RESEARCH ON RESEARCH INSTITUTE

SSHRC 16 October 2025

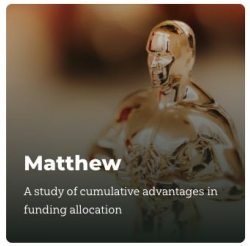
Tom Stafford, AFIRE programme lead

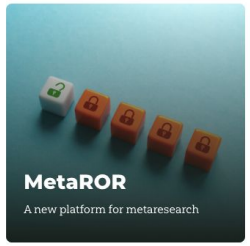
t.stafford@researchonresearch.org

tomstafford.github.io

RoRI RESEARCH ON RESEARCH INSTITUTE

10 codesigned projects





18 Core Partners

Australian Government
Australian Research Council

Michael Smith Health Research BC

NRF National Research Foundation

Fundació "la Caixa"

Swiss National Science Foundation

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada

novo nordisk foundation
Benefitting people and society

NWO

CWTS Meaningful metrics

DIGITAL science

VolkswagenStiftung

wellcome

King Baudouin Foundation
Working together for a better society

Social Sciences and Humanities Research Council of Canada
Canada

GORDON AND BETTY MOORE FOUNDATION

Luxembourg National Research Fund

UKRI Research England

UCL

Our definition of experiment

Principled: a research design that allows inference about what causes what (before/after, shadow experiments, true experiment/RCT)

Planned: primary outcome measure and analysis plan declared in advance

Public: a commitment to sharing the results regardless of outcome

AFIRE: Accelerator for Funder Experimentation

Forum — **Sharing work by funders, for funders**

Capacity building — Sprints on **AI/ML in reviewer selection**

Experiments — **Distributed Peer Review, Partial Randomisation, Desk Rejection, and more!**

RoRI RESEARCH ON RESEARCH INSTITUTE

IGL Innovation Growth Lab by nesta

These slides:
bit.ly/tomstafford

Partial Randomisation Trials Catalogue

Funder	Dates
Health Research Council of New Zealand	2013-
VolkswagenStiftung	2017-2020
Austrian Science Fund (FWF)	2019-
Swiss National Science Foundation (SNSF)	2018-
Novo Nordisk Fonden	2022-2025
British Academy	2022-2025
UKRI / NERC	2022-
Wellcome	2023-
Nesta	2019-2020
University of Leeds	2023
UMC Utrecht/Ministry of OCW	2023

bit.ly/PRtrials

Desk Rejection Shadow Experiment

Can agency staff predict those proposals with the least likelihood of success?

- a shadow experiment, not an intervention
- supports optimal use of external review
- UKRI leading participation
- recruiting schemes which will complete by end of 2026

Evaluating Distributed Peer Review at the Volkswagen Foundation

Anna Butters, Melanie Benson Marshall, Tom Stafford & Stephen Pinfield (Research on Research Institute and University of Sheffield);
Hanna Denecke, Alexander Bondarenko, Barbara Neubauer, Robert Nuske & Pierre Schwidlinski (Volkswagen Foundation)

RoRI RESEARCH ON RESEARCH INSTITUTE

VolkswagenStiftung

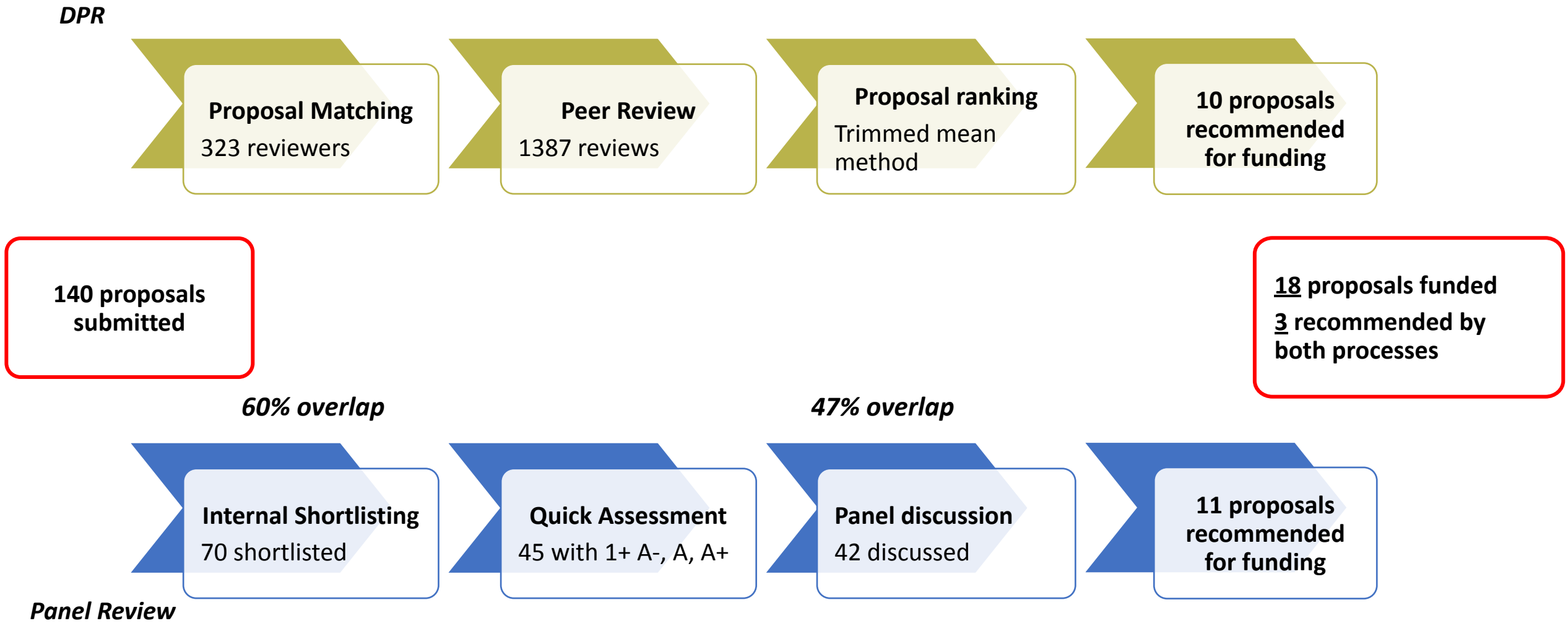
Distributed Peer Review (DPR)

- Applicants review other applications submitted for the same funding call
- Has been used at the European Southern Observatory (ESO), Netherlands Research Council (NWO) and more recently by UK Research and Innovation (UKRI)
- Potential (being tested)
  - → Builds on accepted mechanism: peer review
  - → Solves reviewer recruitment
  - → Incentivises timely submission by reviewers
  - → Aligns reviewer understanding of call criteria
  - → Trains participants in grant reviewing (and by extension grant writing)
  - → Provides more feedback to applicants
  - → Diversified and democratised grant review
  - → Scalable: more applicants, more reviewers
  - → Accelerated process – time saving
  - → Cost savings
- Concerns (being tested)
  - → Lack of expertise
  - → Bias
  - → Gaming the system
  - → Scooping
  - → Time commitment for applicants
  - → Confidence of applicants

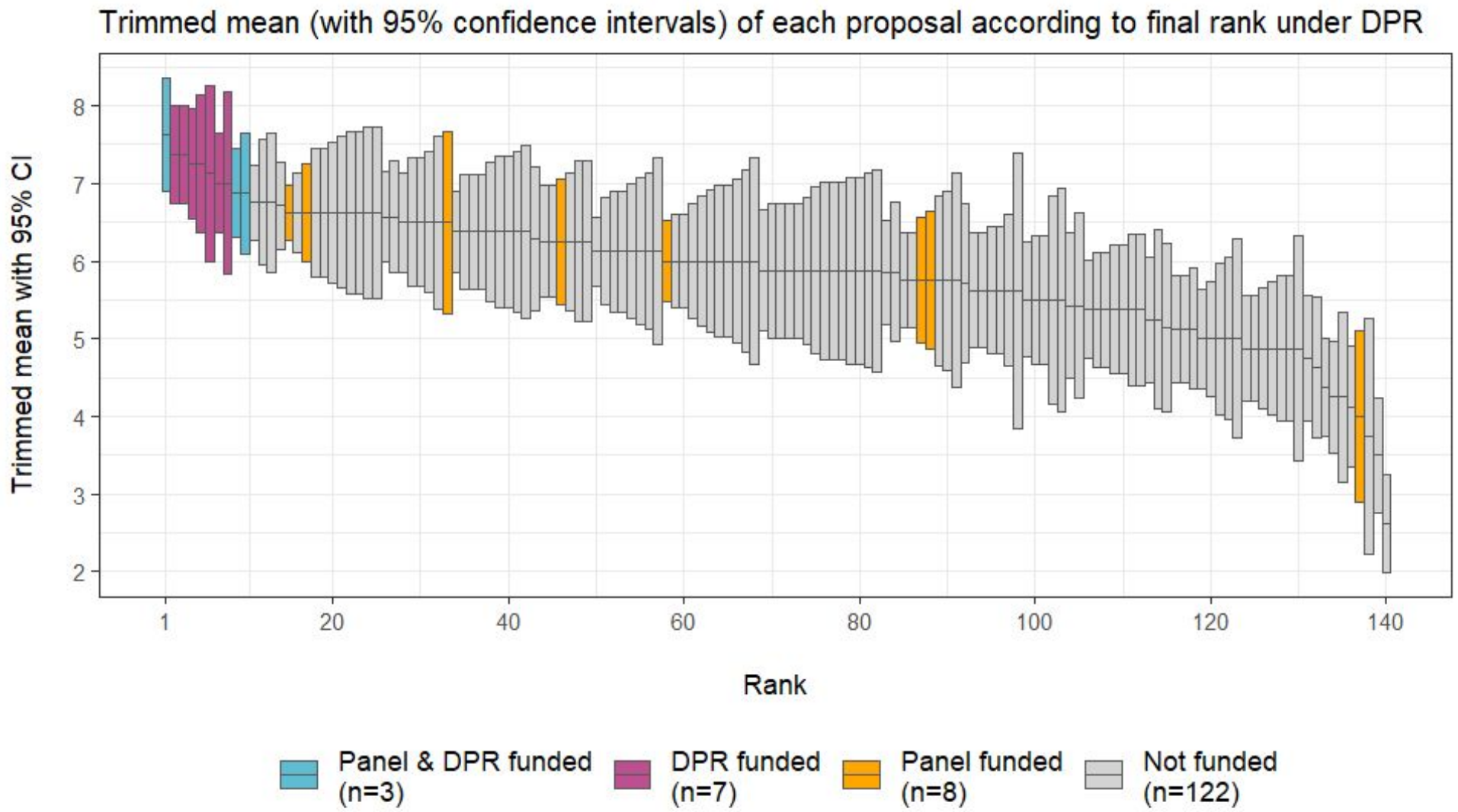
DPR Experiment at the Volkswagen Foundation

