

# Introduction to Open Research

Dr Andrew J. Stewart

Fellow of the Software Sustainability Institute

E: [drandrewjstewart@gmail.com](mailto:drandrewjstewart@gmail.com)

T: [@ajstewart\\_lang](https://twitter.com/ajstewart_lang)

G: [ajstewartlang](https://github.com/ajstewartlang)



Let's travel in my time machine...

Back to 2011...

## 'Proof' of Extrasensory Perception



# Have scientists really discovered proof of ESP?

*The Week Staff*



Beyond palm reading: The fact that a credible scientific journal is publishing Dr. Bem's research has some wondering if ESP is more than a new-age hobby or a scam. Corbis

January 7, 2011

**T**he rigorous, widely respected *Journal of Personality and Social Psychology* will publish a paper later this year offering "strong evidence" that extra-sensory perception (ESP) exists. Although Daryl J. Bem, an emeritus professor at Cornell University, claims his tests of over 1,000 college students over eight years have yielded proof of ESP, his findings have provoked "amusement and scorn" from the scientific community. Should we believe Bem, or do his claims give serious science a bad name?

hobby or a scam. Corbis

Ad closed by Google

[Report this ad](#)

[Why this ad?](#) [I ♡](#)

# Bem (2011)

The *Journal of Personality and Social Psychology* published a paper by Daryl Bem showing that the future can influence the present - in one study, using a standard priming participants responded **faster** to targets that were then followed by a prime word, than to targets that weren't.

Bem had a distinguished track record, the paper came out in a top tier journal, went through rigorous peer review, and used standard statistical and scientific methods.

---

# Daryl Bem Proved ESP Is Real

Which means science is broken.

MAY 17, 2017 • COVER STORY

---

<https://slate.com/health-and-science/2017/06/daryl-bem-proved-esp-is-real-showed-science-is-broken.html>

# Simmons, Nelson, and Simonsohn (2011)

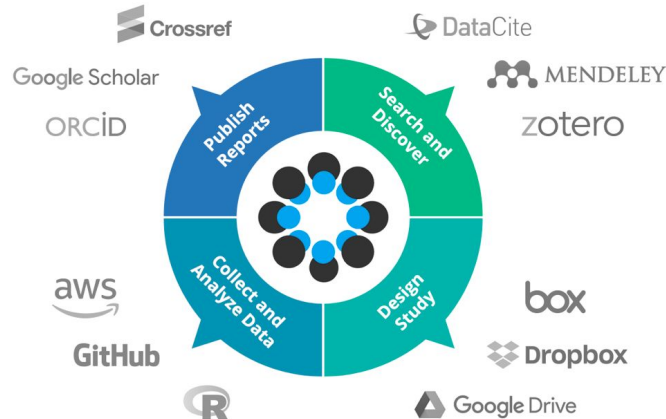
“False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant” published in *Psychological Science*.

Simmons et al. (2011) show that selectively reporting data (e.g., dropping participants, ‘problematic’ trials) and selectively reporting analyses (e.g., only reporting comparisons that are significant) results in vastly inflated false positives.

Later termed *p*-hacking.

# Open Science Framework (OSF)

Brian Nosek set up a series of replication studies to try to determine how big a replication issue psychology might be facing. This resulted in the establishment of the Centre for Open Science (CoS) in 2013.









# Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance

Psychological Science  
21(10) 1363–1368  
© The Author(s) 2010  
Reprints and permission:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0956797610383437  
<http://pss.sagepub.com>  

