

Fraud Shows Peer-Review Flaws

The peer-review system is supposed to guarantee that published research is carried out in accordance with established scientific standards. Yet recently, an internal report from Lucent Technologies' Bell Laboratories concluded that data in 16 published papers authored by researcher Hen-

review system without raising alarms. The fraud was discovered only after journal readers started pointing it out.

What went wrong? Does the Schön affair indicate major flaws in the peer-review system? In its aftermath, many people are asking these questions, and some are suggest-

representing data, substituting mathematical functions for data, and creating false data. Bell Labs immediately dismissed Schön, which ended a career of apparently extraordinary productivity. From 1998 to 2002, Schön authored or co-authored no fewer than 100 papers, an average of one every other week. These were no ordinary papers but claimed significant advances in a variety of fields—organic semiconductors, organic superconductors, inorganic superconductors, and fullerenes. Schön's productivity peaked in mid-2001, when he submitted several papers only a few days apart. In a period of 10 weeks, from late September to early December, Schön published 12 papers. Many of them, although not all, were co-authored by Bertram Batlogg, a senior Bell Labs researcher with a long record of accomplishment.

Yet in case after case, efforts to duplicate the results failed. By late 2001, researchers were pointing to obvious discrepancies in Schön's data. In response, Bell Labs convened its inquiry committee in May 2002, which concluded that in paper after paper, data had been duplicated, with the same data ascribed to different experiments. In the most glaring cases, in a paper published in *Science* (2000, 289, 599) and another in *Nature* (2001, 410, 189), Schön presented the same data in the same paper as coming from different experiments. In addition, other data were in reality mathematical functions or were impossibly perfect, varying from theoretical predictions by less than a tenth of a standard deviation. When confronted with these accusations, Schön admitted some substitutions but insisted he had done the experiments. However, he kept no systematic logbooks, and he claimed that the raw data had all been erased because of a lack of computer storage space.

Bell Labs' failure

Clearly, the first defense of the integrity of scientific results (other than researchers' own morality) lies in the collaboration of co-authors and colleagues and in the review

drik Schön were fraudulent. The papers were reviewed and accepted by several prestigious scientific journals, including *Nature*, *Science*, *Physical Review*, and *Applied Physics Letters*. Yet, in many of the papers, the fraud was obvious, even to an untrained eye, with data repeated point-for-point and impossibly smooth or noise-free. All the papers passed through internal review at Bell Labs, one of the world's foremost industrial research institutions, and the journal peer-

ing reforms. The implications may extend beyond the relatively limited problem of preventing scientific fraud to the broader question of ensuring the fairness and efficacy of peer review itself.

On September 24, 2002, a Bell Labs committee of inquiry chaired by Malcolm R. Beasley, professor of applied physics and of electrical engineering at Stanford University, concluded that Schön had committed scientific misconduct by manipulating and mis-



Mike Stones

procedures of a research laboratory. At Bell Labs, these defenses failed for several reasons. First, although Schön had co-authors on all of his papers, in actuality “all device fabrication, physical measurement, and data processing ... were carried out by Hendrik Schön alone.... None of the most significant physical results was witnessed by any co-author or colleague,” the Bell Labs’ report concluded.

“In our view, this was an isolated, anomalous incident,” says Saswato Das, a spokesman for Bell Labs. “Many of the experiments in question were done when Schön was in Germany working at the University of Konstanz, waiting for his visa, so it was not possible for colleagues to participate in those experiments.” However, many of the fraudulent papers, including one of the more egregious cases of copying data within a paper, were submitted for publication in 2000. At that time, Schön was working continuously at Bell Labs’ main facility in Murray Hill, New Jersey, in a laboratory only steps away from Batlogg’s office. According to Das, Schön met frequently with collaborators Christian Kloc and Batlogg. Yet at no point did either man look at the raw data or ask to participate in one of Schön’s claimed experiments.

“This is certainly not the way things used to be at Bell Labs,” says John M. Rowell, a former director of chemical physics at Bell Labs who worked there from 1961 to 1983. “In the good old days, experiments would be immediately witnessed by one or sometimes even two levels of management, and by collaborators.”

