

Unfamiliar citations breed mistakes

Sir—Methods for assessing the quality of research are increasingly based on impact factors and citation analysis. But there is concern about the accuracy of science citation^{1–7}, and suspicion has been cast on the impartiality of citation analysis^{2–4}.

As well as conscious bias^{5,8}, it has been suggested that there is a tendency to make more citation errors in names from unfamiliar languages⁴. We have examined the tendency to make errors when typing strings of familiar and unfamiliar names. These experiments, together with an analysis of the numbers of authors per paper in different journals, show why there is a positive association between journal quality (impact factor) and rate of apparent undercitation⁴.

To see whether language familiarity and the length of name strings influenced typing accuracy, we asked 36 undergraduate zoology students to type name strings that consisted of one, three or six Finnish (unfamiliar) or English (familiar) names that were projected on a screen for 8, 24 or 48 seconds, respectively. All subjects were native English speakers with no knowledge of Finnish. The error rate in unfamiliar names increased significantly faster with the length of the name string than the error rate in familiar names (Fig. 1).

We then calculated the mean number of authors per article for each of 65 journals, and plotted this number against the journal's impact factor. We found a strong positive correlation (Fig. 2). Our experimental results (Fig. 1) show that typing errors occur at a greater rate in long strings of non-English names. Therefore, the positive correlation between journal impact factor and number of authors

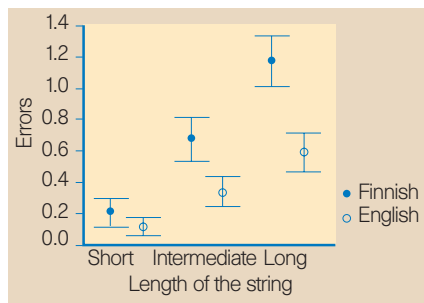


Figure 1 Number of errors (mean \pm 1 s.e.m.) in short, intermediate and long name strings for Finnish and English names. Errors are ln-transformed. There is a significant interaction between language and length of the name string on error rate (repeated measures ANOVA $F_{1,50,52,33} = 6.44$, $P = 0.007$). (To correct for a violation of the sphericity assumption, the degrees of freedom are Greenhouse–Geisser adjusted.) Short name string was one name totalling 5 to 8 characters; intermediate string, three names totalling 15 to 24 characters; and long string, six names totalling 30 to 48 characters. There was no difference in the total number of characters between Finnish and English name strings (ANOVA $F_{1,54} = 0.00$, $P = 0.953$). Finnish names with non-English characters (Å and Ö) were not used. All subjects received all combinations 20 times and each name was used only once.

(Fig. 2) will inflate the effect of mis-citation, and contribute to the positive association between journal impact factor and rate of apparent undercitation⁴.

Our results confirm suspicions that citation analysis is biased. Mis-citation of unfamiliar names is serious, because almost half of all scientific publications come from English-speaking countries¹, so non-

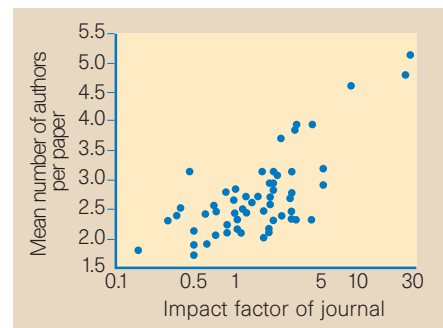


Figure 2 Correlation between a journal's impact factor and the mean number of authors per paper. Pearson's $r = 0.706$, $n = 65$, $P < 0.001$. Journals are a random sample of biological/ecological/evolutionary journals from the Biological Sciences Library of the University of Western Australia. The first 30 papers in 1998 in each journal were analysed. When there were fewer than 30 papers during that year, additional papers from 1997 were included. Impact factors are from the 1997 Science Citation Index.

English-speaking countries will suffer more mis-citations. Our results suggest that undercitation^{3,4} could be an artefact of mis-citation rather than true undercitation, and so cast doubt on the objectivity of citation analysis.