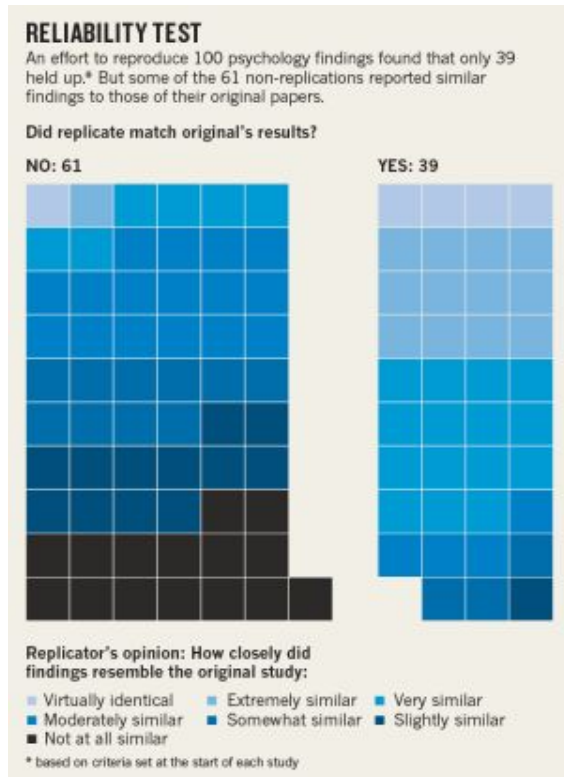

Dana R. Carney<sup>1</sup>, Amy J.C. Cuddy<sup>2</sup>, and Andy J. Yap<sup>1</sup>

<sup>1</sup>Columbia University and <sup>2</sup>Harvard University

## Abstract

Humans and other animals express power through open, expansive postures, and they express powerlessness through closed, contractive postures. But can these postures actually cause power? The results of this study confirmed our prediction that posing in high-power nonverbal displays (as opposed to low-power nonverbal displays) would cause neuroendocrine and behavioral changes for both male and female participants: High-power posers experienced elevations in testosterone, decreases in cortisol, and increased feelings of power and tolerance for risk; low-power posers exhibited the opposite pattern. In short, posing in displays of power caused advantaged and adaptive psychological, physiological, and behavioral changes, and these findings suggest that embodiment extends beyond mere thinking and feeling, to physiology and subsequent behavioral choices. That a person can, by assuming two simple 1-min poses, embody power and instantly become more powerful has real-world, actionable implications.

# Estimating the reproducibility of psychological science (Nosek et al., 2015)



270 authors tried to replicate 100 experiments drawn from high profile Psychology journals - *Psychological Science*, *Journal of Personality and Social Psychology*, and *Journal of Experimental Psychology: Learning, Memory, and Cognition*.



# Power Posing – 2010 vs. 2016

**Appearance:** Big ... very big. Spread your hands and legs wide, argued the authors, and you will both exude power and - this was the new finding - feel great. Adopt a power pose and your testosterone rises and your stress levels fall. Or, as columnist David Brooks neatly **put it**: “If you act powerfully, you will begin to think powerfully.”

**And now?** Well, that’s the odd thing. One of the original report’s three authors, Dana Carney, **says** it was all nonsense. “I do not believe that ‘power pose’ effects are real,” she wrote in a **blog** that detailed the original research’s methodological failings. Standing like John Wayne in a gunfight does not make you feel like a successful gunslinger. It just makes you look silly.

<https://www.theguardian.com/politics/shortcuts/2016/sep/28/george-osbornes-power-pose-the-science-proves-feeble>

# Many replications failed or reported effect sizes much smaller than in the original...

- Power posing
- Ego depletion
- Social priming
- Learning styles
- Stanford prison experiment
- Growth mindset
- and many more...

# We've known about the problems for a while...

Paul Meehl, 1967

114

PAUL E. MEEHL

test, there exists among psychologists (c) a fairly widespread tendency to report experimental findings with a liberal use of *ad hoc* explanations for those that didn't "pan out." This last methodological sin is especially tempting in the "soft" fields of (personality and social) psychology, where the profession highly rewards a kind of "cuteness" or "cleverness" in experimental design, such as a hitherto untried method for inducing a desired emotional state, or a particularly "subtle" gimmick for detecting its influence upon behavioral output. The methodological price paid for this

# We've known about the problems for a while...

Doug Altman, 1994

As the system encourages poor research it is the system that should be changed. We need less research, better research, and research done for the right reasons. Abandoning using the number of publications as a measure of ability would be a start.