Nor did the collaborators or anyone else question Schön’s spectacular productivity until late 2001, when he was asked to slow down and focus on the details. “Actually, Schön was only among the top four in productivity at Bell Labs that year, so it was not considered that strange,” says Das. “Everyone knew he practically lived at Bell Labs and was there at all hours.”

But others find this attitude incredible. “It is clearly impossible to make an experimental device—especially for the first time—take measurements, and write a paper every four

or five days,” comments Rowell. “If three others at Lucent were submitting more than one paper each week as well, a committee probably should look at their papers, too. The collaborators and management had a responsibility to demand more proof of such unbelievable productivity.”

Lucent Technologies, the parent of Bell Labs, has laid off 88,000 workers in the past two years, and as a result, Bell Labs has suffered significant cutbacks. Did this contraction of personnel make it more likely that scientists would work alone on experiments, instead of in pairs or teams, and that collaborators, pressed for time, would give only cursory review to even spectacular results? Would the need to maintain output with fewer researchers give management an incentive to praise extraordinary output rather than see it as a warning flag? Das does not think so. “People at Bell Labs have always worked in about the same manner. As we are smaller today than in the past, we have reduced the number of areas of focus,” he asserts. Thus, research groups are not necessarily smaller than before the cutbacks.

But again, Rowell and others disagree, noting recent changes at Bell Labs, whatever the cause. “Years ago, not only were research teams larger than one person, and first-line supervisors expected to be hands-on researchers, there was a rigorous publication release process that involved circulation of papers to management and other researchers,” says Rowell. “Evidently, that’s not functioning anymore.”

Peer-review breakdown

Once the papers were submitted for publication, how did they get past so many sets of reviewers? Clearly, it was not the fault of one or two reviewers because of the many articles involved. Nor did editors ignore warnings from the reviewers. “After the story broke, we looked back over the reviewer reports,” says Monica Bradford, managing editor of *Science*, “but we did not find any clues that something was wrong.” Although it is common for journal reviewers to critically comment on a paper’s data

and raise questions about noise levels and statistics, not one reviewer at any journal caught the fact that the data was impossibly good or copied from chart to chart.

Some in the scientific community think that the reviewers should not be blamed for missing the flaws in Schön’s papers. “Referees cannot determine if data is falsified, nor are they expected to,” argues Marc H. Brodsky, executive director of the American Institute of Physics, which publishes *Applied Physics Letters*. “That job belongs to the author’s institution, and the readers if they deem the results are important enough. A referee’s job is to see if the work is described well enough to be understood, that enough data is presented to document the authors’ points, that the results are physically plausible, and that enough information is given to try to reproduce the results if there is interest.”

But editors at leading journals take a broader view, and they admit that the reviewers were among those at fault. “Clearly, reviewers were less critical of the papers than they should have been, in part because the papers came from Batlogg, who had an excellent track record, and from Bell Labs, which has always done good work,” admits Karl Ziemelis, physical sciences editor at *Nature*. “In addition, although the results were spectacular, they were in keeping with the expectations of the community. If they had not been, or had they come from a completely unknown research group, they might have gotten closer scrutiny.” Thus, reviewers and editors as a group had a bias toward expected results from established researchers that blinded them to the problems in the data.

The Schön case points to problems in the peer-review system on which considerable discussion has focused recently, and which affect aspects of science far more significant than the infrequent case of fraud. “There is absolutely no doubt that papers and grant proposals from established groups and high-prestige institutions get less severe review than they should,” comments Howard K. Birnbaum, former director of the Frederick Seitz Materials Research Laboratory of the University of Illinois at

Urbana-Champaign. He recently criticized peer-review practices in grant awards in an article in *Physics Today* (2002, 55 (3), 49). “It is not just a problem of fraud,” he says. “I and colleagues have seen sheer nonsense published in journals such as *Physical Review Letters*, papers with gaping methodological flaws from prestige institutions.”

