

Publish, and be damned...

Recent controversies over scientific fraud and other disputed findings have raised questions over the way in which journals select papers for publication. Is there a problem? And what more could be done to weed out dubious results? David Adam and Jonathan Knight investigate.

You browse through the latest issue of a journal and find a paper describing work from a competing group that you know to be riddled with holes. Your hackles begin to rise. Were the referees asleep? What was the editor thinking of?

Sometimes, it's only with hindsight that such feelings kick in. When a prominent researcher is accused of fabricating data, for instance, you might look back over the contested publications and see warning signs in almost every paper. In retrospect, those data really were too good to be true. So why did no one question their veracity when the papers were being reviewed?

Over the past few months, a series of high-profile controversies has brought such questions to the fore, throwing a spotlight on the workings of the journals that published the contentious work. Competition between scientists can tempt some individuals to conduct 'quick and dirty' experiments, rather than doing the job properly, in the hope of being the first to unveil startling new data.

In extreme cases, less scrupulous researchers may commit outright fraud. But are leading journals exacerbating the problem by competing to rush 'sexy' findings into print?

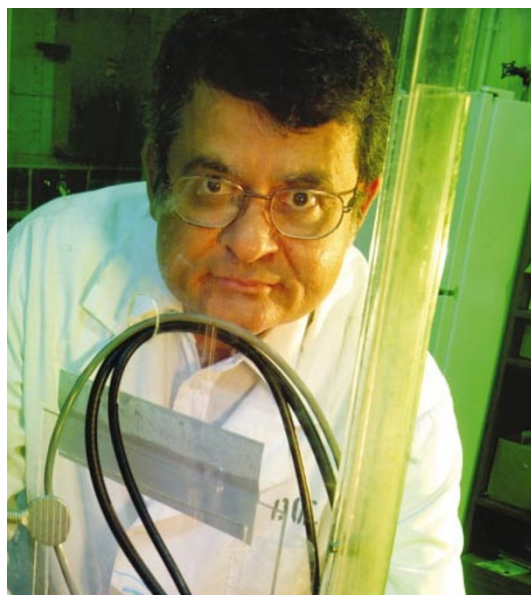
Accusations began to fly in March, when *Science* published a report¹ from scientists led by Rusi Taleyarkhan at the Oak Ridge National Laboratory in Tennessee who claimed to have triggered nuclear fusion in a beaker of organic solvent. The paper appeared to howl of protest, both from leading physicists who were sure that the authors were mistaken and from other researchers at Oak Ridge who had examined the work and claimed to have uncovered serious flaws.

A month later, *Nature* printed a brief statement² effectively disowning a paper³ it had published the previous year, which suggested that DNA from genetically modified maize had invaded the genomes of native Mexican varieties of the crop. The original paper, by David Quist and Ignacio Chapela of the University of California, Berkeley, provoked a political storm in Mexico. But

after publication, other experts argued that the findings were probably experimental artefacts.

In those two cases, researchers are arguing over whether papers' conclusions are justified by the data they contain — there is no suggestion of any misconduct. But it is the scandal surrounding the work of Jan Hendrik Schön of Bell Laboratories in Murray Hill, New Jersey, that has really set tongues wagging. Schön's research on molecular-scale electronic devices and induced superconductivity in carbon 'buckyballs' led to an avalanche of stunning papers — many in leading journals including *Nature* and *Science*. But we now know that he was the perpetrator of the biggest fraud ever to taint the physical sciences, fabricating and misrepresenting data on a massive scale⁴. And some researchers argue that the journals must shoulder some of the blame, for failing to scrutinize more closely the extraordinary claims coming from Schön's lab.

Each of these controversies has its partic-



L. FRENY/DOE

Too hot to handle? The conclusions of papers on nuclear fusion by Rusi Taleyarkhan (above) and on transgenic maize by David Quist (above, far right) and Ignacio Chapela have been challenged. Meanwhile, Jan Hendrik Schön (below) has been found guilty of fabricating data in several papers.

ular set of circumstances. But collectively, they are causing scientists and interested observers to ask whether erroneous or downright fraudulent papers are becoming more likely to be published; and whether editors and referees are doing enough to prevent the pollution of the scientific literature. After the Schön verdict, for instance, an article in *The Wall Street Journal* alleged that *Science* and *Nature* “are locked in such fierce competition for prestige and publicity that they may be cutting corners to get ‘hot’ papers”.

