal feasibility or potential societal impacts (beyond education and outreach). Because most environmental conservation and management policies involve human beings, one potential area in which NOAA's merit review system may lack theoretical adequacy is in linking scientific understanding to society (especially human behavior). Is evaluating education and outreach enough to cover such considerations?

Do proposers and reviewers for NOAA evidence preference for any of NOAA's review criteria over others? How do they respond to the "Outreach and education" Criterion? Which criteria do they consider most important in their ranking of various proposals? How does NOAA's practice of maintaining the independence of reviewers compare to NSF's and NIH's uses of review panels? Should NOAA consider adding lay reviewers? How would that affect the NOAA model of merit review?

NSERC—Peer Review/Évaluation par les pairs

NSERC is essentially Canada's version of the US's NSF (although the social sciences are supported in Canada not through NSERC, but rather

through the Social Sciences and Humanities Research Council of Canada). NSERC refers to their model of reaching funding decisions simply as peer (not "merit") review. NSERC uses Grant Selection Committees (GSCs) consisting of 10 members of the scientific and engineering community. Although all members of a GSC are responsible for reading all proposals they are to review, for each proposal, two members of the GSC who are closest in expertise to the proposed research are selected as Internal Reviewers. The first Internal Reviewer also selects (based both on a list of possible reviewers provided by each proposer and on her or his own knowledge of the specialized research community) three External Reviewers (or referees). The GSC then meets as a group to discuss the proposals under review, as well as the reviews of the External Reviewers. During the meeting of the GSC, the Internal Reviewers of a proposal essentially act as its "champions" and argue its merits to the rest of the members. For all proposals submitted to NSERC for their Discovery Grants (which are the major source of funding for research, as opposed to funding for fellowships or for infrastructure), the following criteria are used in the peer review process: scientific or engineering excellence of the researcher(s); merit of the proposal; contribution to the training of highly qualified personnel; and need for funds.

NSERC's mission is quite specific: NSERC invests in people, discovery, and innovation through programs that support post-secondary research in the natural sciences and engineering on the basis of national, peer-reviewed competitions. NSERC expands on its mission by describing its "role" as pursuing its mission "for the benefit of all Canadians" and enumerating the students and postdoctoral fellows (28,000), university professors (11,800), and Canadian companies (1,500) in which it invests each year. NSERC also provides a statement of values it pursues: excellence, leadership, teamwork, open communication, integrity, flexibility, openness, and accountability—all of which it explicitly connects to the process of peer review: "The practice of basing all decisions regarding the awarding of grants or scholarships on the results of a thorough assessment of detailed proposals by experts in the field is fundamental to NSERC's values." Interestingly, NSERC provides extensive

guidance to reviewers to clarify how the agency expects its peer review criteria to be interpreted (22). These guidance policies reveal that NSERC asks proposers to fill out a Personal Data Form (Form 100) in which they describe their most important recent contributions to research, including the impact on the “end users” of that research. Essentially, then, NSERC reviewers are asked to judge *past*, if recent, impacts, including societal impacts, of the work produced by a researcher rather than only predicting potential (possible *future*) impacts. Proposers to NSERC are also given the opportunity to provide evidence of past impacts, including copies of entire publications, technical reports from industry partners, and so forth.

Assessing TA for NSERC. NSERC’s mission, values, and peer review system appear to be remarkably well aligned. Nevertheless, their goal of benefitting “all Canadians” is somewhat diffuse. Interestingly, however, while it places the onus on the individual proposers to make the case for their own individual impacts to reviewers, NSERC takes full responsibility for demonstrating the impact of overall investments in S&T research to the public, a position linked, perhaps, to Canada’s S&T Strategy: Mobilizing Science and Technology to Canada’s Advantage.

Given their similar roles in their respective national contexts, how does NSERC’s peer review process compare with NSF’s merit review process? How does always using a Grant Selection Committee (rather than sometimes using only mail reviews) affect the peer review process? What about the fact that Internal Reviewers have more expertise in the specialized area of the proposals they “champion,” and how does this compare with having competing reviews (one favorable, one unfavorable) from experts in the same field? Would adding lay reviewers enhance NSERC’s peer review process? Is demonstrating the results of S&T funding a matter for peer review of grant proposals?

