Open Science, Open Source and R

Free software will save psychology from the Replication Crisis.

By Andy Wills

"Study reveals that a lot of psychology research really is just 'psycho-babble'".
—*The Independent.*

Psychology changed forever on the August 27, 2015. For the previous four years, the 270 psychologists of the Open Science Collaboration had been quietly re-running 100 published psychology experiments. Now, finally, they were ready to share their findings. The results were shocking. Less than half of the re-run experiments had worked.

When someone tries to re-run an experiment, and it doesn't work, we call this a *failure to replicate*. Scientists had known about failures to replicate for a while, but it was only quite recently that the extent of the problem became apparent. Now, an almost existential crisis loomed. That crisis even gained a name: the Replication Crisis. Soon, people started asking the same questions about other areas of science. Often, they got similar answers. Only half of results in economics replicated. In pre-clinical cancer studies, it was worse; only 11% replicated.

Open Science

Clearly, something had to be done. One option would have been to conclude that psychology, economics and parts of medicine could not be studied scientifically. Perhaps those parts of the universe were not lawful in any meaningful way? If so, you shouldn't be surprised if two researchers did the same thing and got different results.

Alternatively, perhaps different researchers got different results because they were doing different things. In most cases, it wasn't possible to tell whether you'd run the experiment exactly the same way as the original authors. This was because all you had to go on was the journal article—a short summary of the methods used and results obtained. If you wanted more detail, you could, in theory, request it from the authors. But, we'd already known for a decade that this approach was seriously broken—in about 70% of cases, data requests ended in failure.

Even when the authors sent you their data, it often didn't help that much. One of the most common problems was that when you re-analysed their data, you ended up with different answers to it! This turned out to be quite common, because most descriptions of data analyses provided in journal articles are incomplete and ambiguous. What you really needed was the original authors' source code—an unambiguous and complete record of every data processing step they took, from the raw data files, to the graphs and statistics in the final report. In Psychology in 2015, you almost never could get this.

If you did eventually manage to replicate the authors' analysis, could you be confident that their results were real? Not necessarily. Perhaps they tested only a few people, who were not particularly representative of the population as a whole. In this case, you might want to re-run the experiment yourself, testing a lot more people. Or perhaps the problem was not with their analysis, or their data, but with the method by which they collected their data. For the last 20 years, psychology experiments largely have involved computer-based testing. So, for very many experiments, there is a complete and unambiguous specification of the methods used—the source code for the testing program. But in 2015, this almost never was publicly available either.

In other words, psychology research at the beginning of the Replication Crisis was like closed-source software. You had to take the authors' conclusions entirely on trust, in the same way you have to trust that closed-source software performs as described. There essentially was no way to audit research properly, because you could not access the source code on which the experiment was based—the testing software, the raw data and the analysis scripts.

A growing number of scientists felt this had to change. The year before, in 2014, I had read Stephen Levy's *Hackers*, and from there, I went on to read more about Richard Stallman, Eric S. Raymond and Linus Torvalds. For me, it was a revelation. The Free and Open Source Software community, I felt, showed how science could be different. The pervasiveness of Linux showed that tens of thousands of people with different views and goals could collaborate on a complex project for the common good. Just as important, they could do so without necessarily even having to like each other all the time. That was good news for science. Discussions between academics can get...well, let's just say "heated".

So, in the same way that computing has its advocates of open-source software, psychology and other sciences started gaining advocates for Open Science. The phrase Open Science had been coined back in 1998 by Steve Mann, but once the Replication Crisis hit psychology, a lot more of us began to sit up and take notice. Early on, the Centre for Open Science, a non-profit company started in 2013, had set up the Open Science Framework (OSF). The OSF is a web-based public repository for experiment-related data and code. It's built entirely from free and open-source software.

As awareness of the Replication Crisis grew, peer reviewers started insisting that data and code be made publicly available. Peer review in research is a bit like a code review in IT. Scientists send their articles to a journal for consideration. The journal sends the article out to experts in the field for comment, and the work is accepted for publication only when the journal editor thinks those comments have been adequately addressed. In 2015, Richard Morey and colleagues started the Peer Reviewers' Openness Initiative, a declaration that they would not recommend any paper for publication unless it met certain basic standards of open science. Within three years, more than 500 peer reviewers in psychology had signed that declaration.

Open Platforms and R

There's still one major problem to solve. Publishing your scientific source code is essential for open science, but it's not enough. For fully open science, you also need the platforms on which that code runs to be open. Without open platforms, the

future usability of open-source code is at risk. For example, there was a time when many experiments in psychology were written in Microsoft Visual Basic 6 or in Hypercard. Both were closed-source platforms, and neither are now supported by their vendors. It is just not acceptable to have the permanent archival records of science rendered unusable in this way. Equally, it's a pretty narrow form of Open Science, if only those who can afford to purchase a particular piece of proprietary software are able to access it. All journal articles published in psychology since around 1997 are in PDF format. Academic libraries would not tolerate these archival files being in a proprietary format such as DOCX. We can and must apply the same standards of openness to the platforms on which we base our research.