DOUGLAS G ALTMAN  
Head

Medical Statistics Laboratory,  
Imperial Cancer Research Fund,  
London WC2A 3PX



# We've known about the problems for a while...

Personality and Social Psychology Review  
1998, Vol. 2, No. 3, 196–217

Copyright © 1998 by  
Lawrence Erlbaum Associates, Inc.

## **HARKing: Hypothesizing After the Results are Known**

**Norbert L. Kerr**

*Department of Psychology  
Michigan State University*

*This article considers a practice in scientific communication termed HARKing (Hypothesizing After the Results are Known). HARKing is defined as presenting a post hoc hypothesis (i.e., one based on or informed by one's results) in one's research report as if it were, in fact, an a priori hypotheses. Several forms of HARKing are identified and survey data are presented that suggests that at least some forms of HARKing are widely practiced and widely seen as inappropriate. I identify several reasons why scientists might HARK. Then I discuss several reasons why scientists ought not to HARK. It is conceded that the question of whether HARKing's costs exceed its benefits is a complex one that ought to be addressed through research, open discussion, and debate. To help stimulate such discussion (and for those such as myself who suspect that HARKing's costs do exceed its benefits), I conclude the article with some suggestions for deterring HARKing.*

# We've known about the problems for a while...

## PLOS MEDICINE

 OPEN ACCESS

ESSAY

### Why Most Published Research Findings Are False

John P. A. Ioannidis

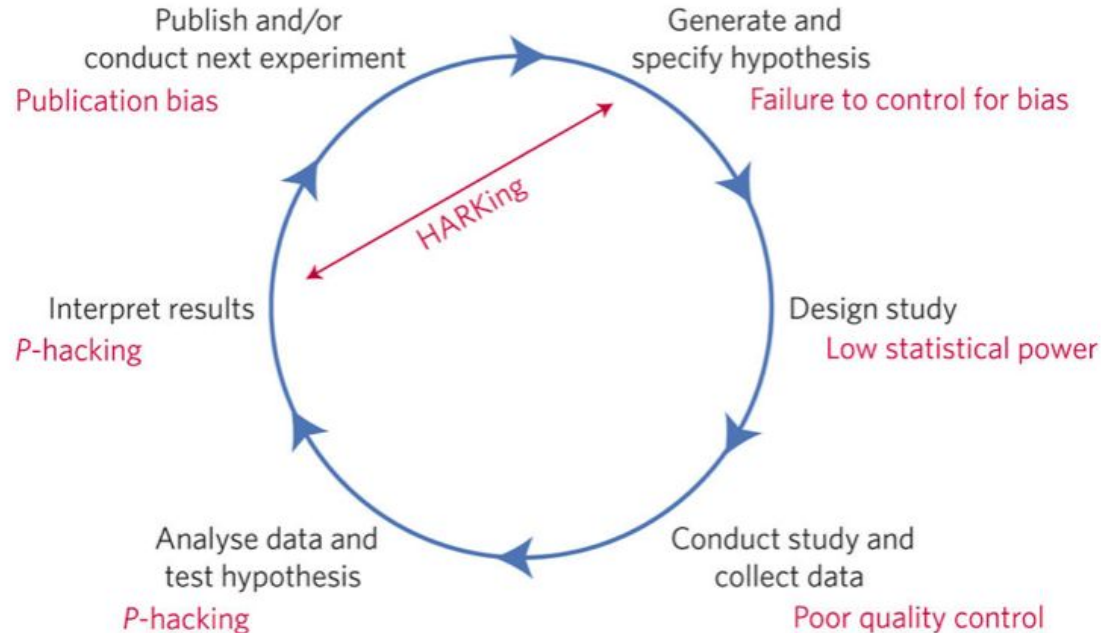
Published: August 30, 2005 • <https://doi.org/10.1371/journal.pmed.0020124>

Is there not just “good science” and “bad science”?

Without realising it, good scientists have been engaging in questionable research practices (QRPs) partly driven by an incentive structure that doesn't incentivise good scientific practice...

Problems include *p*-hacking, lack of power, HARKing, failing (refusal) to share data and code, too many researcher degrees of freedom...

