



THE TROUBLE WITH REPLICATION

The idea that readers should be able to replicate published scientific results is seen as the bedrock of modern science. But what if replication proves difficult or impossible? **Jim Giles** tracks the fate of one group of papers.

There is nothing particularly unusual about the 6,893rd issue of *Nature*. Published four years ago, it covers the usual mix of disciplines. One paper describes a development in quantum computing¹, another contains the genome sequence of a slime mould². The subsequent impact of the papers has also been within the normal range: a handful have been referenced hundreds of times each, but most have notched up only a few tens of citations.

A third feature of the edition, though just as normal, is more surprising: at least two of the papers we published on 4 July 2002 contain results that may not be replicable. There is nothing suspicious about the papers, nor any suggestion that their authors are anything other than excellent scientists. Nor was that week particularly odd, and there is no reason to think that other journals publish fewer problematic papers. It is simply the case that the replication of results, a process absolutely central to science, is not always possible.

“If you want to know whether a duck is

crossing the street, you look twice,” says Harry Collins, a social scientist at Cardiff University, UK, who cheerfully describes himself as the world expert on replication. Replication in science, he says, is the same: it is a way of being sure that something really exists, and the process by which tentative discoveries acquire textbook status. If, on the other hand, attempts to replicate a result meet failure again and again, that result will end up being discounted.

But, as Collins would be the first to point out, the situation is more complicated than that. For a start, many papers, especially in minor journals, go unreplicated simply because of lack of interest (a third of all papers are never again cited in the scientific literature). Their replicability thus becomes moot.

More concerning for scientific progress is what happens when attempts to replicate an interesting experiment fail. At what point does ‘unreplicated’ give way to ‘unreplicable’? And what is the best way to find out where on this scale a particular result might sit — by gossiping in the bar? By reading another paper? Or

by some semi-formal mechanism in between?

To see replication and its absence in practice, and to ask whether journals and scientists are doing enough to monitor it, I tracked the fate of the 19 papers in issue 6893, asking their authors and others in the field whether the results had been reproduced. In a large majority of cases they had. For example, the procedure with which Ron McKay and his colleagues at the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, treated rats suffering from an analogue of Parkinson’s disease with embryonic stem cells³ has since been repeated in rats elsewhere and in other animals. Papers on the formation of silica films⁴ and on yeast genetics⁵ have quietly racked up citations as the results have been replicated and built upon.

In other disciplines, results are corroborated rather than reproduced. If an identical description of the fossil found by palaeontologist Jennifer Clack⁶, based at the University of Cambridge, UK, was published elsewhere, for example, it would look more like plagiarism

than replication. Her interpretation, though, has undergone something like replication; similar fossils that date from the same period have since been found and described in a way that conforms with her conclusions. In the case of the genome of *Dictyostelium discoideum*², an amoeba, few researchers would see the need to repeat the sequencing from scratch; in any case, the genome stored on the dictybase.org website can be updated should errors be identified.

Giant's signature

But for other papers from the 4 July issue, textbook status looks a long way off. One of those was authored by Sean Brittain and Terrance Rettig, both then based at the University of Notre Dame in Indiana. Their finding was an exciting one: they claimed, for the first time, to have seen H_3^+ ions in the disk of gas and dust surrounding a young star⁷. H_3^+ is seen in the atmospheres of Jupiter and Saturn, suggesting that the astronomers had spotted a gas giant in the act of formation.

Yet right from the start, other researchers wondered whether Brittain and Rettig really had seen H_3^+ . The evidence was in the form of distinct frequencies of infrared radiation: H_3^+ emits at three particular frequencies, and Brittain and Rettig reported detecting emissions in only two of those three. Takeshi Oka of the University of Chicago in Illinois wrote a cautiously optimistic News and Views commentary on the finding in the same issue⁸ but had his doubts about the result. "We used our earliest observation time to check," he says now. "We couldn't see it."

Over the next year, the two sets of authors exchanged their raw data in a bid to resolve their contradictory results. Such exchanges are never easy, given that the scientists are to some extent putting their reputations on the line. In this case, neither side seems to have completely accepted the other's conclusions. But Oka has since published a paper describing how he failed to find any of the three H_3^+ lines⁹. Rettig says that he and Brittain no longer "promote our very tentative interpretation that the unidentified lines might be H_3^+ ". But a second key result from the original paper — the detection of carbon monoxide in the same disk — has since been confirmed.

