

a chaotic pile. No artist directs its creation; random perturbations are caused by variations in the consistency of the wax, the gun pressure and in how the deposits accumulate. Yet it is stage-managed. The art is not in the wax mound but in the whole performance.

Another artful machine generated a set of extruded grey concrete sculptures called *Greyman Cries, Shaman Dies, Billowing Smoke, Beauty Evoked* (2008–09). These were produced by a scaled-up version of a rapid prototyping machine. Such technology is normally

used by engineers to build accurate models from fine threads of molten plastic. Kapoor's larger version extrudes a thick concrete sausage that builds up layers of soft coils, ropes and worms under computer control.

Whereas engineers seek precision with their models, Kapoor delights in his products' imperfections. Yet the appearance of randomness involves technical sleights of hand. To achieve each particular texture, his contraption must be finely tuned. Kapoor deliberately finds a point of balance between

opposites — between perfection and imperfection, softness and firmness, movement and repose — to tantalize the viewer.

Kapoor knows he is treading a fine line between artist and entertainer. He says: "It's a short trip from Disneyland to something truly mysterious." But that mystery is delivered only through precise engineering. ■

Neil Dodgson is professor of graphics and imaging at the University of Cambridge, UK. e-mail: neil.dodgson@cl.cam.ac.uk

BIBLIOMETRICS

Measure for measure

A useful guide to citation analysis shows that counting publications is harder than it looks, finds **Ton van Raan**.

Citation analysis offers a means to quantify the impact of a scientist's work. One tool for tracking citations is the Publish or Perish (PoP) software program developed by Anne-Wil Harzing, professor of international marketing at the University of Melbourne, Australia. Her guide describes how her program generates citation analyses from Google Scholar and gives an overview of bibliometric methods and sources. She champions the practical use of citation measures, yet also recognizes that calculating them reliably is a difficult task.

The Publish or Perish Book focuses on citation analysis of individual researchers, not groups or institutes. Several metrics may be calculated for scientists and for journals, including their number of publications and citations, average number of citations per publication and per author, and the h-index, a widely used characterization of citation impact. Harzing argues using practical examples that such indicators are good markers of a researchers' influence, and are useful in assessing applications for jobs, promotion and tenure, and for literature research and choosing a journal in which to publish.

Using Google Scholar as a data source is advantageous as it retrieves publications not covered by Thomson Reuters' Web of Science: books, edited volumes and 'grey' literature such as conference proceedings. Harzing explains how to analyse citations with Google Scholar, and discusses ways that citation patterns of early reports can be used to predict the later impact of journal articles derived from them. But there are inevitable problems in tying together varied data, such as matching conference proceedings with the subsequently published paper.

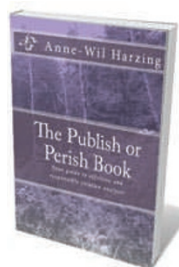
Harzing considers the main downside of the Web of Science to be its limited coverage

of different disciplines, particularly of engineering, the social sciences and the humanities. In my view, however, its coverage of well-funded fields, such as the natural sciences and medicine, is very good. For novice users, the Web of Science does have trouble identifying ambiguous author names, especially those in which the order of the first name and surname is unclear. It

also struggles to aggregate articles with variations of the same title and to identify self-citations. But professional bibliometricians such as myself build and work from Web of Science reconstructions — usually proprietary to their institutes — in which such sources of error are fixed.

The book underplays the ethical issues that arise when performing a citation analysis for a person other than yourself. Verification of research output is important — missing just one highly cited paper can distort the results dramatically. This highlights the necessity of cleaning raw data. For instance, incorrect referencing will lead to cited publications being missed. It is a huge effort to correct for these 'homeless' citations. In this respect, Google Scholar is a black box.

Harzing discusses both the perspective of the person to be evaluated, and that of the evaluator. This is important because evaluators of tenure



The Publish or Perish Book: A Guide to Effective and Responsible Citation Analysis
ANNE-WIL HARZING
Tarma Software
Research: 2010.
250 pp. \$29.95

promotions might apply home-made metrics that are not transparent and may incorporate unknown mistakes. Citation metrics are attractive because they have the potential of objectivity, but evaluators may put too much faith in quantitative aspects of research performance. Simple metrics then become a shortcut to decision-making.

Many scientists are concerned that citation analysis, particularly that done in an amateurish way, is having detrimental effects on science. They fear that researchers are driven to pursue citation quantity instead of scientific quality. Statistical reliability may become a serious problem when dealing with individuals rather than groups, as Harzing recognizes. Field-specific normalization is also necessary if research impact is to be compared across disciplines.

Further statistical factors limit metrics. Indicators often concern arithmetic mean values, yet the distribution of citations across publications is skewed. Averages are thus not the best statistic. Although this problem is discussed, Harzing's book doesn't offer indicators that are related to the distribution of impact across a field, which would answer the question 'Does he or she belong to the top 10% of his or her field?'

Harzing explains that the problem of the skewed distribution can be removed using the h-index: for instance, a researcher has an h-index of ten if ten of his or her papers have at least ten citations and the other papers have no more than ten citations each. But in my view, the h-index is inconsistent. For example, suppose that researcher A has three publications with five citations each ($h=3$) and researcher B has four with four citations each ($h=4$). Both obtain one additional publication with five citations. Researcher A's h-index then increases to four, whereas researcher B's h-index remains equal to four. This makes no sense.

With these caveats in mind, *The Publish or Perish Book* is a useful resource for scientists, particularly in fields in which Google Scholar is a major source of citations. ■

Ton van Raan is professor of science studies in the Centre for Science and Technology at Leiden University, the Netherlands.

➔ **NATURE.COM**
For our Science
Metrics special, see:
go.nature.com/nj2xqk