From: *A manifesto for reproducible science*



Munafo et al. (2017), *Nature Human Behaviour*

# Why do so many studies not replicate?

- There are too many studies with experimental power too low to detect the effect size of interest.
- One of the consequences of a low powered study is that when real effects are detected their magnitude is likely to be over-estimated.
- Studies which find the effect are published and studies that don't are not published - due to a bias to publish positive results.
- Future work may use the published effect size during a *a priori* power analysis (and then fail to find the effect as the new study is effectively under-powered for what it's looking for).

# Low Statistical Power

Button et al. (2013), Nature Reviews Neuroscience, small sample size undermines the reliability of neuroscience. Nord et al., (2017), Journal of Neuroscience, highlight wide heterogeneity in power in neuroscience studies.

**Table 2. Median, maximum, and minimum power subdivided by study type**

Group	Median power (%)	Minimum power (%)	Maximum power (%)	2.5 <sup>th</sup> and 97.5 <sup>th</sup> percentile (based on raw data)	95% HDI (based on GMMs)	Total N
All studies	23	0.05	1	0.05–1.00	0.00 – 0.72, 0.80 – 1.00	730
All studies excluding null	30	0.05	1	0.05–1.00	0.01 – 0.73, 0.79 – 1.00	638
Genetic	11	0.05	1	0.05–0.94	0.00 – 0.44, 0.63 – 0.93	234
Treatment	20	0.05	1	0.05–1.00	0.00 – 0.65, 0.91 – 1.00	145
Psychology	50	0.07	1	0.07–1.00	0.02 – 0.24, 0.28 – 1.00	198
Imaging	32	0.11	1	0.11–1.00	0.03 – 0.54, 0.71 – 1.00	65
Neurochemistry	47	0.07	1	0.07–1.00	0.02 – 0.79, 0.92 – 1.00	50
Miscellaneous	57	0.11	1	0.11–1.00	0.09 – 1.00	38

ROBERT TAYLOR



## Rein in the four horsemen of irreproducibility

Dorothy Bishop describes how threats to reproducibility, recognized but unaddressed for decades, might finally be brought under control.

More than four decades into my scientific career, I find myself an outlier among academics of similar age and seniority: I strongly identify with the movement to make the practice of science more robust. It's not that my contemporaries are unconcerned about doing science well; it's just that many of them don't seem to recognize that there are serious problems with current practices. By contrast, I think that, in two decades, we will look back on the past 60 years — particularly in biomedical science — and marvel at how much time and money has been wasted on flawed research.

How can that be? We know how to formulate and test hypotheses in controlled experiments. We can account for unwanted variation with statistical techniques. We appreciate the need to replicate observations.

Yet many researchers persist in working in a way almost guaranteed not to deliver meaningful results. They ride with what I refer to as the four horsemen of the reproducibility apocalypse: publication bias, low statistical power, *P*-value hacking and HARKing (hypothesizing after results are known). My generation and the one before us have done little to rein these in.

In 1975, psychologist Anthony Greenwald noted that science is prejudiced against null hypotheses; we even refer to sound work supporting such conclusions as 'failed experiments'. This prejudice leads to publication bias: researchers are less likely to write up studies that show no effect, and journal editors are less likely to accept them. Consequently, no one can learn from them, and researchers waste time and resources

be adequately powered. Other disciplines have yet to catch up.

I stumbled on the issue of *P*-hacking before the term existed. In the 1980s, I reviewed the literature on brain lateralization (how sides of the brain take on different functions) and developmental disorders, and I noticed that, although many studies described links between handedness and dyslexia, the definition of 'atypical handedness' changed from study to study — even within the same research group. I published a sarcastic note, including a simulation to show how easy it was to find an effect if you explored the data after collecting results (D. V. M. Bishop *J. Clin. Exp. Neuropsychol.* **12**, 812–816; 1990). I subsequently noticed similar phenomena in other fields: researchers try out many analyses but report only the ones that are 'statistically significant'.

This practice, now known as *P*-hacking, was once endemic to most branches of science that rely on *P* values to test significance of results, yet few people realized how seriously it could distort findings. That started to change in 2011, with an elegant, comic paper in which the authors crafted analyses to prove that listening to the Beatles could make undergraduates younger (J. P. Simmons *et al. Psychol. Sci.* **22**, 1359–1366; 2011). "Undisclosed flexibility," they wrote, "allows presenting anything as significant."

