

Peer review and the *ex ante* assessment of societal impacts

J. Britt Holbrook and Robert Frodeman

Funding agencies and research councils around the world rely on peer review to assess the potential impacts of proposed research. This article compares the procedures of two major public science agencies – the US National Science Foundation and the European Commission’s 7th Framework Programme – for evaluating *ex ante* the potential societal impact of research proposals. In this paper we survey the state of the art and discuss some of the conceptual questions that arise in using *ex ante* peer review to assess the societal impact of scientific research.

J. Britt Holbrook is Assistant Director of the Center for the Study of Interdisciplinarity and Research Assistant Professor in the Department of Philosophy at the University of North Texas, 1155 Union Circle #310920, Denton, Texas 76203-5017, USA; Email: britt.holbrook@unt.edu.

Robert Frodeman is Director of the Center for the Study of Interdisciplinarity and Professor in the Department of Philosophy at the University of North Texas, USA; Email: frodeman@unt.edu.

Introduction

Peer review is the multi-dimensional evaluation tool of the academy. It is used for everything from determining the quality of articles and books, to whether individual academics are worthy of tenure or promotion, to the viability of entire academic programmes. Moreover, peer review is also used in quasi-academic contexts, including the review of grant proposals and the appraisal of science to be used in regulatory contexts (Jasanoff, 1990).

The peer review of grant proposals requires that we make judgments about research yet to be performed. The difficulty of knowing that one is judging accurately ahead of time prompted Ziman (1983) to characterize peer review as a “higher form of nonsense.” In the view of some, the *ex ante* use of peer review by funding agencies and research councils around the world requires reviewers to become “unwilling futurologists” (Rip, 2000).

Despite the difficulties surrounding predicting the future, the *ex ante* character of peer review is not the central conceptual quandary plaguing grant proposal peer review today. Rather, citing the increasing incorporation of ‘relevance’ as a review criterion at research councils, Rip (2000: 468) claimed that grant proposal peer review is moving toward “still higher forms of nonsense”. Asking scientists to predict the future outside their own areas of expertise – that is, to judge the potential societal impact of scientific research – is seen as even more difficult than asking them to predict the future within their own areas of expertise. Or as Bozeman and Boardman (2009: 189) put it:

In the case of peer review related to BIC [the US National Science Foundation’s Broader Impacts Criterion], the problem is that scientists (and all of us) have no special claim to knowledge of ‘the social good’. Why is the scientist who does research on the genetics of grasses any more qualified to judge social good than the person who mows the grass?

This article explores the incorporation of ‘relevance’, or as we will say, broader societal impact, within the decision-making process of public science funding agencies.

A number of interesting issues are raised by the inclusion of societal impact criteria within peer review:

- What do questions of broader impact imply about what it means to be a peer?
- What is the proper balance between the values of intellectual merit and broader impact?
- Is the peer review of individual proposals the best way to incorporate considerations of relevance into funding decision-making?

This paper, however, focuses on an issue that lies prior to, and to a degree includes, all of these questions: the assumption that assessing impact is more difficult or more problematic than assessing the intrinsic merit of proposed research. Does asking scientists to assess potential societal impact turn peer review into an even higher form of nonsense?

We pose this question via an examination of two prominent public science agencies, the US National Science Foundation (NSF) and the European Commission (EC) 7th Framework Programme (FP7). In doing so we draw on research from our NSF-funded project, the Comparative Assessment of Peer Review (CAPR). CAPR is a four year project (2008-2012) that examines the changing nature of the peer review processes across six public science agencies, three US, two European, and one Canadian, with a particular focus on how these agencies integrate broader societal impacts issues into the review of grant proposals.¹

We limit our discussion to the two agencies mentioned above because of the clear contrast they offer. NSF has increasingly emphasized that proposers and reviewers must address broader impacts. On the other hand, the EC has progressively deemphasized proposer and reviewer judgment regarding societal impact. A comparison of these two agencies allows us to focus on the issue of whether incorporating questions of 'relevance' into peer review makes any less sense than asking reviewers to assess the intellectual merit of proposed research.

