# Peer Review Under Fire

## *Science's gatekeeping system may need an overhaul*

By BRUCE BOWER

Perhaps nothing arouses scientists' professional passions as much as peer review. After all, this process — a gatekeeping and quality-control system run by journal editors, independent referees and research program directors — determines which projects receive funding and which studies get published in what journals. When institutional bouncers rough up your cherished theory and dump your career aspirations on the doorstep, the wounds go deep. And even if the bouncers show respect, they may still inspire fear and doubt with each new grant proposal or manuscript submission.

Despite its crucial role in this era of "publish or perish," scientific peer review today limps along with its own disabling wounds, asserts Domenic V. Cicchetti, a psychologist with the Veterans Administration Medical Center in West Haven, Conn. In his comparative review of peer-review studies conducted over the past 20 years by various researchers, Cicchetti finds consistently low agreement among referees about the quality of manuscript submissions and grant proposals in psychology, sociology, medicine and physics.

These rampant reviewer rifts represent more than an inevitable clash of perspectives on the worth of any particular study, Cicchetti maintains. Scientists must act quickly to improve the quality of peer review, particularly "the rather arbitrary rejection of grant submissions [that] may prevent or seriously delay the implementation of worthy research endeavors," he argues.

Cicchetti examined studies of peer recommendations and publication decisions made by 18 scientific journals, including the JOURNAL OF ABNORMAL PSYCHOLOGY, AMERICAN SOCIOLOGICAL REVIEW, NEW ENGLAND JOURNAL OF MEDICINE and PHYSICAL REVIEW. Additional data came from grant reviews in chemistry, economics and physics conducted during the 1970s and 1980s at the National Science Foundation and the National Academy of Sciences.

Cicchetti's findings and his recommendations for pumping fresh life into peer review appear in the March BEHAVIORAL AND BRAIN SCIENCES, along with 34 responses solicited from researchers and editorial decision-makers. Many support Cicchetti's proposed reforms, although they differ in their interpretations of the importance of disagreements among reviewers evaluating the same paper or proposal. Others reject the reforms, arguing that referee agreement — known as "reliability" when analyzed with statistics — has nothing to do with quality peer review.

Cicchetti remains undaunted. In addition to noting generally poor reliability among referees, he finds them more likely to agree about which grant proposals do *not* deserve support than about which proposals have scientific value. The same pattern applies to manuscript submissions in disciplines with a general focus, such as general physics, cultural anthropology, social psychology and broad fields of medicine such as cardiology and psychiatry, he says.

Conversely, in disciplines with a specific focus — including nuclear physics, physical anthropology, experimental psychology and biological specialty areas such as physiological zoology — Cicchetti finds that reviewers agree more often about worthy manuscript submissions and less often about rejects.

Journals dealing with specific areas of scientific inquiry have higher article acceptance rates and use fewer referees than journals covering general disciplines, Cicchetti points out. Editors of the more specialized publications usually rely on only one initial reviewer; a "thumbs up" review sparks publication approval from the editor, pending suggested revisions, while a "thumbs down" leads to solicitation of one or more additional reviews. General-focus journals often solicit two independent reviews; editors reject a manuscript with two negative reviews, solicit a third evaluation for split reviews and accept papers with two favorable reviews, pending revisions.

Grant proposals receive at least two independent reviews, and sometimes five or more evaluations. Reviewers rate the quality of a proposal on a numerical scale that varies from one funding outlet to another. All the scales contain a strict cutoff score for grants worthy of funding.

Scores often cluster just above or below the cutoff, Cicchetti notes. The inability of referees to agree on the relative quality of these borderline proposals justifies concerns that many worthy projects never win funding, victimized by the draw of reviewers or the proposal's emphasis on a relatively new or unfashionable realm of research, he says.

As for publishing research results, Cicchetti harbors less concern for rebuffed manuscript authors in both general and specific areas of social and medical science, since these researchers invariably succeed in publishing their rejected articles in other journals, often with few or no changes. In contrast, authors in the physical sciences usually do not submit rejected articles elsewhere, apparently regarding the initial peer verdict as decisive, Cicchetti says.

To improve the peer review process, he offers the following suggestions to editors and funding organizations:

- Send manuscripts to at least three independent reviewers carefully chosen for their area of expertise, such as content specialists or biostatisticians.
- Allow authors to request anonymity from reviewers.
- Encourage referees to take more responsibility for their comments by signing their names to each review.
- Solicit author reviews of referees for periodic evaluation by the editor or program director.
- Reward referees who provide consistently high-quality evaluations by inviting them to serve as consulting editors or members of key funding panels.
- Develop systems for peer-review appeals, particularly for grant submissions.
- Train reviewers by distributing guidelines for what constitutes a good paper, by showing them other reviewers' remarks on previously evaluated papers and by organizing practice sessions with feedback on reviews of sample articles.

The new findings and recommendations dovetail with recent reforms in peer review at the National Science Foundation (SN: 4/14/90, p.234) and find strong support in some quarters.

"This study holds a mirror up to peer review that provides a distinctly unflattering picture," says psychiatrist Kenneth M. Adams of the University of Michigan in Ann Arbor. "The remedial steps proposed by Cicchetti require urgent attention."

Cicchetti documents "a system that is seriously flawed," says psychologist Thomas R. Zentall of the University of Kentucky in Lexington. But asking reviewers to reveal their identities would not promote fairer reviews, he contends. Many referees, fearing retribution from a powerful senior author in the event of a negative review, would simply refuse to evaluate a large portion of submissions, Zentall says.