Attempts to replicate another result from issue 6893 — one that, like the H_3^+ , was considered worthy of mention on the cover — have created far more controversy. In July 2002, stem-cell biology was national news in the United States. Because some stem cells are pluripotent — they can develop into many different cell types — they offer a chance of replacing the tissues lost to old age or disease. But for opponents, most notably those on the influential religious right, the fact that cells for research are extracted from human embryos renders the process unethical, whatever its promise (see pages 486 and 491).

As the two sides battled over plans to regulate the field, a team from the University of



The test of time: most papers in this 2002 issue are looking good for their age — but who could tell which would stand up to testing?

Minnesota pitched in with a paper that seemed to offer a peaceful solution. Catherine Verfaillie and her colleagues described how they had isolated fully functioning stem cells from adult human bone marrow¹⁰. If the results were correct, all the benefits of stem cells could be realized by taking samples from the patient involved — no embryos, no cloning. To those with moral objections this sounded vastly preferable; to others it simply sounded easier. It looked like a win-win situation.

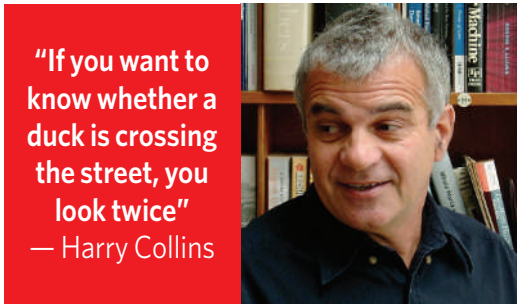
But four years later, the implications of the paper are still far from clear. "People found the paper amazing," says Stuart Orkin, a stem-cell biologist at Harvard University. "But there has been very little published literature since. There

Verfaillie counters that the procedure takes up to six weeks to master and that those who stayed for long enough have cracked it. Some, indeed, have published results¹¹.

Shy journals

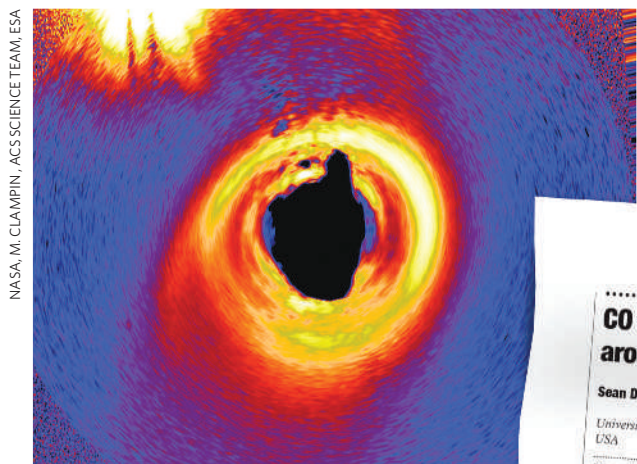
Verfaillie adds that her own team has since ironed out problems with the serum it used and will soon publish a comprehensive methods paper describing the new protocol. But researchers who think they have derived pluripotent stem cells from human bone marrow using related but not identical techniques to Verfaillie's, including Dominique Bonnet at Cancer Research UK's Lincoln's Inn Fields Laboratories in London, have still found it difficult to get their results into print. They say that referees for top journals, aware of the difficulties independent groups have had in replicating Verfaillie's results, are now so sceptical that it is hard to publish their findings.

Verfaillie's protocol maybe indeed be unusually complicated, but it illustrates a common problem: it is often hard to tell whether an inability to replicate a result is due to a group's failings or a flaw in the original paper. The reason is often the countless tiny details of experimental method that are omitted from the methods sections of papers but can influence results. "Things are different in different labs for very subtle reasons," says Gillian Murphy, a cell biologist at the University of Cambridge, UK. "The water can be different. We're about to move labs, and my group is very concerned that delicate cells might hate something in the new pipes."

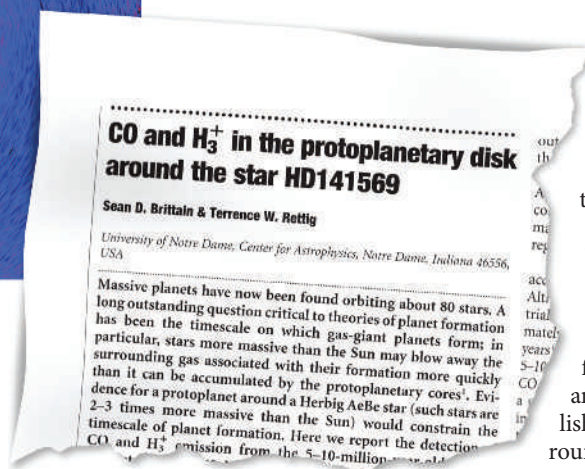


has been no clarification of what those cells are." The story will be a familiar one to many biologists. After publication, rival labs fell over themselves to reproduce the results. Many contacted Verfaillie with requests for the cells and the reagents used to create them, or to ask for more details of the experimental protocol involved. Yet progress was not smooth. Several high-profile groups sent researchers to Minnesota to learn how to extract and culture the cells, and then brought the result into doubt by stating that they could not repeat the process back in their own labs.