- Experiment at the Volkswagen Foundation for the “Open Up” programme – focus on innovation in the Humanities and Social Sciences
- Parallel implementation of DPR and established panel review
- Additional funding provided: funding recommendations from both panel review and DPR
- Mixed methods analysis of results: quantitative analysis of data from submissions and surveys of participants, and qualitative analysis of interviews with a sample of participants
- Rich datasets to gain insight into dynamics of grant peer review e.g.:
  - → Comparisons between review processes
  - → Reviewer uncertainty
  - → Consistency between reviewers
  - → Stability of funding decision
  - → Attitudes of actors

Distributed Peer Review and Panel Review - Parallel Processes

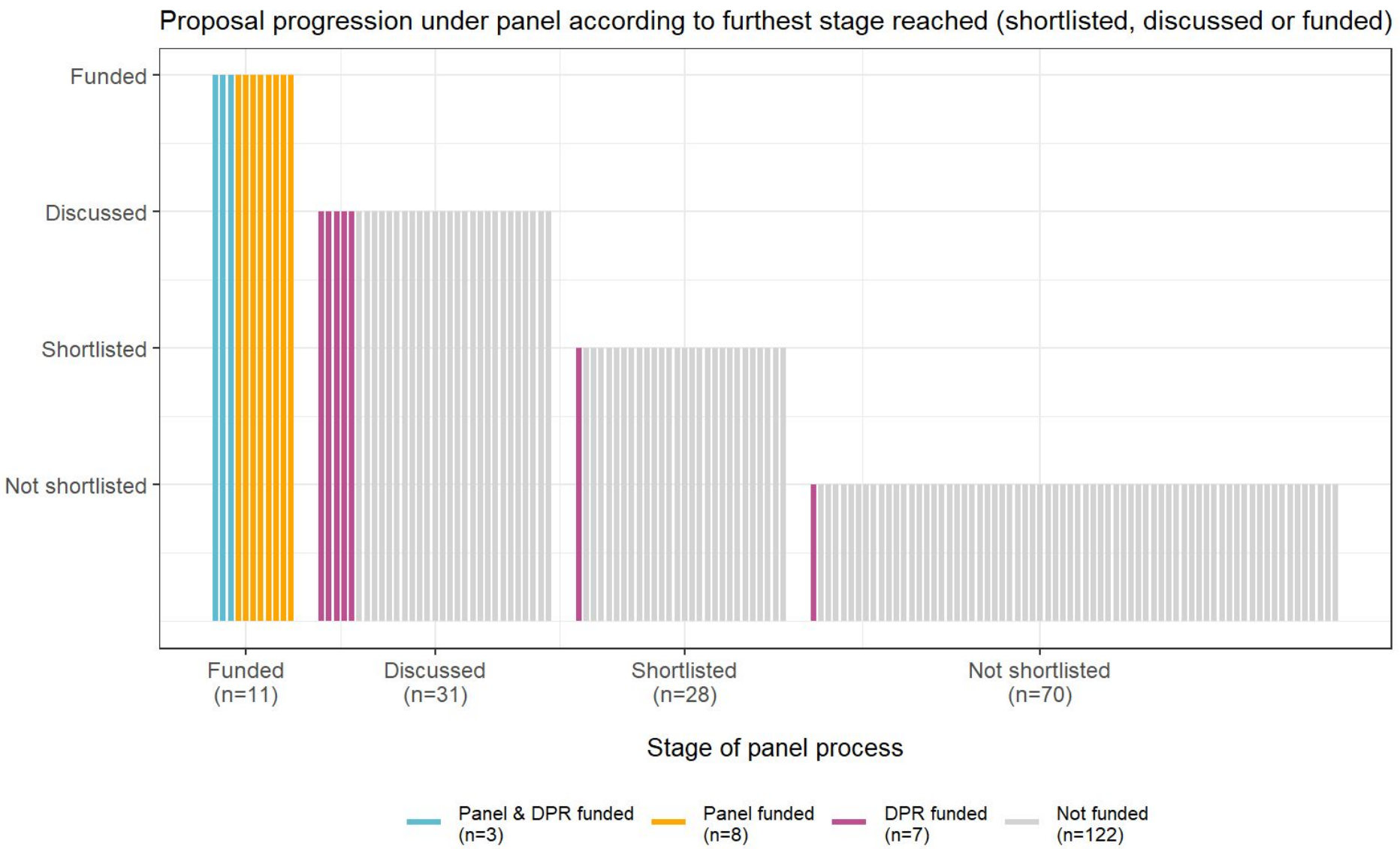


Panel selected proposals are found across the full range of DPR scores



Note: For the purpose of this visualization, where ranks were tied proposals were ordered by standard error. Any remaining ties were broken by proposal order

DPR selected proposals are found across all Panel stages



Some headlines and moving forward

- In DPR, more time is spent reviewing but distributed more equally between more people (each applicant completed 4 or 5 reviews)
- DPR could reduce the duration of the funding allocation process
- DPR and panel reviewers used criteria similarly
- Stability increases with more reviews per proposal but no optimal number of reviews
- The majority of DPR participants felt positive about the process but positivity higher amongst those who were funded
- Comparisons difficult – conventional systems often seen as “tried and tested” but commonly a “black box” (with little feedback) compared with more transparent DPR (each applicant received 9 or 10 review reports)
- Important not to see one system as normative but recognise trade-offs
- Implications for peer review more widely: From the ‘wisdom of the gatekeeper’ to the ‘wisdom of the (expert) crowd’

Areas of concern, particularly:

- → Gaming
- → Workload
- → Review quality
- → …

Our work is focusing on how these concerns can be addressed

Please let us know your thoughts!

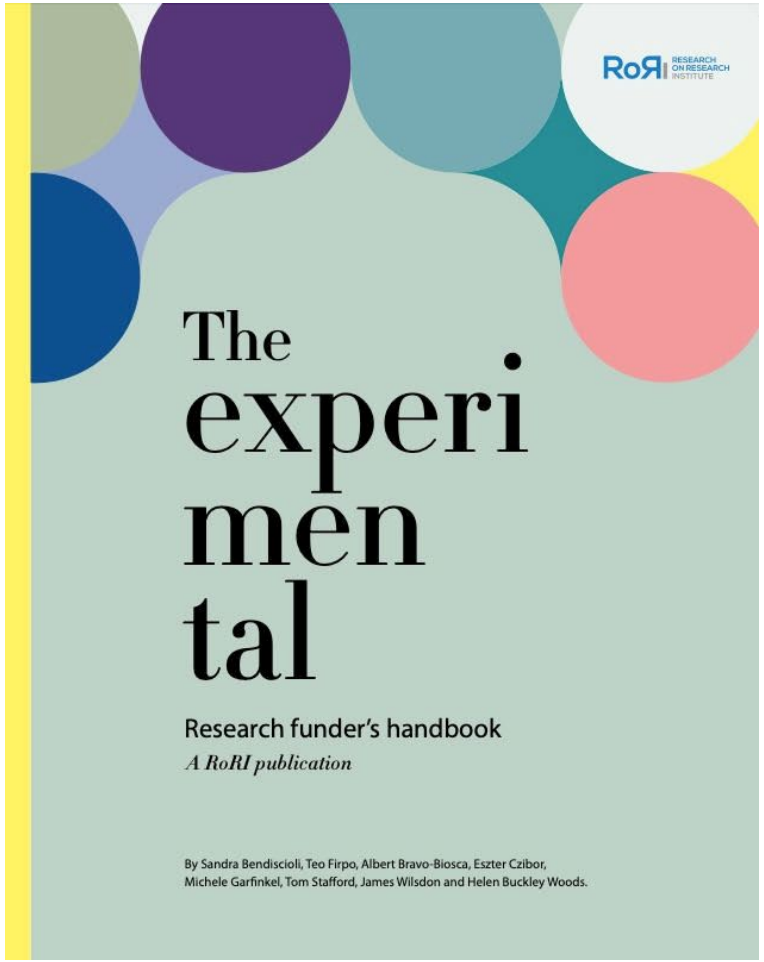
researchonresearch.org
@RoRInstitute

t.stafford@sheffield.ac.uk
a.l.butters@sheffield.ac.uk
s.pinfield@sheffield.ac.uk
m.benson-marshall@sheffield.ac.uk

https://doi.org/10.6084/m9.figshare.29270534.v1

RoRI RESEARCH ON RESEARCH INSTITUTE

Get in touch!



