Sitting in Judgment: Myths and Realities of Peer Review

April 05, 2002

It's burdensome, it's time-devouring, it plays havoc with your life and your research, it stifles innovation, it over-values flaws and undervalues potential, and the pay is somewhere between paltry and nonexistent.

Why in the world would anyone, let alone our nation's most successful scientists, subject themselves to anything like that? Yet, literally tens of thousands do so each year, including a sizeable cadre of psychological investigators.

"It," of course, is peer review, the process by which grant proposals are evaluated anonymously en route to life-or-death funding decisions.

"The first time you do it, it's an honor, and there's curiosity," says APS Charter Member Lauren Adamson, Chair of the Department of Psychology at Georgia State University. "By the tenth time you do it, you kind of scratch your head and wonder why."

But, she quickly concedes, she does know why: "For me, it's my contribution to the community of scientists. It's a very central way in which work gets done, getting feedback at different points. Also, there's a finite amount of resources being distributed, and people who care a lot about these topics should be a voice for what they think is the best quality work."

APS Fellow Toni Antonucci, of the University of Michigan's Institute for Social Research, agrees: "I honestly believe it's one of the most important things we do. If you're funding lousy research, your field is the worse for it, and if you're funding high quality work, your field is the better for it."

The Center for Scientific Review (CSR), overseer of most of the peer review process at the National Institutes of Health (NIH), has about 150 standing study sections, plus special emphasis panels (SEPs) that are convened to meet new or emerging needs. Together, these panels use some 10,000 scientists as reviewers. And that does not include those who sit on review panels for the individual NIH institutes.

The population that devotes time to deciding the future of scientific investigation reaches awesome proportions when you consider that another 50,000 individuals are evaluating the 30,000 proposals received each year by the National Science Foundation (NSF), not to mention an untold number of others who are involved in the review processes of other federal research agencies.

It should not be surprising that such a vast network of highly honed and highly motivated minds, invited to pick apart others' ideas and work plans, has spawned some criticism. And it should be equally unsurprising that some of the complaints about the process are myth, some are quite real, and some are a little of both.

REALITY: IT'S A LOT OF WORK

"It's very time consuming, no question about that," says Constance Atwell, the Associate Director for Extramural Research at the National Institute of Neurological Disorders and Stroke (NINDS).

Here's how the process works at NIH: In the recent past, an intimidatingly-large box containing 70 or so proposals would arrive on the doorstep of each reviewer three times a year. Nowadays, they usually arrive on a compact disk, except that the three or four reviewers assigned primary responsibility for a proposal still get the hard copies of those proposals. Over the ensuing six or seven weeks, each panelist will spend the equivalent of a solid week reading proposals and writing comments.

"It's a lot of work and takes a lot of time," says Antonucci. "It wreaks havoc with the rest of your life. The problem is your competing duties. It's really tough."

Then they meet at a hotel – typically for two days – in or near Washington DC, to discuss their findings and score the applications. It should be noted that not every application will receive detailed discussion and scoring. In a process referred to as "streamlining," if reviewers agree that a grant application doesn't warrant further consideration, then the reviewer's individual critiques of the strengths and weaknesses of the application are sent to the applicant. Although streamlining has reduced the length of study section meetings from 2.5 or 3 days to 1.5 days and has had other positive effects, the process is not without controversy. While NIH indicates that having an application streamlined and returned does not constitute disapproval, the reality is that it can discourage new applicants who may feel otherwise. In addition, there may be inconsistencies among review panels in their use of streamlining. But even the seeming rejection of being streamlined has a positive side: the critiques can provide valuable feedback about the strengths and weaknesses of a proposed project, which the applicant can use for purposes of revision. As with the non-streamlined applications, if the applicant believes that the review panel made a significant mistake in critiquing the application, there is an appeals process.

At NSF, review panels are organized by scientific program areas and typically have between five and 10 members each. Most psychological proposals are channeled into one of five program areas – social psychology, cognition and perception, developmental and learning sciences, linguistics, and cognitive neuroscience. (Psychologists also have seats on other panels that review some psychology-related proposals.) An estimated 40 or so psychologists in all are members of NSF review panels.

Each NSF panel meets twice a year to review anywhere from 60 to 120 proposals, with each panelist assigned primary responsibility for 10 to 20 of them. In the end, the workload is about the same as at NIH, says Steven J. Breckler, NSF program director for social psychology. At NSF, however, the process is now almost entirely paperless – handled start to finish either via the Internet or on compact disks. NIH is heading in that direction, as well.

REALITY: THE PAY IS POOR

"You don't do it for the money," says Antonucci. NIH gives an honorarium of $200 a day for the meetings – plus expenses – but the bulk of the work is done in advance.