These are tough issues to address, because hard facts are difficult to come by. The US Office of Research Integrity did find more cases of misconduct in the biomedical sciences in 2001 than at any time since monitoring began in 1997, but this may simply be down to increased vigilance. And the frequency with which flawed, rather than fraudulent, papers are entering the literature is almost impossible to quantify. Certainly, few become a matter of public record. In the year to the end of September 2002, for example, *Nature* published just one retraction and two corrections that significantly altered a paper's conclusions.

Editors of leading journals reject the suggestion that standards are slipping in the face of heightened competition. “*Nature* has nothing to gain by the pursuit of glamour at the expense of scientific quality, considering, not least, the criticisms, corrections and retractions we would then habitually be forced to publish,” argued this journal in an editorial comment earlier this month⁵. And many researchers who were interviewed for



this article agree that the system is still working tolerably. If it ain't broke, they say, don't try to fix it.

But in the wake of the Schön scandal, critics of the status quo are speaking out. One of the most vocal is Nobel physics laureate Robert Laughlin of Princeton University in New Jersey. “In this case the editors are definitely culpable,” claims Laughlin. “They chose reviewers they knew would be positive.” Laughlin suspects that it was because of his open criticism of Schön and his co-authors, for failing to provide sufficient information about their methods to others who wanted to replicate the work, that he was not asked to review any of the papers. “It was well known that I was angry at these guys for not allowing their experiments to be reproduced,” Laughlin says.

Due process

Such charges are difficult to prove or disprove, given the confidentiality of the review process. But Karl Ziemelis, chief physical sciences editor at *Nature*, denies that referees were cherry-picked to ensure Schön's papers safe passage. “I can absolutely guarantee that we did not choose reviewers on the basis that we knew they would be positive,” he says. *Science*'s editor-in-chief Donald Kennedy has also rejected accusations that the Schön case reveals any shortcoming in the review process^{6,7}. But candidly, Ziemelis admits that *Nature*'s editors — like many physicists — were enthralled by Schön's work. “Given the exciting nature of the claims made by the papers, we were certainly hoping that the outcome would be positive,” Ziemelis says. But that is not unusual, he says, and tough refereeing ensures that “we are often disappointed”.

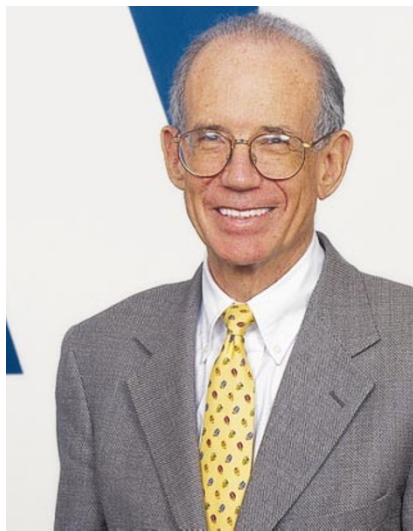
The success rate of Schön's submissions to *Nature* was “far from 100%”, Ziemelis adds. Of those that were eventually pub-

lished, he says, the referees delivered fundamentally positive reports, although they did take issue with some aspects of the papers — which were revised accordingly. But given that *Nature*'s editors were enthused by the claims coming from Schön's lab, would they ever have been willing to gloss over negative reviews and publish a paper anyway?

“It depends entirely on the basis of the negative comments,” says Ziemelis. There are essentially two types: technical and editorial. Ziemelis says that referees pointing out technical flaws that undermine the work will sink a paper — unless other reviewers whose expertise is sought on this specific point disagree. But when it comes to deciding whether a paper is interesting or important enough for *Nature*, editors might overrule a reviewer's negative comments. “The distinction between editorial and technical decisions is an important one, but it is often misunderstood,” Ziemelis says. “On editorial matters we have overruled



BELL LABS



Donald Kennedy denies that recent controversies reveal problems with the peer-review system.

both negative and positive referees.”

