(Don't be) Bullied into Bad Science

Stephen J Eglen
https://sje30.github.io
sje30@cam.ac.uk

Cambridge Computational Biology Institute
University of Cambridge
@StephenEglen

Slides: http://bit.ly/eglen-riot (CC-BY license)

Declarations

1. Member of *Bullied into Bad Science* campaign
2. Affiliate editor of *bioRxiv*
3. Senior editor of *Scientific Data*

Acknowledgements

1. Corina Logan and Laurent Gatto.
2. Joined by Ross Mounce, Adrian Curry and myself. Joined by Lauren Maggio.

http://bulliedintobadscience.org has most of the key background

Do ask questions!

Thanks to Ben Farrar and Yvonne Nobis for material.

What does it mean to be bullied into bad science?

1. Science is a competitive arena; long work hours culture
2. Race to publish first
3. Not being allowed to shared resources
4. . . .

Problems

We are evaluated on where, not what, we publish

Evaluation of research by impact factor or journal title

$$I_y = \frac{C_{y-1} + C_{y-2}}{P_{y-1} + P_{y-2}}$$

C_y = number of citations in year y, P_y = number of publications in year y...

How could this possibly go wrong?

Strong pressure e.g. in China to pressure in journals with impact factor over some threshold (5).

So many ways to game the system: *A user's guide to inflated and manipulated impact factors* (Ioannidis and Thombs, 2019).

Current status of Open Access

Two academics walk into a bar. They bring their own drinks, pay $5000, and leave feeling both proud and ashamed. It's a publishing metaphor.

11:28 AM · Mar 13, 2015 · Twitter for iPhone

1K Retweets **802** Likes

Reproducibility crisis

Crisis? What crisis? Science is getting better, right?

"Negative results are disappearing from most disciplines and countries" (Fanelli, 2012). Study of ~4000 papers:

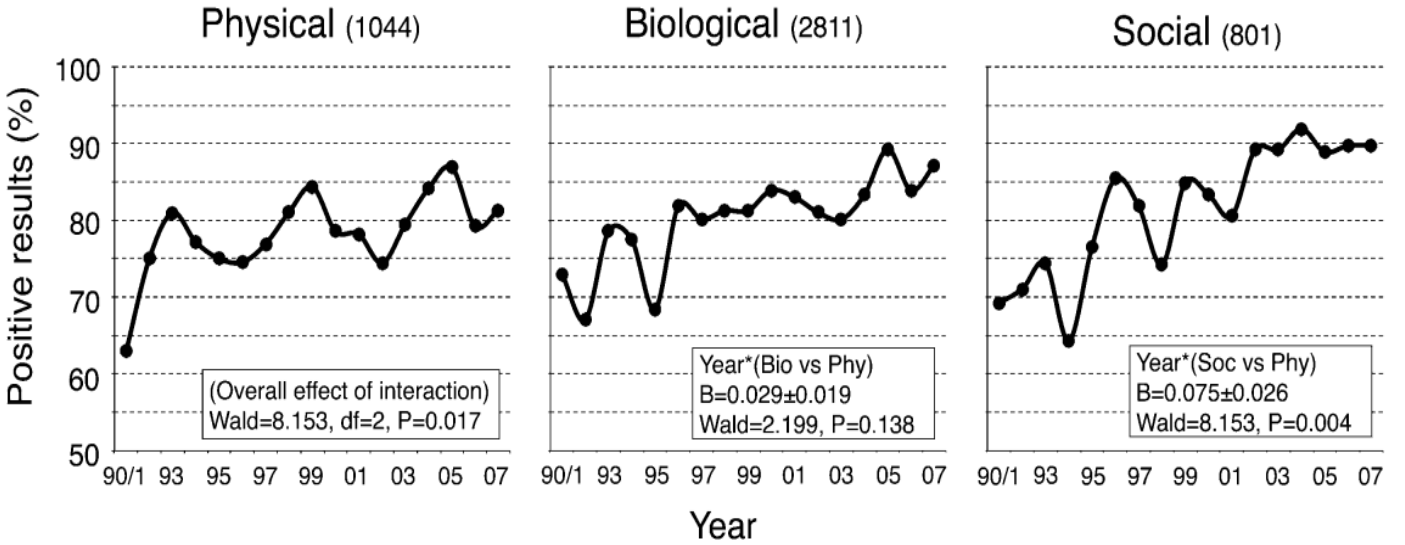


Fig. 3 Percentage of papers reporting a support for the tested hypothesis plotted against year of publication and divided by scientific domain of the journal (physical, biological and social sciences). Logistic regression estimates are interaction effects in a hierarchically well-formulated model. The main effects of this model, calculated with interaction components removed, are reported in Table 1. Numbers in brackets are sample size

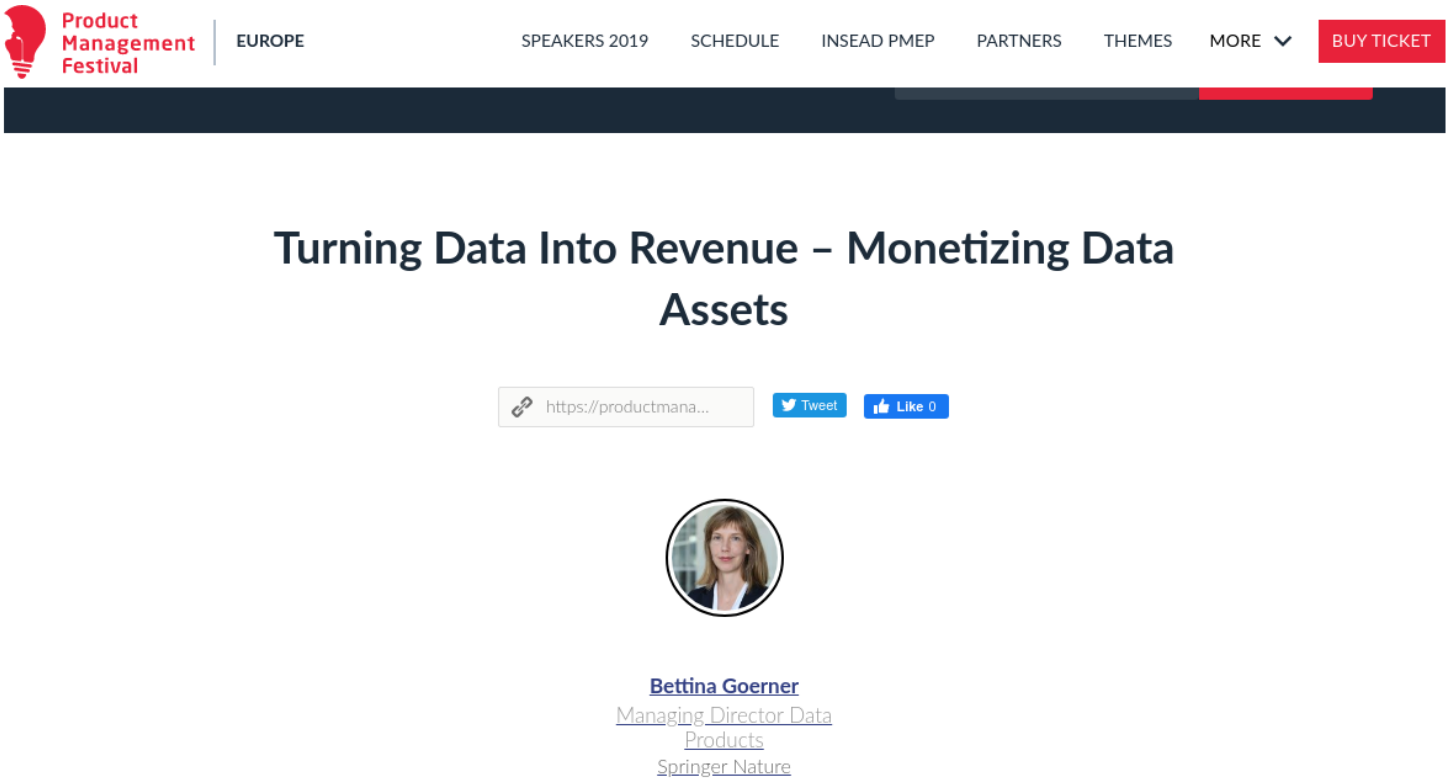
So what if we are wrong?