**The Dutch Technology Foundation (STW):
The Dialogical Model, Utilization,
and the Users’ Committee**

The major Dutch research funding agency (Netherlands Organization for Scientific Research NWO) is divided into eight divisions along disci-

plinary lines: Earth and Life Sciences (ALW), Chemical Sciences (CW), Physical Sciences (EW), Humanities (GW), Social Sciences (MaGW), Medical Sciences (ZonMw), Physics (N), and Technical Sciences (STW). Each division judges research proposals and supports research according to discipline-specific criteria. The Technical Sciences Division, or the Dutch Technology Foundation (STW), is unique, however, in that it supports research from any disciplinary field, as long as that research is aimed at what the Technology Foundation terms “utilization,” which it defines as “the embedding of results in society.” As STW’s website puts it, “Anything is welcome, as long as utilization is provided for.” A closer examination of their peer review model, however, reveals that “utilization” is merely a necessary, but not sufficient, condition for funding.

STW requires research proposals to meet two criteria: a) it must have high scientific quality, and b) it must be directed towards practical application. In support of STW’s second criterion of “utilization,” each funded research project meets regularly with a Users’ Committee. Interestingly, many eventual members of the Users’ Committee of a funded project were directly involved with the writing of the research proposal itself. As STW puts it, “A proposer of an STW project usually has a network of relations with people in industry. These contacts brought him inspiration for a new research topic. As he wrote his proposal he discussed his ideas, the design and planning of the research work, with them. In the utilization chapter of his application, he mentioned their names as possible members of the users’ committee.” If a proposal receives a grant, the first step is to provide funding to begin the initial phases of the research. Once a project has passed the initial stages and is up and running, STW forms a Users’ Committee: “Candidates mentioned in the original proposal and other suitable representatives will then be invited to join the users’ committee and a first meeting is planned. Members represent their company. Every meeting will have two main themes: reporting about the progress of the work and utilization. There will be ample opportunity for discussion. Viewpoints from industrial practice may well pose new perspectives to university scientists. Scientific results can bring new insights into industrial prob-

lems.” This “opportunity for discussion” is essentially an extension of a conversation that began during the initial stages of proposal preparation.

The actual proposal review process represents a sort of midpoint in this conversation. Once a proposal is submitted to the STW, a Program Officer reads the proposal to determine whether it meets the necessary criteria. If so, the Program Officer selects a group of experts to serve as referees for the proposal: “Experts are approached both in universities and in the business world. Always an effort is made to find those experts who are naturally interested in these new developments, be it for their own research or for possible application in their own business.” Once referees are selected, they read the proposal and send written comments judging the proposal according to an extended list of criteria to the Technology Foundation. The Technology Foundation then analyzes the referee comments and reorganizes them according to the criteria on which they comment; it then sends the reorganized comments to the proposer, who is “asked to formulate a separate answer to every remark.” These answers are then sent to STW, which then gathers the initial proposal, referees’ comments, and proposer’s responses together in what is referred to as the “protocol” of the project: “The final result is a document that contains a point by point discussion of the merits of the proposal, a discussion conducted between experts.” The completion of a project protocol represents the end of step 1 of the review process.

Step 2 begins when 20 projects have a finished protocol (STW accepts proposals continuously, but proceeds to step 2 as soon as 20 protocols are finished). STW then forms a new jury made up of 12 members of the general research community (including both academic and nonacademic researchers), without regard for their specific area of expertise: “Their own specialties are not important for the process: they are asked to judge the 20 proposals on the basis of the discussions between the experts contained in the protocol.” The jury does not meet, and in fact their identities are kept hidden until the publication of STW’s annual report. Instead, each member of the jury is sent 20 proposals and asked to score each proposal according to a) scientific merit and b) utilization. After all members of the jury have communicated their scores to

STW, each member is informed of the average scores for each project, after which they are allowed to adjust their scores. Thereafter, each proposal is given a final score for scientific merit and for utilization, with both weighted equally. These scores are then forwarded to the Board of STW, which makes the final determination of award (which usually happens about 6 months after the application, with roughly 40% of proposals receiving grants).

STW refers to its peer review procedure as “*a continuing public debate*” (emphasis theirs). For our study, however, we refer to it as the Dialogical Model of peer review.

Assessing TA for STW. Given STW’s mission of funding utilization-oriented research, its Dialogical Model of peer review seems remarkably theoretically adequate. For contrast, one could imagine an agency that aimed to fund utilization-oriented research that failed to incorporate users into the process, which would be much less theoretically adequate. The fact that the peer review process is actually the midpoint of a process that incorporates users at the proposal-writing stage and while the research is actually being conducted is also well-aligned with the mission of funding utilization-oriented research.