Comparing how different funding agencies incorporate the assessment of societal impact into the peer review process helps us come to grips with crucial issues concerning *ex ante* assessment. Each agency is unique, and there are often significant differences even within a particular agency in terms of how the peer review process actually works. Nevertheless, based on our research, and to preview our argument, we offer a set of summary statements on the state of the art regarding the use of peer review to render *ex ante* judgments of the potential societal impact of proposed research:

1. Science agencies around the world are placing increasing emphasis on funding research that has clearly identified potential benefits to society.
2. Many agencies use the peer review process to assess the potential societal impacts of the research they fund.
3. Agencies often encounter resistance from both proposers and reviewers to the incorporation of societal impacts considerations into the peer review process.²
4. Despite the well-documented resistance on the part of the scientific community to including impacts criteria in peer review, there is little evidence to suggest that peer review is any less effective at *ex ante* assessments of societal impact than it is at *ex ante* assessments of scientific, technical, or intellectual merit.

5. In response to pressures from both the scientific community and society at large, agencies continue to experiment to find better ways to include societal impacts considerations in *ex ante* research evaluation.

We conclude this introduction with a few points concerning methodology. Our research attends more to the process than to the outcomes of assessment. It is natural to want to know whether the process of peer review is successful at predicting the outcomes of research impact. But answering this question would involve a study orders of magnitude larger than our own. It would require access to information that funding agencies either generally do not make publicly available (for instance, the actual content of research proposals, the reviews those proposals receive, and so forth) or information that many funding agencies do not carefully track. Although the academic outputs of research, such as publications or patents resulting from grant-funded research are often tracked by agencies, reporting requirements for broader societal impacts are generally much less stringent, if they exist at all.

In any case, the review process itself, including both the procedures set up by the agencies and the reactions of proposers and reviewers to impact considerations, reveals important underlying philosophical commitments concerning the nature of research, expertise, and the relation between science and society. It is these philosophical commitments that we see as most determinative of the overall science-society relationship – and least attended to by the parties involved. These underlying commitments factor into the proposal and review process, as well as into the ultimate funding decisions, in important and generally unrecognized ways. In this paper, we hope to throw some light on why and how these underlying philosophical commitments influence both the process of peer review and attitudes toward the inclusion of impact considerations in the process.

In the following sections we summarize the peer review process as it functions at NSF and the EC. This is followed by a discussion that draws points of comparison between the two agencies, and some final concluding remarks. The descriptions of the peer review processes at both agencies are drawn from administrative documents obtained from each agency, from presentations by and discussions with participants at two CAPR workshops (the first held in Washington, DC in April, 2010 and a second held in Brussels in December, 2010), site visits by CAPR team members to both agencies, and a survey of stakeholders in the peer review processes at these agencies.³

US National Science Foundation – Broader Impacts Merit Review 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 in the US by the National Institutes of Health). With an annual budget of more than \$7 billion, NSF is the funding source for approximately 20 percent of all federally supported basic research conducted by US colleges and universities. Over 95 percent of the funding proposals submitted to NSF (now routinely over 40,000 annually) undergo the process of what NSF terms ‘merit review’ rather than ‘peer review’ – this to indicate that review by peers is merely part of the larger funding decision-making process, which also includes NSF staff. In other words, at NSF, peer reviewers do not make funding decisions.

Peer review takes place in one of three ways at NSF: via mail-only, panel-only, or combination (mail and panel) methods. In mail-only reviews, proposals are distributed to reviewers (typically via email) who then submit their comments to NSF without having had any contact with other reviewers. In panel-only reviews, NSF convenes a panel of reviewers at NSF headquarters who meet to discuss a group of

proposals. Combination reviews send proposals out to individual reviewers for comment, followed by panels convened to discuss the group of proposals. Since 1997, the percentage of proposals that receive combination reviews has remained relatively steady at around 30 per cent. However, the percentage of mail-only reviews has decreased during that same period from around 25 per cent to under 10 per cent, while the percentage of panel-only reviews has risen from around 40 per cent to near 60 per cent (Lightbourne, 2010).

