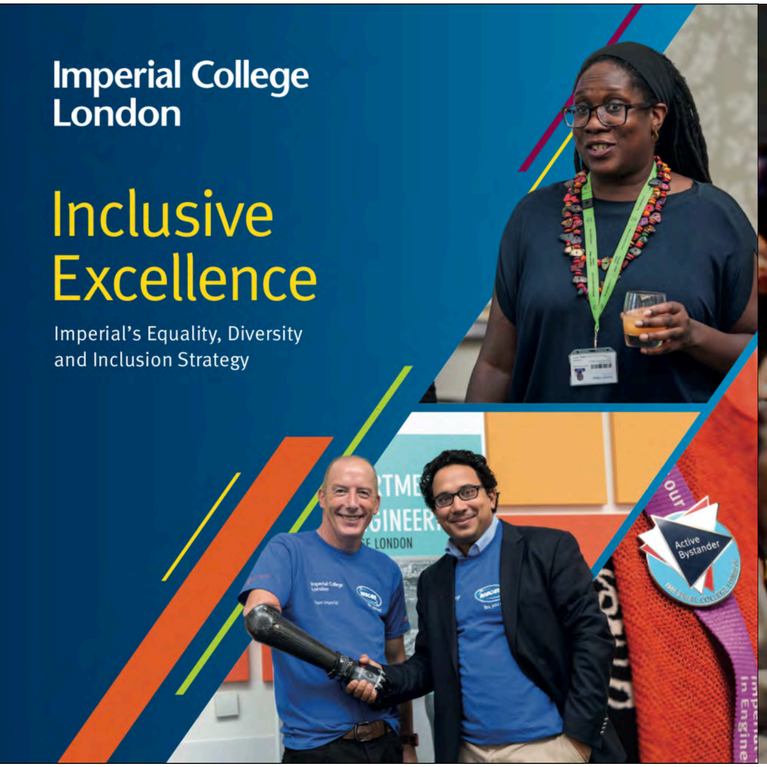
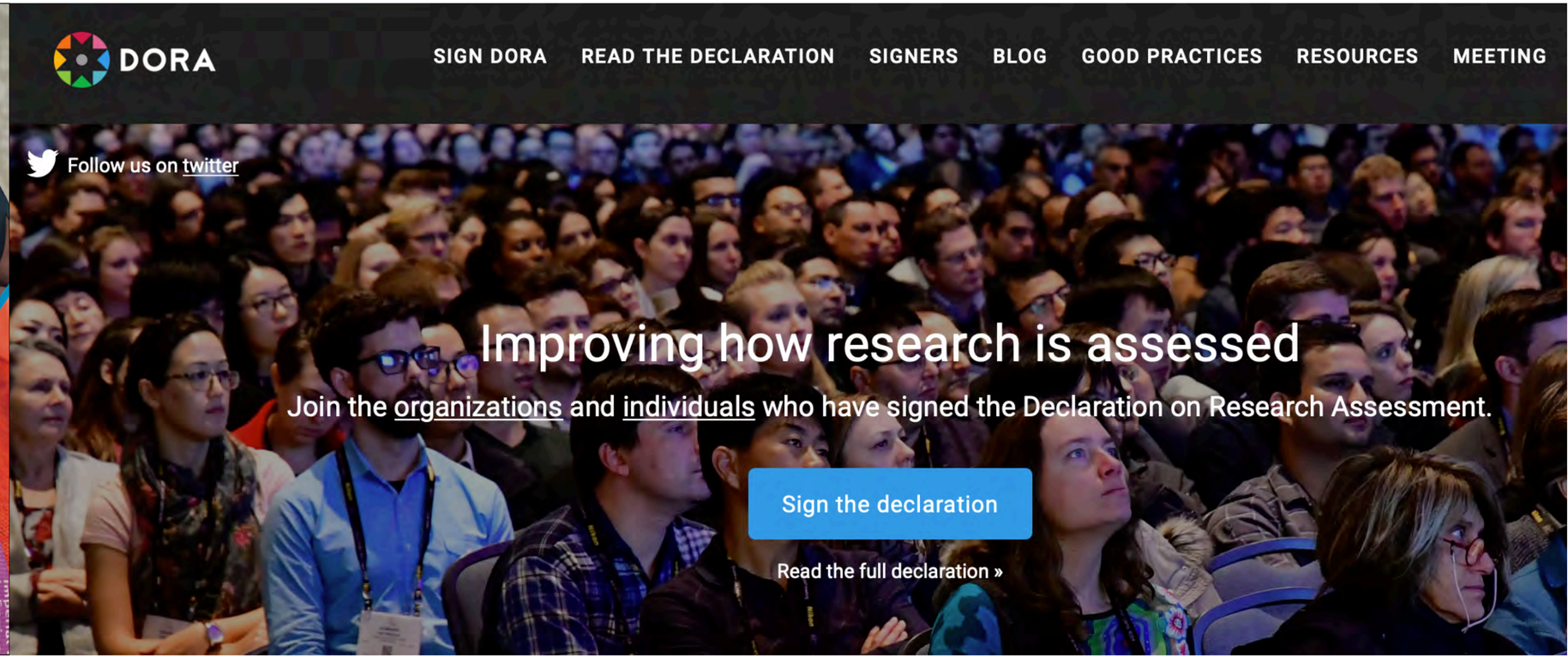
Re-focusing how the scientific community assesses research