But the controversy surrounding Tale-yarkhan’s fusion paper blew up because of accusations that *Science*’s editors had overruled referees’ technical criticisms. Four of the referees asked by the journal to review the paper have taken the highly unusual step of speaking publicly about the confidential reports they supplied. “My problem with the paper was clearly technical — that the authors failed to provide sufficient evidence for what they were claiming,” says William Moss, a physicist at the Lawrence Livermore National Laboratory in California, who reviewed two different versions of the paper.

Another of the referees, Seth Putterman of the University of California, Los Angeles, goes further. “The earlier version of the paper had information that I claimed disproved their results and I pointed that out,” he says. But he claims that the offending data were simply removed from the final version.

Putterman and two other referees — physicist Larry Crum of the University of Washington in Seattle and chemist Ken Suslick of the University of Illinois at Urbana-Champaign — have even released their own critique of the work on the arXiv physics preprint server⁸. “Somewhere out there is a positive report from someone,” Putterman says. “*Science* should publish that report because then we’ll see what kind of information they went on to overrule four negative reviewers.”

That isn’t going to happen, says Kennedy. “We maintain our end of the confidentiality bargain about peer review,” he says, “so I can’t discuss the process specifically, except to say that the positive reviews outweighed the negative ones. Why else would we publish the paper?” Critics suggest that the kudos to be gained by the journal if the paper’s findings had been shown to be valid might have been a factor in the decision — a view that draws a forthright denial from Kennedy: “We at



Karl Ziemelis rejects claims that referees were chosen to ensure safe passage for Schön’s papers.

Science emphatically reject such charges.”

The top journals, rarely reticent in publicizing their latest hot paper or soaring impact factor, make easy targets for criticism when a paper subsequently comes under attack. But editors who have worked for these journals argue that scientists must share responsibility for any problems that exist. “There is sometimes pressure on referees to be quick rather than thorough,” says Philip Ball, a London-based science writer and a former physical sciences editor for *Nature*. “That’s an issue for the journals but also for the scientific community in general,” he says, noting that some authors try to play one journal off against another to get their papers published as quickly as possible.

Despite such concerns, standards of edit-

ing and refereeing are generally agreed to be higher at the elevated end of the scientific-publishing food chain. Further down the scale, fewer questions are asked. “There is an awful lot of literature pollution,” says Carole-Beth Stewart, an evolutionary geneticist at the State University of New York in Albany. “If people get rejected at *Nature*, they just work their way down the ladder, often ignoring all the reviewers’ comments.”

Damage limitation

Ultimately, it is to everyone’s advantage to keep ‘bad’ results out of the literature. Both scientists and journals rely on their reputations, and no one wants to have to retract a paper. So, in the light of this year’s controversies, could the system be tweaked to protect all involved from such embarrassment?

Journals manage their editorial processes in subtly different ways — with one important issue being whether they employ professional editors, or get working scientists to fulfil the role (see “Who should sit in the editor’s chair?”, below). But journals rely heavily for quality control on the efforts of the specialist referees asked to review individual papers. They are “unsung heroes”, says Nicholas Cozzarelli, editor-in-chief of *Proceedings of the National Academy of Sciences*.

Most scientists agree that the feedback from a careful referee is invaluable, regardless of the final decision on publication. “Many scientists don’t have the ability to step back from their own work,” says Anne Weil, an anthropologist at Duke University in Durham, North Carolina. But the refereeing process is meant also to spot crucial flaws in a paper that might affect its conclusions.

That apparently failed to happen with

Who should sit in the editor’s chair?

One topic of regular debate in scientific publishing is whether the job of editor is best done by a full-time professional or by a working scientist. Journals such as *Nature* and *Cell* employ a large number of editorial staff, usually lured from top research labs. These editors have considerable power — papers that fail to pass their initial screen don’t make it to peer review. But at many other journals, including most of those run by scientific societies, the editing role is handed to active researchers. *Science*, meanwhile, operates a hybrid system, employing a cadre of professional editors but also having papers screened by a board of reviewing editors — working scientists who have a major say in whether papers are sent out for review.