S. GELDER



Now you see it: one paper (below) reported a tantalizing hint that gas giants might be forming in the dust and gas around a young star (left).



This issue is not confined to the biological sciences, as Collins's research reveals. In his 1985 book *Changing Order*¹², he quotes a physicist on the difficulties of recreating an exact copy of a piece of experimental kit: "It's very difficult to make a carbon copy. You can make a near one, but if it turns out that what is critical is the way he glued his transducers, and he forgets to tell you that the technician always puts a copy of *Physical Review* on top of them for weight, well, it could make all the difference."

A paper can never be a foolproof recipe for the replication of its results because this sort of information, which the chemist and philosopher Michael Polanyi called "tacit knowledge", can never be entirely captured in a scientific paper. It is thus not in principle possible to tell whether a failure to replicate is down to a lack of this tacit knowledge or a flaw in the original result. In practice, researchers compensate by exchanging tips by e-mail and at conferences. Replication is a social phenomenon, which accounts for the interest of sociologists such as Collins. But because the social interactions are not recorded anywhere, it is hard to consult or build on them.

Fraud and its fallout

This becomes a particularly vexed problem in cases of fraud. Just a few weeks before issue 6893 was published, a scandal hit nanotechnology. Jan Hendrik Schön of Bell Labs in Murray Hill, New Jersey, was one of the brightest prospects in the field, his string of high-profile papers in top journals a remarkable feat for a researcher in his early thirties. But in May 2002 Schön's world began to unravel: people noticed that data in some of his papers had been manipulated. Later that year he was fired, with many of his papers withdrawn as fraudulent.

Allen Goldman's lab at the University of Minnesota in Minneapolis was one of many that wasted much time trying to replicate Schön's work — specifically his findings on the superconductivity of spherical carbon molecules known as buckyballs. A postdoc and two graduate students spent more than a year trying to make the things Schön had described in the papers happen anew in their own lab,

convinced that they were failing because they could not replicate Schön's procedures. Others recount similar experiences: "A postdoc of mine burned up a couple of years of his life," says Robert Dynes of the University of California, Berkeley.

The experience was not all bad: Goldman notes that the work they did inadvertently led his group towards more discoveries. He also says he is now far more critical about the papers he reads in journals. But while his team was trying and failing, others were having

"Rumours go round when part of a study doesn't work in other people's hands" — Chris Surridge



similarly frustrating experiences — experiences that were only discussed at meetings and during the odd phone call between friends in rival labs. Had something appeared in the peer-reviewed literature, Goldman says he would probably have realized more quickly that something was wrong.

One obvious solution is for journals to publish more papers that describe failures to replicate results. Most top publications have procedures for dealing with data that conflict with previously published results, although they demand that authors amass a comprehensive data set before allowing them to question a published result. *Nature* does this through Brief Communications, and like most journals publishes only a handful a month.

Yet few scientists contacted by *Nature* suggested that the journals expand this activity, or

lower the thresholds they require for questioning a result. "Most failures to replicate exhibit incompetence," says Collins, whose feelings sum up those of many scientists. "It would be misleading to publish each one." At *Cell*, editor Emilie Marcus says she would be reluctant to publish a statement saying that someone had simply failed to replicate a paper. "Thorough attempts to reproduce a result should be published," she says, "if you want to claim publicly that someone is wrong, that takes a certain degree of evidence."

This leaves editors in a dilemma. The findings in papers are often hard to reproduce. Readers want to know if they should bother trying or instead dismiss the results as flawed. But the data that could help answer that question often lie unpublished in lab notebooks. "Rumours go round when part of a study doesn't work in other people's hands," says Chris Surridge, an editor at the open-access publisher the Public Library of Science in Cambridge, UK. "But it's very difficult to get the information into the public domain."

Electronic coffee

Scientific publishing's move online may be a big help here, opening up new possibilities for sharing and commenting on papers and methods. If these can be harnessed, coffee-break conversations about replication could be shared with the entire scientific community, allowing problems to be cleared up more easily and frauds discovered more quickly.

The most obvious first step towards this is simply to allow comments to be posted on published papers. Several publishers, including the open-access journals hosted by BioMed Central, already have this sort of comment function but have found that it is not widely used. That could simply be because the facilities need promoting, as BioMed Central publishers say they are now trying to do. Another reason could be that only the most successful researchers are confident enough to criticize others in this public way, even if journals allow it.