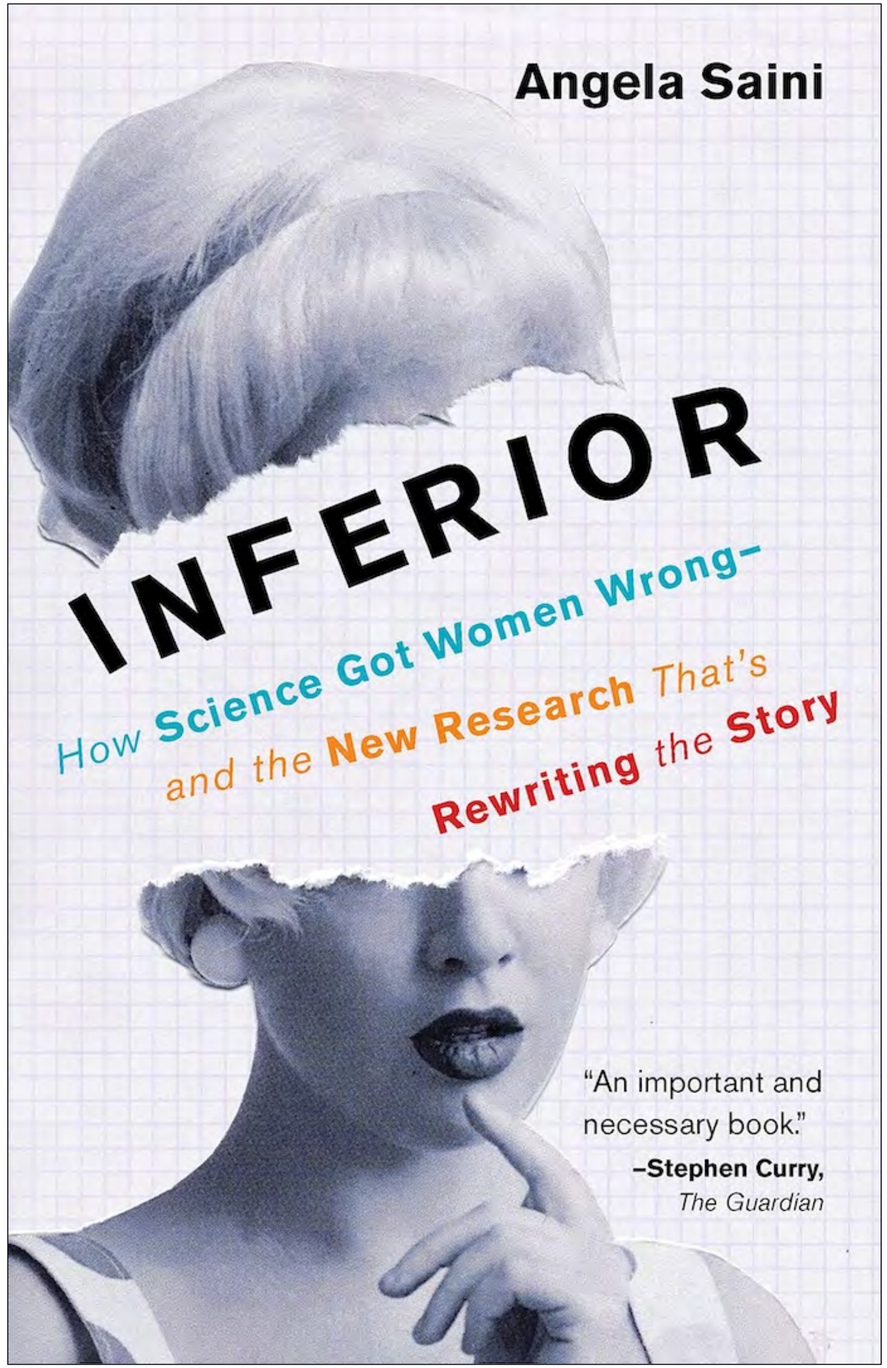




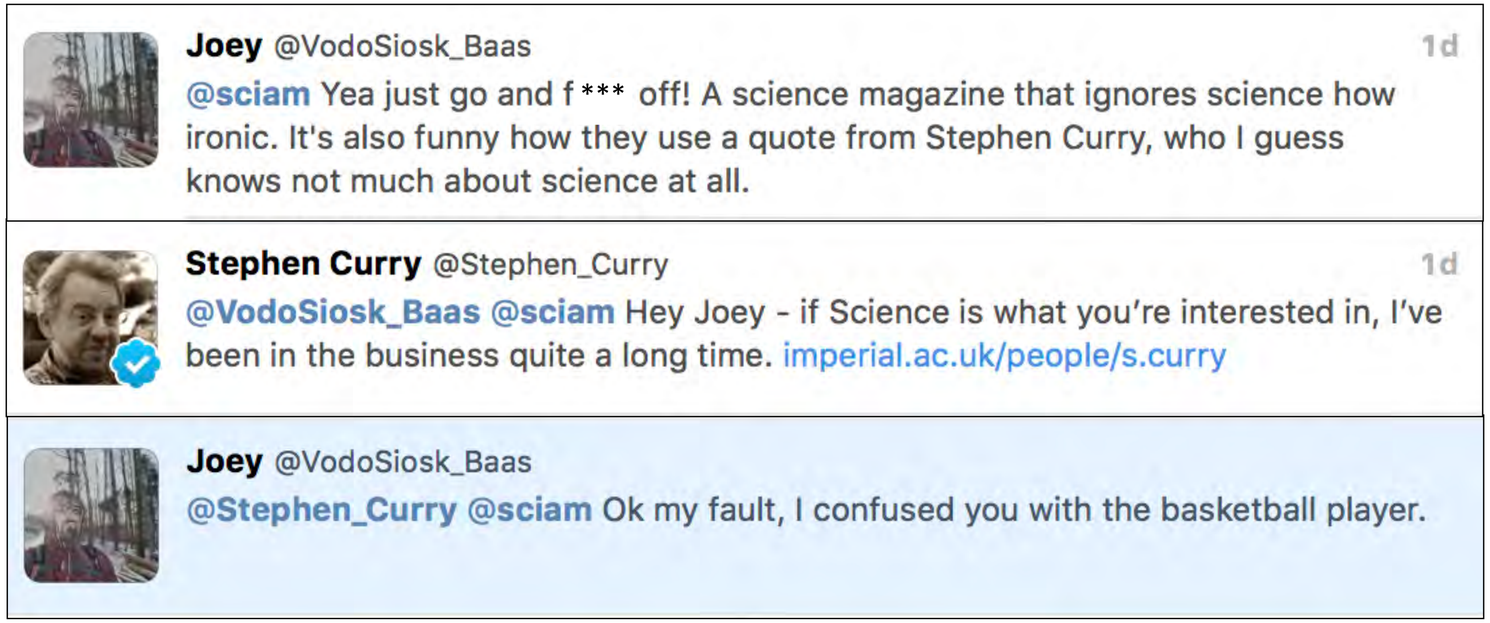
Prof. Stephen Curry

Assistant Provost (Equality, Diversity & inclusion), Imperial College
Chair, San Francisco Declaration on Research Assessment (DORA)