Full-time editors argue that they are not only able to devote their full attention to manuscripts, but are also more readily able to share information about the strengths and weaknesses of the reviewers they ask for advice. What’s more, full-time editors claim also to be unencumbered by professional bias. Working researchers may find themselves facing decisions that have ramifications for their own careers, such as whether to publish data that challenge a hypothesis that they have championed. And although it’s likely that most scientist-editors take the high road, professional editors never face the dilemma in the first place.

In response, proponents of editors as active researchers argue that their continuing hands-on expertise better equips them to decide whether a paper represents a significant advance in a given field. “We are proud of the fact that we don’t use professional editors,” says Nicholas Cozzarelli, a biochemist at the University of California, Berkeley, who serves as editor-in-chief of *Proceedings of the National Academy of Sciences*.

But no one has produced any solid evidence that one system is inherently superior to the other. Ultimately, any journal that applies rigorous editorial standards will continue to attract high-quality submissions. And with the top tier of journals including examples of both professional and working-scientist editorship, the debate shows no sign of reaching a definite conclusion.

Quist and Chapela's paper. The Berkeley researchers used two techniques to look for transgenic contamination in Mexican varieties of maize. The first was the standard polymerase chain reaction (PCR), a method for amplifying and detecting tiny quantities of a specific DNA sequence. The second, a variant called inverse PCR (i-PCR), was used to examine the DNA flanking these sequences to reveal their location in the genome. And it was the authors' claim that i-PCR had shown the transgenes to have fragmented and scattered throughout the genomes of the native maize varieties that caused such consternation. Groups opposed to agricultural biotechnology seized on the result, whereas scientists who are supportive of the technology began poring over the results in search of flaws.

Before long, the doubters found a problem. Close examination of the sequences amplified by i-PCR, deposited in the GenBank database, suggested that some were probably not associated with transgenes after all^{9,10}. Quist and Chapela then produced other evidence to support their original claim¹¹. But when this failed to convince one of the referees consulted by *Nature*, the journal published the exchange with its controversial note concluding that "the evidence available is not sufficient to justify the publication of the original paper"².

Crucial details

Johannes Fütterer, a plant scientist at the Swiss Federal Institute of Technology in Zurich, was a co-author of one of the critiques pointing out the problem with the sequences amplified by i-PCR, and claims that *Nature's* referees were remiss in letting the original paper through. "In this case, the sequences were the only piece of data that could prove what the authors claimed and they should have been checked," he says.

Ritu Dhand, *Nature's* chief biology editor, is reluctant to criticize the referees. "It was an unfortunate incident that bypassed us all," she says. Maybe so, but given the political sensitivity of the work, an awful lot of trouble and embarrassment would have been saved if at least one of the referees had followed the path that Fütterer suggests.

Quist and Chapela's paper was bound to become a focus for the contentious debate over genetically modified crops. Less obvious to those outside the Berkeley campus was the fact that the pair had campaigned against a multimillion-dollar deal between the university and the Swiss-based multinational Novartis — subsequently inherited by the spin-off agribiotech firm Syngenta — that gave the company privileged rights to exploit Berkeley's plant-science research. And the authors of the two published critiques of Quist and Chapela's i-PCR analysis included current or former Berkeley colleagues who backed the Novartis deal. In retrospect, given these circumstances, the paper was always likely to be combed for any flaws that

You can't refuse to publish something just because it doesn't feel right.

David Lindley

the referees had missed.

Whatever checks are applied, some papers are published when the referees, editors and even the authors suspect problems. David Lindley, a science writer in Virginia who previously worked as an editor for both *Nature* and *Science*, recalls one such *Nature* paper in 1991 — the first discovery of a planet beyond our Solar System¹². The planet appeared to complete a full orbit around a pulsar in exactly half the time taken by Earth to orbit the Sun. "The authors said: 'We know this is odd, but we can't find an obvious mistake.' One of the referees agreed there was something fishy," Lindley says. "But you can't refuse to publish something just because it doesn't feel right." Six months later, the red-faced authors discovered an error in their calculations. There was no planet, and the paper was retracted¹³.