The term HARKing was coined in 1998 (N. L. Kerr *Pers. Soc. Psychol. Rev.* **2**, 196–217; 1998). Like *P*-hacking, it is so widespread that researchers assume it is good practice. They look at the data, pluck out a finding that looks exciting and write a paper to tell a story around this result. Of course, researchers should be free to explore their

MANY RESEARCHERS  
PERSIST IN WORKING  
IN A WAY ALMOST  
GUARANTEED  
NOT  
TO DELIVER  
MEANINGFUL  
RESULTS.

How do we make our science more reproducible?



# The UKRN

## The power of networks

A group of researchers recently launched the [UK Reproducibility Network](#), supported by Jisc and a range of other stakeholders, including funders and publishers.

Our aim is to bring together colleagues across the higher education and research sector, forming local networks at individual institutions to promote the adoption of initiatives intended to improve research.

This is very much a peer-led, grassroots initiative that will allow academics to coordinate their efforts and engage with key stakeholders.

# The UKRN



The UK Reproducibility Network (UKRN) is a peer-led consortium that aims to ensure the UK retains its place as a centre for world-leading research.

This will be done by investigating the factors that contribute to robust research, providing training and disseminating best practice, and working with stakeholders to ensure coordination of efforts across the sector.

It is led by Marcus Munafò (Bristol), Chris Chambers (Cardiff), Laura Fortunato (Oxford), Alexandra Collins (Imperial), and Malcolm Macleod (Edinburgh).

# ReproducibiliTea

[Home](#)[About](#)[Journal Clubs](#)[Blog](#)[Calendar](#)[...](#)

## Welcome to ReproducibiliTea

We are a grassroots [journal club initiative](#) that helps young researchers create local Open Science journal clubs at their universities to discuss diverse issues, papers and ideas about improving science, reproducibility and the Open Science movement. Started in early 2018 at the University of Oxford, ReproducibiliTea has now spread to 81 institutions in 22 different countries. We are completely volunteer run, and provide a unique and supportive community for our members.

# ReproducibiliTea



# ReproducibiliTea

The ReproducibiliTea parent organisation is run by a group of ECR volunteers:

Amy Orben @OrbenAmy: high-level organisation and UKRN liaising

Sam Parsons @Sam\_D\_Parsons: podcast production

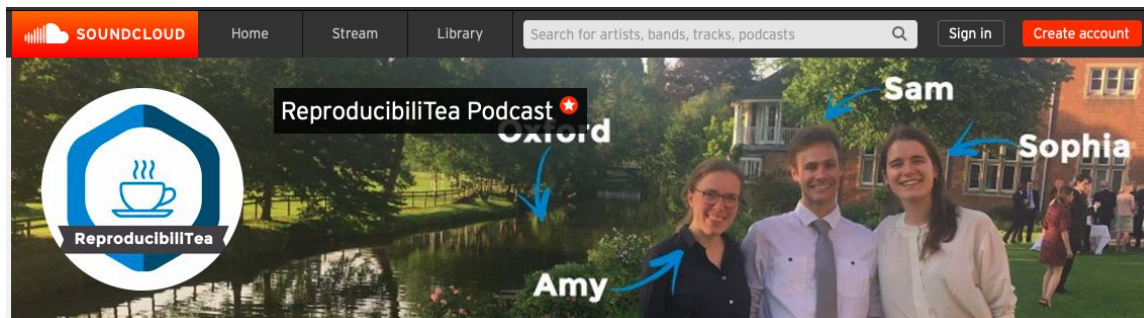
Sophia Crüwell @cruwelli: podcast scheduling and webinars

Matt Jaquierey @MJaquierey: website design and maintenance

Katie Drax @katiedrax: external communications

Jade Pickering @Jade\_Pickering: community building and merchandising

They even have a  
podcast!



# Institutional Academic Leads for Research Improvement and Integrity

Institutional academic leads should be independent of the grassroots Open Research Working Groups.