Gender Summit | Amsterdam | 04 Oct 2019



Researcher assessment

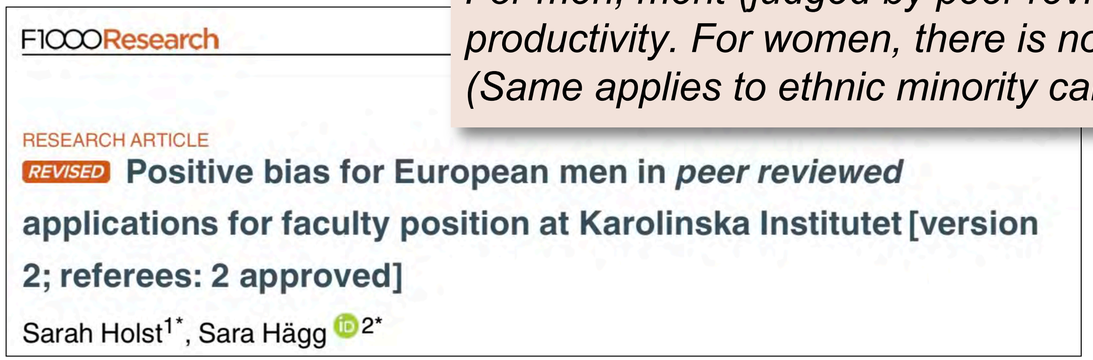


How do we solve the problem of bias in research assessment?

Methodically…

1. Define the nature and scale of problem
2. Understand its root causes – share that understanding convincingly
3. Design and test tools and processes to fix it
   - Education, training
   - Methods for credible, time-efficient assessment of the qualities of research
   - Methods to counter the biases rooted in human psychology and society

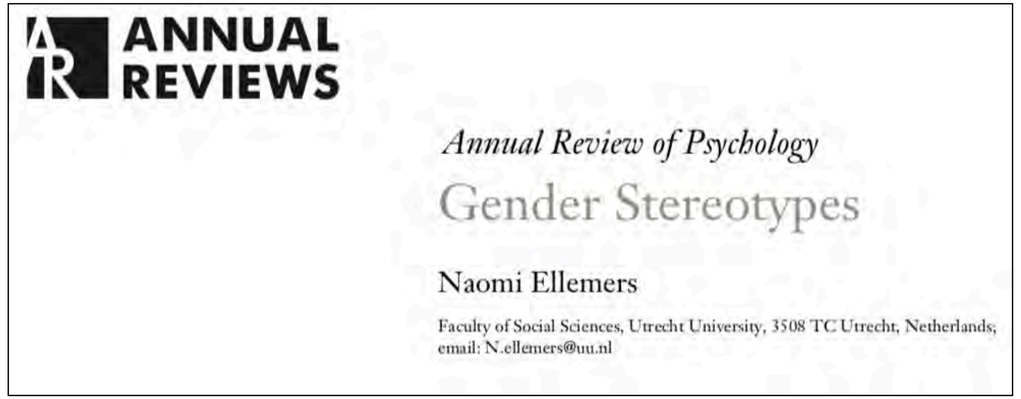
When there are joint first authors, reflecting equal contributions, men are more likely to be listed first.



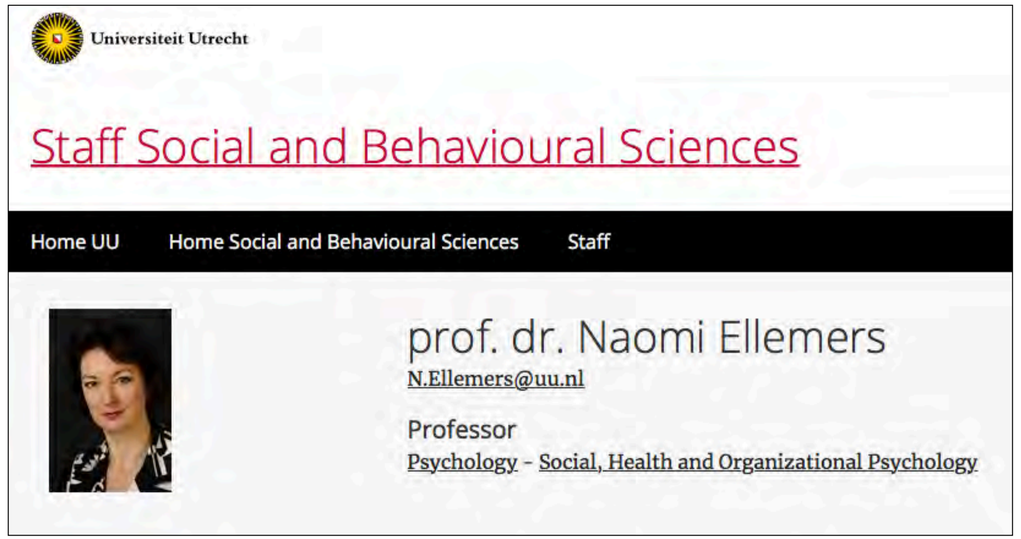
For men, merit (judged by peer review) is associated with productivity. For women, there is no such association. (Same applies to ethnic minority candidates).

Gender gaps in grant funding are attributable to less favourable assessments of women as PIs, not of the quality of their proposed research.

