DGPs | Deutsche Gesellschaft für Psychologie

CoARA

EINSTEIN FOUNDATION AWARD

IN COOPERATION WITH BIH QUEST Center for Responsible Research

Responsible Research Assessment

Evaluating research quality beyond *h*-index and JIF with the RESQUE framework

(**RES**earch **QU**ality **E**valuation)

Felix Schönbrodt, Anne Gärtner, Daniel Leising

Alp Kaan Aksu, Maximilian Ernst, Maximilian Frank, Nele Freyer, Mario Gollwitzer, Malika Ihle, Jens Lange, Dorothee Mischkowski, Philipp Musfeld, Aaron Peikert, Le Vy Phan, Manfred Schmitt, Anne M. Scheel, Anna-Lena Schubert, Ulf Steinberg

Prof. Dr. Felix Schönbrodt
Ludwig-Maximilians-Universität München

2024-06-10

www.osc.uni-muenchen.de
www.researchtransparency.org
@nicebread@scicomm.xyz

A typical hiring committee

A caricature.

(?)

A typical hiring committee …

- You get unstructured PDFs from which you have to laboriously (and error-prone) search for the relevant information

A typical hiring committee …

- A secretariat creates an xlsx file with an applicant overview, with the “usual” indicators.
  - Does not find some applicants in the Web of Science, then uses Google Scholar
  - It is unclear whether third-party funding only counts as “PI”, often it is not indicated at all.

A	B	C	D	E	F	G	H	I	J	K	L	M
Name	**Vorname**	**Anrede**	**Habilitation bzw. equivalente Leistung**	**Publikationen (Peer-Review Artikel)**	**Erstautorenschaften**	**Jahre der Publikationstätigkeit**	**Publikationen/Jahr**	**Drittmittelprojekte (> 20.000 Euro)**	**Summe Drittmittel**	**Forschungsansatz (gewünshct: fMRT, EEG, tMS)**	**Anzahl Zitationen (Google Scholar)**	**h-Index**
Mustermann	Max	Dr.	2019	31	11	8	3,9	4	640.000	fMRT, EEG	8.239	22
Musterfrau	Eva	Prof. Dr.	2015	13	4	13	1,0	2	1.100.000	fMRT, EEG	10.265	13
Einstein	Albert	Prof. Dr.	1895	8	8	7	0,5	0	0	(nicht empirisch)	172.073	8

A typical hiring committee …

- Very few people make the first selection relatively subjectively
  (so as not to overburden the committee)

- In the committee meeting …
  - … it is unclear how elements from the job ad should be measured/evaluated (cf. “Open Science Statement”)
  - … it is unclear how the various dimensions should be weighted
  - … 1 paper from the shortlist is read by 1 committee member, who presents his/her summarizing judgement
    - More of a gut feeling - apart from the external member, no one really knows the subject, as it is precisely this subject that is not covered.
  - … a list is compiled on the basis of an intuitive weighting.

Aspects of a good diagnostic process

- **Standardization** of available information: All relevant information should be available from all candidates in parallel
  - To this end, the Commission must determine which evaluation dimensions are important before the announcement of the position and request them
- **Efficiency**: The Commission should be able to concentrate on the important questions
  - With many applicants: Efficiency and validity in the first selection (from long-list to short-list).
- A preparation of the applicant information that encourages discussion about research **quality** (less quantity) and multidimensional profiles
  - Encourage a substantive discussion, especially in the shortlist phase.

Areas of research assessment

- Graduation (PhD, habilitation), hiring, tenure track
- Funding
- Awards, Rankings
- Performance-oriented payments and rewards

What is the goal of research assessment in academic hiring?

- To find a colleague who …
  - brings a lot of third-party funding to the university?
  - boosts the university's THE ranking by having a lot of papers, which are also highly cited?
  - excels in teaching and mentoring?
  - contributes to scientific progress and credible knowledge?
  - is a nice and agreeable person who makes no trouble in the department?