Every proposal that goes through NSF's process of merit review is judged according to two generic merit review criteria (NSF 1):

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, original, or potentially transformative 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?*

Mail reviewers receive the following instructions from NSF (NSF 2):

Comment in detail on the quality of the proposal:

- Provide an overall rating of the proposal
- Identify the proposal's strengths and weaknesses for each NSF Merit Review Criterion:
 - What is the intellectual merit of the proposed activity?
 - What are the broader impacts of the proposed activity?
- Provide a summary statement that includes the relative importance of the two criteria in assigning your rating. (You do not have to weigh the criteria equally.)

In addition to the two generic criteria of intellectual merit and broader impacts, reviewers are asked to consider any additional program-specific criteria outlined in the request for proposals. NSF's Broader Impacts Criterion is specifically focused on societal impacts. However, what specific broader impact a proposer chooses to address is left to the proposer. That is, the proposer may make a case that the activity they propose falls under the rubric of broader impacts based on any sub-section of the multi-part broader impacts criterion.

Reviewers are asked to comment on the strengths and weaknesses of a proposal relative to the Broader Impacts Criterion, as well as to provide an indication of the relative importance they place on broader impacts and intellectual merit. Reviewers are explicitly told that they need not weigh the criteria

equally, and reviewers are asked to comment only on aspects of the proposal they feel qualified to judge. Proposers and reviewers of NSF proposals are thus granted a great deal of leeway in terms of their specific use of the Broader Impacts Criterion. We discuss the intentional vagueness of the Broader Impacts Criterion in the Discussion section, below.

European Commission Framework Programmes – Expected Impacts

EC funding for scientific research is organized via multi-year Framework Programmes. The EC is currently in the midst of its 7th Framework Programme (FP7), which runs from 2006-2013, with funding of approximately 50 billion Euros. Each Framework Programme is agreed on by the Council of European member states and the European Parliament. The general areas of research to be funded under the rubric of a specific FP are determined top-down – that is, by policymakers rather than by members of the scientific community. Objectives and topics for research are set out in annual work programmes and addressed through specific calls for proposals. The generally top-down determination of research priorities by the EC has been recently counter-balanced by the creation of the European Research Council (ERC) in February 2007 – a bottom-up, investigator-driven element, the Ideas Programme – as part of FP7.

The EC appoints roughly 10,000 independent expert reviewers per year and receives over 16,000 proposals annually. The peer review process itself consists of multiple steps. Minimally, proposals undergo review first by individual reviewers acting independently. This is followed by a consensus discussion, which takes place at the Covent Garden facility in Brussels, of those same experts (the ERC follows a different procedure). Finally, a panel of reviewers establishes a priority order for funding. Within the constraints of budget, the EC follows the advice of the peer reviewers, who thereby essentially make funding decisions.

Instructions to reviewers, whom the EC refers to as ‘evaluators’, are extensive. They not only offer instructions in the evaluation procedure and descriptions of the basic evaluation criteria, but also outline the entire evaluation process as well as the role of the evaluator and the sorts of virtues evaluators ought to embody in their reviews (CORDIS 1). Despite the detailed instructions to evaluators, however, most of the details about particular evaluation criteria are contained in the annual work programmes. Thus, although the instructions to evaluators say that ‘impact’ is one of the criteria for evaluation, reviewers are also instructed to “refer to the applicable list of impacts specified in the Workprogramme [*sic*]” (CORDIS 1, 2008: 7).

Each annual work programme contains not only an account of the sort of activities to be funded by means of a specific call for proposals, but also a detailed description of the review criteria to be used by reviewers of calls relevant to the specific work programme. The criteria typically include the scientific and technical quality of the research, the resources for implementing the research project, and a description of the specific ‘expected impacts’ of each of those particular actions. Proposers are asked to address and reviewers to judge the expected impacts of the proposed research with reference to the set of expected impacts outlined in the specific work programme. As a result, what impacts are expected will vary according to the specific call a proposal is responding to – and the expected impacts will sometimes be concerned with broader societal issues, sometimes not. For instance, in the case of the ERC, scientific excellence is the sole criterion, and impacts are expected only on science.