2. Understand the root causes of the problem



https://www.annualreviews.org/doi/10.1146/annurev-psych-122216-011719



https://www.uu.nl/staff/NEllemers/0

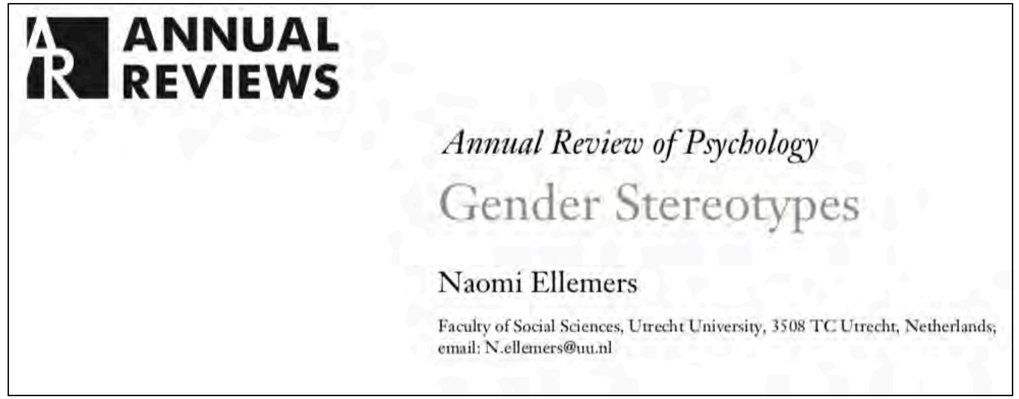
"Gender stereotypes affect the way people attend to, interpret, and remember information about themselves and others. … They reflect general expectations about members of particular social groups."

"he is smart" (stable attribute); "she did well on the test" (temporary achievement)

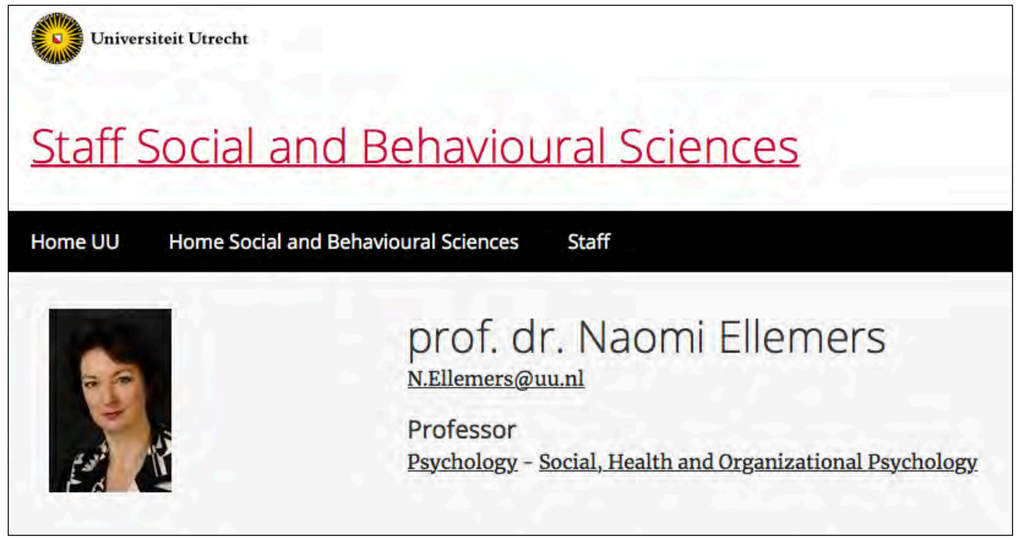
he is "outstanding"; she is "hardworking"

he is "assertive, ambitious"; she is "caring, conscientious"

3. Design and test tools and processes to fix it



https://www.annualreviews.org/doi/10.1146/annurev-psych-122216-011719



https://www.uu.nl/staff/NEllemers/0

Stereotyping "leads people to over-emphasize differences between groups and underestimate variations within groups"

- **Training:**
  - **Acknowledge** the pervasive nature and impacts of gender stereotyping
  - **Accept** that we are all subject to gendered expectations - and that these bias our judgements
  - **Educate** people to recognise the implicit effects that gender stereotypes have for themselves and others
- **Process:**
  - **Lift the burden** of proof from those who may be disadvantaged;
  - **Support** employees in reconciling male and female role expectations regarding work and family life

San Francisco Declaration on Research Assessment (DORA)

Campaigning for reform of research evaluation: we can do better than journal impact factors (JIFs)

- *sfdora.org*
- 6 years old; >14,500 individuals and >1500 organisations have signed
- International steering group; a global advisory board
- Roadmap for action:
  - Increase awareness of the need to develop alternatives to the JIF
  - Research and promote best practice in research assessment
  - Extend the global and disciplinary impact of DORA

https://www.nature.com/articles/d41586-018-01642-w

WORLD VIEW *A personal take on events*

DAVE GUTTRIDGE

Words were a good start — now it is time for action

Five years ago, the Declaration on Research Assessment was a rallying point. It must now become a tool for fair evaluation, urges **Stephen Curry**.

Declarations are bound to fall short. The 240-year-old United States Declaration of Independence holds it self-evident that "all men [*sic*] are created equal", but equality remains a far-off dream for many Americans.

The San Francisco Declaration on Research Assessment (DORA; https://sfdora.org) is much younger, but similarly idealistic. Conceived by a group of journal editors and publishers at a meeting of the American Society for Cell Biology (ASCB) in December 2012, it proclaims a pressing need to improve how scientific research is evaluated, and asks scientists, funders, institutions and publishers to forswear using journal impact factors (JIFs) to judge individual researchers.

DORA's aim is a world in which the content of a research paper matters more than the impact factor of the journal in which it appears. Thousands of individuals and hundreds of research organizations now agree and have signed up. Momentum is building, particularly in the United Kingdom, where the number of university signatories has trebled in the past two years. This week, all seven UK research councils announced their support.

Impact factors were never meant to be a metric for individual papers, let alone individual people. They're an average of the skewed distribution of citations accumulated by papers in a given journal over two years. Not only do these averages hide huge variations between papers in the same journal, but citations are imperfect measures of quality and influence. High-impact-factor journals may publish a lot of top-notch science, but we should not outsource evaluation of individual researchers and their outputs to seductive journal metrics.

Most agree that yoking career rewards to JIFs is distorting science. Yet the practice seems impossible to root out. In China, for example, many universities pay impact-factor-related bonuses, inspired by unwritten norms of the West. Scientists in parts of Eastern Europe cling to impact factors as a crude bulwark against cronyism. More worryingly, processes for JIF-free assessment have yet to gain credibility even at some institutions that have signed DORA. Stories percolate of research managers demanding high impact factors. Job and grant applicants feel that they can't compete unless they publish in prominent journals. All are fearful of shrugging off the familiar harness.