He adds that editors should carefully select reviewers who do not hold strong biases against the kind of research or direction of findings contained in a particular manuscript.

Fred Delcomyn, a neuroscientist at the University of Illinois in Urbana, says the best prospect for refining peer review lies in providing uniform, explicit review criteria and reviewer training. The peer-review process reliably classifies documents into broad categories, such as excellent, fair and poor, but counting on it to reveal finer distinctions "is like expecting to be able to measure the diameter of a nerve cell with a meter stick," Delcomyn maintains.

Extracting meaning from low reviewer reliability also proves tricky, says University of Illinois sociologist Lowell L. Hargens. Editors often choose referees with different specialties, and sometimes solicit evaluations from scientists representing both sides of a controversy — tactics that undermine referee agreement but hardly serve as an indictment of peer review, Hargens observes.

Moreover, editors reject many papers themselves before sending the survivors on for independent evaluations. If this practice weeds out the weakest or most inappropriate submissions prior to peer review, it also accentuates disagreement among referees struggling to rank the highest-quality reports, Hargens says.

Some scientific observers contend that even if referees viewed the entire spectrum of submissions received by journals and funding institutions, neither the reliability nor the validity of their decisions would likely improve. Peer-review validity lies in the ability to predict which articles or proposals will stimulate the most progress in a given field of study. The number of citations an article garners in the years following publication serves as one common marker of validity, although a report's citation record may largely reflect the degree to which it fits into a fashionable field of research.

The belief that basic research deserves generous funding because new understanding springs from unexpected, serendipitous sources — a cherished argument in scientific circles — implies that no one can accurately forecast which work most needs financing and publication, points out J. Barnard Gilmore, a psychologist at the University of Toronto in Ontario.

In addition, an individual's constantly shifting frame of reference greatly limits the consistency of human judgments and the level of agreement attainable in peer review, says psychologist Donald Laming of the University of Cambridge in England. He cites psychophysical experiments indicating that a volunteer's judgment about even a simple stimulus that occurs as part of a successive chain of stimuli — say, the frequency of a tone — varies depending on the qualities of the preceding stimuli. For example, as the number of different auditory frequencies in the chain increases, participants make more errors in identifying new frequencies, and those errors further cloud subsequent frequency judgments.

Laming says these findings suggest that two different peer reviewers, each with a unique and shifting frame of reference regarding a submission, stand little chance of agreeing any more often than the referees studied by Cicchetti.

Gilmore envisions a future in which journal and grant submissions reach a far-flung jury of scientific peers through computerized electronic mail. Rather than jostling for space in prestigious journals, authors would vie for the attention of prestigious reviewers and other readers who subscribe to the electronic peer network. Reviewers' computerized suggestions and ratings would determine a submission's funding or publication destiny.

In the meantime, some of Cicchetti's reviewers do not subscribe to his reform effort. They see no problem with low reliability among referees, and they claim Cicchetti misunderstands how a good editor or grant manager makes decisions.

"The purpose of peer review is not reliability, but to improve decisions concerning publication and funding," says epidemiologist John C. Bailar of McGill University School of Medicine in Montreal. Bailar served as editor of the JOURNAL OF THE NATIONAL CANCER INSTITUTE (JNCI) from 1974 to 1980.

Thorough peer review relies on comments from reviewers with a variety of perspectives and encourages disagreement, he contends. The editor pores over the responses, seeks opinions from editorial board members, and sometimes solicits the views of special consultants. At JNCI, Bailar says, this process led to the publication of some papers that all three reviewers had recommended for rejection, and vice versa.

Bailar argues that Cicchetti has no evidence to support his claims that larger numbers of reviewers, given training and instructions for reviewing, will improve decisions. Editors typically look beyond a submission's scientific and technical merit, evaluating originality, the topic's suitability for a given journal, the journal's need for a balance of topics, the importance of the findings to readers and the appropriateness of an article's length and style, Bailar observes.

"This notion of the editor having a very active role in the judgment of a manuscript seems lost on Cicchetti," remarks Charles A. Kiesler, a psychologist with Vanderbilt University in Nashville and an associate editor of AMERICAN PSYCHOLOGIST.

Cicchetti responds that the active editor who skeptically sifts through referee comments and consults other sources probably represents the exception, not the rule. Extensive reviews of publication decisions at the JOURNAL OF ABNORMAL PSYCHOLOGY, AMERICAN SOCIOLOGICAL REVIEW and PHYSICAL REVIEW point to reviewer recommendations as the major factor in an editor's ruling to accept or reject, he says.

Kiesler argues that reformers should focus instead on grant reviews, where the ability of a single grant reviewer to "blackball" a submission proves particularly worrisome. For instance, he notes, if a funding agency averages reviewer ratings in making its grant decisions, and if funding for a project hinges on an average score near 100, one reviewer can deny funding to a grant proposal by assigning it an extremely low score. This problem plagues controversial and innovative research proposals, says Kiesler, who recommends that granting agencies reserve some funds for innovative projects and disregard the worst rating of any grant proposal.

Though serious flaws mar the current system, the recent findings and proposals seem unlikely to spur a new wave of peer-review research and reform, says Michigan psychiatrist Adams. Everyone complains about peer review, but no one does anything about it, he contends.

"One may find this more amusing later in one's [scientific] career than earlier," he adds. □