AFIRE:

t.stafford@researchonresearch.org

josie.coburn@ucl.ac.uk

The experimental research funder's handbook (Revised edition, June 2022, ISBN 978-1-7397102-0-0). https://doi.org/10.6084/m9.figshare.19459328.v2

These slides:

bit.ly/tomstafford

END
(reserve slides follow)

References

Barnett, A., Allen, L., Aldcroft, A., Lash, T. L., & McCreanor, V. (2024). Examining uncertainty in journal peer reviewers' recommendations: a cross-sectional study. Royal Society Open Science, 11(9), 240612. https://doi.org/10.1098/rsos.240612

Bendiscioli, S. (2019). The troubles with peer review for allocating research funding. EMBO reports, 20(12), e49472. https://doi.org/10.15252/embr.201949472

Graves, N., Barnett, A. G., & Clarke, P. (2011). Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. BMJ, 343, d4797. https://doi.org/10.1136/bmj.d4797

Guthrie, S., Ghiga, I., & Wooding, S. (2018). What do we know about grant peer review in the health sciences? F1000Res, 6, 1335. https://doi.org/10.12688/f1000research.11917.2

Merrifield, M. R., & Saari, D. G. (2009). Telescope time without tears: a distributed approach to peer review. Astronomy & Geophysics, 50(4), 4.16-14.20. https://doi.org/10.1111/j.1468-4004.2009.50416.x

Prelec, D., Seung, H. S., & McCoy, J. (2017). A solution to the single-question crowd wisdom problem. Nature, 541(7638), 532-535. https://doi.org/10.1038/nature21054

In DPR, more time is spent reviewing <u>but</u> distributed more equally between more people:

DPR: Total of **1763 hours** across **323 reviewers** to review 140 proposals

On average, reviewers spent **4 hours** reviewing all allocated proposals

Panel: Preselection + quick assessments = **195 hours** across **8 panellists + VWS staff**

Panel meeting 9am - 4pm = 12 people x 7 hours = **84 hours**

On average, panellists spent **19.5 hours** completing all quick assessments + attending the panel meeting

More time is spent considering proposals <u>but</u> distributed more equally across all proposals:

- On average the total time spent on all reviews for each proposal was **11.7 hours**
- **At least 70% of proposals received more attention under DPR** than is possible under panel review

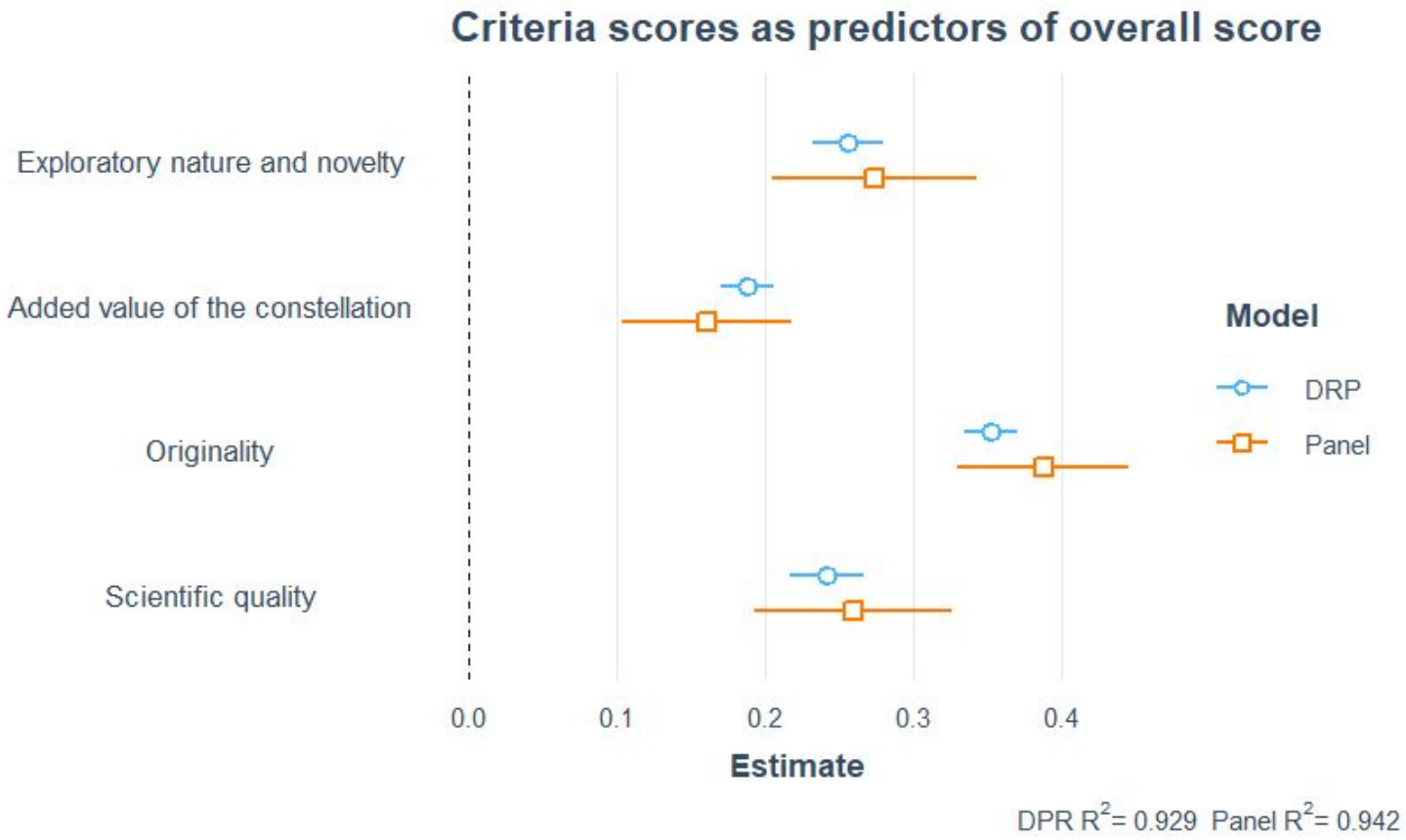
DPR could reduce the duration of the funding allocation process

		A	M	J	J	A	S	O	N	D
General	Submission deadline									
	Initial processing									
	Applicants informed									
DPR	Reviewer: proposal matching									
	Peer review									
	Ranking									
	Preparing feedback									
Panel	Pre-selection									
	Panel quick assessments									
	Panel meeting									

DPR peer review:
~ 6 weeks

Panel review:
~ 8 weeks for pre-selection
~ 5 or 6 weeks for panel quick assessments