So, DORA's job now is to accelerate the change it called for. I feel the need for change whenever I meet postdocs. Their curiosity about the world and determination to improve it burns bright. But their desires to pursue the most fascinating and most impactful questions are subverted by our systems of evaluation. As they apply for their first permanent positions, they are already calculating how to manoeuvre within the JIF-dependent managerialism of modern science.

There have been many calls for something better, including the Leiden Manifesto and the UK report 'The Metric Tide', both released in 2015. Like DORA, these have changed the tenor of discussions around researcher assessment and paved the way for change.

It is time to shift from making declarations to finding solutions. With the support of the ASCB, Cancer Research UK, the European Molecular Biology Organization, the biomedical funder the Wellcome Trust and the publishers the Company of Biologists, *eLife*, F1000, Hindawi and PLOS, DORA has hired a full-time community manager and revamped its steering committee, which I head. We are committed to getting on with the job.

Our goal is to discover and disseminate examples of good practice, and to boost the profile of assessment reform. We will do that at conferences and in online discussions; we will also establish regional nodes across the world, run by volunteers who will work to identify and address local issues.

IT'S WORTH DOING THE EXPERIMENT TO PROPERLY EVALUATE EVALUATION.

This week, for example, DORA is participating in a workshop at which the Forum for Responsible Metrics — an expert group established following the release of 'The Metric Tide' — will present results of the first UK-wide survey of research assessment. This will bring broader exposure to what universities are thinking and doing, and put the spotlight on instances of good and bad practice.

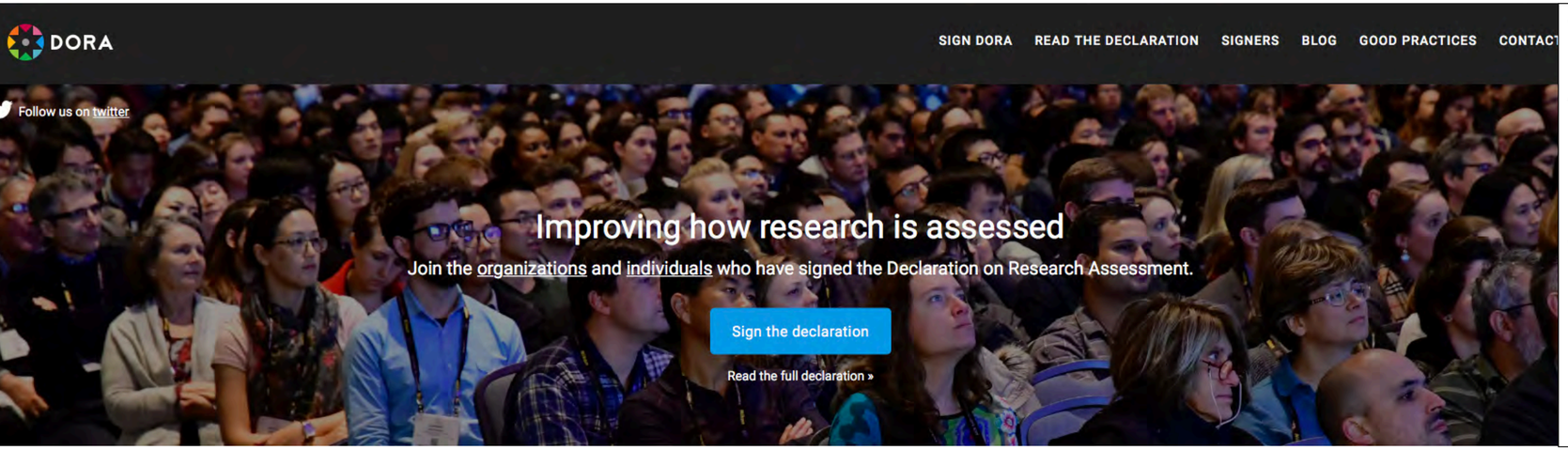
We have to get beyond complaining, to find robust, efficient and bias-free assessment methods. Right now, there are few compelling options. I favour concise one- or two-page 'bio-sketches', similar to those rolled out in 2016 by the University Medical Centre Utrecht in the Netherlands. These let researchers summarize their most important research contributions, plus mentoring, societal engagement and other valuable activities. This approach could have flaws. Perhaps it gives too much leeway for 'spin'. But, as scientists, surely we can agree that it's worth doing the experiment to properly evaluate evaluation.

This is hard stuff: we need frank discussions that grind through details, with researchers themselves, to find out what works and to forestall problems. We need to be mindful of the damage wrought to the careers of women and minorities by bias in peer review and in subjective evaluations. And we need to join in with parallel moves towards open research, data and code sharing, and the proper recognition of scientific reproducibility.

Declarations such as DORA are important; credible alternatives to the status quo are more so. True success will mean every institution, everywhere in the world, bragging about the quality of their research-assessment procedures, rather than the size of their impact factors. ■

Stephen Curry *is a professor of structural biology and assistant provost for equality, diversity and inclusion at Imperial College London. He is also chair of the DORA steering group.*
e-mail: s.curry@imperial.ac.uk

8 FEBRUARY 2018 | VOL 554 | NATURE | 147



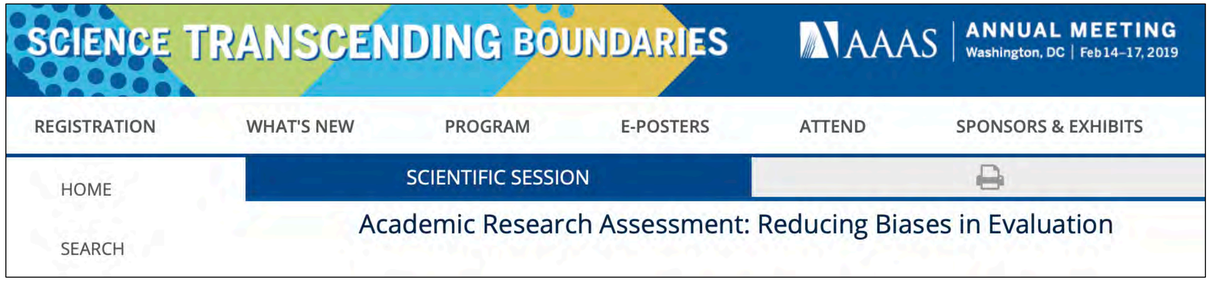
ascb an international forum for cell biology
CANCER RESEARCH UK
The Company of Biologists
eLIFE
EMBO excellence in life sciences
F1000Research Open for Science
Hindawi
hhmi Howard Hughes Medical Institute
PLOS
wellcome