The following excerpts will give the reader some idea of the range of expected impacts across FP7:

From the Science in Society Work Programme (Part 5 of Capacities):

Expected impact: Research outcomes will contribute to a better insight on how to use ethical expertise in the policy-making process at the European and national levels and in particular how public policy should address the issue of human enhancement. (CORDIS 2, 2009: 10)

From the Nanosciences, Nanotechnologies, Materials and New Production Technologies –NMP Work Programme (Theme 4 of Cooperation):

Expected impact: (i) Solutions going well beyond the state-of-the-art in terms of cost (target of far below 1 €/Wp) and efficiency; (ii) Stimulation and acceleration of the industrial take-up of promising results beyond laboratory scale; (iii) New competitive industrial processes. (CORDIS 3, 2010: 14)

From the Marie Curie Action: Intra-European Fellowships for Career Development (IEF) Work Programme (Activity 2 of People):

Expected impact of the action: Rather than merely providing employment opportunities for *experienced researchers*, this action aims to catalyse significant development in researchers' careers, specifically by adding different and/or complementary research competences at an advanced level, in the process of reaching and/or reinforcing a position of professional maturity or to permit them to resume a research career. Projects are therefore expected to add significantly to the career development of the best and most promising researchers active in Europe, in order to enhance and maximise their contribution to the knowledge-based economy and society. (CORDIS 4, 2010: 15)

Annual work programmes also lay out the predetermined weighting of the criteria, as well as thresholds for each criterion (for instance, to be considered for funding a proposal must obtain a minimum score of 3 out of 5 for expected impact, regardless of the weighting of the expected impact criterion). These weightings can vary from one specific activity to another even within a particular work programme (see the People program for examples).

Comparison of NSF and the EC

A comparison of the peer review procedures of NSF and the EC reveals similarities and differences that illuminate factors to be considered in any modification of the peer review process. Both NSF and the EC have explicit societal impacts criteria. Moreover, in the case of both agencies, these criteria go beyond the scope of concern for education and public outreach.⁴ These similarities bring the differences between the two agencies into greater relief.

One difference between NSF and the EC involves their approaches to the issue of weighting their impact criteria. NSF has explicitly chosen not to assign a particular weight to either the Intellectual Merit Criterion or the Broader Impacts Criterion. The EC, on the other hand, assigns a specific weight to the expected impacts of the proposed research, as well as to the other peer review criteria. If no weighting is specified, the EC weights each criterion equally; whereas for NSF, no weighting means that the decision is left to the individual reviewer. In terms of weighting, then, the EC is much more directive

than NSF. Indeed, while the EC is intentionally clear about how and how much considerations of impact should influence each review, NSF is intentionally vague on this point.

In addition to the fact that NSF does not assign a specific weight to considerations of intellectual merit or broader impacts, NSF asks reviewers to comment only on those areas of the proposal they feel competent to judge. This leads to 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 (Holbrook, 2005). This also points to a potential downside in NSF's intentional vagueness regarding impacts. NSF has recognized this tendency and has – without altering the Broader Impacts Criterion itself – increasingly emphasized that broader impacts must be addressed in both proposals and reviews (Holbrook, 2009).

The EC, on the other hand, has taken a different pathway to impacts, as the following historical account shows. In addition to reviewers, the EC includes independent observers during review panel meetings, who offer their recommendations to the Commission regarding the peer review process itself (Cross, 2010). In large part on the basis of recommendations from these independent observers, the EC's incorporation of societal impacts considerations as part of the peer review of proposals has undergone significant change over the course of successive Framework Programmes.

The 5th Framework Programme (1998-2002) used five review criteria, three of which concerned societal impacts: (a) Community added value and contribution to EU policies; (b) contribution to Community social objectives; and (c) economic development and science and technology prospects (Cross, 2010). In addition to these criteria that received a score, FP5 also included additional unscored criteria that reviewers could comment on regarding ethical considerations and Community policies. Observers suggested that the FP5 criteria were too complicated. There were too many criteria, and the inclusion of both scored and unscored criteria created confusion among reviewers, leading to the need for both scientific and “strategic” experts, the latter of which were difficult to identify (Cross, 2010).

