The cargo cult of scientific computing (and a cure)

Harold Thimbleby
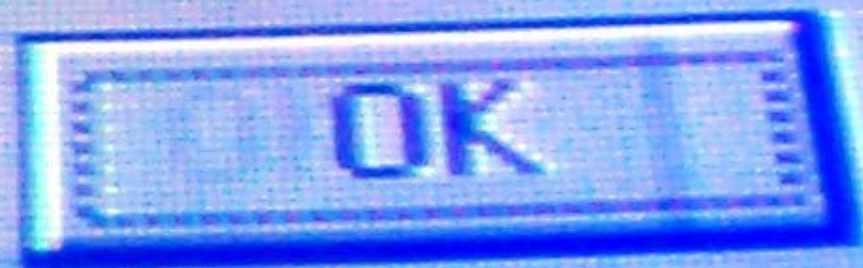
University of Wales Swansea

Network Connections

error occurred while reconnecting F: to

osoft Windows Network : The netwo

o not try to restore the connection

onnection has not been restored

[OK]

```vba
Private Sub Document_New()
ActiveDocument.SpellingChecked = True
ActiveDocument.GrammarChecked = True
ActiveDocument.ActiveWindow.Caption = ActiveDocument.ActiveWindow.C
" - Thesis"
ActiveDocument.ActiveWindow.View.ShowFieldCodes = False
'
'Need to come up with real solution to handle this before shipping
'
'Dim strProperty As String
'Const strREG_SETTINGS_BASE_KEY As String =
"HKEY_CURRENT_USER\Software\Microsoft\Office\8.0\Word\Wizards\"
'strProperty=System.PrivateProfileString("",strREG SETTINGS BASE K
"PageSize")
' If strProperty = "A4" Then
'     ActiveDocument.PageSetup.PaperSize = wdPaperA4
'Else
'     ActiveDocument.PageSetup.PaperSize = wdPaperLetter
' End If
End Sub
```

Setting the new multimeter standard

The new 170 series are the new benchmarks for multi-purpose multimeters

All models measure True RMS for correct measurements independent of the signal waveform (other multimeters could measure non-sinusoidal signals up to 50% wrong). They have new features including a backlight display (177, 179), min-max-average recording and temperature measurements (179) in a familiar shape and size. The 6000-count digital display with analog bargraph offers readings twice as fast as other multimeters.

The 175, 177 and 179 meters allow calibration adjustments to be made via the front panel. Battery replacement is done through a battery access door without opening the complete case and voiding calibration seals.

The Fluke 170 Series tools meet the latest EN61010 CAT III 1000V and CAT IV 600V safety ratings. The meters are compatible with the optional ToolPak accessory which allows the user to hang the meter with a strap or magnet and keep both hands free to move the test leads around for optimum safety and control.

Click on the rotary knob and buttons to get a virtual demonstration of the different functions.

QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.

- M+ – MRC M+ MRC

- = – MRC M+ MRC

- *if* … 2 conditions apply

- (demo later)

No Signal

The *Journal of Machine Learning Research* was founded in 2001, in its words "as an international forum for the electronic and paper publication of high-quality scholarly articles in all areas of machine learning". Machine learning is a fascinating and important area of computer science: basically, can computers learn? It is a theoretical and practical area, where computer programs are developed and run on data to see what they can learn. There are enormous and widespread practical applications for machine learning, from labour negotiation, medical treatment, agriculture, to generally making computers more "intelligent". The slant of the journal will make it interesting for theoreticians, biologists and psychologists interested in how animals and humans learn and what the theoretical limits to learning are. There are even applications in machine learning for countering terrorism. And as the worldwide web fills up with vast amounts of unstructured information, we need all the help we can get to learn how to use it effectively. It says something about the relevance of machine learning that one managing editor of this journal works at Google and the other has published papers on financial markets.

Machine learning makes a difference, and makes a lot of money worldwide. Yet the *JMLR* has a free website, and it costs only $75 (£47) — $111 outside the US — for an annual individual print subscription, a fraction of the cost of subscribing to a conventional science journal. The journal runs like a collective, with MIT Press taking just the paper print rights, so costs are minimised. Turnaround time for authors is dramatically reduced. If somebody, say, in the third world wants to know anything up to date and rigorous about machine learning, it is the definitive place to reference. This is an excellent model for all journals to copy, especially in science. As one of the editors says:

Harold Thimbleby

Give your computer's IQ a boost

RONALD GRANT ARCHIVE

HAL in *2001*: can computers learn?

