

The state of the art in research on research evaluation & peer review

Theories and models,
formats and examples,
future directions

Prof.dr. Sarah de Rijcke - CWTS, Leiden University

Introduction



Hackett & Chubin, 2003:

“Peer review challenges whether new ideas are truly new and worth pursuing, in principle distinguishing between sound innovation and reckless speculation.”

The Royal Society:

‘The only effective way of properly assessing the quality of research proposals’

- Peer review has become the main mechanism for distributing resources from governments and philanthropic funders
- More than 95% of public funding for medical research in the UK allocated by peer review
- Last year's budget for the US National Institutes of Health (NIH) was US\$39 billion; it awards more than 80% to more than 300,000 researchers using peer review
- ERC distributed 1.8 billion Euro budget using peer review in 2017

(Guthrie et al, 2018)

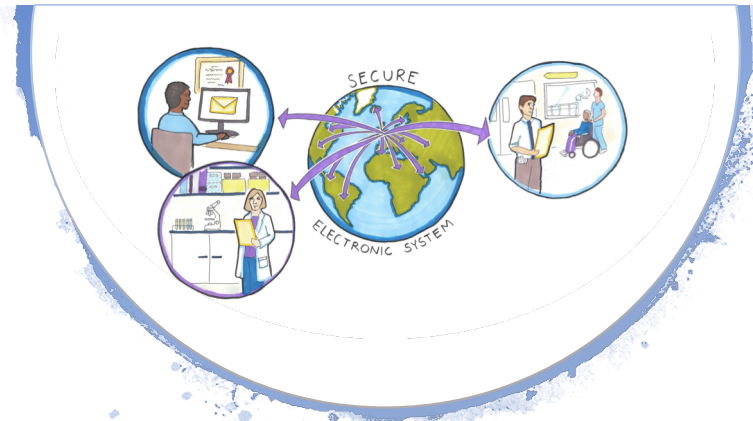
Outline talk

1. Different approaches to organising peer review
2. Grounds for support and challenges raised
3. Alternative approaches and models

1. Peer review: different approaches

Peer review: different approaches

- *Ad hoc* only
- Panel only
- *Ad hoc* + panel review
- Two stages:
 - Review outlines
 - Review full applications



2. Grounds for support and challenges raised

Available evidence

1. Does peer review support valuable science?
2. Can the design of peer review systems be improved to increase effectiveness and reduce burden?

Support is relatively strong

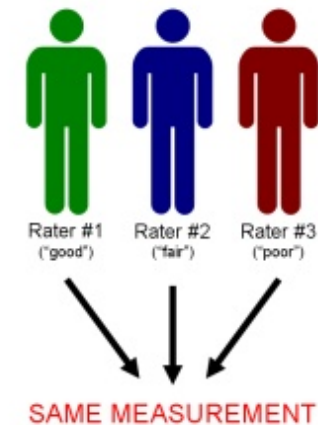
- Support for peer review amongst the academic community remains strong (Bornmann, 2011; Wooding & Grant, 2003)
- Dominance of peer review across international funding system → confidence of institutional stakeholders
- A recent review of literature about the NIH peer review processes found a firm belief in the transparency and objectivity of peer review amongst grant reviewers (Miner, 2011)

Concerns: is peer review reliable?

- Inter-rater reliability
- overall consistency of decision-making
- Biases
- Burden on the research system

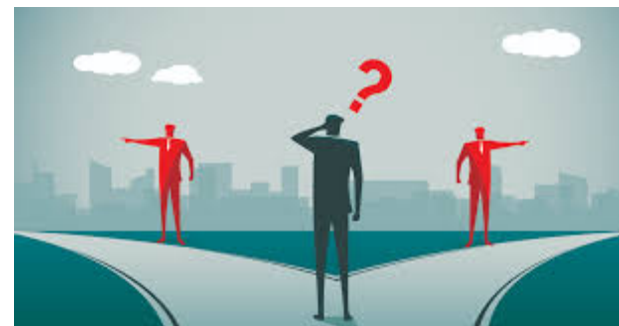
Inter-rater reliability

- Jayasinghe *et al.* (2003): single-rater reliability correlation of 0.21 for SSH, and 0.19 for the sciences.
- Increasing diversity of background and discipline of peer reviewers also reduces rating consistency (Lobb et al 2013)
- Peer review processes may not work well for transdisciplinary teams integrating both academic and non-academic experts



Overall consistency of decision-making

- Conflicting evidence; existing studies offer mixed assessments
- Bornmann et al. (2008) identified a threshold of 80–90 per cent as the expectation for agreement for this kind of decision-making
- Two early studies (Cole *et al.*, 1981; Hodgson, 1997) found reliability rates across funding boards of 75 and 73 per cent respectively for funding decisions which they felt was a satisfactory level of agreement
- More recent evidence is mixed



Correlation metrics and peer review

- No association between percentile application rankings and the research's subsequent bibliometric performance (Danthi *et al.*, 2014; Danthi *et al.*, 2015; Doyle *et al.*, 2015; Fang *et al.*, 2016; Kaltman *et al.*, 2014; van den Besselaar & Sandström, 2015).
- Grant review outcomes only weakly predict bibliometric performance (Lauer *et al.*, 2015; Reinhart, 2009).