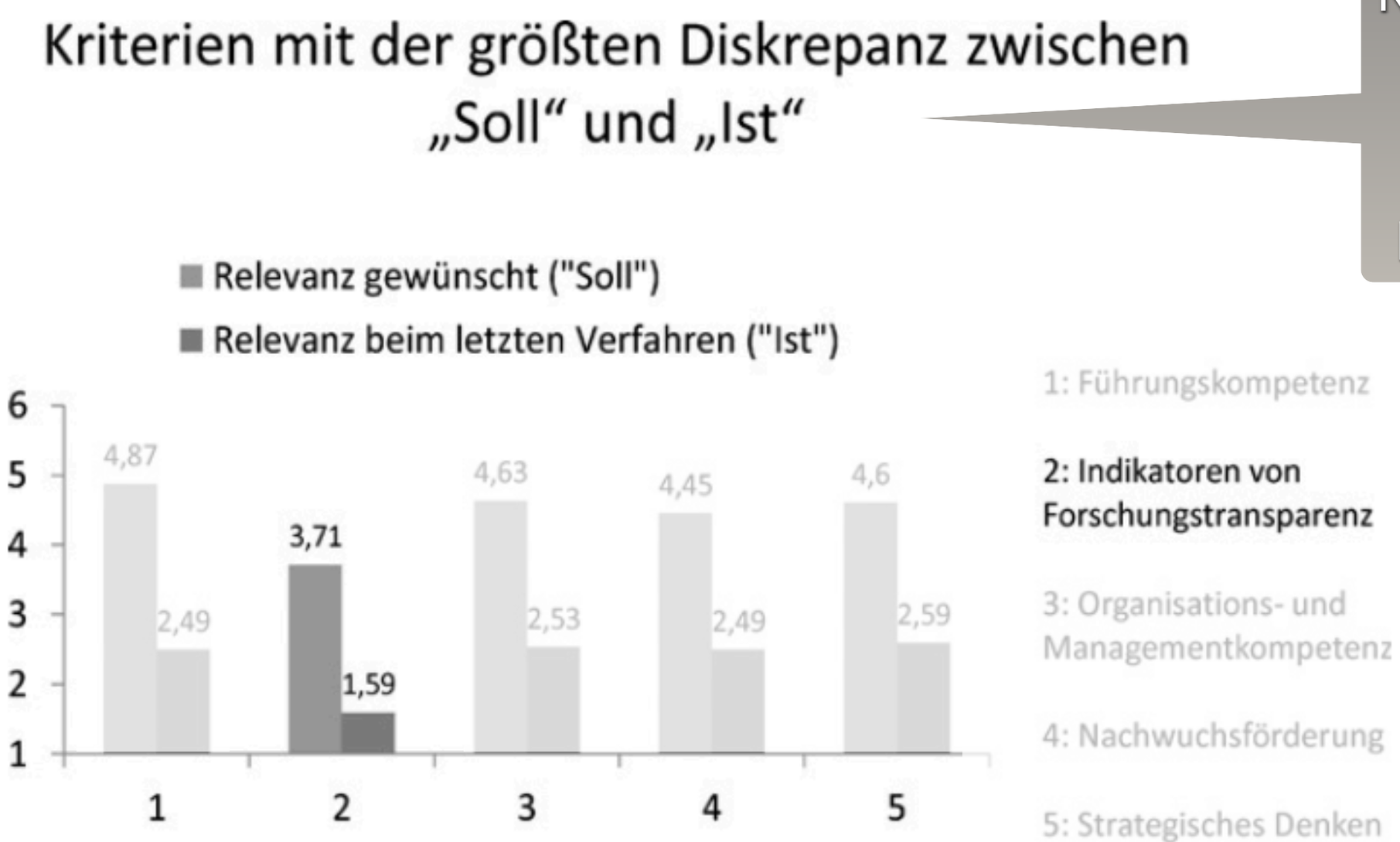
Quantity, not quality

Actual (not desired) relevance at professorship hiring committees:	Rank
Number of peer-reviewed publications	1
Fit of research profile to the advertising institution	2
Quality of research talk	3
Number of publications	4
Volume of acquired third-party funding	5
Number of first authorships	6
…	…
Quality assessment of the best three publications	17
…	…
Indicators of research transparency	**41 (of 41)**

Abele-Brehm, A. E., & Bühner, M. (2016). Wer soll die Professur bekommen? Psychologische Rundschau, 67(4), 250–261. http://doi.org/10.1026/0033-3042/a000335

