Peer Review Comes Under Peer Review

A meeting on biomedical publishing heard lots of gripes about the review system but reached no consensus on reforms

Chicago

WHAT DO JOHN DINGELL, Stanley Pons and Martin Fleischmann, and Jacques Benveniste have in common? They've sparked an explosion of interest in peer review.

Should the vaunted peer review system have cleansed the errors from the paper at the heart of the John Dingell–David Baltimore battle? Could the public brouhaha over cold fusion have been avoided if Pons and Fleischmann had followed standard publishing practice? Should *Nature* have published Benveniste's "unbelievable" results while intending to discredit them later?

These recent events provided a dramatic backdrop and a source of steady gossip for the first international meeting on peer review in biomedical journals, held in Chicago on 10 to 12 May. The fact that such a meeting was held at all demonstrates how uncomfortable the scientific community has become with some parts of the peer review process. The fact that it attracted reporters from the *New York Times*, the *Washington Post*, and the *Los Angeles Times*, as well as over 250 editors of biomedical journals and researchers, shows that the collective soul-searching that went on has a broad audience now.

Few, if any, of the participants went away from the meeting ready to scrap peer review. But many said there is need for considerable fine-tuning, if not outright reform. And some called for more research to understand how the system really works.

That there has already been research on this topic may surprise many scientists. A clutch of new findings presented at the conference lent credence to long-standing suspicions and complaints about peer review and publication practices—the system is riddled with bias, particularly on the part of reviewers; journals favor studies with "positive" rather than negative results; and that researchers pad their résumés by spreading the results of a single experiment over several publications.

According to one study, which captured the most interest at the meeting, masking the names of authors on manuscripts improved the quality of evaluations by reviewers. The study was led by Robert McNutt of the University of North Carolina, who is associate editor of the *Journal of General Internal Medicine*. Manuscripts submitted to the *Journal* were sent to two outside reviewers. One reviewer was told who the authors and their institutions were, the other was not. Distribution to the reviewers was randomized.

The quality of the reviews was evaluated by the journal's staff. They judged how well the reviewer paid attention to key issues in the paper, the methodology, and presentation, such as the clarity of writing and the use of graphs.

McNutt reported that masking the names improved the quality of the reviews by 7%. He said it is not clear whether reviewer bias accounts for the difference, although he suspects it does. Reviewers may simply spend more time to evaluate anonymous papers, he said.

Arnold Relman, editor in chief at the *New England Journal of Medicine*, while praising the study, questioned the importance of the 7%. In any event, he contended that masking names was impractical for his journal because of the large volume of manuscripts it reviews and because epidemiological studies are difficult to mask. The patient population in such studies usually gives away the authors' identities. Masking should be done on all papers or not at all, he said. The journal's executive editor, Marcia Angell, said, however, that if masking made a significant difference, it would be worth doing it on as many studies as possible.

Angell also noted that reviewer bias is not necessarily a bad thing. She said she sometimes purposely sends a paper to a reviewer whom she knows will be a strong critic in order to ascertain the biggest weaknesses of the study.

Stephen Lock, editor of the *British Medical Journal* and Thomas Chalmers of the Harvard School of Public Health urged that reviewers be required to sign their names to

Following the Royal Society's Lead

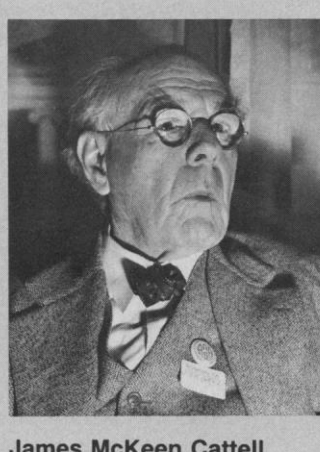
Peer review is standard practice for scientific journals these days, but it did not come into widespread use until fairly recently.

As near as historians can tell, peer review's roots date back to the mid-1600s, when the Royal Society of England sought advice from its members about what papers to publish in its journal, *Philosophical Transactions*. In the mid-1700s, the Royal Society president set up an external committee to evaluate manuscripts, according to Stephen Lock in his book, *A Delicate Balance: Editorial Peer Review in Medicine*.*

But the practice of using outside experts to scrutinize scientific findings before publication apparently did not catch on until late in the 19th century. During the 1800s, science editors in Europe and the United States relied almost entirely on their own judgment, says science historian John Burnham of Ohio State University. The model for medical and scientific publications was newspaper and magazine journalism, and editors considered their publications personal journals.