Fernanda Ferreira, of Michigan State University, is typical. Ferreira estimates she spends four to eight hours reviewing each of about 24 grant proposals a year for a study section on language, plus another 18 hours in each of three two-day committee meetings a year. That's 150 to 246 hours a year for $600 –

between $2.44 and $4 an hour. And at NSF, it is even more a labor of love: only the expenses are covered.

MYTH: PSYCHOLOGISTS "EAT THEIR OWN"

"I know that there is a common perception that psychologists are much harder on each other than are people in other disciplines," says APS Fellow and Charter Member Keith Rayner, of the University of Massachusetts-Amherst. "As far as I know, this is one of the reasons that NIH went to a system in which the percentile score is more important than the absolute score assigned to a proposal. My own opinion is that reviewers are not out to 'get' anyone and that there is a remarkable amount of fairness associated with the review process."

Most of the others the *Observer* talked with seem to agree.

"I have never seen any evidence of that," says the NSF's Breckler. "I suspect if you go to any field of science, they'll say the same about their field. When I arrived at NSF six years ago, I worried a lot about this. I sat in on panel meetings in a lot of fields – mathematics, chemistry, biology, engineering – and I didn't see that psychologists were being any more or less hypercritical of their own than you'll find in any other discipline."

Rafael Klorman, an APS Fellow who specializes in psychopathology at the University of Rochester, says he doesn't know what it's like in other fields, but "I have sat on committees made up exclusively of psychologists as well as committees with members from other disciplines, but always dealing with psychological processes, and I think that the committees were generally quite critical, though always trying to be fair."

John D. Gabrieli specializes in behavioral neuroscience at Stanford University. Appointed to an NIH committee that reviews animal and basic neuroscience, as well as human psychological research, he says "it is not obvious to me that the critiques are any more or less stringent for the psychological versus more biological applications."

Adamson agrees that in her specialty, developmental disabilities, review panels she has been on include neurologists, pediatricians and others besides psychologists, but "I don't detect that psychologists are any harder on grants than the others. I don't think there's a disciplinary bias."

APS Member Marlene Behrmann, Carnegie Mellon University, said she has found the review process at NIH to be "judicious, unbiased and fair, and extremely well managed. I am very impressed with the system." Behrmann's perspective on peer review is particularly broad; in addition to serving on a succession of NIH panels, she has been an *ad hoc* reviewer for such other agencies as the Israel Science Foundation, the Canadian Medical Research Council and the Wellcome Trust.

However, Ferreira says she "can directly compare linguists and psychologists and there's no question the psychologists are a bit tougher on each other. The thing about psycholinguistics is that the theory, the methodology, the stimuli, the significance and importance – everything has to be top-notch. I think the same general thing is true for other areas of psychology as well.

"I also think that psychologists are trained to be super-analytical, and that causes them to be super-

critical. I'm not sure this is a bad thing, but I'm pretty confident that often very good work does not end up being funded, both for lack of funds and because of this sometimes excessive harshness."

Atwell, at NINDS, says she has heard the complaint, too, and agrees that "you'll probably hear the same comment from virtually every field, that they feel their own is unusually competitive." She explains that, at least in behavioral sciences, sometimes the "tough love" approach is because "reviewers feel they have to establish scientific rigor by being very critical."

Antonucci also believes psychological scientists "are less forgiving than we need to be. It's important to go into this looking for the pearl of wisdom we might get (from a grant proposal) than to look for the fatal flaw. If there is a real fatal flaw, it needs to be addressed, but I think we jump to that conclusion a little too often."

Ultimately, she says, it depends on a study section's Scientific Review Administrator (SRA), the person responsible for recruiting reviewers and setting the tone. "I've been impressed. The people I've worked with have been very good at setting the tone. They'll tell you, 'We have an important job here, and we need to do it in a scholarly, serious way.' It is very respectful, if done well. It is agreed in advance that you are allowed to and expected to disagree. The question is with what tone do you disagree.

"I have been in a couple of meetings where people have felt their reputation was at stake, and it was a personal insult if the committee didn't agree with them. And I've been on others in disagreement where the tone was, instead, 'Oh I'm so sorry. I must have missed something,' as opposed to 'You're wrong and I'll prove it to you.' You can just feel the difference."

It also depends on who is recruited – or accepts appointment – as a reviewer. "If you know you have trouble being fair, I wouldn't recommend it," Antonucci says. "If you have an open mind it's good, but if you only believe in your own approach, you're not going to be a good reviewer. Some believe their view is the only view. We all believe that a little bit. After all, you have to believe strongly in yourself to devote so much of your life to a particular research effort. But there has to be room for other opinions. The rule of thumb has to be not whether you agree with the hypothesis, but whether the mechanism for testing the hypothesis is valid. We should not be deciding whether it's right or wrong before they get to test it."