PLoS ONE, a new online-only journal from the Public Library of Science, will take the comment model further than anyone else when it launches later this year, with various functions for promoting discussion. Users will, for example, be able to annotate papers, including methods sections, with their own comments.

In a bid to tackle information overload, comments — or perhaps the people making them — will be rated by visitors to the site. Users will then be able to see only comments above a certain rating — a system pioneered by Slashdot.com, a technology website. Authors will also be able to correct mistakes or misleading statements, and others can link to improved methods published elsewhere. The aim, says

Surridge, who is managing *PLoS ONE* during its launch, is to recreate the kind of discussion that takes place in front of conference posters. To help keep things informal, the publication is considering allowing commenters to use nicknames — but only if they provide the site with academic bona fides.

This is not the only way in which the new journal hopes to shake things up. *PLoS ONE* also has an unusual publication policy: it will not consider the novelty of a result when deciding whether to publish a paper. As long as a result is judged by referees to be methodologically sound, and the author can provide the open-access publication fee, it will be published. That will, says Surridge, make it easier to publish papers that cast doubt on previous results, as well as those that confirm them. Sceptical editors at other journals say it will be hard to attract submissions to a journal that sets such a low bar for acceptance. Referees might, for the same reason, also be reluctant to review papers for the journal.

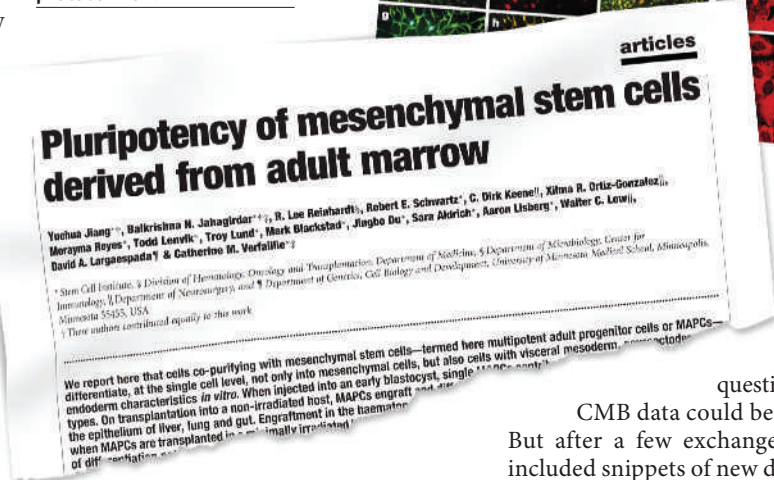
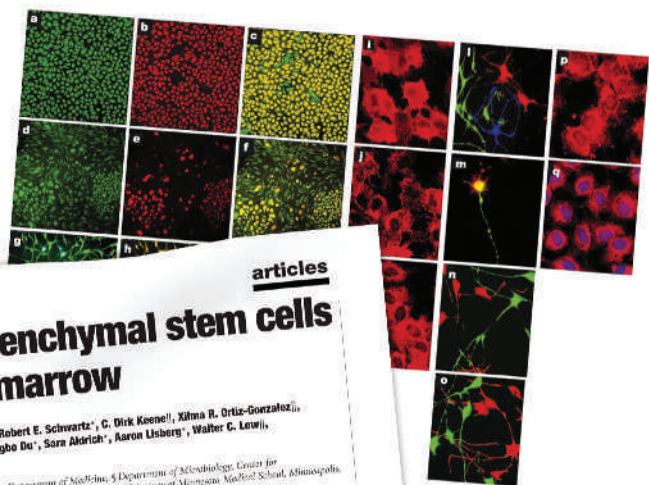
Here's how

Another possibility is to pay more attention to the methods sections of papers. The fact that publishers see this as a potentially important market can be judged by the launch in June 2006 of two journals devoted exclusively to publishing experimental protocols. By giving methods sections the same editorial care as a full scientific paper, from peer reviewing to archiving in a dedicated and searchable website, the journals should allow authors to detail the often critical minutiae of their experimental methods. "When I started in science I was told that I should be able to repeat an experiment by reading the paper, but that is almost never the case," says Michael Ronemus, editor of *Cold Spring Harbor Protocols*, published by Cold Spring Harbor Laboratory Press. "Editors tend to ignore supplementary information. It doesn't get the same scrutiny."