Slowly, various journals began relying on outside reviewers—the *British Medical Journal* being one of the first. Burnham says that "peer review has been adopted by journals in no particular pattern." Even in the 1930s, the *Journal of the American Medical Association* relied mainly on a small, in-house staff to judge papers and rarely relied on outside help, Burnham said.

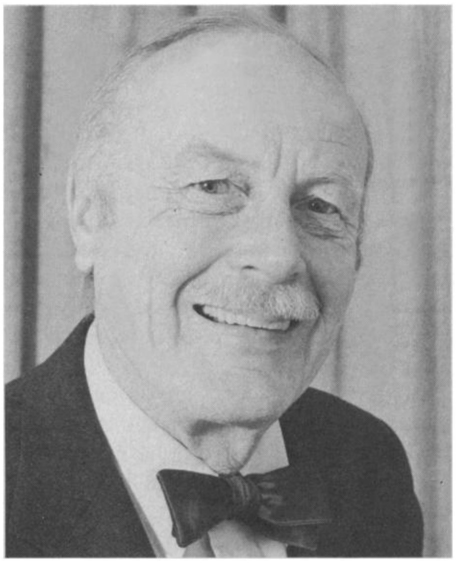
At *Science*, James McKeen Cattell, who edited the journal from 1894 to 1945, depended heavily on his son, a Harvard graduate in physiology, to judge papers, according to historian Michael Sokal of Worcester Polytechnic Institute. It was only after Cattell died in 1945, when the American Association for the Advance of Science took control of *Science*, that the journal adopted external peer review as a regular practice. ■ M.S.



Herbert Gehr/Black Star

James McKeen Cattell

*ISI Press, 1986.



Thomas Chalmers. *Reported that the source of funds can influence the outcome of a study.*

evaluations. But Angell vigorously objected. Reviewers for the *New England Journal of Medicine* are given the option to sign and about 20% choose to do so. Angell remarked that it was her impression that many of those who do sign "are patently trying to curry favor or want to avoid hassle. These reviews become useless to us." Unconvinced by this, Lock urged that the matter be put to a rigorous test.

Whatever a reviewer may recommend, several editors stressed that they make the final decision. No single opinion from a reviewer kills a paper, Relman said. Daniel E. Koshland, Jr., editor of *Science*, said, "Peer review identifies the very good and the very bad papers. It's the middle ones that are difficult" to judge. "That's what editors are for," Koshland said.

Authors pose their own share of bias problems. Some journals, including the *New England Journal of Medicine* and the *Journal of the American Medical Association*, already require authors to list financial interests that may be related to their research. (*Science* and *Nature* do not.)

For good reason, if the results of research by Chalmers are widespread. Chalmers told attendees that he and associates had investigated the published opinions of surgeons and non-surgeons regarding the benefits of coronary bypass surgery. They found that the surgeons were more inclined to be "enthusiastic" that surgery reduced mortality and improved angina than the non-surgeons. In another study, they found that scientists funded by the manufacturers of a drug used to treat diabetes tended to publish more favorable reviews of the drug than those who did not receive company support.

To tighten up the process, conference participants floated a variety of proposals to make authors, reviewers, and editors more accountable for their actions. Drummond Rennie, deputy editor of the *Journal of the American Medical Association*, suggested that publications conduct random audits of raw data of studies accepted for publication. The audit, which editors could not use to block publication, would help determine the prevalence of gross error and fraudulent work "as a basis for making institutional and journal policy, but not to police the system," Rennie said in a 5 May journal editorial. Audits would be conducted randomly and performed by people with some research experience, but not by the editors themselves.

To improve the accountability of authors, Rennie also suggested that journals insist that each author of a paper sign a statement that he or she has not only read and approved the paper, but is also "responsible" for the work described. There was also a proposal that journals require that manuscripts contain footnotes describing the contribution of individual authors. Participants voiced no strong reaction.

With regard to potential bias among editors, one study attempted to evaluate the extent to which journals may be more likely to publish studies with "positive" rather than "negative" results. Anecdotal evidence shows that "selective suppression of negative results may lead to the adoption of ineffective or hazardous treatments," said

Setting the Record Straight

Detecting fraud in science may be difficult, but purging the literature of bad papers is even harder. At least that has been the frustration of Paul Friedman, an official at the University of California at San Diego.