"Most models are wrong, but some are useful". (George Box)

"Most papers are wrong, but some are useful". (Me).

Research life cycle being commercialised

See seminar given *this afternoon* in Zurich



The Lancet

Longo DL, Drazen JM (2016) Data Sharing. N Engl J Med 374:276–277 http://dx.doi.org/10.1056/NEJMe1516564.

Data sharing sounds great ... but

1. Someone not involved in collecting data may misunderstand it.

2. "a new class of research person will emerge --- people who had nothing to do with the design and execution of the study but use another group's data for their own ends, possibly stealing from the research productivity planned by the data gatherers, or even use the data to try to disprove what the original investigators had posited. There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as **research parasites''**."

Response from the community was to embrace the term "resesarch parasite", e.g. https://twitter.com/dataparasite

This is just good science. I think there are few valid reasons for not sharing data.

Data sharing

Often seen as giving away competitive advantage.

Funders allow data citations and can often lead to new collaborations / data papers.

Solutions?

What can you as a researcher do?

- Forget about the journal title and its impact factor, share with your colleagues interesting research that you find (Twitter).
- Evaluating research: use the Sydney Brenner "trick" (Brenner, 1995)

  "we should remind ourselves that what matters absolutely is the scientific content of a paper and that nothing will substitute for either knowing it or reading it"
- Find a mutual network (reproducibility crowd)
- Investigate preprints . . .