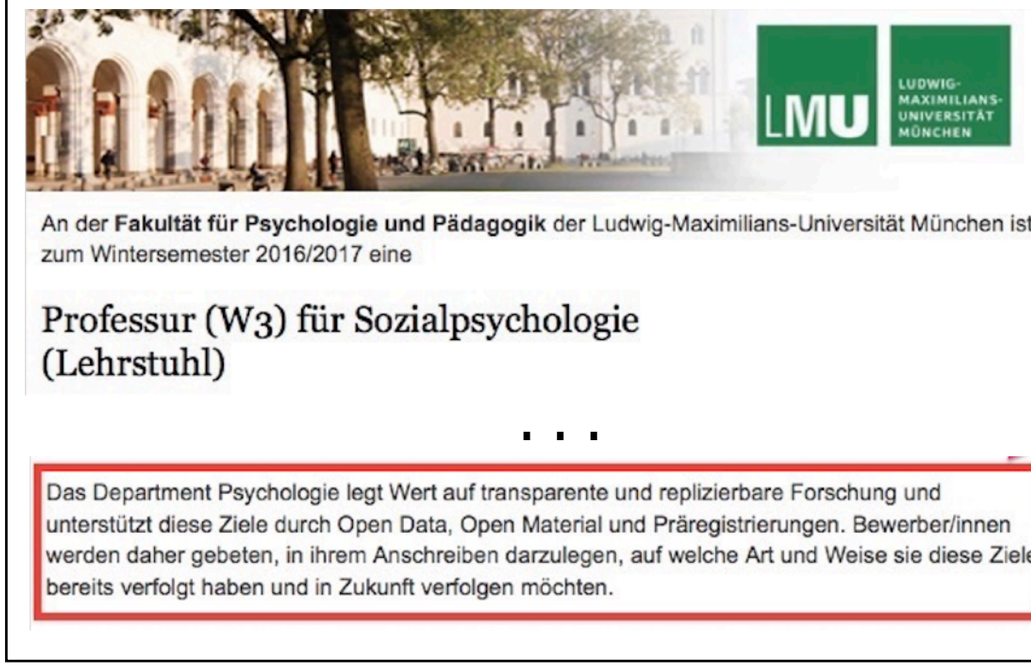
Quality, not quantity

Indicators with the largest discrepancy between „desired“ and „actual“: Researchers **want** to have indicators of research transparency in hiring committees!



Abele-Brehm, A. E., & Bühner, M. (2016). Wer soll die Professur bekommen? Psychologische Rundschau, 67(4), 250–261. http://doi.org/10.1026/0033-3042/a000335

Change of incentive structures: Hiring policy



An der **Fakultät für Psychologie und Pädagogik** der Ludwig-Maximilians-Universität München ist zum Wintersemester 2016/2017 eine

Professur (W3) für Sozialpsychologie (Lehrstuhl)

. . .

Das Department Psychologie legt Wert auf transparente und replizierbare Forschung und unterstützt diese Ziele durch Open Data, Open Material und Präregistrierungen. Bewerber/innen werden daher gebeten, in ihrem Anschreiben darzulegen, auf welche Art und Weise sie diese Ziele bereits verfolgt haben und in Zukunft verfolgen möchten.