In a move to correct the deficiencies observers found in FP5, FP6 (2002-2006) included only one scored ‘potential impact’ criterion. All other scored criteria related to either scientific and technical merit or relevance to the specific call for proposals. FP6's ‘potential impacts’ included considerations of whether the proposed research reinforced competitiveness or solved societal problems, the extent to which plans for innovation, exploitation, and dissemination of results were adequate to ensure optimal use of the research results, and the ‘European added-value’ of the research. FP6's ‘potential impacts’ criterion was in fact quite similar to NSF's Broader Impacts Criterion, in terms of both its form (one criterion with many facets) and its content (dissemination of results, benefits to society).

FP6 still included several unscored criteria that allowed reviewers to comment on ethical, safety, or gender issues, public engagement, education, and ‘3rd country’ – non EU – involvement. Observers found the FP6 criteria an improvement over FP5, although they continued to note that the unscored criteria produced confusion (Cross, 2010). With the transition to FP7 the meaning and scope of impact became considerably narrower. As described above, ‘impact’ was now restricted to the potential contribution of the proposed research to a list of ‘expected impacts’ outlined in the annual work programmes and calls for proposals.

The progression from FP5 to FP7 can be seen as marking the progressive limitation of the scope of societal impacts criteria as part of the EC's peer review process. This interpretation of events is favoured

by EC Policy Officer Alan Cross (2010), who claims: “This limitation of the meaning and scope of ‘impact’ is a reflection of the following ‘lesson learned’: ‘Don’t try to make EU policy via evaluation criteria.’” On the other hand, it is also plausible to interpret the progression from FP5 to FP7 as a further specification, rather than a limitation or even phasing out of impacts considerations. Moreover, the trajectory from FP5 to FP7 heightens consideration of impacts insofar as it moves judgments about what impacts are to be considered toward the top (policymakers) and away from the scientific community.

During the same period, NSF can be seen as staying the course with the Broader Impacts Criterion, which has not been modified since its introduction in 1997.⁵ Indeed, a case can be made that the scope of the broader impacts criterion has steadily increased since its introduction. Broader impacts at NSF has both become more central to the review process and expanded beyond its early focus on education and public outreach (Frodeman and Briggie, under review). The steady increase of attention to broader impacts on the part of the scientific community has not, however, decreased policymakers’ attention to NSF’s Broader Impacts Criterion. Indeed, the America COMPETES Reauthorization Act of 2010 includes specific suggestions by lawmakers as to how the Criterion ought to be applied (Holbrook, 2010).

In terms of both weighting considerations of impact, and of specifying considerations of impact, the EC is much more directive than NSF. As noted above, the EC is more top-down in managing the review of grants: each work programme specifies the expected impact and the weighting of impact relative to other criteria. NSF, on the other hand, leaves both the weighting and scope of assessing broader impacts in the hands of the reviewers, using a generic impacts criterion that applies to all grant proposals and allowing consideration of impacts to be determined largely bottom-up, by members of the scientific and engineering community who serve as reviewers.

Conclusion

Peer review that combines considerations of intellectual merit and societal impact is the state of the art for conducting *ex ante* assessment of research proposals. But the inclusion of broader impacts considerations within peer review remains controversial. Scientists generally dislike impacts considerations, which they often see as both challenging their authority and undermining the autonomy of the scientific enterprise (see, for instance, NAPA, 2001; APSN, 2007; *Times Higher Education*, 2009; *Nature*, 2010). On this view, integrating societal impacts into the peer review process threatens to erase the boundary between science and politics. It mixes categories and creates conceptual confusion. And it takes scientists beyond the bounds of their disciplinary expertise.