"What is the role of the scientist in academic publishing? Doing the publishing!"

The bulk of the journal's papers are devoted to discussing and evaluating learning methods. I was interested to see how ideas talked about in the journal actually worked, because that is really the whole point. So, as the journal is available online, I looked at every paper and then emailed the authors to ask them about their ideas. After a few weeks I had more than 100 replies. I drafted this review, and then bounced it off the editorial board and the authors again. The enthusiasm of authors for their work was impressive: I had replies covering every paper published.

I asked whether the system described in each paper was available. Of course, some papers were theoretical; I had a few replies saying my question was irrelevant. Of the remaining, about a third specifically said their systems were unavailable. Their systems were private, commercial confidential or incomplete in some way. Consider some of the replies I got: "Unfortunately, I do not have the system in a state where I can give it away right now" and "we don't have the data ready to be published". Other quotes are revealing about authors' attitudes. "The system is a research prototype developed in my group, and is not appropriate for public dissemination" and "the implementations we had were very much 'research code', and not suitable for public consumption".

My survey suggests some authors have a relaxed regard for scientific virtues: reproducibility, testability and availability of data, methods and programs — the openness and attention to detail that supports other researchers. It is a widespread problem in computer science generally. I am guilty, too. We programmers tend not to keep the equivalent of lab books, and reconstructing what we have done is often unnecessarily hard. As I wrote elsewhere (see www.uclic.ucl.ac.uk/harold/ warp) there can be real problems with publishing stuff that is not rigorously supported. It is the computer-science equivalent of fudging experimental data — whether this really matters for scientific progress is another controversial debate.

Then there is the problem of who owns the work. As one author puts it: "We have not had the time to turn our experimental code into something other people can use (and anyway our employers wouldn't like to see things given away)." Certainly there needs to be a balance between science and protecting intellectual property. It is a big problem, as turning research ideas into code that really works might involve a company that then owns it. On the other hand, there is no reason why open-source code cannot be made freely and immediately available, at least to the depth to which the ideas are discussed in the papers. And it is possible: look at sites such as the GNU-licensed open-source Weka machine-learning project (www.cs.waikato.ac.nz/ml), which provides a framework in which people can share work. Many other sites have papers, code, demos and data too.

The *Journal of Machine Learning Research* does try to encourage authors to add electronic appendices with source code, data, demonstrations: anything, as the journal puts it, that will make life easier or more interesting for readers and researchers who follow in the authors' footsteps. Some authors do an excellent job, but spreading the good practice is an uphill struggle. Machine learning will change our uses of computers dramatically, so let us hope the journal achieves its goals with more and more success.

Harold Thimbleby is director of the University College London Interaction Centre.

Different sorts of knowledge

- Objective
- Relative
- Subjective
- Imaginative

Reliable knowledge

- "Independent of context"
- True for me, true for you
- True any where, any time

… *reproducible*

Some science criteria

- Publish facts
- Provide sufficient methodology
- Concise

- *Key idea:*

Reproducibil

Cargo cult science

- "but they don't land!"
- **Radical integrity**
- Leaning over backwards
- Explicit knowledge type

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

CS is different?

- 'Artificial science' — creative
- Envisions the future
- Close to industry and IPR worries

- IS, animation, simulation…
- IR v DL…
- AI v ML…
- *Etc*

Chemistry v CS

- Test tube
  - Production in milligrams
- Chemical engineering
  - Production line, quality control, safety, in the tonnes
- Projections of infinite objects
- Clearer distinctions
- *Very little imagination*

Artificial science & irreproducibility

Constantine v. Feynman

Some time with Feynman

– Leonard Mlodinow, Penguin, 2003

Constantine's claim to fame was his computer calculation… but there was a rumour going around that Constantine did not translate the problem to the computer in an honest way. "What's the big deal?" Constantine said. "I used what I knew to improve my computer model. Everybody does that."