MYTH OR FACT? IT STIFLES INNOVATION

The story is told that Johannes Kepler once wrote Galileo to suggest a theory he was working on. The senior astronomer dismissed Kepler's idea as youthful nonsense. Kepler's "nonsense," of course, was that the moon causes the tides, a discovery that eventually led him to such firsts as correctly explaining celestial mechanics.

That tale is sometimes used to buttress the notion that senior scientists should not be passing judgment on the work of their younger peers, whose ideas may clash with established notions. Antonucci believes otherwise, that senior scientists do make good reviewers as long as they're "willing to allow some divergence from what might be their traditional approach."

"We certainly hear this all the time, that old people are set in their ways and don't appreciate new ideas," concedes Atwell, but she says there's a counter-complaint as well. "The other version is that

reviews now are dominated by younger people who don't have the seasoning and the broad picture, who don't yet know what's important, so they aren't willing to take a chance. There's obviously a kernel of truth in any legend or myth, but it also obviously doesn't characterize the whole system."

The critics themselves, she says, are often guilty of what they criticize. "I've seen those who level that criticism, when they are themselves in the role of reviewers, drop right back into the same behavior as everyone else. Even though everybody salutes the flag of innovation, not everybody acts on it."

Reviewers who find it difficult to select from among so many top-notch applications, she says, fall back on the tried-and-true because it's easier to predict which studies will be successful than to gamble on those that might not be.

"That's not what we in the institutes want," Atwell says. "In my opinion, if all the grants we fund are successful, we haven't done a good job in selecting the ones we should fund. The very nature of research is that it's not always successful. We learn as much from what goes wrong or from unexpected results as we do if everything works exactly as anticipated. There's no impetus from the institutes for this conservative approach. The problem is to get reviewers themselves to take more of a risk."

Ferreira says she believes that "sometimes the process does end up keeping good work from being funded" even though overall it helps advance science. "For instance, if an innovative proposal goes to a panel that is theoretically hostile to its perspective, then that research will probably not be funded."

Something else is also at work, she says. "It has to be acknowledged that a powerful social dynamic happens at these meetings. If I really like a proposal and give it a high score but the two other (primary) reviewers don't, I'm pretty likely to come to some middle ground and moderate my views. This happens partly because I'm persuaded and partly because no change at all would be fairly confrontational. On rare occasions I've seen a reviewer not budge, and it makes the room somewhat uncomfortable."

For her part, Adamson says, she believes the current system does support younger scientists with new ideas, "in part because if you look at review panels, they are not all that aged, they're made up of a lot of people who are in mid-career, who are there in part because it's a good way to learn about the process."

"Generally, the peer review process does err a little bit on the conservative side," concedes the NSF's Breckler, "but it is not a dramatic bias, in my opinion. We rely on so many sources of review at NSF that it counteracts the tendency toward conservatism."

That's one reason why diversity among reviewers is a high NSF priority, Breckler says. "In my program, for example, I feel it's important to have somebody who knows the group dynamics area, others who know the stereotypes area and the emotions area. We know we will get proposals in those three areas. In addition to balancing content areas, though, we also want people who reflect the diversity of our society, in gender and ethnic groups, and we try to achieve a balance of representation geographically.

"I also like to have at least one relatively junior person. Very often I'll recruit a junior faculty member,

a post-doctorate, and have him or her sit on a panel once or twice to get the experience and exposure to the process. I feel strongly about giving people the opportunity to see what the review process is like from the inside. I also try to recruit people from places where there isn't a lot of grant-making activity, for example people on the faculty of small colleges."

Breckler says NSF tries to find younger scientists as *ad hoc* reviewers, too, to introduce them to the process without committing them to "the tons of time you need to sit on a panel."

FINAL JUDGMENT

Reviewers that the *Observer* spoke with seemed generally to agree that, as Klorman says, peer review may be "imperfect" but it also is "the best way of evaluating research. The alternative would give too much power to the agencies that control the funding. Peer review, in principle, makes it more likely that judgments of quality will be made fairly and less likely that a single point of view on what is important will prevail."

Its primary function, after all, is "to make sure the public is protected" in two ways, says Ferreira: by assuring that the work scientists propose to do is "sound, logical, and safe, particularly science that involves human subjects or that has implications for humans," and by assuring that taxpayers' money "is not squandered on silly research projects."

"I think the fact that researchers get feedback on their research ideas is very important," adds Rayner, "as is the fact that research is funded solely on the merits of the proposed research and not simply as a function of one's reputation. Of course, prior track record is taken into account in evaluating proposals, but it really comes down to the equitable distribution of research funds based on the quality of the proposed research."

"The issue is how do we invest in scientific research," says Atwell. "We all have ideas on how to improve the peer review system and make it more efficient, but the overall goal is the very best science."