While more sympathetic to the new requirements, policymakers at science agencies such as NSF and the EC have also found that the inclusion of societal impacts concerns within peer review presents difficulties. Typically scientists themselves, they may share the reactions of the scientific community to the criterion. But inhabiting the worlds of science and policy, they recognize the necessity of implementing this dual mandate to fund the best science that will also have the biggest impact on society.

The central fact today concerning the peer review process at both NSF and the EC is that societal impact is viewed as more difficult for peers to evaluate than scientific merit. In our experience, however, no evidence is offered in support of this claim. It may seem intuitively obvious that peer review is less effective at the *ex ante* assessment of societal impact than it is at *ex ante* assessment of intellectual merit. Moreover, there is widely reported resistance on the part of proposers and reviewers to

addressing societal impacts. But these failures and resistance seem more a matter of group psychology or philosophic presumptions than epistemology.⁶

Imagine the following scenario. A proposal states the goal of increasing our understanding of the dynamics underlying vertebrate mortality in high altitude streams under stress from a variety of factors such as pollution, fishing, and climate change. The grant also proposes to train graduate students in the connections between ecology and policy, and to inform local communities of the results of this research. Is it really easier to judge *ex ante* the intellectual merits of this proposal as compared with its broader impacts?

The claim that intellectual merit is easier to judge gains its plausibility from academic assumptions concerning expertise. Experts within a discipline are seen as competent to judge proposals within their area of expertise. But judging intellectual merit is a tricky process. The decision-making process within peer panels – or within the peer evaluation of articles and monographs, or for hiring and promotion – is susceptible to a variety of influences. Strong personalities or prestigious individuals can dominate the process, and differences in personal and professional outlook often result in quite different judgments (Lamont, 2009). The claim that the process is objective because everyone present shares a common disciplinary framework obscures the facts that every biologist differs in training and background, specializes in a different field, and comes from a different generation, circle, or tradition. Agreement among biologists may be enough to keep Intelligent Design theory out of the peer reviewed literature, but it does not reach down far enough to guarantee consensus concerning proposals that exist within the bounds of standard biological practice.

NSF and the EC today present two positions in the ongoing negotiation between the goals of accountability and autonomy.⁷ The EC removes some of the responsibility from reviewers, placing it in the hands of policymakers by specifying expected impacts and predetermining the weighting that considerations of impact can have in the review. The EC also emphasizes consensus judgments of reviewers. The point is to reach a judgment that everyone can agree to as efficiently as possible. Once that consensus judgment is reached, the EC follows it – but only after largely predetermining the terms of the debate. The EC process emphasizes accountability to the concerns of policymakers at the expense of the autonomy of the scientific community (both proposers and reviewers) to determine impact.

NSF, on the other hand, emphasizes the discretion of the reviewer in the consideration of impacts. The responsibility for articulating impacts is placed on proposers, and much of the responsibility for judging those impacts is placed on reviewers. Neither the scope nor the weighting of impacts is tightly controlled by NSF. Since NSF officials make the actual funding decisions, consensus among reviewers is not required. This leaves more room for individual judgment regarding the broader impacts of proposed research. It also leaves room for proposers and reviewers to pay little attention to broader impacts. NSF thus emphasizes the autonomy of the scientific community at the expense of accountability to policymakers.⁸

NSF and the EC have each designed their review processes in order to balance the competing values of autonomy and accountability. Each agency has also reacted – albeit in different ways – to the perception that assessing impacts is somehow more difficult than assessing intellectual merit. If we are correct, however, that assessing impacts is in principle no more difficult than assessing intellectual merit, then the explanation for resistance to its implementation lies beyond simple epistemology. Scientists have

preserved their autonomy by claiming a special ability to judge that has been institutionalized in the disciplinary nature of peer review. If accountability demands that others have a voice in the decision-making process, then peer review becomes interdisciplinary and transdisciplinary as well as disciplinary in nature. Who, then, ought to count as a peer?

Scientists ought to play a central role in determining what research gets funded. But this will only continue to be possible if scientists also embrace the fact that their research can be judged on its potential societal impacts as well as its intrinsic intellectual merit. And unless scientists embrace their own ability to judge impacts, their role in the decision-making process will increasingly be transferred to others.