DPR and panel reviewers use criteria similarly



Criteria scores very strongly predicted overall score for both sets of reviewers

Originality most important

Added value of the constellation least important

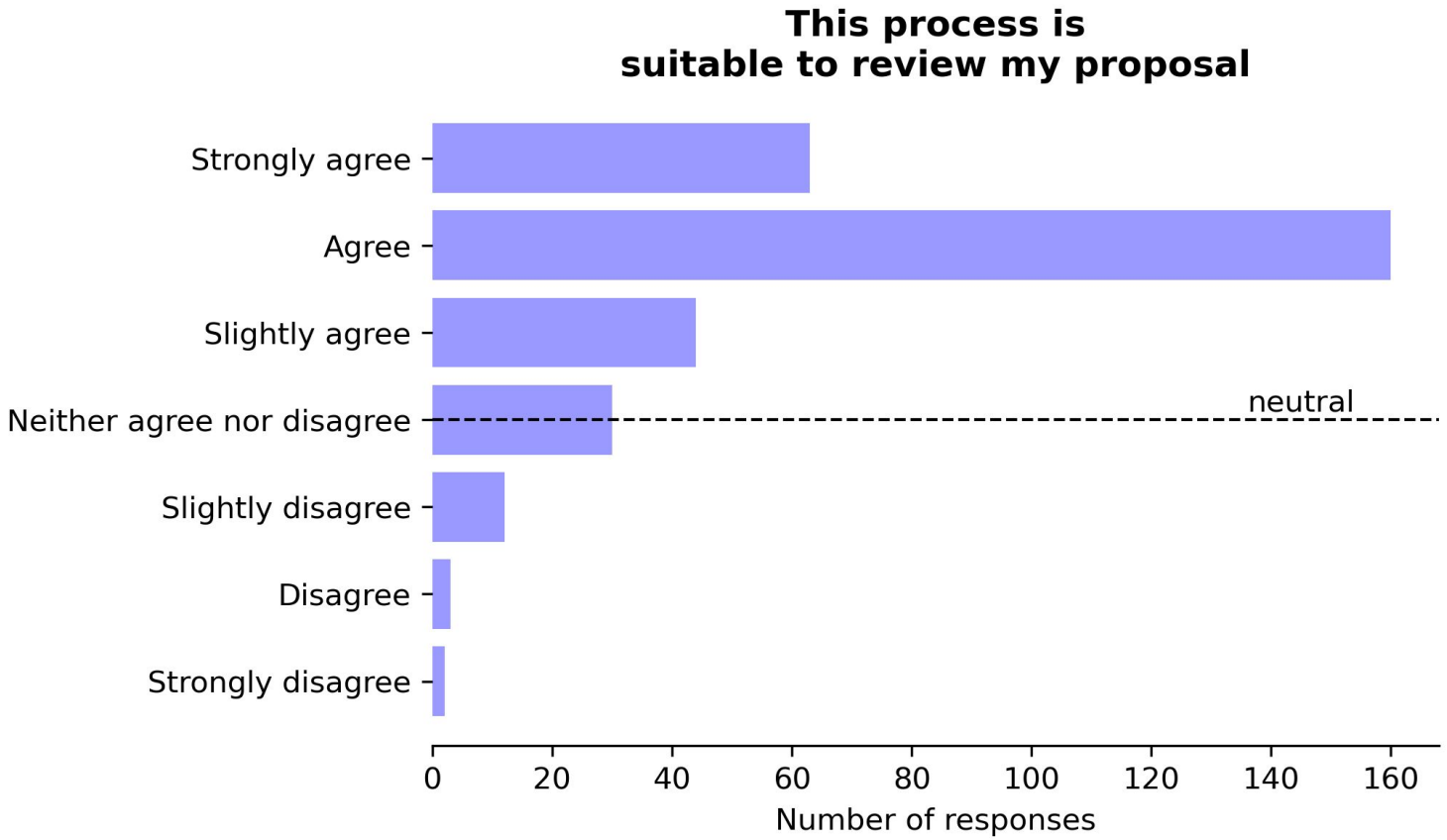
Expectations of DPR were generally positive

Expectations of DPR process:

- Suitability; fairness; identifying appropriate reviewers; selecting best proposals

Expectations of how DPR would compare to panel review

- Identify similar set of proposals; more adventurous proposals



- **85%** thought DPR was suitable
- **74%** trusted it to be fair & fund best research
- **70%** thought would select more adventurous proposals

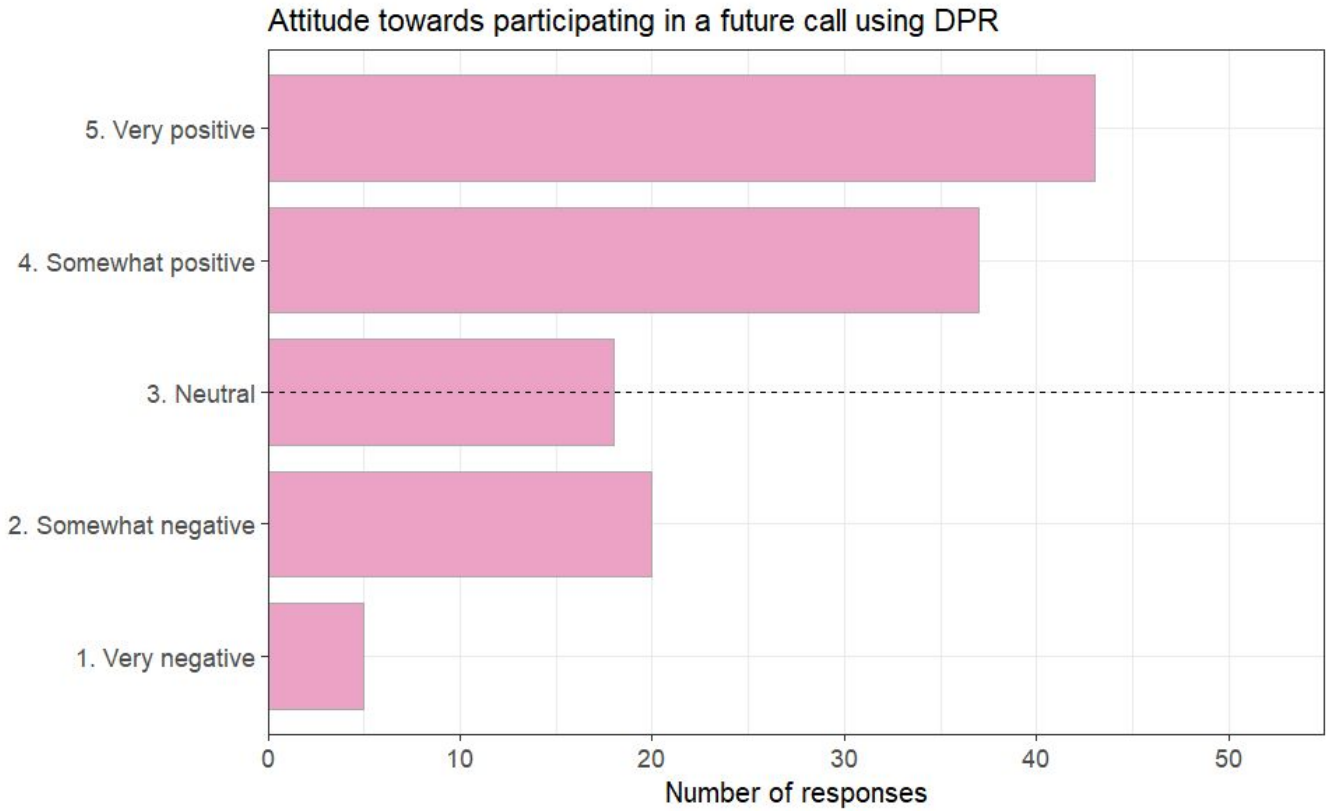
Applicant Feedback

Feedback from 127 applicants across 84 proposals

- 97 not funded, 30 funded

Feedback survey

- Constructiveness of each review
- Overall helpfulness, politeness, expertise, attitude to future DPR



Taking part in future calls using DPR:

83% of **funded** applicants felt somewhat or very positive

60% of **unfunded** applicants felt somewhat or very positive

Caution warranted: applicants' view of review comments is inconsistent

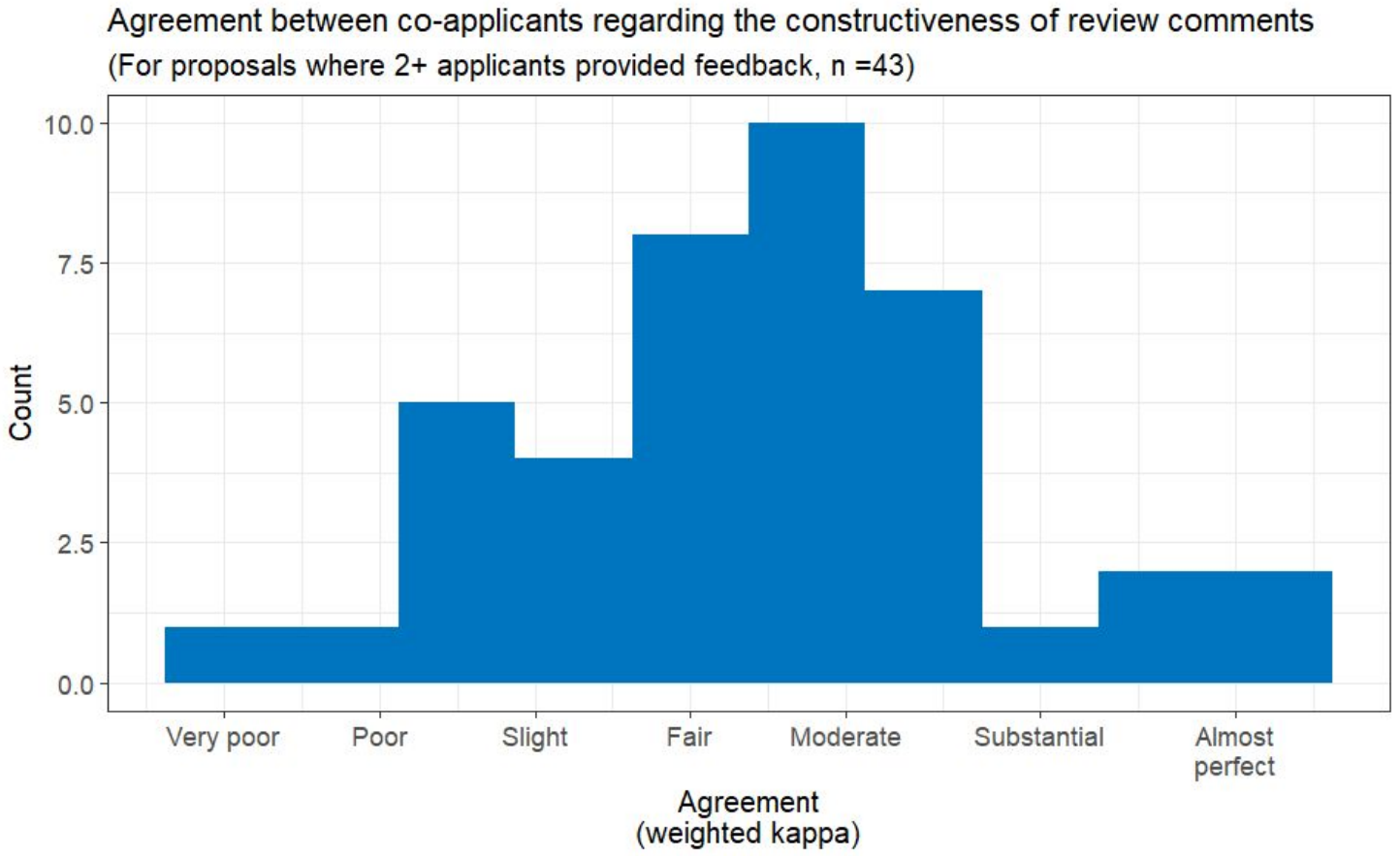
58% of respondents provided text comments