Would something like the Dialogical Model of peer review work in contexts other than that of an agency devoted to applied research and “utilization”? Could the idea of the Users’ Committee be adapted to the peer review of basic research? How does the Users’ Committee compare with NIH’s use of lay reviewers?

CONCLUSIONS

Although more work remains to be done to complete our study, it is possible to conclude at this point not only that different agencies employ different models of peer review (which is obvious—although the actual range of possible models is perhaps surprising), but also that some models of peer review are more theoretically adequate than others. The models employed by NSERC and STW, in particular, stand out as examples of high theoretical adequacy. In comparison, the US funding agencies in our study are less theoretically adequate. This is not meant to condemn the peer review processes at

all US S&T funding agencies or to malign the peer review processes at NSF, NIH, and NOAA. Indeed, in many ways, US agencies lead the way in incorporating societal impacts considerations as part of peer review (11). It is to suggest, however, that attaining a high degree of theoretical adequacy for the ostensibly theoretically simple process of peer review is more complicated than it at first appears. Public S&T funding agencies spend a great deal of time and effort evaluating and attempting to improve their own processes of peer review. Perhaps, however, they have something to learn from each other.

CAPR's preliminary conclusions are being used to ground the remainder of the study, which includes finishing the survey and conducting interviews with relevant stakeholders in these agencies, such as program officers, proposers, reviewers, grantees, and members of the lay-public. Because our preliminary assessments of TA have allowed us to find areas of theoretical strength and weakness in the five models of peer review, we were able to construct survey questions that will reveal areas in which the models perform particularly well or poorly in practice, as well. The survey results will allow for the development of interview questions that will shed additional light on the TA of different models. By combining our initial TA assessments with our empirical data regarding what works and what does not, we aim to develop a taxonomy of different models of peer review for different agency contexts, thereby improving the peer review process at public S&T funding agencies worldwide.

ACKNOWLEDGMENTS: This material is based upon work supported by the National Science Foundation under Grant No. 0830387. The author also acknowledges contributions from members of the CAPR team, especially Robert Frodeman and Kelli Barr, and thanks several anonymous reviewers and Alan Moghissi for helpful comments. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF), the CAPR team, or anyone else who helped bring the article to fruition.

ABOUT THE AUTHOR



J. Britt Holbrook is Assistant Director of the Center for the Study of Interdisciplinarity (CSID) at the University of North Texas (UNT). CSID conducts research into the theory and practice of inter- and transdisciplinary approaches to knowledge. Interdisciplinary centers focused on a particular theme or problem is common; CSID is the first center devoted to developing a philosophy of interdisciplinarity. Since January 2005 Holbrook has also served as Research Assistant Professor within the Department of Philosophy and Religion Studies at UNT. Prior to joining UNT, Holbrook held teaching positions in philosophy at Emory University and at Georgia State University, both located in Atlanta, Georgia. Holbrook has interdisciplinary teaching experience as a member of the faculty of the 2007 Texas Governor's School and as co-instructor of a field course in socioecological conservation in the Cape Horn Biosphere Reserve in southernmost sub-Antarctic Chile. Holbrook's current research focuses on interdisciplinarity, peer review, and the relationship between science, technology, and society.

REFERENCES

1. Center for the Study of Interdisciplinarity. Comparative Assessment of Peer Review (CAPR) digital repository. Retrieved from <http://csid-capr.unt.edu/repository>
2. Center for the Study of Interdisciplinarity. Comparative Assessment of Peer Review (CAPR) survey. Retrieved from <http://www.surveymonkey.com/s/BT3KM9R>
3. Center for the Study of Interdisciplinarity. Comparative Assessment of Peer Review (CAPR) homepage. Retrieved from <http://csid-capr.unt.edu/>
4. Chubin, D.; Hackett, E. Peerless science: Peer review and U.S. science policy. Albany, NY: State University of New York Press; 1990.
5. Cole, S.; Rubin, L.; Cole, J. R. Peer review in the National Science Foundation: Phase I of a study. Washington, DC: National Academy of Sciences; 1978.
6. Frodeman, R.; Holbrook, J. B. Science's social effects. *Issues Sci. Technol.* 23:3; 2007.
7. Holbrook, J. B.; Frodeman, R. Policy dimensions of NSF's criterion 2. *Ogmios: Newsletter of the Center for Science and Technology Policy Research at the University of Colorado, Boulder, No. 13; 2005.*
8. Holbrook, J. B. Assessing the science—society relation:

- The case of the U.S. National Science Foundation's second merit review criterion. *Technol. Soc.* 27(4):437–451; 2005.
9. Holbrook, J. B., ed. Special issue on US National Science Foundation's broader impacts criterion. *Social Epistemol.* 23(3–4); 2009.
 10. Holbrook, J. B. Peer review. In: Frodeman, R.; Thompson Klein, J.; Mitcham, C., eds. *The Oxford handbook of interdisciplinarity*. Oxford, UK: Oxford University Press; 2010.
 11. Holbrook, J. B. Accountable science: The COMPETES Act needs to demonstrate an accountability attitude. *Science Progress*; September 16, 2010. Retrieved from <http://www.scienceprogress.org/2010/09/accountable-science/>
 12. Horrobin, D. F. Something rotten at the core of science? *Trends Pharmacol. Sci.* 22(2):51–52; 2001.
 13. Horrobin, D. F. The philosophical basis of peer review and the suppression of innovation. *J. Am. Med. Assoc.* 263(10):1438–1441; 1990.
 14. Kostoff, R. *Research program peer review: Purposes, principles, practices, and protocols*. Arlington, VA: Office of Naval Research; 2004.
 15. Kronick, D. Peer review in 18th-century scientific journalism. *J. Am. Med. Assoc.* 263(10):1321–1322; 1990.
 16. Lamont, M. *How professors think: Inside the curious world of academic judgment*. Cambridge, MA: Harvard University Press; 2009.
 17. Lamont, M.; Mallard, G. *Peer evaluation in the social sciences and the humanities compared: The United States, the United Kingdom, and France*. Ottawa: Social Sciences and Humanities Research Council of Canada; 2005.
 18. Langfeldt, L. The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Soc. Stud. Sci.* 31(6):820–841; 2001.
 19. Langfeldt, L. The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. *Res. Eval.* 15(1):31–41; 2006.
 20. Laudel, G.; Origgi, G., eds. Special issue on interdisciplinarity. *Res. Eval.* 15(1); 2006.
 21. National Institutes of Health (NIH). *Directory NIH federal advisory committees and HHS advisory committees managed by NIH staff*; 2008. Retrieved from <http://ofacp.od.nih.gov/committee/DirectoryofNIHFedlAdvCmtesNov2008.pdf>
 22. Natural Sciences and Engineering Research Council of Canada (NSERC). *Policy and guidelines on the assessment of contributions to research and training*. Retrieved from http://www.nserc-crsng.gc.ca/NSERC-CRSNG/Policies-Politiques/assesscontrib-evalcontrib_eng.asp
 23. Nature.com. Peer-to-peer: For peer reviewers and about the peer review process. Retrieved from <http://blogs.nature.com/peer-to-peer/>
 24. Orszag, P. Memorandum for the heads of executive departments and agencies: Increasing emphasis on program evaluations (M-10-01). US Office of Management and Budget; 2009. Retrieved from http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-01.pdf
 25. Pielke, Jr., R. A.; Byerly, Jr., R. Beyond basic and applied. *Phys. Today* 51(2):42–46; 1998.
 26. Royal Society of London. *History*. Retrieved from <http://www2.royalsociety.org/campaign/strategic/increase.htm>
 27. Rothwell, P.; Martyn, C. Reproducibility of peer review in clinical neuroscience: Is agreement between reviewers any greater than would be expected by chance alone? *Brain* 123(9):1964–1969; 2000.
 28. Roy, R. Alternatives to review by peers: A contribution to the theory of scientific choice. *Minerva* 22(3–4):316–328; 1984.
 29. Roy, R. Funding science: The real defects of peer review and an alternative to it. *Sci. Technol. Hum. Values* 10(3): 73–81; 1985.
 30. Sandström, U.; Hällsten, M. Persistent nepotism in peer-review. *Scientometrics* 74(2):175–189; 2008.
 31. Shatz, D. *Peer review: A critical inquiry*. Lanham, MD: Rowan and Littlefield Publishers, Inc.; 2004.
 32. Spier, R. The history of the peer-review process. *Trends Biotechnol.* 20(8):357–358; 2002.
 33. Tregenza, T. Gender bias in the refereeing process? *Trends Ecol. Evol.* 17(8):349–350; 2002.
 34. Wennerås, C.; Wold, A. Nepotism and sexism in peer-review. *Nature* 387:341–343; 1997.