Acknowledgements

The authors gratefully acknowledge the support for this research provided by the US National Science Foundation under Grant No. 0830387. The authors also wish to express their appreciation to NSF and EC officials who contributed to this research, while emphasizing that any opinions, conclusions, and recommendations expressed here are those of the authors and do not necessarily reflect the views of NSF or the EC. The authors also thank Brunel's Health Economics Research Group (HERG) for hosting the international two-day workshop on 'State of the Art in Assessing Research Impact' and for including our presentation in the workshop. Comments received during the workshop were very helpful in formulating this paper. Special thanks are due Claire Donovan for inviting us to participate in the workshop and to Martin Buxton for his insightful criticisms. Finally, the authors thank the anonymous reviewers for their helpful comments.

Notes

¹ The six agencies are: the US National Science Foundation, National Institutes of Health, and National Oceanic and Atmospheric Administration; the European Commission 7th Framework Programme, and the Dutch Technology Foundation; and the Canadian Natural Sciences and Engineering Research Council. More can be learned about CAPR at <<http://csid-capr.unt.edu/>>.

² We discuss this point in more detail, below, in comparing NSF and the EC. For a discussion of resistance to Broader Impacts at NSF, see especially Holbrook (2005).

³ Concerning the survey, participants from the six funding agencies that are part of the CAPR project (NSF, NIH, NOAA, NSERC, STW, and the EC) and the general public were invited via individual e-mailed invitations, listserv postings, website and Facebook advertisements throughout 2010. 490 usable responses were collected, which included 145 NSF participants and 77 EC participants. Publication of results is in progress (Holbrook and Hrotic, under review).

⁴ For a discussion of NSF's Broader Impacts Criterion and its relation to education and outreach, see Holbrook, 2005; Frodeman and Holbrook, 2007; Holbrook and Frodeman, 2007; and the essays contained in the 2009 special edition of *Social Epistemology*, (Holbrook, 2009).

⁵ Though, the criteria are likely to change in 2012. See note 8, below.

⁶ Participants in the CAPR survey were asked pairs of questions regarding "intrinsic" and "instrumental" value of research. Participants were asked to rate how strongly they agreed with statements on a 7-point Likert scale (1 = "strongly agree," 7 = "strongly disagree"). The first pair of statements asked how comfortable they felt assessing intrinsic and instrumental value. Results showed significantly more comfort reported for assessing intrinsic value: "comfortable assessing INTRINSIC" (M=2.12, SD=1.34); ". . . INSTRUMENTAL" (M=2.78, SD=1.55); Z=-6.42, p<0.01. However, the effect is quite small, and both statements were between "slightly agree" and "agree" on the scale used. When asked a related pair of questions assessing if participants thought intrinsic and instrumental *could* be

measured, there was no significant difference: “INTRINSIC cannot be measured” (M=4.77, SD=1.71); “INSTRUMENTAL . . .” (M=4.87, SD=1.58); $Z=-1.07$, $p=0.29$ (see Holbrook and Hrotic, under review).

⁷ For a related comparison of 16 national mechanisms for evaluating government-funded research, see Coryn *et al*, 2007.

⁸ In June of 2011 NSF issued new draft merit review guidelines that remove much of the flexibility described here. As this article goes to press, the comment period is still open, and no final determination concerning the merit review criteria has been made. See Frodeman and Holbrook (2011) for a discussion.