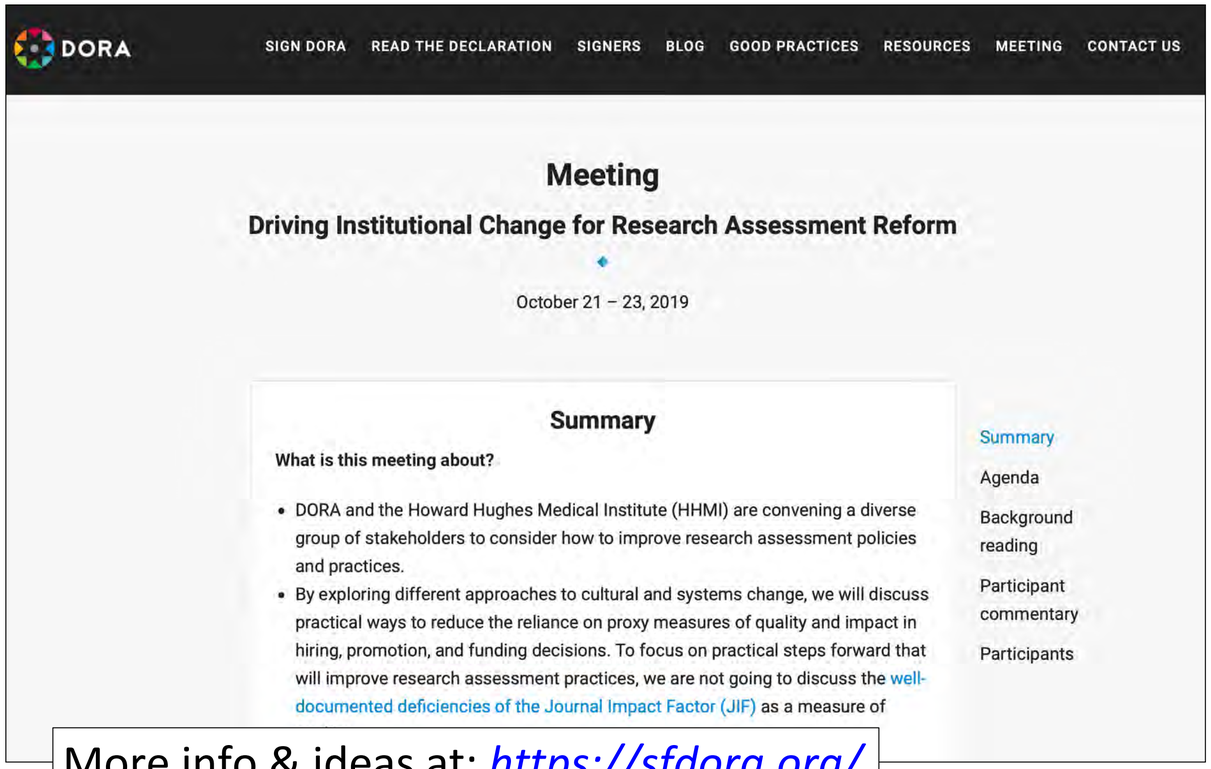
DORA: building, promoting and testing new tools and processes for evaluation

https://elifesciences.org/inside-elife/1fd1018c/research-assessment-reducing-bias-in-the-evaluation-of-researchers

DORA session at ASCB|EMBO (Dec 2018)

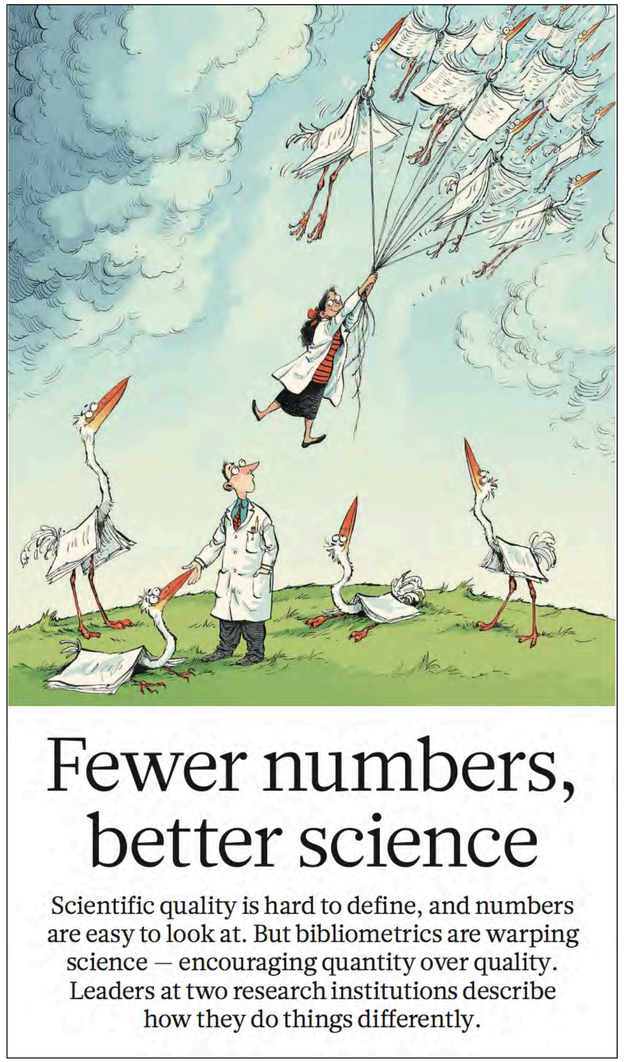
DORA session at AAAS (Feb 2019)





More info & ideas at: *https://sfdora.org/*

New tools to break old habits - structured narratives



Fewer numbers, better science

Scientific quality is hard to define, and numbers are easy to look at. But bibliometrics are warping science — encouraging quantity over quality. Leaders at two research institutions describe how they do things differently.