In this case, Lindley says that it was impossible for him or the referees to discover what was an honest mistake. But what if authors are less than honest? Many editors have encountered papers that look 'too good'. Gene Wells, editor of *Physical Review Letters* for two decades until his retirement in 2001, recalls one case in which the referees all noted that the data seemed unnaturally free of noise. The manuscript was rejected, and Wells never heard about the paper again.

Wells' referees were alert to the possibility of data manipulation. But many scientists argue that referees cannot and should not be expected to view each paper they receive as a crime scene awaiting investigation. Unless



Robert Laughlin was critical of the selection process for reviewers of Schön's papers.

there are obvious grounds for suspicion, "I never question the authenticity of the data", says Bert Vogelstein, a cancer researcher at Johns Hopkins University in Baltimore, who has a reputation among editors as a thorough referee. "If a clever person wants to manipulate the data, it is hard enough to catch in the lab, let alone in the paper."

But according to researchers who have investigated scientific misconduct, fraudsters aren't always so clever. Ulf Rapp, a cell biologist at the University of Würzburg in Germany, headed the inquiry into one of the most spectacular cases of scientific fraud in recent times: that perpetrated by cancer researchers Friedhelm Herrmann and Marion Brach, who worked at the Max Delbrück Centre for Molecular Medicine in Berlin in the early 1990s. Rapp's task force examined 347 publications and concluded that data in 94 of them had either definitely or "highly probably" been manipulated¹⁴. He now says that, in many of these cases, referees and editors should have spotted the apparent duplication of data. "Some cases were clearly a failure on the part of the reviewer and the journals. They were very superficially evaluated," Rapp says.

Paper copies

Schön's manipulation and fabrication of data was less obvious. But still, had the reviewers of his papers, or the editors who handled them, placed the figures they contained alongside those from work that he had published previously, they might have noticed suspicious signs of data duplication. After all, when the news first broke that several of Schön's papers were under investigation, it didn't take long for physicists to identify more publications containing questionable data.

Although many journals issue guidelines to referees detailing particular sections of the paper that should receive scrutiny, few offer specific advice about how and what to check



Limits of detection: Bert Vogelstein says that clever fraudsters are almost impossible to catch.

Z. KAREEM/JOHNS HOPKINS

and in what level of detail. So, in light of the Schön case, should journals require referees to check the data in papers under review against those from previous publications? And should *Nature's* experience with Quist and Chapela's manuscript encourage journals to demand that referees make more stringent checks on DNA-sequence data to ensure that they support a paper's conclusions?

Most scientists and editors believe that such prescriptive approaches are unlikely to be of much help, given the huge diversity of data in the papers under review. Leo Kouwenhoven, a nanotechnologist at the Technical University of Delft in the Netherlands, believes the Schön case will make researchers in his field take more care in comparing figures with previously published work. But he doubts that this, in itself, will do much to prevent future scandals. "This time it was duplicated figures to look out for, but next time something else will be the problem," he says.

Rather, most experts believe that it would be more fruitful to consider ways to encourage improved diligence in general. Almost all editors interviewed for this article said that, in their experience, the standard of refereeing is extremely variable. "The refereeing process is hit-and-miss, and part of an editor's job is to understand that and choose the right referees," says Ball.

Some referees go to incredible lengths and will even re-plot data to check that a paper's conclusions are correct. In crystallography, for example, some reviewers run structural coordinates through their own computer programs to check the resulting molecular structure against the one described in the manuscript. Others will return a brief paragraph from which it is clear, to an experienced editor, that they have given the paper only a superficial read.

"I think one reason that the quality of refereeing often isn't very good is that it's not

Some referees will even re-plot data to check that a paper's conclusions are correct.

important to the referee," says Simon Wain-Hobson, an AIDS researcher at the Pasteur Institute in Paris. "It's not a priority and people don't break their back over it. I've had colleagues say they can review a paper in ten minutes."