Positives & negatives of DPR experience

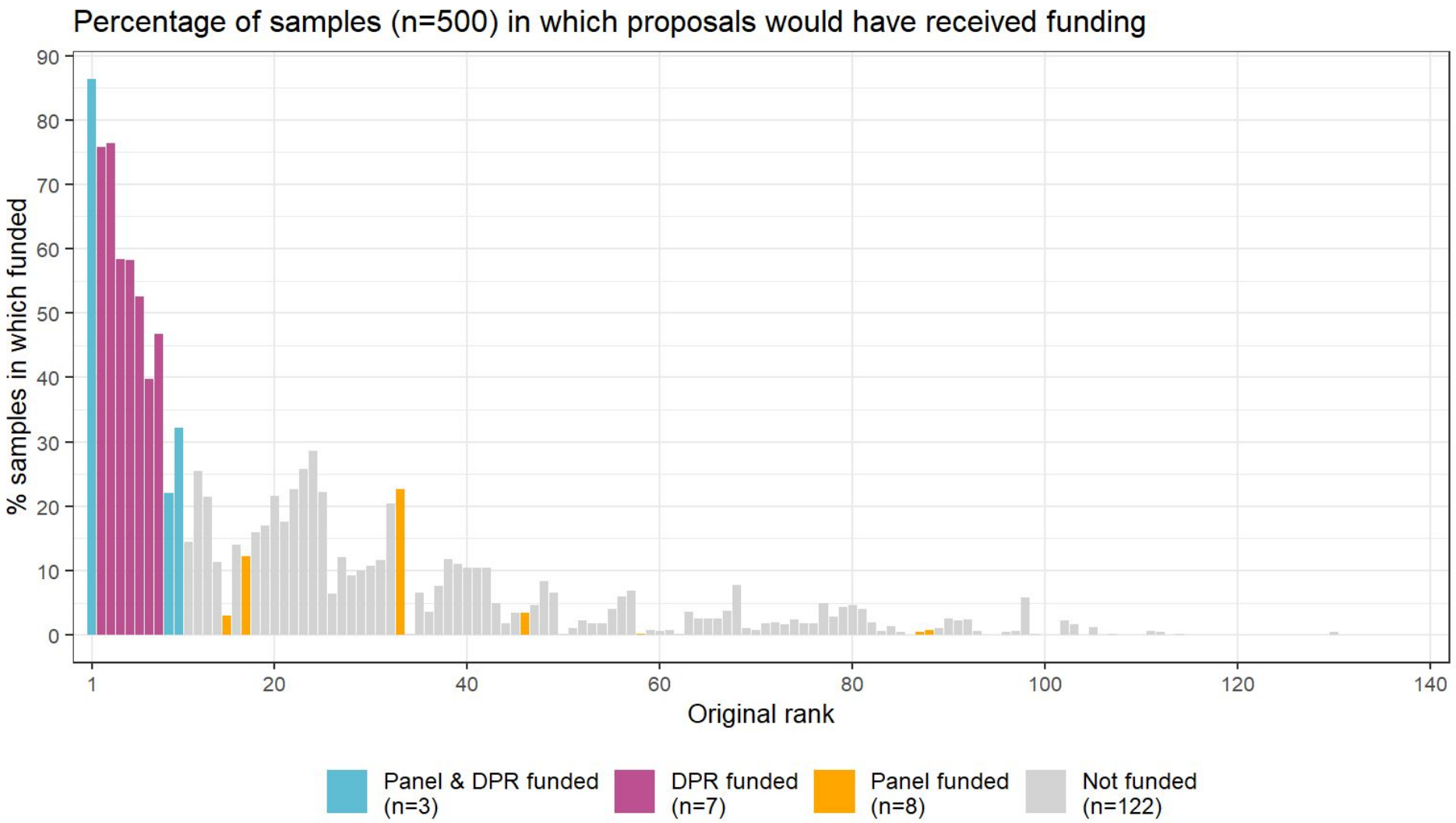
Concerns:

- Gaming
- Additional workload
- Lack of expertise

Agreement about the constructiveness of reviewer comments ranged from very poor to almost perfect.



Estimating stability: What would happen if we did it again?



To examine stability of the rankings we used the observed data to simulate 500 samples.

Across all samples, on average **5.49 proposals** funded under DPR would still be funded

Estimating stability

		DPR B			
		1	2	3	4
DPR A	[1]	**0.34**	0.29	0.19	0.17
	[2]	0.25	**0.29**	0.24	0.21
	[3]	0.21	0.22	**0.31**	0.25
	[4]	0.19	0.19	0.25	**0.37**

VWS DPR

Proportion of overlap in rankings in each quartile.

Greatest agreement for top ranked 25% and lowest ranked 25% .

Agreement in each quartile is better than would be expected by chance

Stability in VWS experiment is similar to previous trials of DPR

		DPR B			
		1	2	3	4
DPR A	[1]	**0.34**	0.29	0.19	0.17
	[2]	0.25	**0.29**	0.24	0.21
	[3]	0.21	0.22	**0.31**	0.25
	[4]	0.19	0.19	0.25	**0.37**

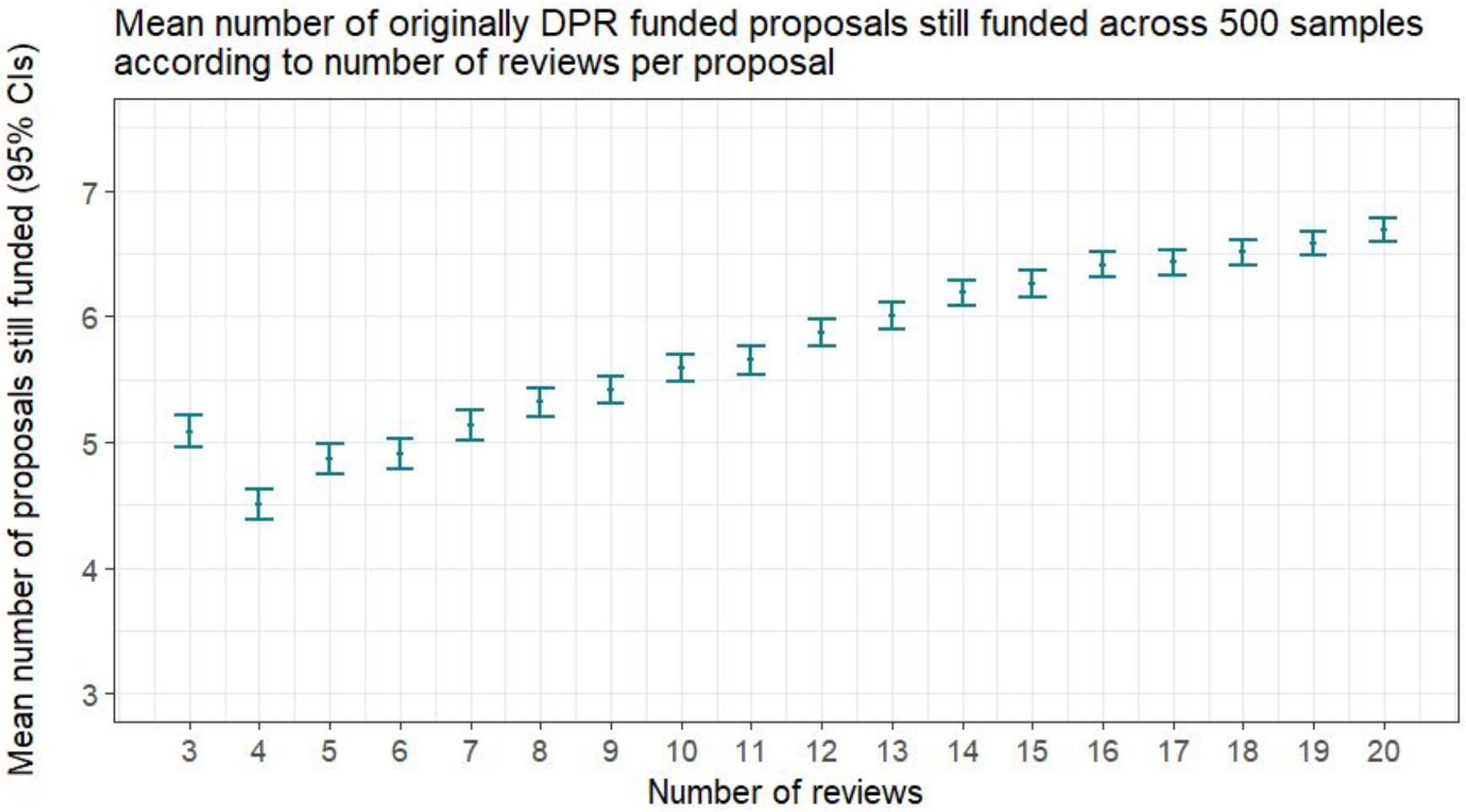
VWS DPR

		DPR B			
		1	2	3	4
DPR A	[1]	**0.33**	0.26	0.24	0.18
	[2]	0.26	**0.26**	0.25	0.23
	[3]	0.24	0.25	**0.25**	0.26
	[4]	0.18	0.23	0.25	**0.34**