Janne S. Kotiaho, Joseph L. Tomkins, Leigh W. Simmons

Evolutionary Biology Research Group, Department of Zoology, University of Western Australia, Nedlands, Western Australia 6907, Australia

1. May, R. M. *Science* **275**, 793–796 (1997).
2. Tregenza, T. *Nature* **385**, 480 (1997).
3. Paris, G. *et al.* *Nature* **396**, 210 (1998).
4. Kotiaho, J. S. *Nature* **398**, 19 (1999).
5. Barreto, G. R. *Science* **276**, 882–883 (1997).
6. Price, N. C. *Nature* **395**, 538 (1998).
7. Brown, N. L. *Nature* **398**, 555 (1999).
8. Gibbs, W. W. *Sci. Am.* **273**, 76–83 (1995).

Ethics of population genomics research

Sir—A recent News article described a new Swedish genomics company, UmanGenomics, as having a higher ethical nature than another genomics company, deCODE genetics, which was portrayed as having questionable practices (*Nature* **400**, 3; 1999). We do not accept this characterization.

The article described deCODE as a “US-based gene-brokering company” with special rights in Iceland. But deCODE and its 260 employees are solely based in Iceland; the company has already delivered to Iceland the much-needed high-tech jobs that UmanGenomics has promised in Sweden.

From early on, deCODE set high ethical and privacy standards that meet or exceed European Union regulations for human research, under the scrutiny of a national bioethics committee and the Data Protection Commission (DPC) of Iceland. The company has not generated any data from biological material without informed consent granted specifically for each study. All information and biological samples arrive at the company with personal identification numbers encrypted by the DPC. To our knowledge, the current system for protection of privacy at deCODE is unsurpassed (see www.decode.is).

Comprehensive genealogy is an essential tool for population genomics research. It does not yet exist in Sweden but, because Icelanders have been uniquely interested in their genealogy for the past 11 centuries,

deCODE has been able to assemble a computerized genealogy database containing all 270,000 living Icelanders, together with 330,000 of their ancestors.

The News article states that a population-based biobank is to be generated in Iceland under the exclusive control of deCODE, and that Icelandic scientists would have to pay deCODE to carry out basic research. This is incorrect. The Icelandic health-care database would contain medical information with irreversibly encrypted identifiers that, according to database law, can be linked to genealogy and genetic information only with informed consent. The bill grants deCODE the exclusive right to market this anonymous database outside Iceland, but not the exclusive right to use the database. According to the law, the database is to be

controlled by a licensee (deCODE) as well as by oversight and ethics committees and the DPC. Icelandic scientists choosing not to collaborate with deCODE will have the same or better access to patients and their medical records, not less. Those doing non-commercial research will have free access to the database. The law dictates that the data are to be stored so that they cannot be linked to individuals, so the database would not be a biobank as defined by the Swedes.

The article incorrectly gives the impression that UmanGenomics' use of its Medical Bank is going to be with the consent of the community, whereas deCODE's use of a database was opposed by the community. The Icelandic poll actually asked whether people were concerned about the protection of privacy in a centralized medical database, and (surprisingly) only 25% said yes. UmanGenomics was granted permission to use its Medical Bank by government committees and bureaucrats, which is not 'community consent'.

In stark contrast, deCODE obtained its licence to construct and run the health-care database through the democratic process. In 1997, the company suggested the database to the Ministry of Health, which then drafted a bill and placed it on its open website for comment. The government submitted the bill to the parliament (Althingi) in early 1998. Vigorous local debate lasted nine months and included hundreds of newspaper articles, radio and television programmes. Icelanders debated the database bill more than any other bill in the history of the republic. On the eve of the vote on the bill, a poll showed that 75% of the population supported it, and Althingi passed it last December by the same margin.

One premise in the News article is that, because government institutions own 51% of UmanGenomics, the proper use of its Medical Bank is assured. But governments have a bad record on violation of privacy. Further, the health-care authorities and the university that own most of UmanGenomics are mainly concerned with attending to diseases and health. Majority ownership of a genomics company that uses health-care information from clients as raw material may be seen as a serious conflict of interest.

Jeffrey R. Gulcher, Kari Stefansson
deCODE genetics, Lynghals 1, Reykjavik, Iceland 110

Striking the right note

Sir — About 50 years ago, a student was working in the library of the Natural History Museum in London one afternoon, and overheard two rather elderly gentlemen discussing *Nature* as they leafed through a

copy on display. "Do you still read this these days?" asked one. "Not really," replied the other. "I put my copy of *Nature* on the music-stand and play it."

Well he might, had he encountered one of those genome pull-outs you offer us these days, although he might not have understood the coloured musical notation (*Nature* 399, 323–329; 1999).