Burden of trust

Skilled editors recognize shoddy refereeing, and tend to ask those scientists who are most conscientious to review more than their fair share of manuscripts. "There is a huge logistical burden, particularly if someone gets a reputation as a thorough reviewer," says Weil.

But can journals provide incentives to encourage improved diligence across the board? One way might be to pay referees, and a few journals have taken this step. The *IBM Journal of Research and Development*, for example, pays a minority of its referees a few hundred dollars per paper. "We do this for reviewers who have submitted a careful, thorough review," says John Ritsko, editor-in-chief of IBM Technical Journals. "It is essentially a 'thank-you' for a job well done."

Paying referees is more common among economics journals. This is partly because — unlike science journals — they regularly publish papers after their results have been widely disseminated, and so need incentives to hurry referees who feel that there is no rush. Referees for the *Journal of Political Economy*, for instance, receive US\$40 if they return their report within three months and \$75 if they finish the task within six weeks. Nevertheless, many referees remain unimpressed by the financial reward and still fail to deliver, says Vicky Longawa, the journal's managing editor — and much of the cost is passed on to authors, who pay \$50 to submit each paper.

Few scientific journal

publishers are enthusiastic about introducing a system of payments that would require the industry's business model to undergo a complete upheaval. And paying referees may not improve the system, says Joshua Gans of the Melbourne Business School in Australia, who is an expert on game theory — which can help to predict people's behaviour when faced with particular reward schemes. The system works at present because academics feel obliged to take responsibility, he argues. Paying referees "can motivate an academic but at the same time take away their feelings of professional responsibility because they know others are motivated by pay, too," Gans claims.

Rewards need not be financial, however. For two decades, the American Geophysical Union (AGU) has run a scheme to honour excellence in refereeing for its journals. Those cited by journal editors get their names and pictures published in the AGU's *EOS* newsletter, as well as a certificate and an invitation to an awards dinner. Judy Holoviak, the AGU's director of publications, says that the system was introduced because some researchers were refusing requests to act as referees or returning one-line reports. "There are still people who will never review a paper," says Holoviak. "But it is probably good for the young people coming in; they see this and hopefully it is setting a standard."

Although such schemes may have some merit, many editors and scientists still feel that the current system should be left essentially unchanged. Vogelstein paraphrases Winston Churchill's quip about democracy: "It's the worst system in the world, except for all the others." Rather than focusing on the problems, suggests Fred Alt, an immunologist and cancer researcher at Harvard Medical School in Boston, we should remind ourselves of how well the system works. "It is impressive that the majority of what is published is reproducible and accurate," he says.

Other scientists argue that it would be impossible to make the system foolproof — and misguided even to try. "You're asking for a judgement, an opinion, and it has to stay that way," says Wain-Hobson. "The most important thing about a paper isn't that it's been peer reviewed but that it's reproducible. The real peer review only starts when it's published." ■

David Adam is a News and Features writer for *Nature* in London; Jonathan Knight is *Nature's* contributing correspondent in San Francisco.

1. Taleyarkhan, R. P. *et al. Science* **295**, 1868–1873 (2002).
2. Quist, D. & Chapela, I. H. *Nature* **414**, 541–543 (2001).
3. *Nature* **416**, 600 (2002).
4. Brumfiel, G. *Nature* **419**, 419–421 (2002).
5. *Nature* **419**, 417 (2002).
6. Kennedy, D. *Science* **297**, 13 (2002).
7. Kennedy, D. *Science* **298**, 495 (2002).
8. Putterman, S. J., Crum, L. A. & Suslick, K. <http://arxiv.org/abs/cond-mat/0204065> (2002).
9. Metz, M. & Fütterer, J. *Nature* **416**, 600–601 (2002).
10. Kaplinsky, N. *et al. Nature* **416**, 601 (2002).
11. Quist, D. & Chapela, I. H. *Nature* **416**, 602 (2002).
12. Bailes, M., Lyne, A. G. & Shemar, S. L. *Nature* **352**, 311–313 (1991).
13. Lyne, A. G. & Bailes, M. *Nature* **355**, 213 (1992).
14. Abbott, A. *Nature* **405**, 871–872 (2000).