http://www.nature.com/news/fewer-numbers-better-science-1.20858

Researcher assessment at UMC Utrecht

1. Research, publications, grants
2. Managerial & academic duties
3. Mentoring & teaching
4. Clinical work (if applicable)
5. Entrepreneurship & community outreach

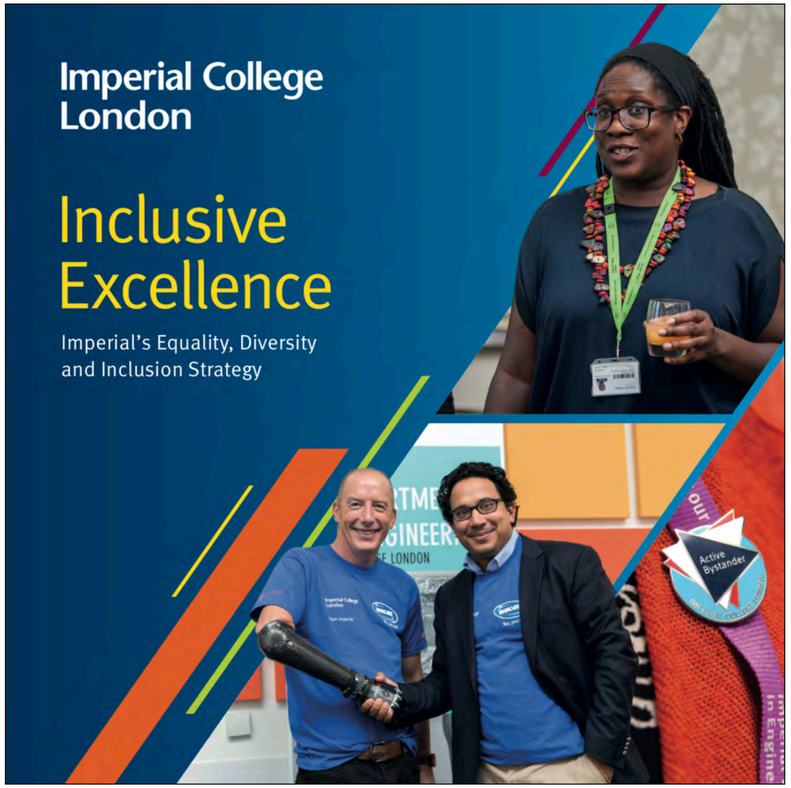
DORA

More examples at: *https://sfdora.org/good-practices/*

Case study (Chem Eng+, Imperial College)

Shift the weight of assessment to more nuanced, qualitative peer review needs built-in processes to reduce bias:
- proactive search for under-represented candidates
- careful use of language in adverts
- anonymisation of candidate info at triage
- bias training for panels
- diverse panels; designated EDI observer at interview
- interview and candidate visit within core hours
- explorative, value-driven rather than aggressive interviews

EDI and culture change at universities: do we have enough evidence?

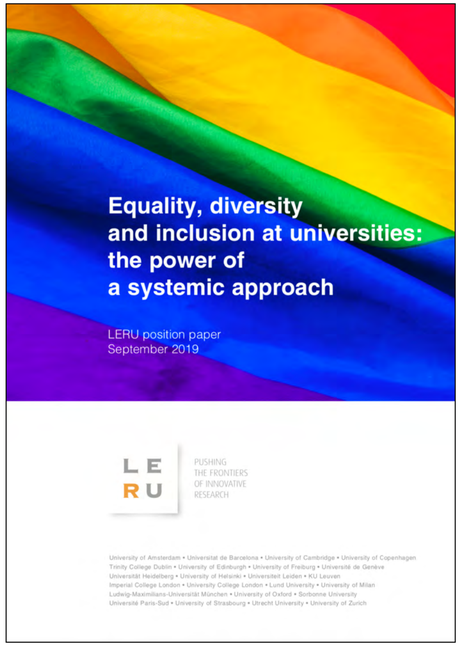


http://www.imperial.ac.uk/equality/governance/strategy/

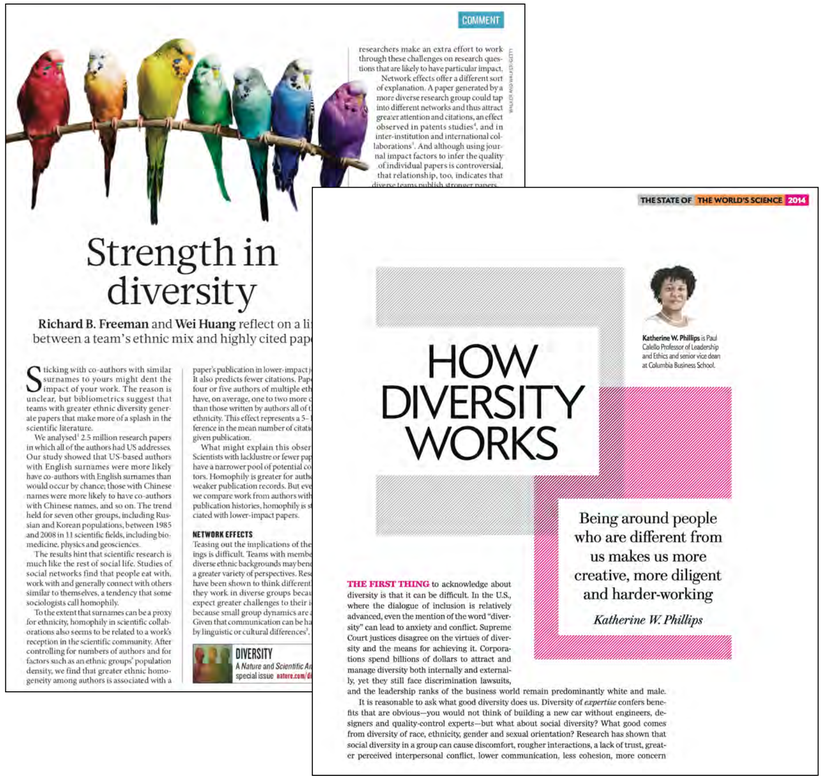
- To ensure our **future success** as a university
- To tap into **pools of talent** that have been neglected for too long – diversity is a strength
- to increase the **quality, relevance and world-changing impact** of our research and education
- to improve the well-being and productivity of the people who work and study here by developing a culture that **values everyone**

"We will integrate EDI into all management processes."