Since 2015: All professorship job descriptions use this requirement

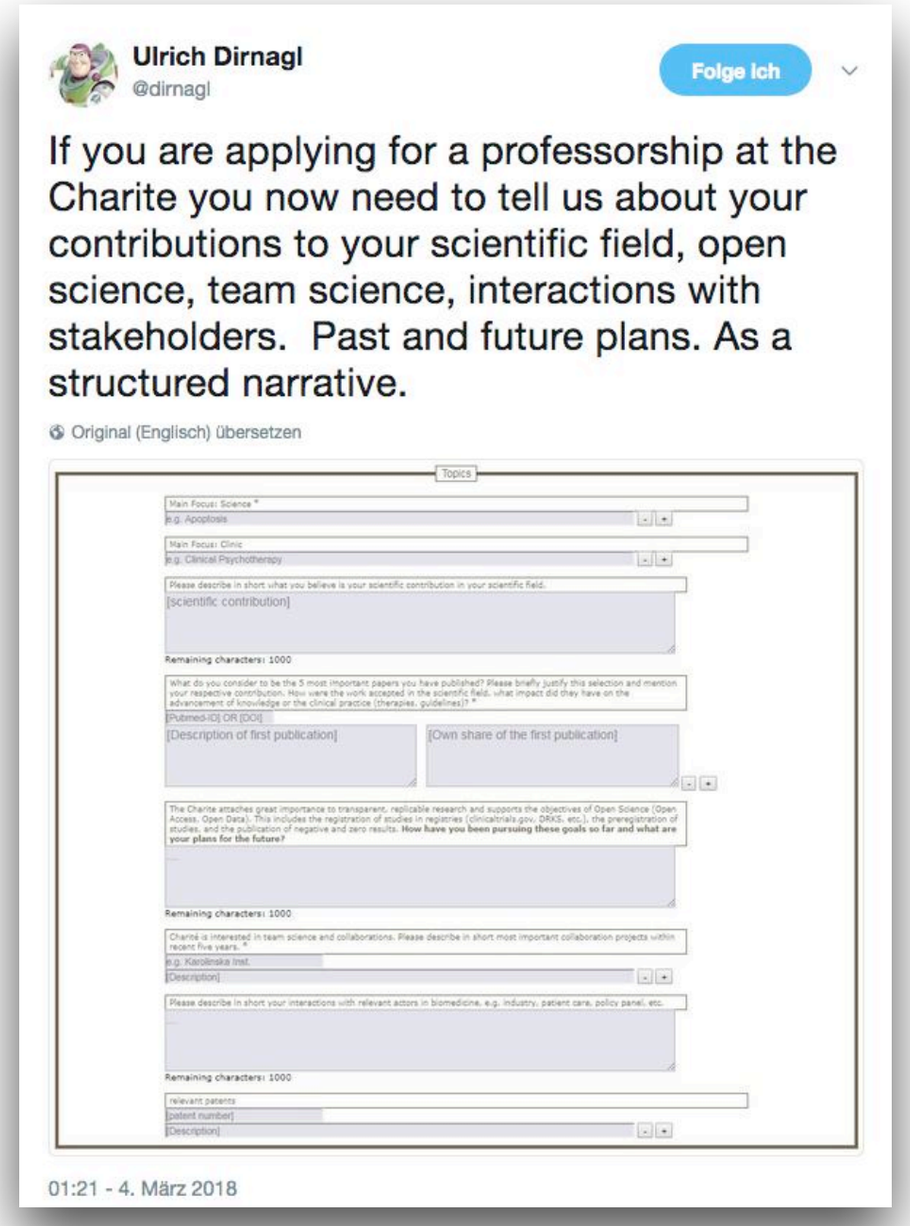
Full Professor (W3)

of Social Psychology

to be filled as soon as possible.

. . .

The Department of Psychology aims for transparent and reproducible research (including