Preprints

Rxivist

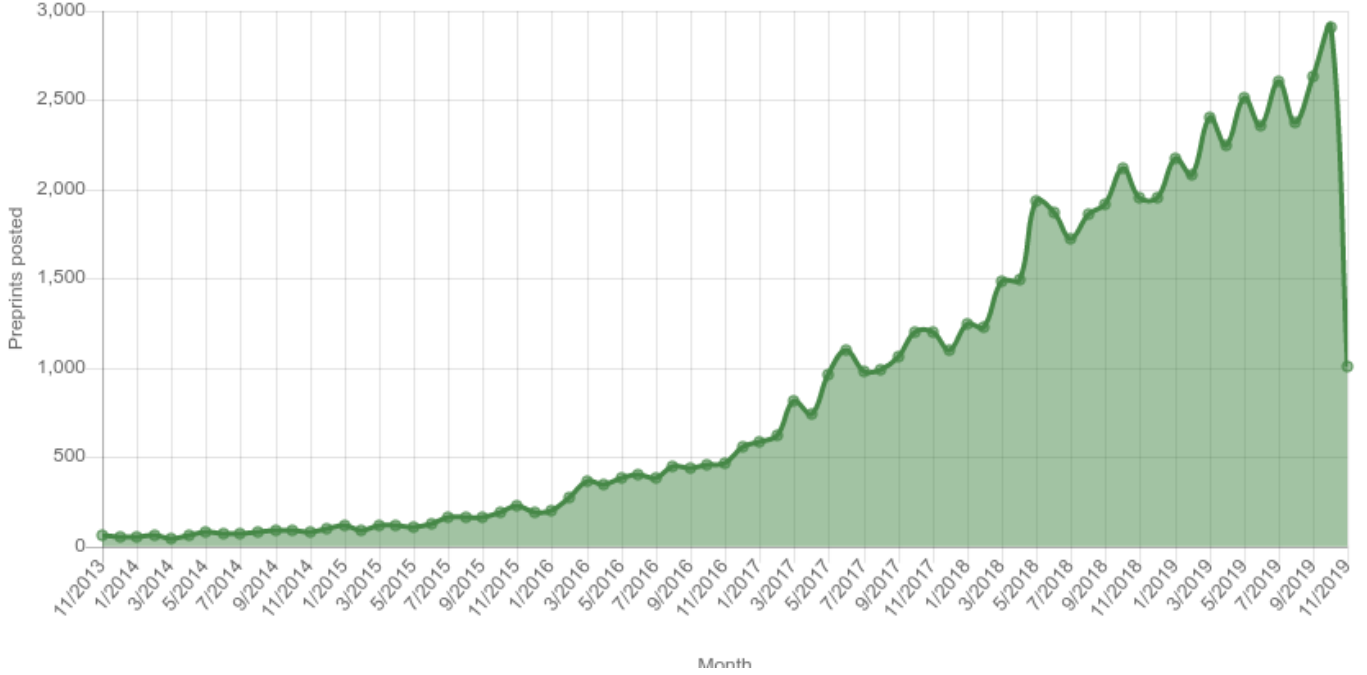
TRENDING OPEN SCIENCE

Rxivist combines preprints from bioRxiv with data from Twitter to help you **find the papers being discussed** in your field. Currently indexing **64,934 bioRxiv papers** from **287,775 authors**.

Site-wide metrics

The numbers below represent the metrics for all papers hosted on bioRxiv.org, based on our indexing of the website.

Monthly submissions, overall



Why preprint?

Advantages (https://www.plos.org/why-preprint):

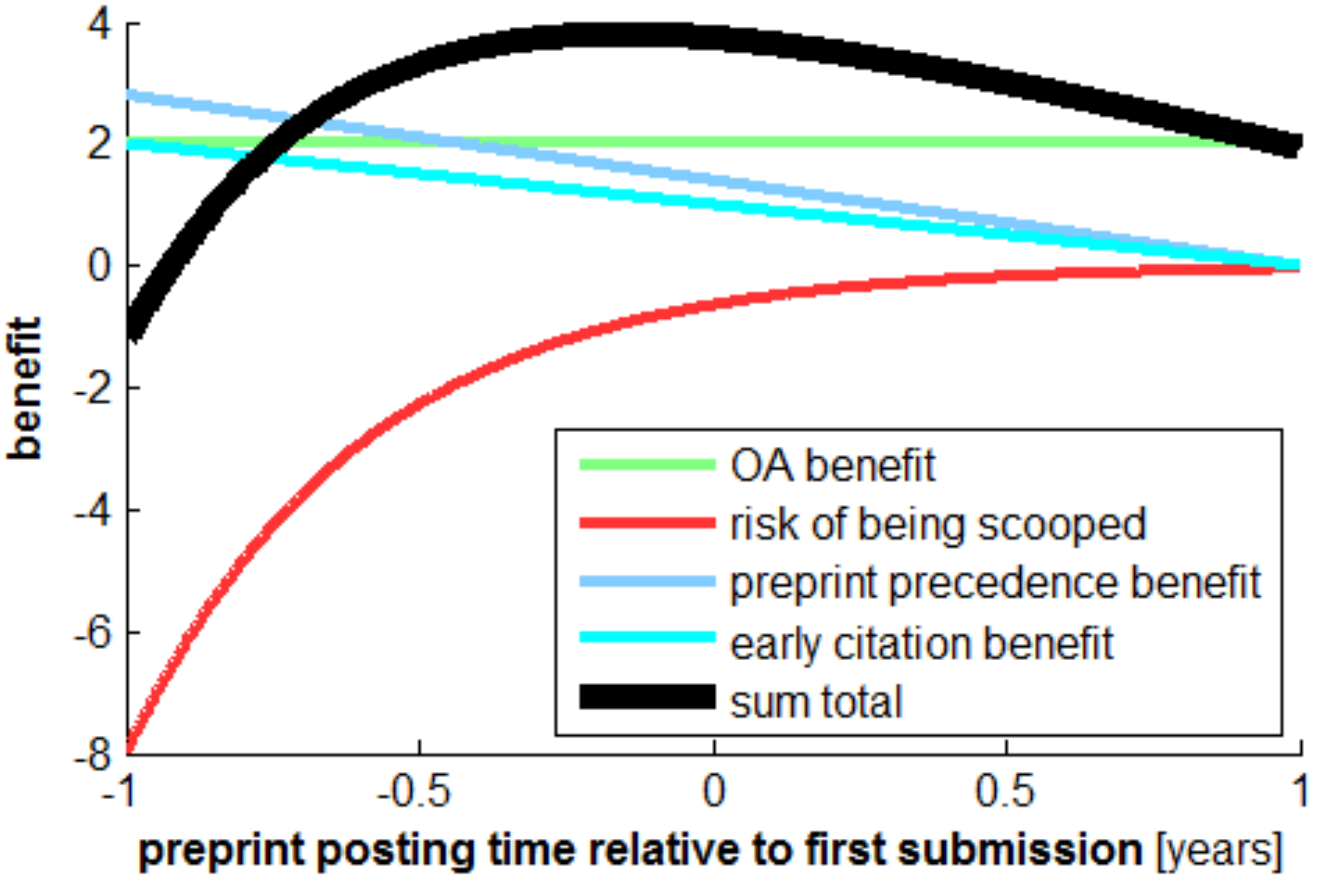
1. Rapid dissemination of results
2. Establish priority
3. Increased attention/citations
4. Career advancement
5. Community
6. Unlimited / timely updates