Psychology has a long history of using closed-source platforms, perhaps most notably the proprietary data analysis software SPSS. SPSS initially was released in 1968, and it was acquired by IBM in 2010. Bizarrely, SPSS is such a closed platform, current versions can't even open SPSS output files if they were generated before 2007! Although it's still the most used data analysis software in psychology, its use has been declining steeply since 2009. What's been taking up the slack?

In large part, it's R. R is a GNU project, and so it's free software released under the GNU General Public Licence. It works great under Linux, but it also works just fine on Windows and Mac OS too. R is a very long-standing project with great community support. R is also supported by major tech companies, including Microsoft, who maintain the Microsoft R Application Network.

R is an implementation of the statistical language S, developed at Bell Labs shortly after UNIX, and inspired by Scheme (a dialect of Lisp). In the 1990s, Ross Ihaka and Robert Gentleman, at the University of Auckland, started to develop R as an open-

source implementation of S. R reached version 1.0 in 2000. In 2004, the R Foundation released R 2.0 and began its annual international conference: _useR!_. In 2009, R got its own dedicated journal (*The R Journal*). In 2011, **RStudio** released a desktop and web-based IDE for R. Using R through RStudio is the best option for most new users, although it also works well with Emacs and Vim. The current major release of R landed in 2013, and there are point releases approximately every six months.

I first started using R, on a Mac, in 2012. That was two years before I'd heard of the concept of Free Software, and it was about three years before I ran Linux regularly. So, my choice to move from SPSS to R was not on philosophical grounds. It also was before the Replication Crisis, so I didn't switch for Open Science reasons either. I started using R because it was *just better* than SPSS—a lot better. Scientists spend around **80% of their analysis time on pre-processing**—getting the data into a format where they can apply statistical tests. R is fantastically good at pre-processing, and it's certainly much better than the most common alternative in psychology, which is to pre-process in Microsoft Excel. Data processing in Excel is infamously error-prone. For example, **one in five experiments in genetics have been screwed up by Excel**. Another example: the case for the **UK government's policy of financial austerity** was based on an Excel screw up.

Another great reason for using R is that all analyses take the form of scripts. So, if you have done your analysis completely in R, you already have a full, reproducible record of your analysis path. Anyone with an internet connection can download R and reproduce your analysis using your script. This means we can achieve the goal of fully open, reproducible science really easily with R. This contrasts with the way psychologists mainly use SPSS, which is through a point-and-click interface. It's a fairly common experience that scientists struggle to reproduce their own SPSS-based analysis after a three-month delay. I struggled with this issue myself for years. Although I was always able to reproduce my own analyses eventually, it often took as long to do so as it had the first time around. Since I moved to R, reproducing my own analyses has become as simple as re-running the R script. It also means that now every member of my lab and anyone else I work with can share and audit each other's analyses easily. In many cases, that audit process substantially improves the analysis.

A third thing that makes R so great is that the core language is supplemented by more than 13,000 packages mirrored worldwide on the Comprehensive R Archive Network (CRAN). Every analysis or graph you can think of is available as a free software package on CRAN. There's even a package to draw graphs in the style of the xkcd cartoons or in the colour schemes of Wes Anderson movies! In fact, it's so comprehensive, in 2013 the authors of SPSS provided users with the ability to load R packages within SPSS. Their in-house team just couldn't keep up with the breadth and depth of analysis techniques available in R.

R's ability to keep up with the latest techniques in data analysis has been crucial in addressing the Replication Crisis. This is because one of the causes of the Crisis was psychology's reliance on out-dated and easily misinterpreted statistical techniques. Collectively, those techniques are known as Null Hypothesis Significance testing, and they were developed in the early 20th century before the advent of low-cost, high-power computing. Today, we increasingly use more computationally intensive but better techniques, based on Bayes theorem and Monte Carlo techniques. New techniques become available in R years before they're in SPSS. For example, in 2010, Jon Kruschke published a textbook on how to do Bayesian analysis in R. It wasn't until 2017 that SPSS supported Bayesian analyses.

Open Science in R: an Example

In my lab at Plymouth University, we work on the psychology of learning, memory and decision making. In many cases, the theories we are testing are expressed in the form of computer models. For example, one of the classic theories of how we learn to group objects into categories (dogs, cats, bagels and so on) is called ALCOVE. This theory takes the form of a neural network model, which makes predictions about how people will classify objects. We compare those predictions to data from real people making those decisions and evaluate the model on that basis.