References

- APSN 2007 NSF's Broader Impacts Criterion gets mixed reviews. *American Physical Society News* **16**(6), June.
- Bozeman, Barry, and Craig Boardman 2009 Broad impacts and narrow perspectives: passing the buck on science and social impacts. *Social Epistemology* **23**(3-4), 183-198.
- CORDIS 1 2008 <<http://csid-capr.unt.edu/fedora/repository/capr:339/-/FP7--European%20Commission%20Instructions%20for%20Evaluators>>, last accessed 30 June 2011.
- CORDIS 2 2009
<http://cordis.europa.eu/fp7/dc/index.cfm?fuseaction=UserSite.FP7DetailsCallPage&call_id=271#infopak>, last accessed 30 June 2011.
- CORDIS 3 2010 <ftp://ftp.cordis.europa.eu/pub/fp7/docs/wp/cooperation/nmp/d-wp-201101_en.pdf>, last accessed 30 June 2011.
- CORDIS 4 2010 <ftp://ftp.cordis.europa.eu/pub/fp7/docs/wp/people/m-wp-201101_en.pdf>, last accessed 30 June 2011.
- Coryn, Chris L S, John A Hattie, Michael Scriven, and David J Hartmann 2007 Models and mechanisms for evaluating government-funded research: an international comparison. *American Journal of Evaluation* **28**(4), 437-457.
- Cross, Alan 2010 Peer review in the EC Framework programme: and assessing impact. Presentation delivered at CAPR Workshop, Washington, DC, April 23, 2010. <<http://csid-capr.unt.edu/fedora/repository/capr:350>>, last accessed 30 June, 2011.
- Frodeman, Robert and J Britt Holbrook 2007 Science's social effects. *Issues in Science and Technology* **23**, 28-30. <http://www.issues.org/23.3/p_frodeman.html>, last accessed 30 June 2011.
- Frodeman, Robert and J Britt Holbrook 2011 NSF and public accountability. *Science Progress*, June. <http://www.scienceprogress.org/2011/06/nsf-and-public-accountability>, last accessed 30 June 2011.
- Holbrook, J Britt 2005 Assessing the science-society relation: the case of the U.S. National Science Foundation's second merit review criterion. *Technology in Society* **27**(4), 437-451.
- Holbrook, J Britt and Robert Frodeman 2007 Answering NSF's question: what are the broader impacts of the proposed activity? *AAAS Professional Ethics Report* **XX**(3), summer. <<http://www.aaas.org/spp/sfpl/per/newper50.shtml#NSF>>, last accessed 30 June 2011.
- Holbrook, J Britt 2009 *Social Epistemology: A Journal of Knowledge, Culture and Policy. Special Issue: US National Science Foundation's Broader Impacts Criterion* **23**(3-4).
- Holbrook, J Britt 2010 Accountable science: the COMPETES Act needs to demonstrate an accountability attitude. *Science Progress*, September. <<http://www.scienceprogress.org/2010/09/accountable-science/>>, last accessed 30 June 2011.
- Jasanoff, Sheila 1990 *The Fifth Branch: Science Advisers as Policymakers*. Cambridge, MA: Harvard University Press.
- Lamont, Michelle 2009 *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.
- Lightbourne, Jim 2010 NSF merit review criteria and process. Presentation delivered at CAPR Workshop, Washington, DC, April 22, 2010. <<http://csid-capr.unt.edu/fedora/repository/capr:354>>, last accessed 30 June, 2011.
- NAPA, 2001 National Academy of Public Administration (NAPA). *A Study of the National Science Foundation's Criteria for Project Selection: A Report by the National Academy of Public Administration for the National Science Foundation*. <<http://csid-capr.unt.edu/fedora/repository/capr:1098>>, last accessed 30 June, 2011.
- Nature 2010 Science funding: science for the masses. *Nature* **465**, 416-18. <<http://www.nature.com/news/2010/100526/full/465416a.html>>, last accessed 30 June 2011.

NSF 1 <http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_3.jsp#IIIA1>, last accessed 30 June 2011.

NSF 2

<https://www.fastlane.nsf.gov/NSFHelp/Printdocs/FastLane_Help/pd_proposal_review/pd_proposal_review.pdf>, last accessed 30 June 2011.

Rip, Arie 2000 Higher forms of nonsense. *European Review* **8**(4), 467-486.

Times Higher Education 2009 Petition decries 'impact' agenda in research. *Times Higher Education*, 11 June. <<http://www.timeshighereducation.co.uk/story.asp?storyCode=406931§ioncode=26>>, last accessed 30 June 2011.

Ziman, John 1983 The collectivization of science. *Proceedings of the Royal Society* **B219**, 1-19.