Open Data, Open Materials, and Preregistrations). Applicants are asked to illustrate how they have pursued these goals in the past and/or how they plan to do so in the future.



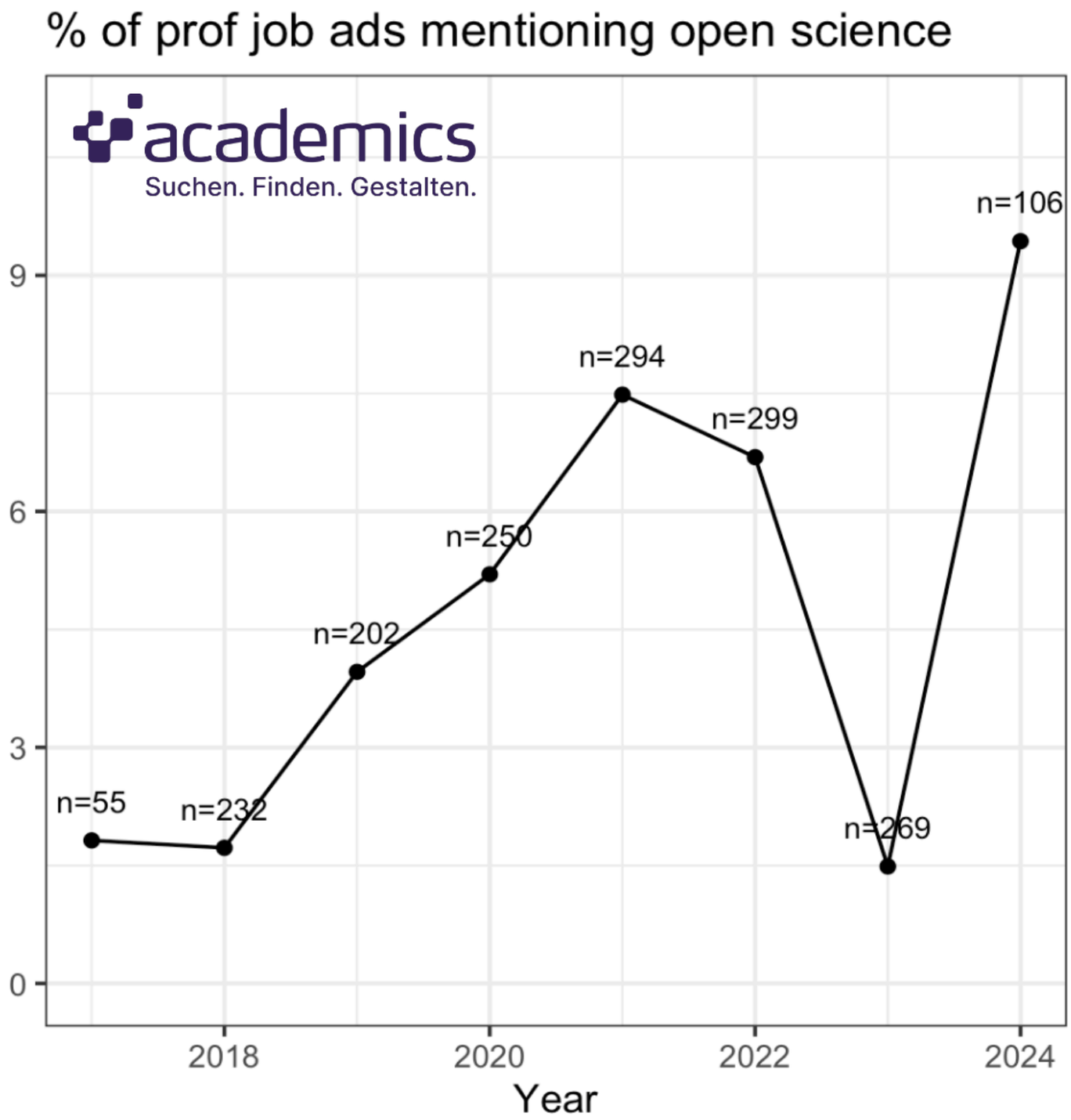
See more such prof job ads at: https://osf.io/7jbnt/