Bias - Gender

- Substantial body of conflicting evidence on whether peer review is gender biased



Bias - Age

- Conflicting evidence



Bias - Cryonism

- A concern for many major funders
- Some evidence that this concern is justified



Bias - Cognitive particularism

- A number of studies confirm that reviewers and panel members will favour proposals in their own fields or that align with their ways of thinking (Travis & Collins 1991; Langfeldt 2006; Li 2015; Wang & Sandström 2015)



Burden on the system

- On applicants
- On reviewers and panel members



Summary on Challenges

- Peer review suffers from biases: strongest evidence is of a bias against innovation; some evidence peer review is influenced by cognitive distance and suffers from cronyism; suggestive evidence that there are age biases.
- Considerable work has been done on gender bias, with conflicting results → illustrates the challenges of accounting for biases outside the scope of the peer review process
- The burden of the peer review system is high and falls primarily on applicants

3. Alternative approaches and models

Randomisation

Completely random

Partial: lottery system

Avin 2015, Barnett 2016, Kurokawa et al.
2015, Fang & Casadevall 2016



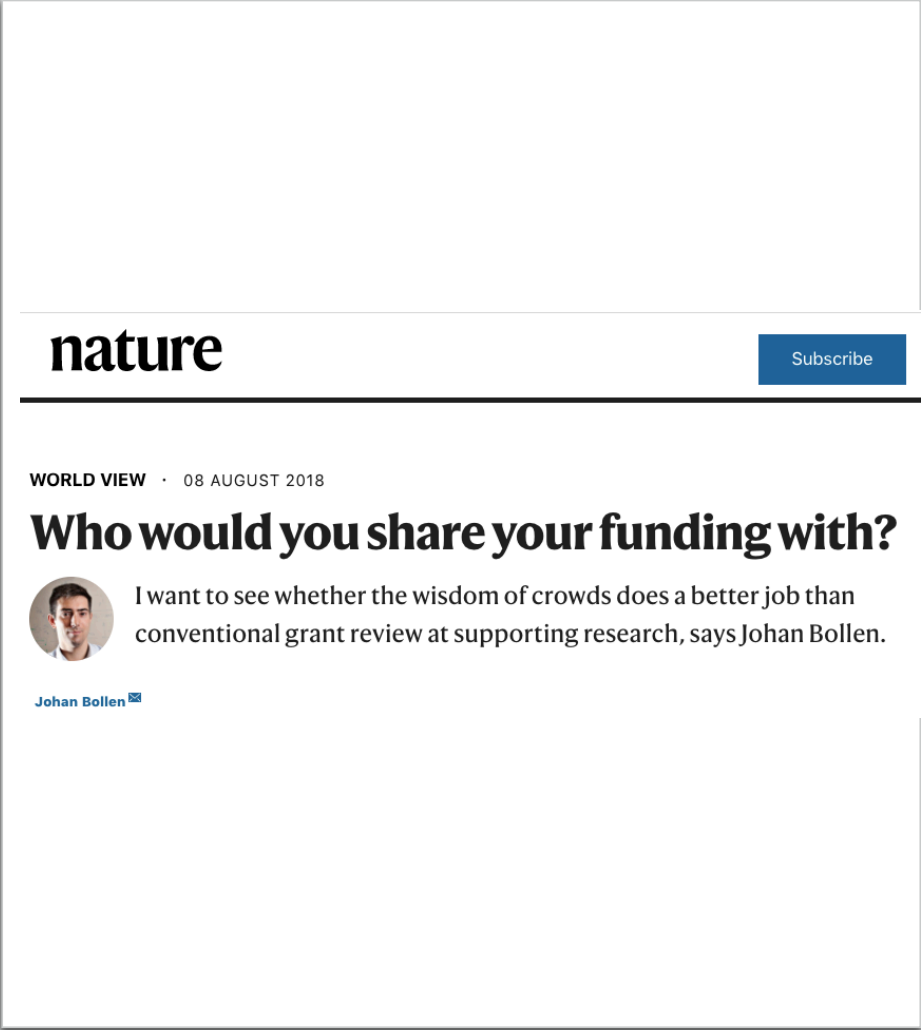
Complete alternatives

Peer to peer allocation

Bollen J, Crandall D, Junk D, *et al.* (2017)

An efficient system to fund science: from proposal review to peer to peer distributions.

Scientometrics, **110**(1): 521–528.



The image shows a screenshot of a Nature article snippet. At the top, the word "nature" is written in a bold, lowercase, sans-serif font. To the right of the logo is a blue rectangular button with the word "Subscribe" in white. Below the logo and button is a horizontal line. Underneath the line, the text "WORLD VIEW · 08 AUGUST 2018" is displayed in a smaller, uppercase font. The main title of the article is "Who would you share your funding with?" in a large, bold, black font. Below the title is a circular profile picture of a man with short dark hair, wearing a light-colored shirt. To the right of the profile picture is the text "I want to see whether the wisdom of crowds does a better job than conventional grant review at supporting research, says Johan Bollen." Below the profile picture and text is the name "Johan Bollen" followed by a small blue icon of an envelope.



International peer-review conference Main outcomes

Amsterdam, 29-30 June 2017

4 | Sandpit and alternative review methods

General conclusions

- Most councils use small adjustments on traditional methods (e.g. using reviewers from different fields, when interdisciplinary projects have to be evaluated; using both a scientific and a civil jury for judging proposals), but not many are yet experimenting with alternative review methods.
- Peers are included in all new methods proposed; experts are needed in almost all methods. Legal obligation: we have to give the applicants understandable reason(s) if their application is not funded. This is also important because of the acceptance of the results. For final decisions we should use strategic consideration instead of a lottery.
- There is resistance to drastic changes in the review system in the scientific community; incremental implementation of innovations. Should it all be peer-reviewed? Should we spend time assessing small proposals? Allocate this to the universities and minimise duplication.

Conclusions