ESO DPR

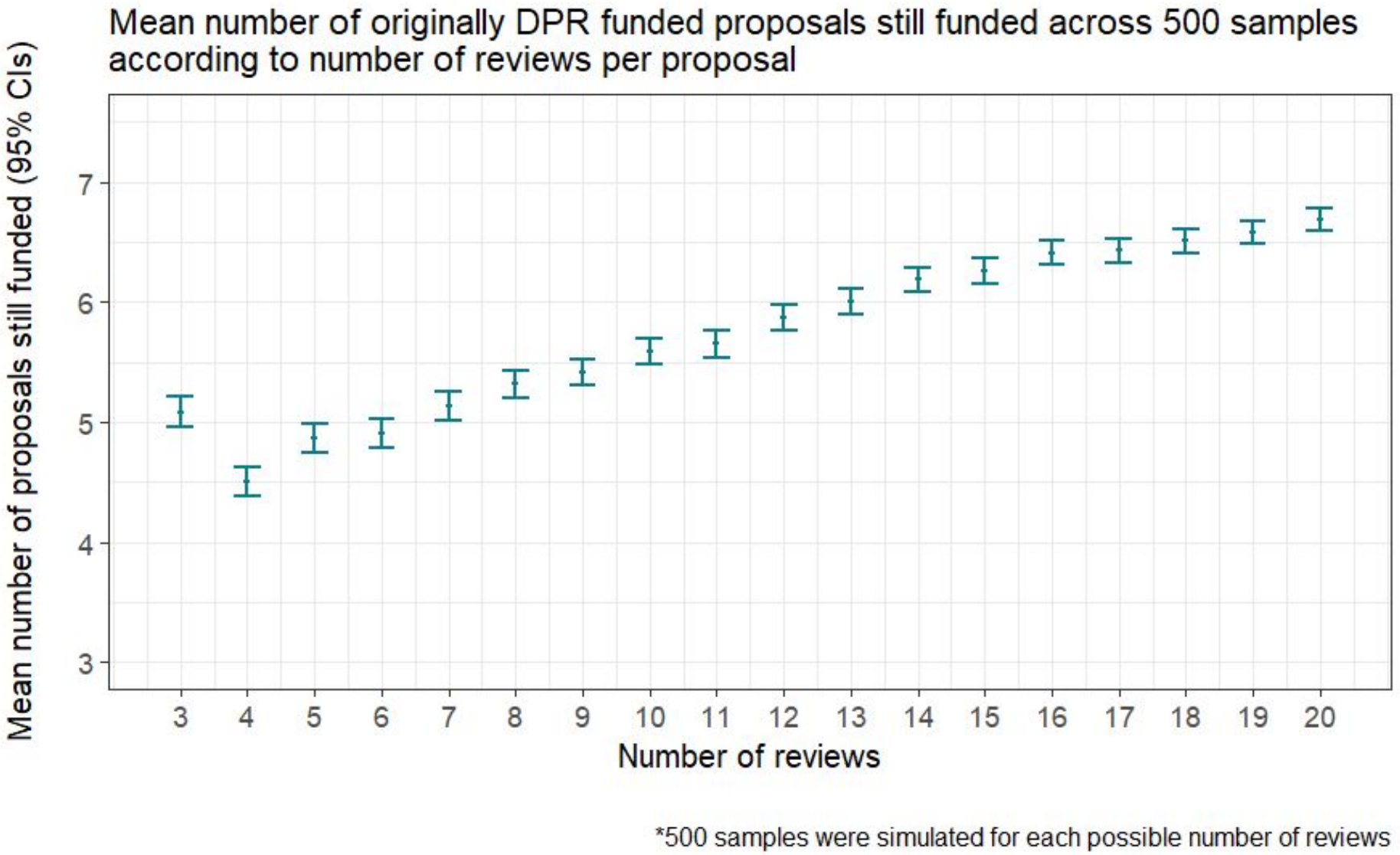
Patat et al. 2019

Stability increases with more reviews per proposal <u>but</u> no optimal number of reviews



*500 samples were simulated for each possible number of reviews

Applicant burden can be reduced



Reducing the number of proposals allocated to each reviewer by 1:

An average of 7.59 reviews per proposal.

Applicant workload reduced by 20-25%

Interviews: methodology

Target population:

- Applicants
  - successful and unsuccessful
  - across different disciplines
  - across different levels of seniority
  - maintain a gender balance
- Panel members
- VWS staff

So far conducted 13 of 20 - aiming for 25 total

- Successful applicants (7)
- Unsuccessful applicants (3)
- Panel members (5)
- VWS staff (5)

Interview topics:

- Expectations and concerns about the panel and DPR processes
- Time and workload commitment of DPR
- Experience as reviewer - criteria, scoring
- Experience receiving feedback - constructiveness
- Fairness of panel and DPR processes
- Advantages and disadvantages of panel and DPR
- Future of this approach
- Other innovations

Interviews: expectations and concerns

- Panel process is established, but lacks transparency - seen as “a black box”
- Initial concerns over gaming in DPR
- Review process being “outsourced” and putting more pressure on ECRs, “deteriorating” the process
- Time and workload commitment (applicants and staff)
- However, short proposal format so happier to spend time reviewing
- Less concern over scooping - projects are often unusual

Interviews: experiences of review

- Experience as reviewer
  - Disciplinary fit sometimes inconsistent
  - Interdisciplinary projects difficult to review, but DPR can be helpful here
  - Some evidence of gaming
  - DPR reviewed ideas; panel reviewed quality of proposal
  - Clear criteria are important and should align closely with call

- Experience receiving feedback
  - Some felt feedback was superficial or lacked understanding
  - Successful applicants tended to feel feedback was constructive!
  - Amount of feedback may be less important than career stage, length/cost of project, or projects that might be submitted elsewhere in future

Panel: *"a process of negotiating… it was not such a clear and strict and criteria-oriented process as people might think or maybe hope"* (panel member)

DPR: *"the pettiness is embedded within the process"* (successful applicant)

"the kind of spirit… in which the DPR was framed and promoted to those of us who were applying, I think that was very positive and [it] feels encouraging to take part in it" (successful applicant)

"communitarian convivial form of academia" (successful applicant)

Interviews: Advantages and disadvantages of panel and DPR

- Panel:
  - Well-established form of review; known advantages and disadvantages
  - Good for discussion, but can also result in being swayed by persuasive arguments or dominant personalities
  - Funder can oversee process better

- DPR:
  - Removes problem of finding reviewers, but is only as 'good' as the applicants (less control, may have gaps, lack of diversity)
  - More and broader feedback
  - Favours innovative, impactful work over traditional "ivory tower discussions"
  - Speed of whole process is seen as a big advantage
  - Concerns over gaming remain

"The reviewing process is a social situation… it has to do with how people interact with each other and who is presenting him or herself in which way. I would say that did make a great difference" (panel member)

"if it comes to emphasising that we really want to fund risky projects and give them kind of this additional push, I would maybe advocate and go for the panel meeting despite all the experiences" (VWS staff)

Interviews: future of DPR

Future of this approach:

- Combination of two systems (merging scores/feedback)
- Two-stage process: DPR, then panel
- Funders may need to retain power of veto
- When to use DPR: calls with “a more or less homogeneous group of disciplines or applicants or topic” (VWS staff), or is the broadness an advantage?
- Could also use DPR for choosing calls/topics; project extensions

Other innovations:

- Involving practitioners/those outside academia
- Presentations by applicants
- Discussion among DPR participants
- Completely open/public peer review
- Successful applicants could review subsequent calls

140 proposals, 323 reviewers, 1387 reviews…..