https://docs.google.com/document/d/1ty43Syw0Flkh8ncjW8MZArIkvYe8hLwwhLIlwbtSk_Y/edit?usp=drive_web&ouid=108982640291853577145

Change of incentive structures: Hiring policy

Analysis of professorship job announcements in psychology:

- 1707 job ads; entire database of academics.de from Feb 2017 to June 2024
- Keyword search for *open science*, *reproduc**, *replication*, *research transparency*, etc.
- Out of 420 advertising institutions, 34 mentioned replicability and transparency at least once (8%) as desired or essential skill of a professor.

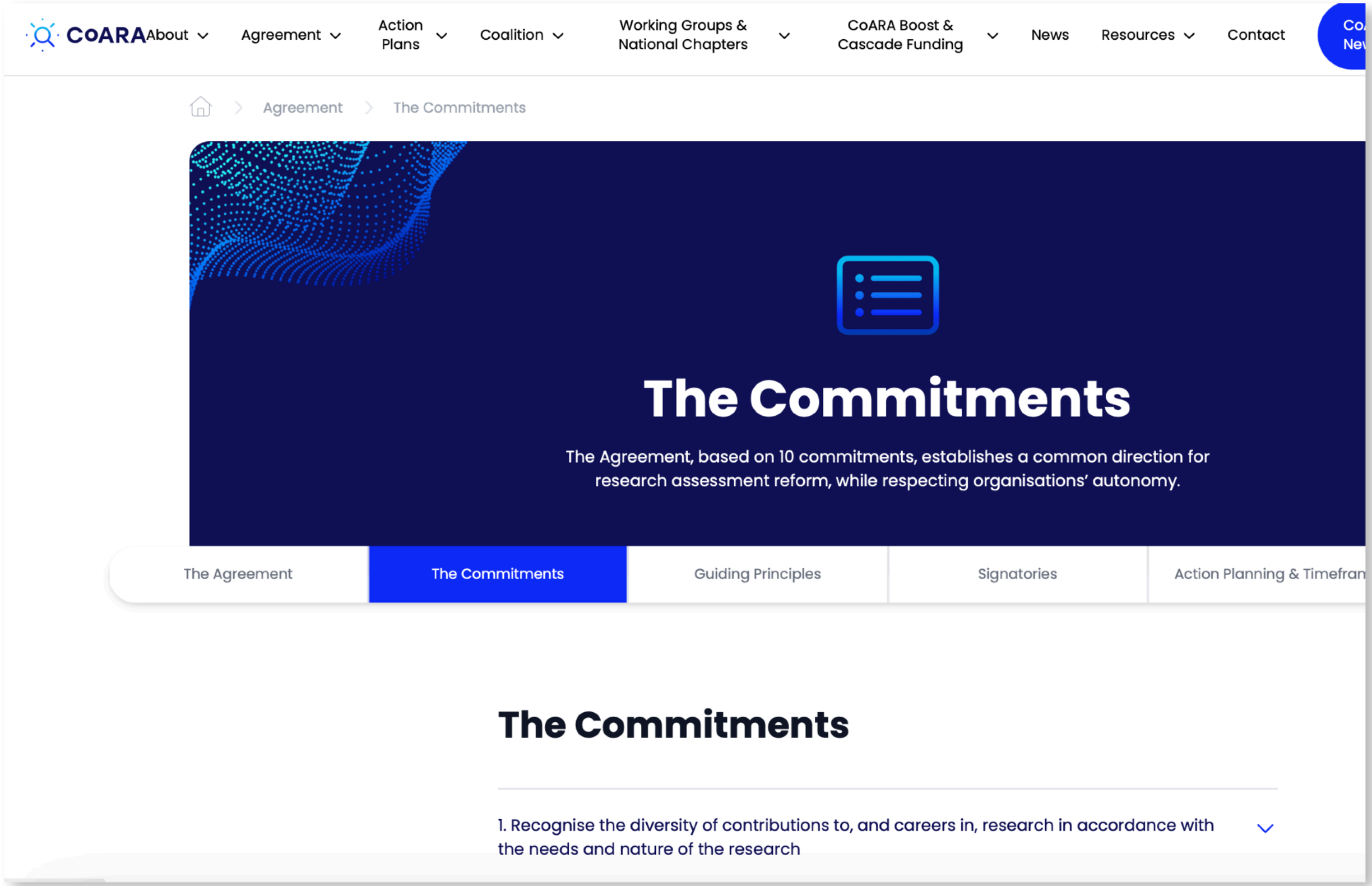


Nosek et al. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology, 73(1)*, https://doi.org/10.1146/annurev-psych-020821-114157; analysis updated in June 2024

What is scientific progress, and can we predict it?

- "**Quality** is primarily an activity-oriented concept, concerning the skill and competence in the performance of some task. **Progress** is a result-oriented concept, concerning the success of a product relative to some goal.
  - All acceptable work in science has to fulfill certain standards of quality. But it seems that there are no necessary connections between quality and progress in science. Sometimes very well-qualified research projects fail to produce important new results, while less competent but more lucky works lead to success.
- Nevertheless, the skillful use of the methods of science will make progress highly probable. *Hence, the best practical strategy in promoting scientific progress is to support high-quality research.*"
(Niiniluoto, 2019, p. 6)
- Overall, very low interrater reliability on quality assessment of publications/grant proposals
  - But: Higher agreement on the low end of the scale (Cicchetti, 1991)

Responsible Research Assessment: A proposal for professorship hiring committees



https://coara.eu

Criteria of research performance

CoARA

„1. Recognise the **diversity of contributions** to research.

➞ „practices that contribute to robustness, openness, transparency,

➞ „Inappropriate uses of journal- and publication-based metrics in research assessment should be abandoned. In particular, this means **moving away from using metrics like the Journal Impact Factor** (JIF), Article Influence Score (AIS) and **h-index** as proxies for quality and impact.“

DFG Deutsche Forschungsgemeinschaft

„das ganze Spektrum wissenschaftlicher Publikationsformen gleichwertig in […] Lebensläufen abzubilden“

„Dies können beispielsweise Artikel auf Preprint-Servern, Datensätze oder Softwarepakete sein. “

„Angaben zu quantitativen Metriken wie Impact-Faktoren und h-Indizes im Lebenslauf oder Antrag werden nicht benötigt und sollen in der Begutachtung keine Berücksichtigung finden.“

Oransky, I., Marcus, A., & Abritis, A. (2023). How bibliometrics and school rankings reward unreliable science. BMJ, p1887. https://doi.org/10.1136/bmj.p1887; Information für die Wissenschaft Nr. 61 (2022): Maßnahmenpaket zum Wandel der wissenschaftlichen Bewertungskultur

RESQUE: The (social) process

2022

Ten Steps Towards better Personality Science

Quality rating scheme for publications

19 commentaries

2023

The Hong Kong Principles for assessing researchers: Fostering research integrity

CoARA

DORA

RRA 1 (Principles)

RRA 2 (Implementation)

Empfehlungen zur Bewertung wissenschaftlicher Leistungen bei Berufungsverfahren in der Psychologie