Also: potential for early feedback, error reduction, improved citations, editors might fish for papers.

Preprints are valid research outputs for REF2021 Naomi Penfold.

When to submit a preprint?

Estimates from https://nikokriegeskorte.org/2016/03/13/the-selfish-scientists-guide-to-preprint-posting/



Selfish approach, adopted by many: submit preprint at time of submissions.

Moral or selfish approach to sharing?

Markowetz *Genome Biology* (2015) 16:274
DOI 10.1186/s13059-015-0850-7

COMMENT Open Access

Five selfish reasons to work reproducibly

Florian Markowetz

Abstract

And so, my fellow scientists: ask not what you can do for reproducibility; ask what reproducibility can do for you! Here, I present five reasons why working reproducibly pays off in the long run and is in the self-interest of every ambitious, career-oriented scientist.

Keywords: Reproducibility, Scientific career

how science actually is. And, whether you like it or not, science is all about more publications, more impact factor, more money and more career. More, more, more... so how does working reproducibly help me achieve more as a scientist.

Reproducibility: what's in it for me?

In this article, I present five reasons why working reproducibly pays off in the long run and is in the self-interest of every ambitious, career-oriented scientist.

Selfish reasons to share

Why not align what is good for science with what is good for scientists?

1. Funding mandates (REF + enforcement from Wellcome Trust)
2. Credit through data papers
3. Fixes data bugs / errors in analysis
4. Prevent data loss (Vines et al 2014). e.g. students have a habit of leaving…
5. Your future self is probably one of the main beneficiaries of sharing.
6. *Now* is a very good time to be an open scientist.
7. Leads to further collaborations
8. Reviewers can do more work…

Rule 1: Data should be shared

Given the cost of generating data, I think data relating to a publication should be shared along with the paper.

Funders (and increasingly many, but not all, journals) agree.

Pregistration studies

Evaluate journal subscription model

Think about which journals you support, by sending papers and reviewing.

Support OA journals as authors and reviewers. (Ever asked a journal for permission to republish one of your own figures?)

Check out status of journal via Sherpa/Romeo service.

What can your institutions & funders do?

- sign up to DORA (Declaration of Research Assessment) (Wellcome Trust have solved this problem). Signing is easy, enforcing is hard.
- Give ECR a voice within institutions.
- Recognise all research artefacts, not just papers, as valuable outputs.
- Give academics more ownership of funds. "Do you really want to spend 5K on an APC, versus other uses in your lab?"
- Routes to OA need to be generate less admin for librarians / funders, not more. Green OA achieves that.
- Support diamond OA initiatives led by academics and academic societies.

Conclusions

- Many reasons to feel bad…
- But there are lots of reasons for optimism at new ways of doing science.
- Learn how to become an open scientist and use modern tools.
- Science improves one preprint submission at a time.
- Read more at http://bulliedintobadscience.org and sign up!