Because journals have a limited number of pages and government agencies have limited funds for research, too lenient reviews of the established and the orthodox can mean too severe reviews of relatively unknown scientists or novel ideas. The unorthodox can be frozen out, not only from the most visible publications but also from research funding. Not only does less-than-sound work get circulated, but also important, if maverick, work does not get done at all. The peer-review system’s biases, highlighted in the Schön case, tend to enforce a herd instinct among scientists and impede the self-correcting nature of science. This is scarcely a new problem. As Samuel Pierpont Langley, president of the American Association for the Advancement of Science, wrote in 1889, the scientific community sometimes acts as “a pack of hounds ... where the louder-voiced bring many to follow them nearly as often in a wrong path as in a right one, where the entire pack even has been known to move off bodily on a false scent.”

Fixing the system

A number of reforms being discussed could reduce the publication of fraudulent or unsound work and make room for better research. *Science* is already considering implementing one of the less drastic steps—requiring that raw data accompany experimental or observational articles and that the data be posted as supplementary material on *Science*’s Web site. Such a step would make simple fraud more detectable and would enable others to use the same data for alternative interpretations.

“Another idea is to have every experimental paper reviewed by a statistician,” says Ann Weller, an expert on peer review and associate professor of library sciences at the University of Illinois at Chicago.

Such a statistical review would presumably have flagged several of Schön’s papers, and would cut back on dubious statistical analysis, a common flaw of many papers.

Bell Labs has introduced one change in procedure. It now requires the posting of all papers for seven days on a prepublication archive before submission to a journal, which allows colleagues to participate in a review process. However, given the ease with which digital data can be fabricated—in ways that are harder to catch than Schön’s were—there seems to be no substitute for collaborations in which more than one researcher participates in experiments or at least looks at the raw data. Such collaborations can also lead to higher-quality research and problem solving.

One way to encourage real collaborations rather than passive co-authoring is to have the responsibility of co-authors listed in the published paper—for example, device fabrication by John Doe, experimental procedure by Jane Smith, data analysis by Tom Harold. Senior researchers would then have to take co-responsibility for specific aspects of an experiment, or remove their names from papers to which they contributed little.

None of these changes, however, directly addresses the bias of reviewers toward prestigious groups and accepted ideas. More drastic reforms aim at fundamental changes in the system of anonymous review. Blind review, for example, involves removing the authors’ names from articles sent to reviewers, while open review requires reviewers to sign their names to reviews seen by authors.

“Blind review can potentially eliminate biases about authors, but only if the reviewer cannot guess who the author is from the references,” explains Weller. “Studies have shown that in about 40% of papers, the reviewer can guess the authors. On the other hand, blind review does not address biases against novel ideas.”

Open review reduces the possibility of bias, argue supporters such as Fiona Godlee, editorial director for medicine at Biomed Central, an online publishing company in London. If authors know reviewers’ names, reviewers must take personal responsibility

for their reviews, and authors can see if editors have chosen reviewers in a balanced manner. If reviewers are also publicly known and their reviews available, editors or funding agencies presumably would not assign papers or proposals from high-prestige groups to reviewers likely to withhold criticism. Authors could also object if only supporters of the mainstream approach review a minority viewpoint.

It is difficult to say in advance whether open review would incline reviewers to be more conscientious about catching fraudulent or sloppy work. So far, no major physical-sciences journal or funding agency has adopted such a radical reform. However, the idea has received sufficient support for *The British Medical Journal* to allow open review of some papers.

Online discussion

Some researchers wonder whether peer-reviewed journals are essential and whether some of their functions could be replaced by online discussion. “If online prepublication archives, such as arXiv, allowed chat-room-style comments on each paper and author’s replies, the community at large would make its own decisions as to the validity of the results,” suggests Rowell. “My bet is that such a chat room for the Schön papers would have been overwhelmed by critical comments because I heard plenty of them informally, but they were not published.”

Whatever reforms eventually emerge, the Schön case has highlighted the need for peer-review improvements, and a vigorous discussion of how to change is timely. “After my article in *Physics Today*, I got a hundred e-mails of support, but almost all of them told me not to mention their names,” comments Birnbaum. Now, such underground criticism of peer review may come out into the open.

Further reading

The Beasley report is available at http://www.lucent.com/news_events/researchreview.html.

Godlee, F. Making Reviewers Visible. *J. Am. Med. Assoc.* 2002, 287, 2762. 