I told Richard Feynman…He just shrugged. I thought he'd say, "What a louse! He did it because he thought what was important was success, not discovery."

Feynman replied, "Hell no. I'm not going to psychoanalyze the guy. But what should bother you as much as whether or not your friend fudged his work is that **a lot of people read it and couldn't tell the difference**. There are so many people out there not being skeptical, or not understanding what they are doing. They're all just following along. That's what we have—too many followers, too few leaders."

A survey

- For papers that describe systems
  - About 1/3 not reproducible:
    - Lost
    - Broken
    - Proprietary
    - Never worked

Does it matter? "No!"

- You wouldn't give away your lab keys
- Science progresses anyway
- Don't give answers to students

Status?

"I know what I'm talking about. It's not a problem. I've published in *Nature*…"

"I know what you're talking about. It *is* a problem…

My group couldn't replicate a paper in *Nature*…"

Does it matter? "Yes!"

- Maybe mistaken

- Maybe could be improved

- "We took 17 [more] months to get our program to work…"

5 years later
… for a 2000 paper

- "The system model and data model are [a] kind of higher level abstraction. We are **in the progress** of building up the multimedia system with the features described in our model. […] we did not think of distribution."

Real issues

- Fraud is a small problem (hopefully!)

- Easier incremental publishing — *but*

- Reproduction helps the reproducer understand issues properly

- Helps improve what was really done — both for author and

Three methodologies

- Informal review
- Refined surveys
- Double-checking

Journals

- *Journal of Machine Learning Research*

- *Computer Journal*

- *Software—Practice & Experience*

Note on *JMLR*

- Founded 2000, by editors of *Machine Learning*

  - A Kluwer journal, that cost $1000 pa

- Free electronic edition, www.jmlr.org

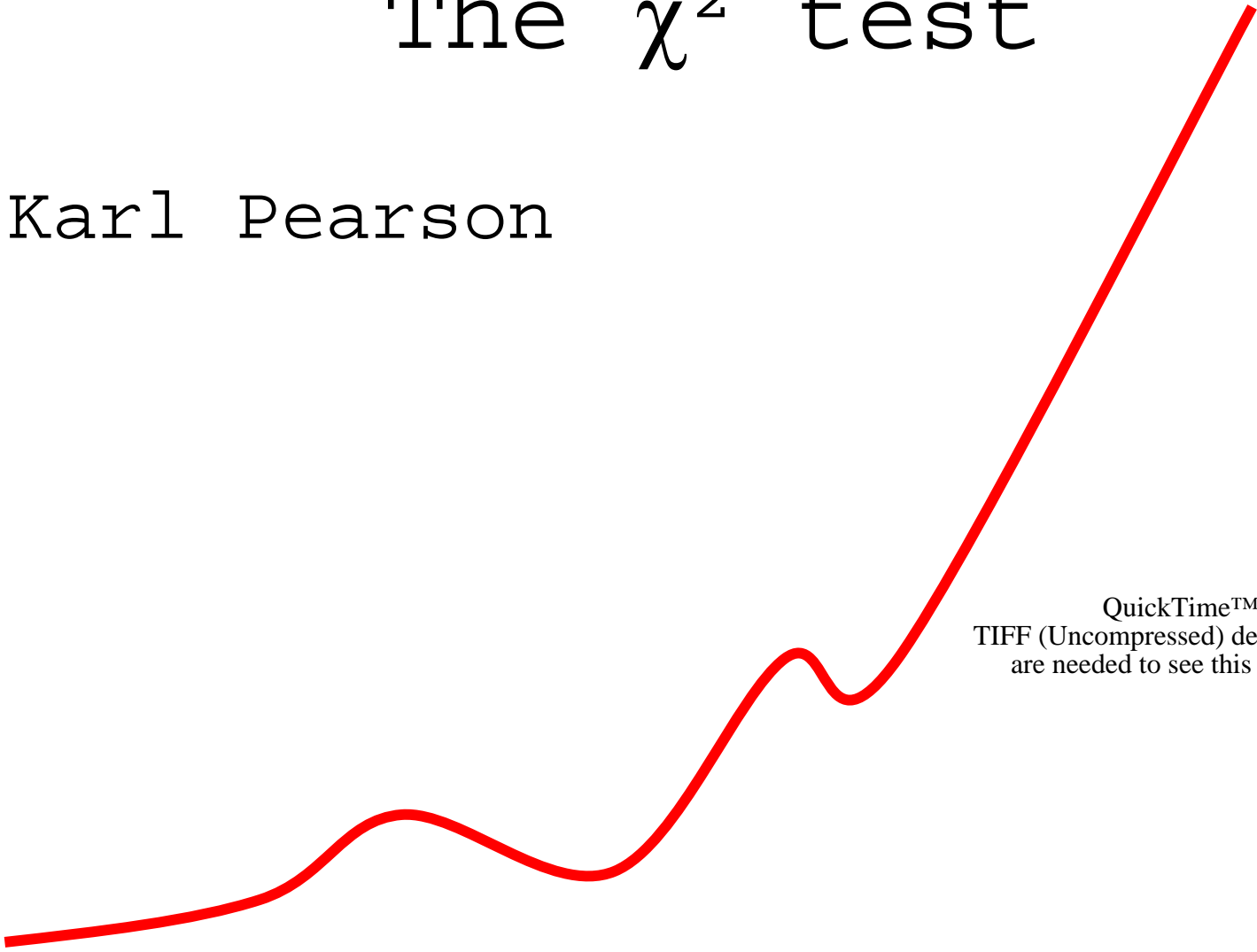- MIT Press does the paper edition

- Highest ISI impact factor for

Preliminary results

	Response	Systems	Broken	**%**
CJ	32 / 201	27	7	**26**
JLMR	111 / 162	62	23	**37**
SP&E	33 (73 papers) / 218	29	13	**45**

The χ^2 test

Karl Pearson

Systematic errors?

- Confessions?

- Charmed?

- Classifications?

- Even my programming!

Other examples

- Porter stemming
- Chinese Postman
- Wayne Gray's 'damaged goods'
- "because our prototype is fragile we cannot make it available"
- "what is described here is neither what we implemented nor what we are implementing"
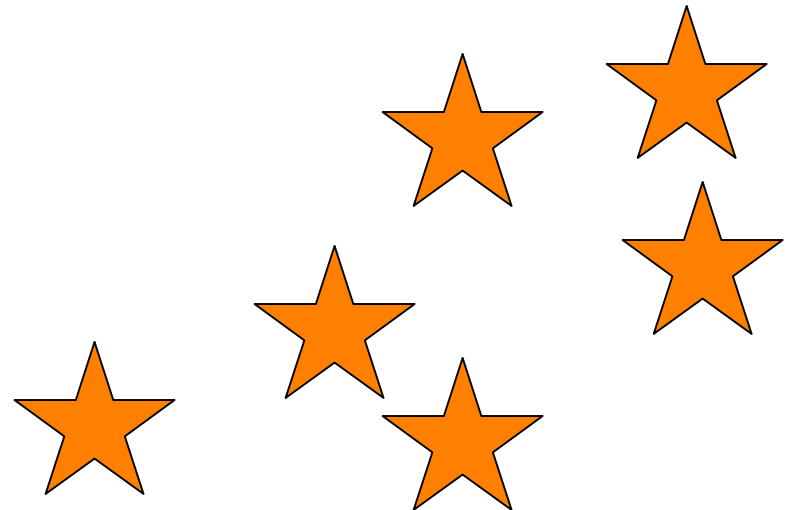
Me

- Lost programs

- Inaccessible programs

- Obsolete programs

- All of my stuff worked (subject to typos)