In 1985, the university discovered that Robert Slutsky, a young cardiologist, had fabricated data in three papers. After further investigation of Slutsky's 135 papers, school officials in the fall of 1986 concluded that 10 more were "fraudulent" and that 55 others were of "questionable validity."

Slutsky then withdrew 15 published papers, saying in a letter to the journals that the results were "subject to serious question," but did not acknowledge fraud.

But the university believed Slutsky did not go far enough and requested that the journals publish a retraction stating that the school had found evidence of "research fraud" and print a general explanation of its review process. The university sent its request to the 30 journals where the 135 papers appeared—even the journals which had published only valid papers.

That was in the fall of 1986. Since then, Friedman, who is associate dean of academic affairs, has been trying to ensure that Slutsky's fraudulent and questionable papers and their references are retracted or flagged in the scientific literature. His success has been mixed.

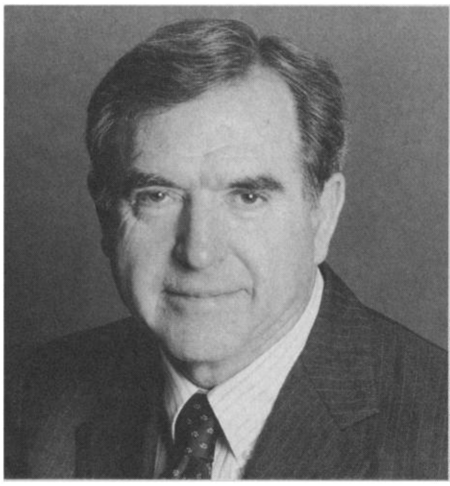
So far, only 9 of the 17 journals that published fraudulent or questionable papers printed everything the university asked. Of the 13 journals that published only valid papers, 5 ran a statement listing the papers retracted from other journals.

The responses of the journals that did not publish the university's statement have run the gamut. Some said they had already printed Slutsky's notice and saw no reason to print anything more. Some said they would not publish a retraction unless all the authors on the paper agreed to it.

One only listed Slutsky's questionable papers, but not the fraudulent one it had published because the lead author (not Slutsky) contended "the paper wasn't that bad," Friedman said. The journal "didn't publish any comment about the [fraudulent] paper at all."

Friedman found that the location and labeling of the retractions was highly variable. Fourteen journals listed a notice in the table of contents under headings including, "Statement," "Notice to Readers," and "Validation of a Study." One wrote an editorial with the retraction under the headline, "A Problem of Deception." Many of the notices were in the letters section. Several were placed at the end of the journal. One journal wrote Friedman that it had printed a retraction, but Friedman could not locate it. Where was it? In the classified ads. "I wouldn't have found it if they hadn't told me," he said.

Friedman argued that journals need to develop a standard written policy about how to deal with retractions and corrections related to fraud and errors. None of the 30 journals had a written policy about how to deal with retractions. "Ten years ago, universities said they didn't have a policy about how to deal with fraud. That now sounds naïve. Journals should have a retraction policy," he said. ■ **M.S.**



Arnold Relman: *"The reviewing process is not meant to achieve perfection."*

Iain Chalmers (no relation to Thomas) of Radcliffe Infirmary, Oxford, England.

Obstetricians, for example, used to advise women pregnant with twins to confine themselves to bed late in their term to prevent premature delivery. But in 1977, doctors in Zimbabwe concluded after an investigation that bed rest instead caused premature births. They didn't publish the results, however, presumably because they thought journals would not be interested in negative results. But during a visit in 1984, Chalmers learned about the findings and later helped the Zimbabwe doctors publish their results in *Lancet* the next year.

Iain Chalmers has tried to figure out how many negative studies related to perinatal medicine have never been published. But he concluded after an extensive survey in which he wrote letters to more than 42,000 obstetricians and pediatricians in 18 countries, "Trying to flush out unpublished trials retrospectively is fruitless." A better way to track unpublished studies, he said, is to require funding institutions, such as government agencies, to keep a registry of all trials they sponsor from the outset. This would help clinicians monitor negative results as well as minimize unnecessary duplication of research.