Alan Longhurst
Place de l'Eglise, 46160 Cajarc, France

Heisenberg, the bomb and the historical record

Sir — Mark Walker's review¹ of my book *Heisenberg and the Nazi Atomic Bomb Project* distorts the record in several ways.

Walker conceals from readers the fact that the book contains documented proof of his own past suppression of crucial evidence. This raises a question of reviewing ethics.

Walker hides the scientific issues relating to Werner Heisenberg's misunderstanding of the critical mass of an atomic bomb which he calculated to be *tons* of uranium-235. Heisenberg's mistake has been fully demonstrated by the publication by Sir Charles Frank in 1993 of the Farm Hall transcripts². Walker has already been reprimanded in an article in *Nature*³ for misleading readers, but he still does not see fit to mention the fundamental Farm Hall evidence in his review. In his own writings, Walker has all along suppressed Frank's explanation to him, in a taped interview of 1985, of the Heisenberg error.

Walker pretends that Heisenberg never seriously considered an exploding reactor-bomb. My book cites German reports analysing such a bomb, including two by a member of Heisenberg's team and a lengthy 1942 synoptic report, as well as Heisenberg's own comments, which show the idea to have been serious.

Any one of these three points should suffice to warn your readers against taking on trust any statement in Walker's review.

Paul Lawrence Rose
Institute for the Arts and Humanistic Studies,
Pennsylvania State University,
University Park, Pennsylvania 16802, USA

Sir — What Walker somehow fails to mention in his dismissive review¹ is that Rose's book includes a devastating critique of Walker's own work on the subject. Re-examining the technical records on which Walker based the central thesis of his 1989 book⁴, Rose shows that Walker misrepresented crucial documents and suppressed essential evidence.

Ten years ago, Walker claimed to have

uncovered archival documents proving that German physicists had "performed the same sort of experiments, had made the same type of calculations, and had come to similar conclusions as the Allies — for example the estimate of explosive critical mass". As Rose demonstrates, this interpretation can be made to seem plausible only by the most convoluted and selective reading of the evidence.

A key 1942 German progress report, for example, cited by Walker to demonstrate that Heisenberg knew the critical mass of a uranium-235 bomb to be small, does contain a parenthetical speculation that with plutonium (then a hypothetical element) a small critical mass might obtain. What is described in the main text of the report, however, is the unworkable multi-ton reactor-bomb first intimated by Heisenberg in 1939. By citing only the parenthetical remark while suppressing the report's substance, Walker transforms contrary evidence into support for his thesis. Inconvenient technical reports by Heisenberg's research assistant Paul Müller elaborating the misguided reactor-bomb concept go unmentioned in Walker's account. An oral history interview with wartime scientific intelligence officer Sir Charles Frank forcefully describing Heisenberg's mistaken estimate of the critical mass is likewise suppressed — Walker cannot have been unaware of it since it was he who conducted the interview. Thus did Walker succeed in "misrepresenting not only the content of the document[s] but also the whole history of the German atomic project".

Reading Walker's review, one would never know any of this, or that such objections had ever been raised before^{3,5,6}. "Rose's book does not really have a conclusion," writes Walker, while suggesting that there is little in it that could possibly interest the readers of *Nature*. Having made the unfortunate decision to review a book that contains discrediting revelations about his own scholarship, Walker might have taken the opportunity to address the disturbing issues it raised. At the least, readers are entitled to a disclosure of self-interest. Instead, Walker chose the dismal principle he has followed before, *si incommoditas est, non est*. Whether a discordant fact, an inconvenient document, or a detailed study putting his claims to the test: if it doesn't fit, it just didn't happen.

Jonathan L. Logan
EPG Research Foundation, 111 East Shore Road,
Manhasset, New York 11030, USA

1. Walker, M. *Nature* 396, 427–428 (1998).
2. Frank, C. *Operation Epsilon* (Institute of Physics, Bristol, 1993).
3. Logan, J. L. & Serber, R. *Nature* 362, 117 (1993).
4. Walker, M. *German National Socialism and the Quest for Nuclear Power: 1939–49* (Cambridge Univ. Press, 1989).
5. Logan, J. *American Scientist* 84, 263–277 (1996).
6. Klotz, I. M. *Nature* 379, 410–412 (1996).