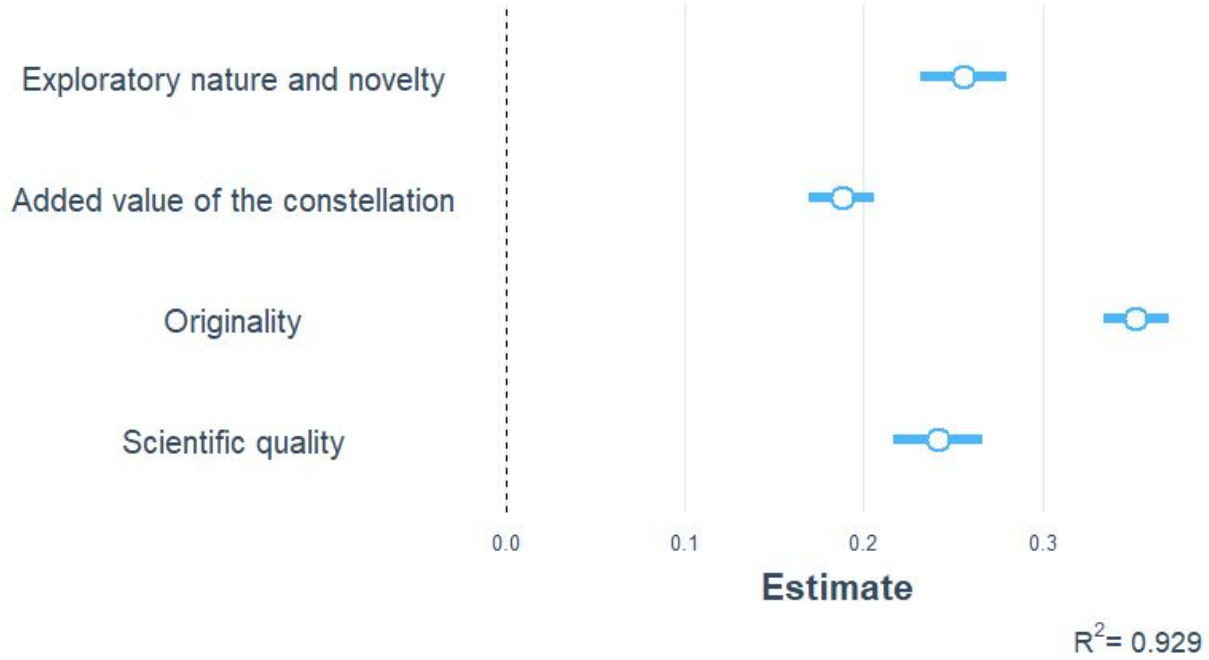
Mean score awarded: 5.81 out of 9

5 = B *Fair: good scientific case but with definite weaknesses*

6 = B+ *Good: minor deficiencies do not detract from strong scientific case*

Scores for all funding criteria were significant predictors of overall score

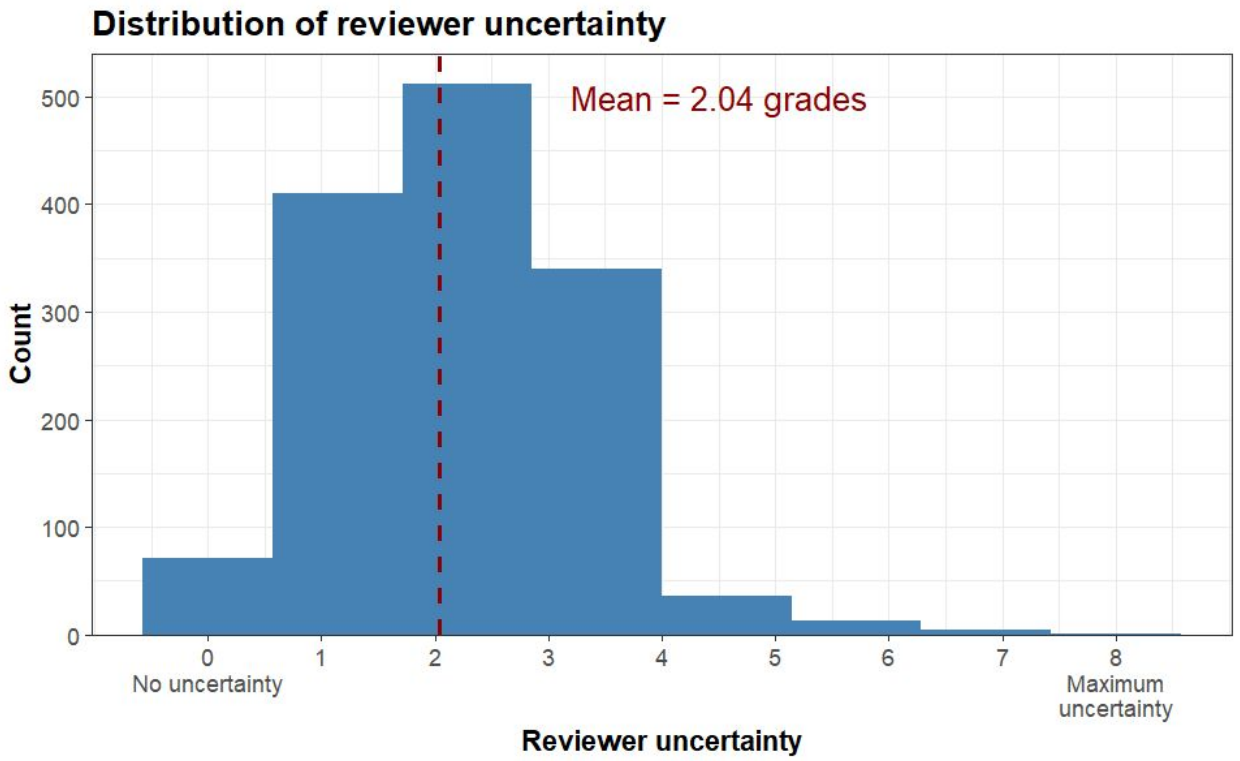
“Originality” was given the most weight



Reviewers are uncertain but uncertainty was not associated with funding likelihood

Reviewer uncertainty: *The distance between the lowest and highest score considered*

- Reviewers reported **no uncertainty** for **5.12% of reviews.**
- **Greater reviewer uncertainty** predicted a **lower score** being awarded (= -0.20; *p*<.001)
- **Average reviewer uncertainty** per proposal **not** **significantly associated with final ranking** (*p* =.083)



Stability of funding decisions is modest

Across 500 bootstrapped samples mean of **5.49** (SD =1.29) **out of 10** previously funded proposals would still be funded.

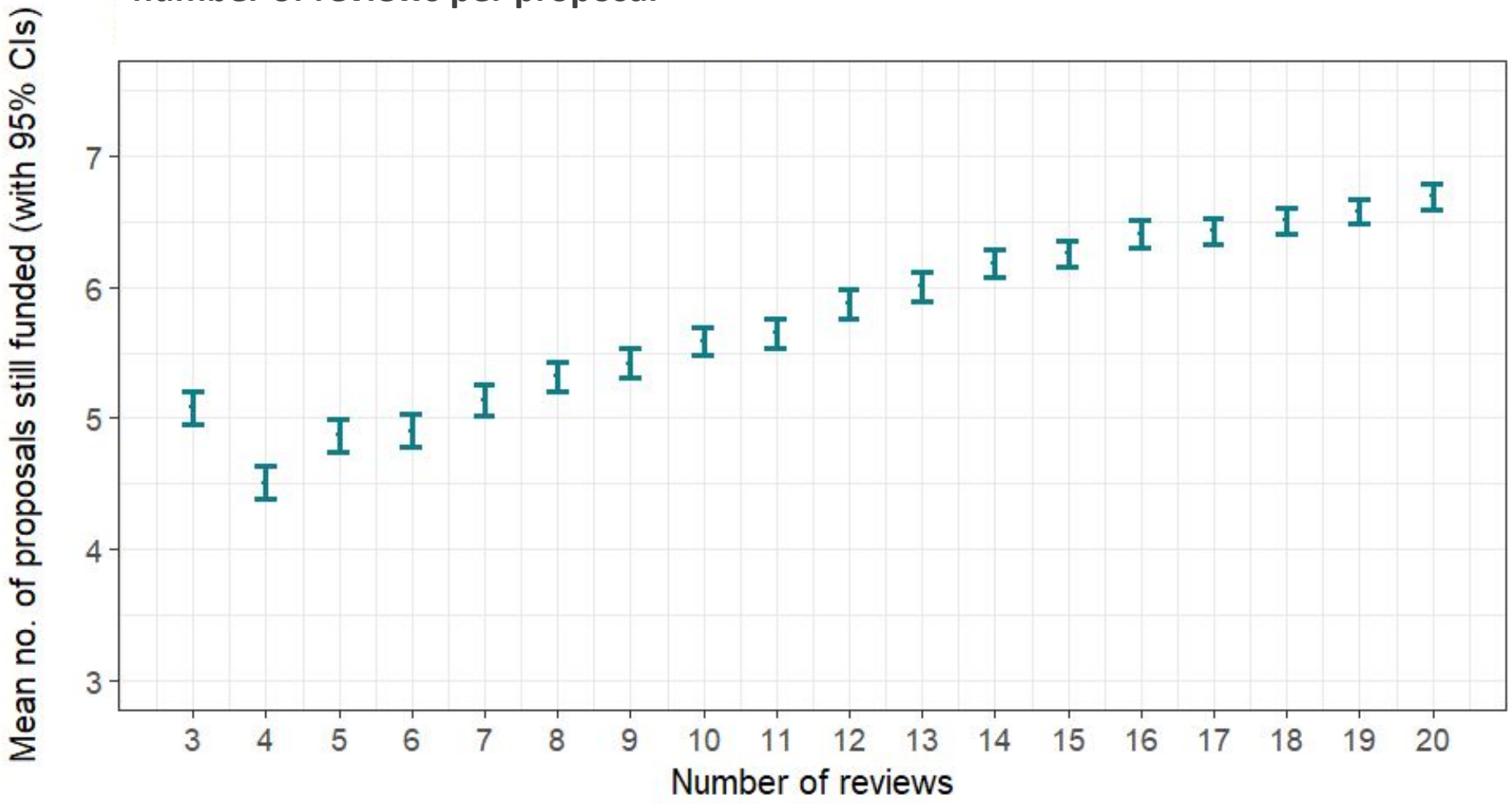
Across all proposals:

- **74.29%** were **sometimes** funded
- **25.71%** were **never** funded
- **No proposal** was **always** funded

As per Graves et al., 2011

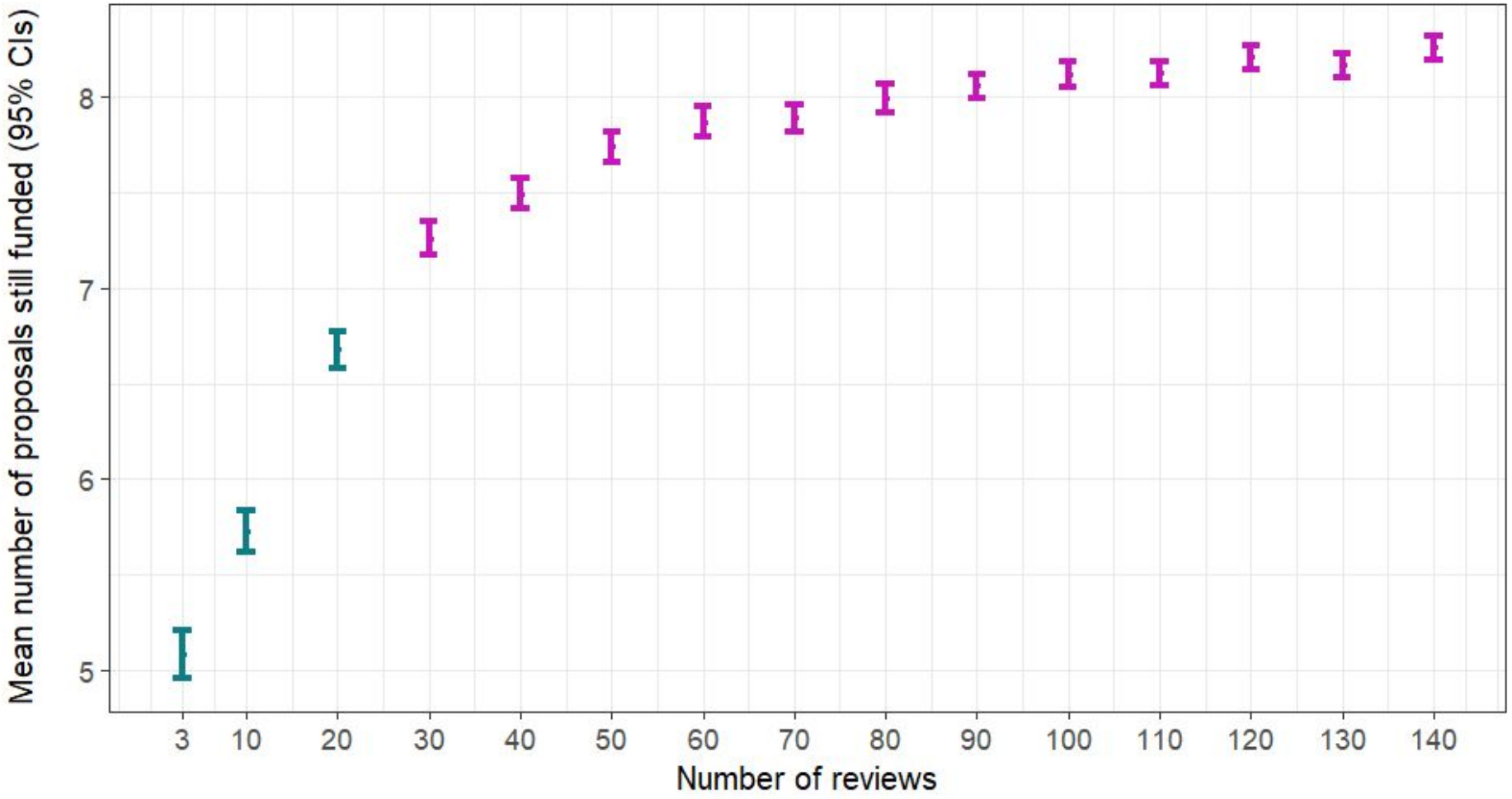
Increasing number of reviews per proposal increases stability <u>to an extent</u>

Mean number of originally DPR funded proposals still funded according to number of reviews per proposal



Increasing number of reviews per proposal increases stability <u>to an extent</u>

Mean number of originally DPR funded proposals still funded according to number of reviews per proposal



VolkswagenStiftung

RoRI RESEARCH ON RESEARCH INSTITUTE

Stability may be impacted by low inter-reviewer consistency

Mixed effects model:

Score = Average score + Effect of proposal + Effect of reviewer + Residual error

More of the variation in scores attributable to differences between proposals than to differences between reviewers:

9.16% attributable to **between-reviewer** differences.

19.14% attributable to **between-proposal** differences

But one third of explainable variance attributable to between reviewer differences

→

Subjective nature of peer review

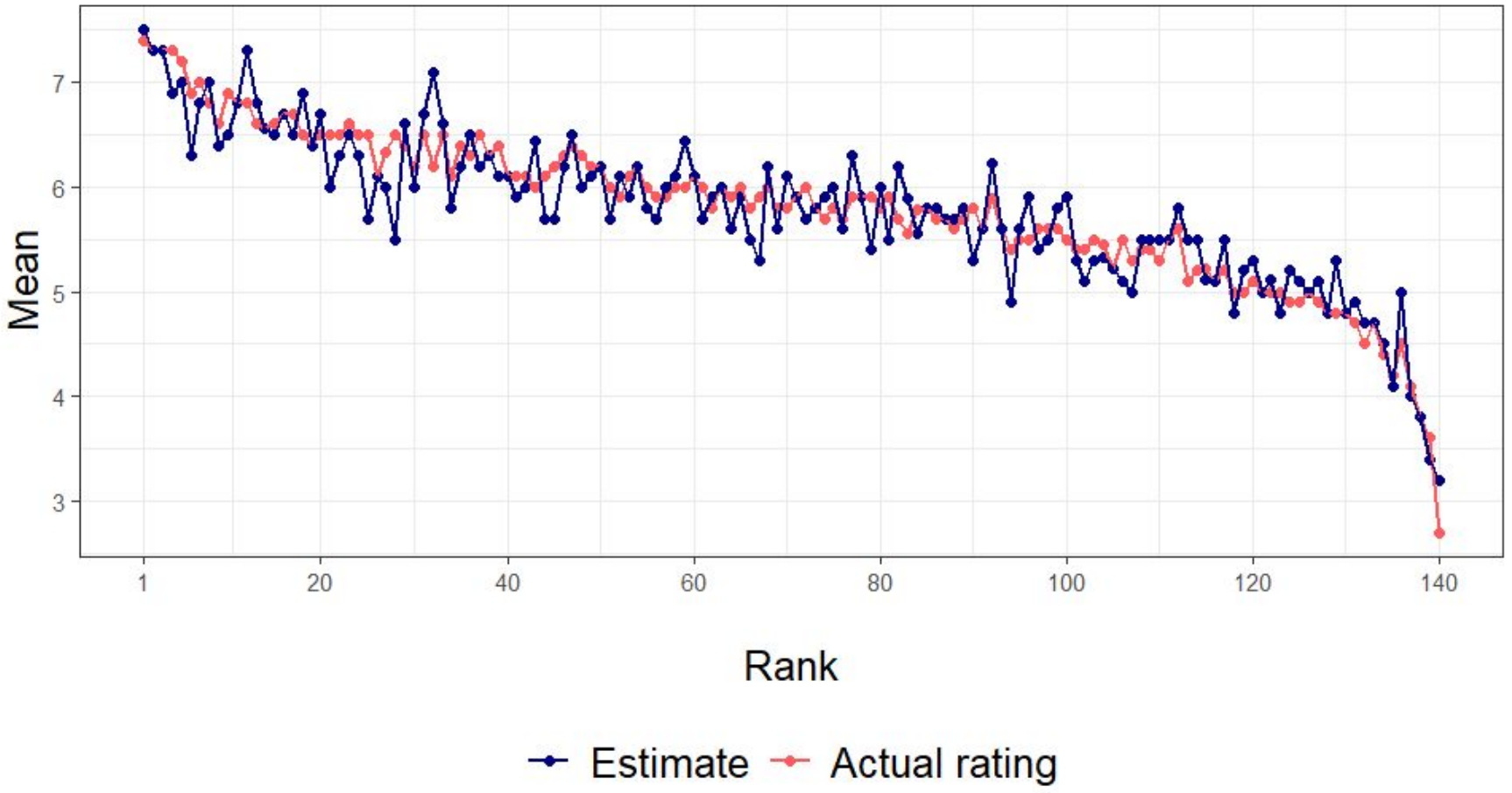
	Fixed Effects					
	β	**SE**	**95% CI**	**t**	***p***	**df**
Intercept	5.86	0.07	5.72, 6.00	82.48	<.001	159.66
	Random Effects					
			Variance		**SD**	
Reviewer (intercept)			0.20		0.45	
Proposal (intercept)			0.42		0.65	

Mixed effects model of the effects of reviewers and proposals on proposal score

Reviewers were sensitive to other reviewers' judgements

Estimate: *"What do you predict will be the average score given to this proposal by the reviewers?"*

Actual and estimated scores per proposal



Average differences between actual and estimated scores per reviewer were also low

(Median difference= 0.10)