- Journals never once asked for a program

- *S-PAE, ACM TOCHI*
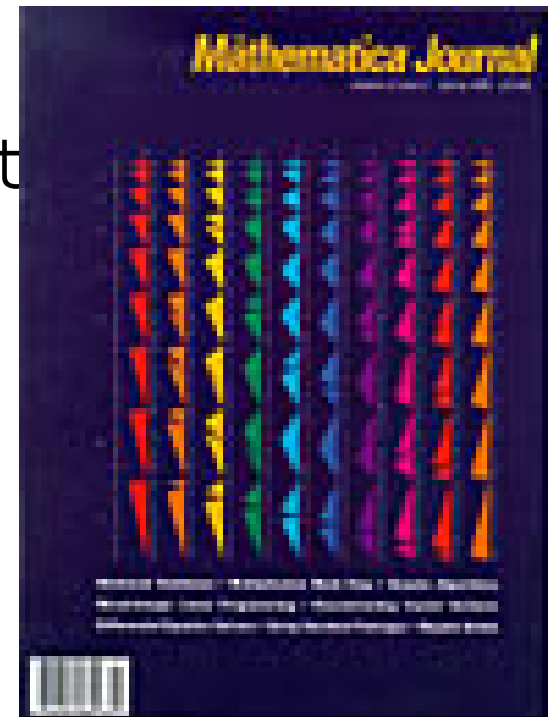
- *I could do better!*

What can be done?

- Awareness

- Open source

- Editorial policy

- *Use computers*!

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Examples

- Chinese Postman
- Genaille's rods
- Mobile phone paper
- Planets and star clust
- **An idea, no product**
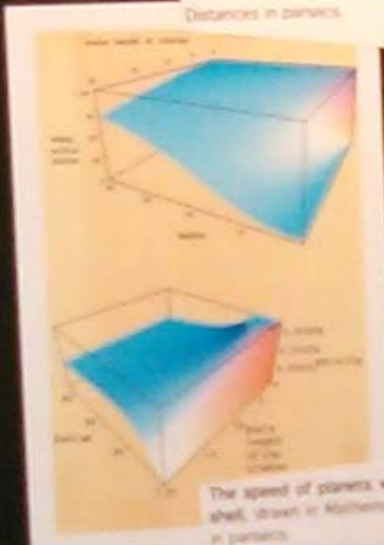
Demo *(1993)*

Isaac Thimbleby

Simon Balle School

in star clusters through computer modelling the**BA**

Model drawn in Mathematica.
Distances in parsecs.

The speed of planets within each
shell, drawn in Mathematica. Distances
in parsecs.

So:

- Build a tool to help…
- A contribution…

Core idea

- Programs
  - They are experiments, results, methods…
  - They are lab notebooks
  - They are text
  - They are easy to reproduce

Some Java

```java
public static void sort(Comparable a[], int sortType)
    {
        // <warp file='identify.xml'>
        int sentinel_position = 0;
        Comparable sentinel_value = a[sentinel_position];
        for( int i = 1; i < a.length; i++ )
                if( a[i].compareTo(sentinel_value) < 0 )
                        sentinel_value = a[sentinel_position = i];
        // </warp>
        :
        :
        :
    }
```

Java → XML → LaTEX

```
% Wed Apr 02 23:03:41 BST 2003 - warped code from Sort/InsertSort.java
% that matches file='identify.tex' -latex
% Define \codeIndent#1{...} if you want to recover original indentation
\expandafter\ifx\csname codeIndent\endcsname\relax\else\codeIndent{8}\fi
% Define \codeFormat to override default code format
{\expandafter\ifx\csname codeFormat\endcsname\relax
    \small \parindent 0pt \parskip -1pt \ttfamily \frenchspacing \obeylines
\fi
\codeFormat
int~sentinel\_position~=~0;
Comparable~sentinel\_value~=~a[sentinel\_position];
for(~int~i~=~1;~i~<~a.length;~i++~)
~~~~if(~a[i].compareTo(sentinel\_value)~<~0~)
~~~~~~~~sentinel\_value~=~a[sentinel\_position~=~i];
}
```

As published

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

Warp

- Part of the solution
- Very simple
- Helps improve quality in many ways:
  - For writing
  - For reproduction
- **Supports writing as transformation**

Don Knuth

"Science is what we understand well enough to explain to a computer."

because *programs run and*
are *reproducible*

Dick Feynman

"I'm talking about a specific, extra **type of integrity** that is not lying, but bending over backwards to show how you're maybe wrong, that **you ought to have** when acting as a scientist. And this is **our responsibility** as scientists, certainly to

Conclusions

- "A certain type of integrity"

- CS is (and its referees are?)
  too close to hype, PoP & IPR

- We can easily improve (e.g., warp)

- **Question for you:
  how reproducible is your work?**

- **(promised demo)**

Question
for you