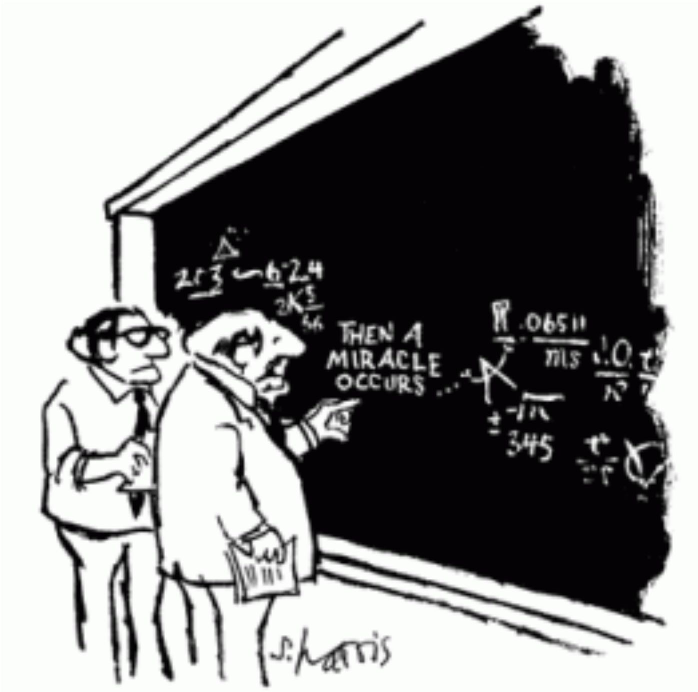
Set the default to “open”

BY DAVID BAILEY, JONATHAN BORWEIN & VICTORIA STODDEN, 28-2-2013

Reproducible science in the computer age Conventional wisdom sees computing as the “third leg” of science, complementing theory and experiment. That metaphor is outdated. Computing now pervades all of science. Massive computation is often required to reduce and analyse data; simulations are employed in fields as diverse as climate modeling and astrophysics. Unfortunately, scientific computing culture has not kept pace.

Experimental researchers are taught early to keep notebooks or computer logs of every work detail—design, procedures, equipment, raw results, processing techniques, statistical methods of analysis, etc. In contrast, few computational experiments are performed with such care. Typically, there is no record of workflow, computer hardware and software configuration, or parameter settings. Often source code is lost. While crippling reproducibility of results, these practices ultimately impede the researchers’ own productivity.



“I think you should be more explicit here in step two.”

The State of Experimental and Computational Mathematics

Experimental mathematics, application of high-performance computing technology to research questions in pure and applied mathematics, including automatic theorem proving---raises numerous issues of computational reproducibility [1,2]. It often pushes the bounds in very high precision computation (hundreds or thousands of digits), symbolic computation, graphics and parallel computation. As with all computational science, one should carefully document algorithms, implementation, computer environments, experiments and results. Even more emphasis needs to be placed on unique aspects of the discipline:

(a) Are precision levels (hundreds or thousands of digits) adequate?
(b) What independent consistency checks were employed to validate results?
(c) If symbolic manipulation software was employed (e.g., *Mathematica* or *Maple*), which version was used? What precise functions were called, what parameter values and environmental settings?
(d) Have numeric spot-checks been performed for derived identities etc.?
(e) Have symbolic manipulations been validated, say using two different packages?

Such checks are crucial, because even the best symbolic and numeric computation packages have bugs and limitations---often exhibited only during hard computations.

The ICERM Reproducibility Workshop on Reproducibility in Computational and Experimental Mathematics Motivated by such concerns, this workshop was held in December 2012 at the Institute for Computational and Experimental Research in Mathematics at Brown University. Participants included computer scientists, mathematicians, computational physicists, legal scholars, journal editors and funding agency officials, representing academia, government labs, industry research, and all points in between.

While different types and degrees of reproducible research were discussed, an overwhelming majority argued the community must move to "open research": research using accessible software tools to permit (a) 'auditing' computational procedures, (b) replication and independent verification of results, and (c) extending results or applying methods to new problems

Of course, the level of validation should be proportional to the importance of the research and strength of claims made.

Workshop Conclusions *First, researchers need persuasion that efforts to ensure reproducibility are worthwhile*, leading to increased productivity, less time wasted recovering data or code, and more reliable conversion of results from data files to published papers.

Second, the research system must offer institutional rewards at every level from departmental decisions to grant funding and journal publication. The current academic and industrial research system places primary emphasis on publication and project results and little on reproducibility. It penalizes those devoting time to developing or just following community standards.

The enormous scale of state-of-the-art scientific computations, using tens or hundreds of thousands of processors, presents unprecedented challenges. Numerical reproducibility is a major issue, as is hardware reliability. For some applications, even rare interactions of circuitry with stray subatomic particles matter.

It is regrettable that software development is often discounted. It is typically compared, say, to constructing a telescope, rather than doing *real science*. Thus, scientists are discouraged from spending time writing or testing code. Sadly, NSF-funded web-projects remain accessible only about a year after funding stops. Researchers are busy running new projects without time or money to preserve the old. Given the ever-increasing importance of computation and software, such attitudes and practices must change.

Finally, standards for peer review must be strengthened. Editors and reviewers must insist on rigorous verification and validity testing, along with full disclosure of computational details [3]. Some details might be relegated to a website, with assurances this information will persist and remain accessible.

Exceptions exist, such as where proprietary, medical, or other confidentiality issues arise, but authors need to state this upon submission, and reviewers and editors must agree such exceptions are reasonable.

Many tools help in replicating past results (by the researcher or others). Some ease literate programming and publishing computer code, either as commented code or notebooks. Others capture provenance of a computation or the complete software environment. Version control systems are not new, but current tools facilitate use for collaboration and archiving complete project histories.

The US has followed the UK, Australia and others in mandating public release of publicly funded research, including data [5]. We hope this brings a cultural change in favour of consistently reproducible computational research. See also the workshop report [6] and Wiki [4].

References:

1. David H. Bailey and Jonathan M. Borwein, "Exploratory Experimentation and Computation," *Notices of the AMS*, vol. 58 (Nov. 2011), pg. 1411-1419.
2. Jonathan M. Borwein and David H. Bailey, *Mathematics by Experiment*, Edition 2, AK Peters, 2008.
3. Daniele Fanelli, "Redefine misconduct as distorted reporting," *Nature*, 13 Feb 2013, available at http://www.nature.com/news/redefine-misconduct-as-distorted-reporting-1.12411.
4. ICERM Reproducibility Workshop Wiki: http://wiki.stodden.net/ICERM_Reproducibility_in_Computational_and_Experimental_Mathematics:_Readings_and_References
5. OSTP, "Expanding Public Access to the Results of Federally Funded Research," 22 Feb 2013, available at http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research.
6. V. Stodden, D. H. Bailey, J. Borwein, R. J. LeVeque, W. Rider and W. Stein, "Setting the Default to Reproducible: Reproducibility in Computational and Experimental Mathematics," 2 Feb 2013, http://www.davidhbailey.com/dhbpapers/icerm-report.pdf.