Byron Bailey of the University of Texas Medical Branch in Galveston and editor of *Archives of Otolaryngology—Head and Neck Surgery* faulted researchers for publishing the same results in more than one journal. The practice is misleading, and may even constitute infringement of copyright, Bailey and others said. Bailey tracked the authors of papers that appeared in his journal during a 7-year period. Out of 1000 authors chosen at random, 201 published 644 articles that duplicated the original manuscript in some form. Here's what he found: A third of the articles are "similar" to the original article, 40% were based on work that included a few more animals or patients than a prior article, and 20% constituted "salami slicing," in which only a portion of work is written up.

Relman and Angell said in a 4 May editorial in their journal that redundant publication "wastes the resources of the peer review system, including time, energy, and expertise as well as money." It "distorts the reward system in academic medicine. . . . [and] is a way of gaining unearned credit." Authors should submit with their manuscripts all published and unpublished articles that may be overlapping, they said.

At the conclusion of the conference, no one even approached a consensus on anything except perhaps a remark by Sheila Jasanoff of Cornell University, who said, "One shouldn't go away depressed about peer review, but one should go away with more humility about it." It was not clear what kinds of changes, if any, journals are likely to adopt. Rennie, who organized the meeting, which was sponsored by the American Medical Association, put a followup questionnaire in the registration packets to ferret out answers, but doesn't expect to report the findings for a couple of months.

With the Dingell investigation fresh on everyone's mind, Relman, not a reserved personality, argued that there are unreasonable expectations about peer review's ability to catch errors or even outright fraud in a scientific paper. He declared, "I don't like the presumption that there's a Holy Grail, that we are seeking truth. The reviewing process is not meant to achieve perfection, but to improve the quality of a paper and eliminate papers that are demonstrably wrong. We don't ensure accuracy, we try to improve it."

It's "impossible for journal editors to know who's cooking data," Relman said. "If a question is raised, editors have to ensure that the institutional process is followed" to evaluate a researcher's work. "We're all interested in the truth, but it's mostly what happens after publication of a study that determines truth." *Lancet* editor David Sharp remarked, "Peer review is achieved by worldwide publication. Peer scrutiny is the very object of publication."

Lock said peer review "is the best we've got, but it's terribly understudied. If we don't put our house in order, the chaps on Capitol Hill and the House of Commons will."

■ MARJORIE SUN

Space Telescope Delayed (Again)

In the National Aeronautics and Space Administration's ongoing game of musical space shuttles, the Hubble Space Telescope is once again the payload left standing. Last year, concerns about overcrowding the launch schedule led NASA officials to postpone the telescope launch from June 1989 until December 1989. Now, citing the priority of classified Defense Department payloads and the need to keep the Galileo mission on schedule for its autumn lift-off for Jupiter, they are postponing Space Telescope until the spring of 1990.

"Hubble is the payload most affected because it is the one that does not have a time-dependent schedule," explains NASA spokesman Charles Redmond. The revised shuttle manifest is neither definite nor official. But the most talked-about date for launching the telescope is 26 March.

Ironically, Space Telescope is paying the price for NASA's recent success in keeping the Magellan spacecraft on schedule for its 30-day "launch window" to Venus. (The window opened on 28 April; the lift-off came on 4 May.) To accomplish that feat with the limited work force available at the Kennedy Space Center, agency officials had to commandeer as many technicians as they could—even though it meant delaying work on the oldest shuttle orbiter, Columbia, which is undergoing a massive refurbishment to give it some of the technical refinements included in the later orbiters, Discovery and Atlantis, and to bring it up to NASA's post-Challenger safety standards.

But that delay, in turn, meant a slip in launching Columbia's first payload: a classified mission originally scheduled for midsummer. And from there, the slippages propagated. A second summertime Defense Department mission had to be moved until after Galileo, which is pegged to the 12 October opening of its launch window to Jupiter. This started crowding the flight that would retrieve the Long Duration Exposure Facility, a boxcar-sized satellite designed to study how materials fare in the space environment. But that flight cannot wait too long because the facility is rapidly spiraling inward from atmospheric drag. And so it went. The upshot: no Space Telescope for Christmas.

One piece of good news, however: since the telescope is already about as ready for launch as it will ever be, the costs of storing it on the ground should soon start declining from about $8 million per month to about $6 million per month.

■ M. MITCHELL WALDROP