(engaged leadership; people-management skills; EDI training; mentoring)

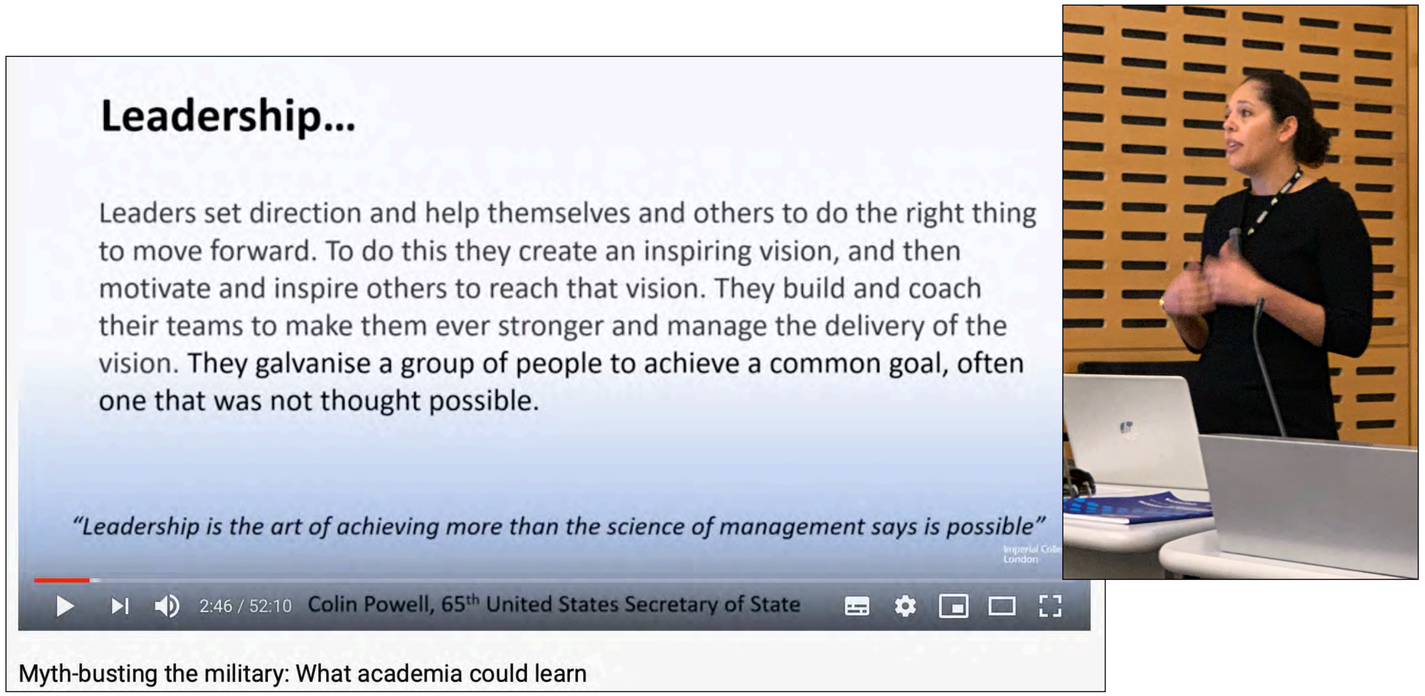


https://www.leru.org/publications/equality-diversity-and-inclusion-at-universities#



"The key to understanding the positive influence of diversity is the concept of informational diversity. When people are brought together to solve problems in groups, they bring different information, opinions and perspectives. This makes obvious sense when we talk about diversity of disciplinary backgrounds…"

How to have competition without hyper-competition?



Myth-busting the military: What academia could learn

https://www.youtube.com/watch?v=b2sgp7Kjjy8

"My one aim is for everyone to understand how important leadership is within the academic setting and how we could do more to incorporate that into the academic culture."

Nadia Soliman
March 2019

"Science is at its best when at its most collaborative, inclusive and diverse."

Margaret Heffernan
June 2019

See also: https://www.imperial.ac.uk/news/190447/from-british-army-imperial-phd-student/

We need to assess research but how should we define success?

ANNALS OF MEDICINE JULY 29, 2013 ISSUE

SLOW IDEAS

Some innovations spread fast. How do you speed the ones that don't?

By Atul Gawande

Why do some innovations spread so swiftly and others so slowly? Consider the very different trajectories of surgical anesthesia and antiseptics, both of which were discovered in the nineteenth century. The first public demonstration of anesthesia was in 1846. The Boston surgeon Henry Jacob Bigelow was approached by a local dentist named William Morton, who insisted that he had found a gas that could render patients insensible to the pain of surgery. That was a dramatic claim. In those days, even a minor tooth extraction was excruciating. Without effective pain control, surgeons learned to work with slashing speed. Attendants pinned patients down as they screamed and thrashed, until they fainted from the agony. Nothing ever tried had made much difference. Nonetheless, Bigelow agreed to let Morton demonstrate his claim.

We yearn for frictionless, technological solutions. But people talking to people is still the way that norms and standards change.

ILLUSTRATION BY HARRY CAMPBELL

http://www.newyorker.com/magazine/2013/07/29/slow-ideas

"We yearn for frictionless, technological solutions. But **people talking to people** is still how the world's standards change."

Atul Gawande

Reliable, rapidly communicated, accessible, high-quality **research** that transforms our understanding of the world and can change it for the better.

Researchers who collaborate, who feel a duty of care to group members & colleagues, and a responsibility to the societies of which they are an integral part.

A **research system** that values the people within it, that cares about their quality of life, and that seeks out the creative vigour of diversity.

“People talking to people is still how the world’s standards change.”

Get the evidence and the message out to those researchers who may still be doubting the value of equality.

The incredulity of Saint Thomas, Caravaggio

http://www.caravaggio.org/the-incredulity-of-saint-thomas.jsp

Thank you

s.curry@imperial.ac.uk
@Stephen_Curry