Teaching R

For more than 20 years, teaching statistics in psychology has been synonymous with teaching people how to use SPSS. However, during the last few years, several universities have switched to R, and many more are considering it. One fear about this change was that psychology students would find R harder to learn than SPSS, and that they would like it less. This turns out to be incorrect. In pioneering work by Dale Barr and colleagues at Glasgow University, psychology undergraduates were taught both SPSS and R. They then got to choose which software to use in their final assessments. Around **two-thirds of the students chose R**. Those who chose R also scored higher on the exam. They also scored lower on a standard measure of statistics anxiety. At Plymouth University, new entrants to our Psychology degrees are now just taught R, with SPSS removed from the curriculum entirely. We've seen an increase in what our students can achieve in statistics, while maintaining high levels of student satisfaction.

One of the side benefits of this change, for the R project, is that psychologists tend to be quite good at writing documentation. Andy Field's textbook, *Discovering Statistics*, much-praised by Psychology undergraduates, has had an R version since 2012. More recently, academics have started developing teaching materials that are as open as R is. For example, my own teaching materials, *Research Methods in R*, aimed at first-year psychology undergraduates, are available under a Creative Commons Licence. *Just Enough R*, written by Ben Whalley and aimed at postgraduate students, is available under the same licence.

Traditionally, this computational modelling side to psychology has been fairly closed-source. These models of the mind, which are moderately complex programs, typically are released only as a set of mathematical equations with some explanatory text. The code required to reproduce the results reported is seldom fully published. The result is that it can take several days to several months to reproduce the results of these computer models. The amount of time this wastes is substantial.

Starting in 2016, our lab decided to do something about this issue. Specifically, we released an R package called catlearn, short for models of CATegorization and LEARNing. In the current version, released in July 2018, we have implemented nine different models. Like all R packages, the code is open source. The package also includes archives of the full code for simulations of specific experiments and the data sets for those experiments. We're beginning to build a community around the world, with people in the USA, UK, Germany and Switzerland all having contributed code. It's a really exciting time, and I'm looking forward to the release of version 0.7 later this year. If you'd like to contribute, we'd love to hear from you—we desperately need more good programmers. Prior experience of psychology is not essential.

A Final Thought

The Replication Crisis might have been one of the best things ever to happen to psychology. It became a catalyst for much-needed change to our scientific processes. If we can build 21st-century psychology on the principles of Open Science, I think great and enduring discoveries await us. Those future successes will owe a lot to the pioneering example of the Free and Open-Source Software community. Thanks in advance, *Linux Journal* readers! ■

Andy Wills is a Professor of Psychology at the University of Plymouth. He published his first source code, in BBC BASIC, in 1984. Since 1997, he's published around 80 research articles on the psychology and neuroscience of learning and memory. He lives on the south-west coast of England with his wife, daughter and two cats. In his spare time, he plays around with synthesizers. You can find him on Twitter @ajwills72 or via his website.

Resources

- "Study reveals that a lot of psychology research really is just 'psycho-babble'" by Steve Connor, *The Independent*

- "Estimating the Reproducibility of Psychological Science" by Alexander A. Aarts, Christopher J. Anderson, Joanna E. Anderson and Peter Attridge, *Science*

- Replication Crisis

- "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say 'Usually Not'" by Andrew C. Chang and Phillip Li, *Finance and Economics Discussion Series 2015-083. Washington: Board of Governors of the Federal Reserve System*

- "Reproducibility: Six red flags for suspect work" by C. Glenn Begley, *Nature*

- "The poor availability of psychological research data for reanalysis" by Jelte Wicherts, Judith Kats, Denny Borsboom and Dylan Molenaar, *American Psychologist*

- *Hackers: Heroes of the Computer Revolution* by Steven Levy, Doubleday, 1984 (Wikipedia)

- Open Science (Wikipedia)

- COS (Center for Open Science)

- Open Science Framework

- Peer Reviewers' Openness Initiative

- SPSS Statistics (Wikipedia)

- "The Popularity of Data Science Software" by Robert A. Muenchen, r4stats.com

- The R Project for Statistical Computing

- The GNU Operating System

- GNU General Public License

- Microsoft R Application Network (MRAN)

- RStudio

- "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says" by Gil Press, *Forbes*

- "Gene name errors are widespread in the scientific literature" by Mark Ziemann, Yotam Eren and Assam El-Osta, *Genome Biology*

- "Microsoft Excel: The ruiner of global economies?" by Peter Bright, *Ars Technica*

- The Comprehensive R Archive Network

- xkcd

- "Calling R from SPSS" by Catherine Dalzell

- Doing Bayesian Data Analysis

- Bayesian statistics (IBM Knowledge Center)

- LTC Workshop

- *Discovering Statistics Using R* by Andy Field, Jeremy Miles and Zoe Field, Sage Publishing

- Research Methods in R by Andy Wills (Teaching Materials)

- Ben Whalley's Just Enough R

- catlearn GitHub Page

- Acorn Programs

Send comments or feedback
via https://www.linuxjournal.com/contact
or email ljeditor@linuxjournal.com.