Ensure training around research improvement and research culture, including promoting the adoption of open research practices and other relevant initiatives, and embedding this into the institutional culture.

<https://www.bristol.ac.uk/psychology/research/ukrn/about/tor/>



Data Viz: Laura Fortunato

# Open Research Working Groups

Currently ~57 institutions with a UKRN local lead - many have set up Open Research Working Groups (ORWs).

The ORWs are grassroots led and should cross academic disciplines - not always easy - many traditional teaching-focused institutions also have ORWs - important that everyone (regardless of research/balance) is part of the conversation.

It's not just about how we do better research, it's also about how we teach our students how to do better research.





# Building A Regional Network



Recent NW Open Research Hub Events - Lancaster, Keele, Chester...

Such hubs may be involved in a Research England Development Fund bid - will be led by institutions with academic UKRN leads (so not Manchester).

Image courtesy of Dermot Lynott (Lancaster UKRN lead).



# Beware: Results May Vary (RMV20). Openness as a Way of Enhancing Research Quality.

---

A North West Open Research Hub event.

---

Friday, 28th February 2020, 14:00 – 17:30 GMT

Dalton Room, University of Manchester, Core Technology Facility, 46 Grafton Street,  
Manchester, M13 9WU.



Software  
Sustainability  
Institute



The University of Manchester

---

If you'd like to Tweet about today's event, please use the hashtag **#RMV20**

---

<https://hackmd.io/@ajstewartlang/r1ec-g5l8>

Start 14:00

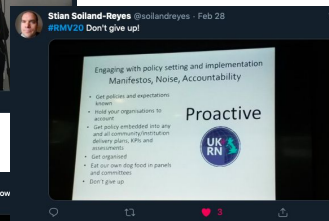
- Overview of Event - Andrew Stewart
- Introduction - Wendy Flavell (Vice Dean for Research in the Faculty of Science and Engineering, University of Manchester)
- The UK Reproducibility Network and starting local with open research - Dermot Lynott
- Better Software, Better Research: How the SSI is helping to promote reproducible research - Rachael Ainsworth
- How my unpublished work ended up on national television - Thomas Richardson

15:10(ish) Tea/Coffee Break

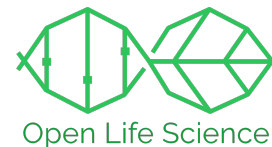
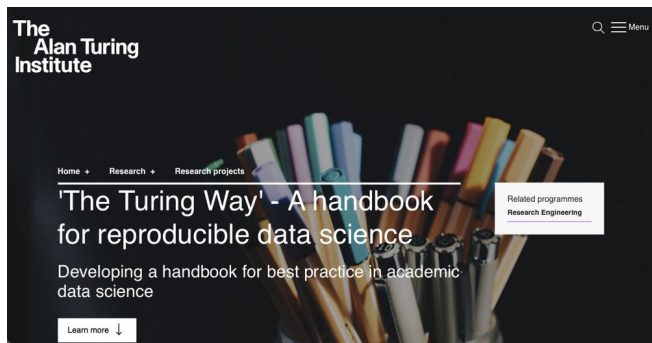
- Empowering early career researchers with ReproducibiliTea journal clubs - Jade Pickering
- Does the system reward scientific fraudsters? - David Eisner
- Wrangling (Inter)national Open Science Policy - Carole Goble

16:15(ish) Comfort Break

- **Keynote** - Four cognitive biases that make it hard to do science well, and how to overcome them - Dorothy Bishop



# Connecting Communities



# Institutions Need to Support/Train/Incentivise Researchers to Adopt Open Research Practices

- The academic incentive structure has to change.
- There should be less focus/reward for people who are doing expensive science, and more focus/reward for people who are doing better science. REF could play a role in this.
- Across disciplines and institutions, there can be a surprising lack of understanding about what open research is (and what it isn't).

# The Biggest Challenge: Education and Training

- We need to be teaching open research practices to our students (the next generation of researchers)..
- There is a huge computational skills gap amongst PhD students, postdoc, and teaching/research academics - without these skills, people simply cannot adopt open and reproducible research practices.