At the other new journal — *Nature Protocols* — editors say they will encourage authors to include a troubleshooting section. Both publications will include discussion facilities that will allow researchers to talk to the original authors and other users of the protocol. These could prove ideal places to resolve problems that crop up during attempts to reproduce a previous result. "The forums will go a long way to resolving conflicts," says Ronemus.

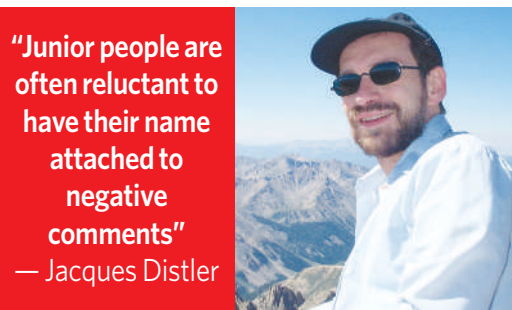
Another possibility is to link to conversation happening elsewhere on the web. ArXiv, a store of online physics preprints now hosted by Cornell University, has from its inception been an online trendsetter. Last August it implemented a 'trackback' function: a system that allows online discussions about a web page, in this case a preprint, to be easily linked to the original page. Following trackbacks from arXiv papers typically leads to physicists' blog posts, in effect

A hard act to follow: this paper said its authors had persuaded bone-marrow cells to turn into different types of body cell (right). Others failed to make the protocol work.



opening up a web of discussion that is something between open peer review and a coffee room at a conference. This might be disturbing for authors, who can now see their papers dissected in public, but it is a great way into the community's views on a paper and offers the benefits of informality and even anonymity.

"Junior people are often reluctant to have their name attached to negative comments," says Jacques Distler, a string theorist at the University of Texas, Austin, who helped to



"Junior people are often reluctant to have their name attached to negative comments"
— Jacques Distler

develop the trackback function at arXiv. "They don't know if it's someone they are going to be relying on later for a job."

Many links from arXiv take users to CosmoCoffee, a forum for discussion of cosmology and high-energy physics in general, and arXiv papers in particular. "Our motivation was to make papers more accessible, but it would be absolutely fantastic if it could resolve controversies as well," says Sarah Bridle, an astronomer at University College London and a co-founder of CosmoCoffee.

In October 2004, for example, an intriguing paper appeared in which the authors claimed to have set limits on the mass of the neutrino using data on the cosmic microwave background (CMB)¹³, a faint glow of radiation left over from the Big Bang. The paper raised some eyebrows on CosmoCoffee, as users

questioned whether the CMB data could be used in this way. But after a few exchanges, one of which included snippets of new data that had been generated to test out the conclusion of the preprint, users decided that the result made sense. For physicists wondering whether a particular result is robust enough to build upon, such discussion could prove a powerful resource. It is also, unavoidably, open to malicious attempts to undermine a particular researcher.

Forty years ago, the Nobel-prizewinning immunologist Peter Medawar declared that all scientific papers were frauds, inasmuch as they describe research as a smooth transition from hypothesis through experiments to conclusions, when the truth is always messier than that. Comments, blogs and trackbacks, by expanding the published realm beyond the limits of the traditional paper, may make the scientific literature a little less fraudulent in Medawar's sense, and in the more general one. They could also help the many frustrated scientists struggling to reproduce claims when, perhaps, they should not be bothering. Replication, for all its conceptual importance, is a messy, social business; it may be that it needs a messy, social medium.

Jim Giles is a reporter for Nature in London. He has just won an award from the Association of British Science Writers for his feature 'The dustiest place on Earth'.

1. Miquel, C. *et al. Nature* **418**, 59–62 (2002).
2. Glöckner, G. *et al. Nature* **418**, 79–85 (2002).
3. Kim, J.-H. *et al. Nature* **418**, 50–56 (2002).
4. Sriraman, S., Agarwal, S., Aydil, E. S. & Maroudas, D. *Nature* **418**, 62–65 (2002).
5. Sun, Z.-W. & Allis, C. D. *Nature* **418**, 104–108 (2002).
6. Clack, J. A. *Nature* **418**, 72–76 (2002).
7. Brittain, S. D. & Rettig, T. W. *Nature* **418**, 57–59 (2002).
8. Oka, T. *Nature* **418**, 31–33 (2002).
9. Goto, M. *et al. Astrophys. J.* **629**, 865–872 (2005).
10. Jiang, Y. *et al. Nature* **418**, 41–49 (2002).
11. Mугuruma, Y. *et al. Exp. Hematol.* **31**, 1323–1330 (2003).
12. Collins, H. *Changing Order* (Sage, London, 1985).
13. Ichikawa, K., Fukugita, M. & Kawasaki, M. *Phys. Rev. D* **71**, 043001 (2005).