15 commentaries

6 commentaries

2024

- ✓ Meetings with Stakeholders:
- ✓ Workshops with 10 DGPs sections („Fachgruppen“)
- ✓ Workshops at conferences
- ✓ Talks at multiple universities, Tenure track Network, …

- ✓ Community-proofed
- ✓ Collaboratively improved with multiple breakout groups focusing on specific indicators (e.g., software, theory development, …)
- ✓ Development backed by DGPs
- ✓ Ready-to-use templates, adaptable to local needs
- ✓ Efficiency & reduced committee workload by automation

RESQUE Framework:

- 4 principles
- 2-phase assessment
- How-tos for hiring committees

Rating schemes:

- RESQUE-Publication
- RESQUE-Data (in progress)
- RESQUE-Software (in progress)

Tools:

- Interactive web form
- R-scripts for profile visualization

Rejoinder

2025

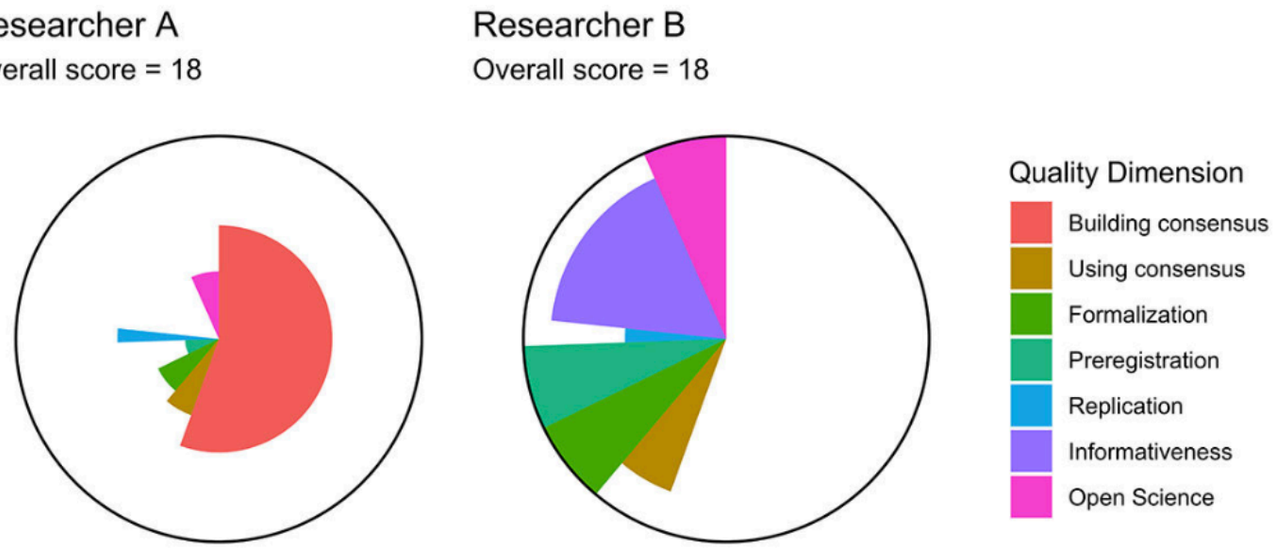
1. Expand the range of academic contributions

Types of academic contributions:

1. Research

Figure 1

Visualizing the Research Quality Profiles of two Researchers (A and B) who Promote Good Science in Different Ways, Through Their Respective Activities



Note. The width of each wedge is proportional to the maximum number of points that may be obtained in each category.

Figure from [Leising et al. (2022)]

2. Move from authorship to contributorship

Types of academic contributions:

1. Research

2. Teaching

3. Leadership
(e.g., mentoring, management and organizational skills, strategic thinking)

4. Service to the academic institution/ field

5. Societal impact
(e.g., science communication/ citizenship)

Contributor roles

Alternative: **Reorder** authors alphabetically in CV; only rely on CRediT

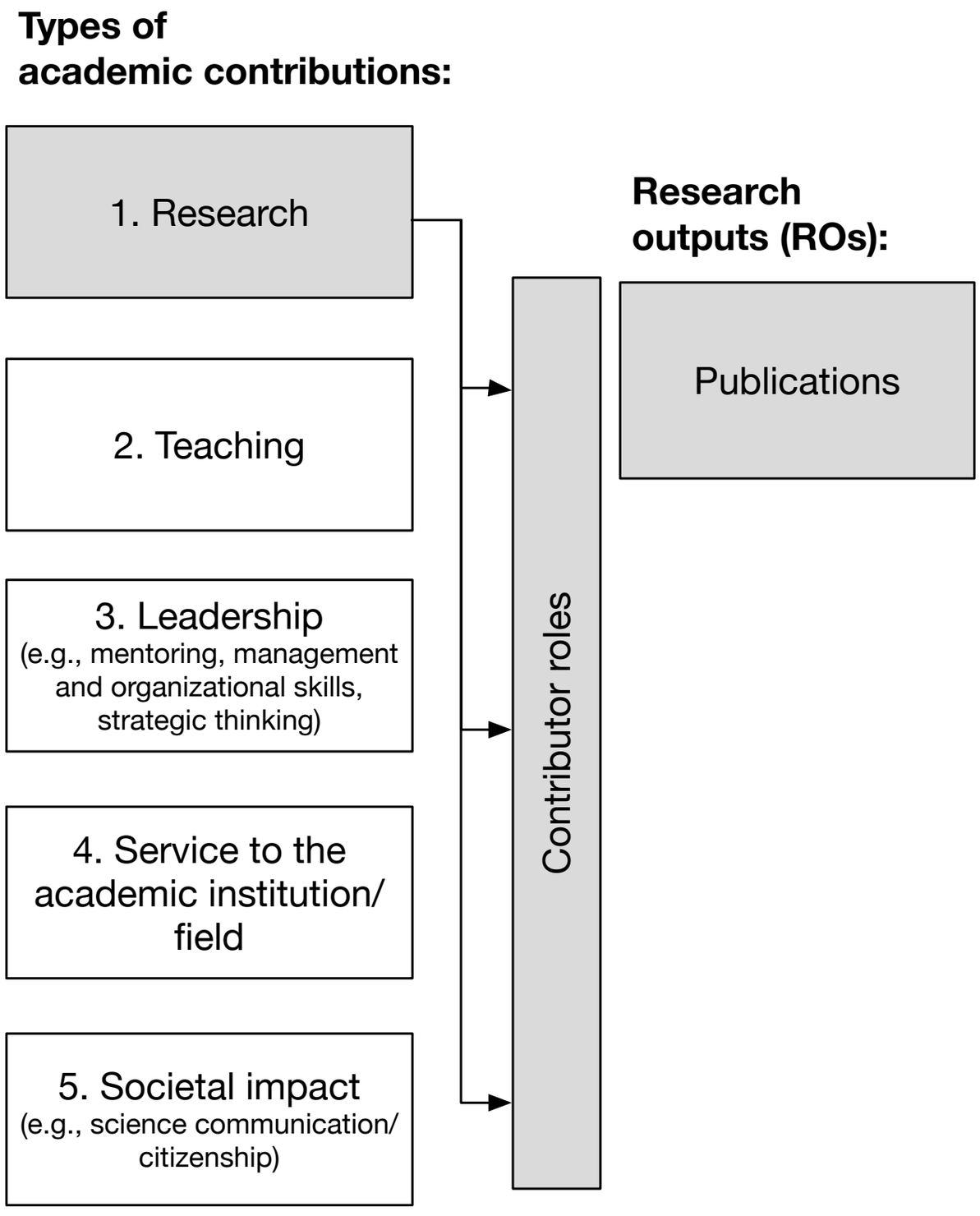
Reference	**CRediT** **L = lead,** **E = equal** **S = supporting**
Schönbrodt, F. D. & Wagenmakers, E.-J. (2018). Bayes Factor Design Analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review, 25*, 128-142. doi:10.3758/s13423-017-1230-y	Conceptualization (L) Formal analysis (L) Methodology (L) Software (L) Writing – original draft (L)
Zygar, C., Hagemeyer, B., Pusch, S., & **Schönbrodt, F. D.** (2018). From motive dispositions to states to outcomes: An intensive experience sampling study on communal motivational dynamics in couples. *European Journal of Personality, 32*, 306–324. doi:10.1002/per.2145	Conceptualization (E) Data curation (S) Formal analysis (S) Funding acq (L) Investigation (S) Supervision (L) Writing – original draft (S)
Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., … **Schönbrodt, F. D.**, ..., & Johnson, V. E. (2017). Redefine statistical significance. *Nature Human Behaviour, 2*, 6–10.	Writing – Review and editing (S)

JIF: 24!

Citations: 3005!

My contribution: Nearly zero.

3. More than publications: Data sets and software as fully-fledged contributions

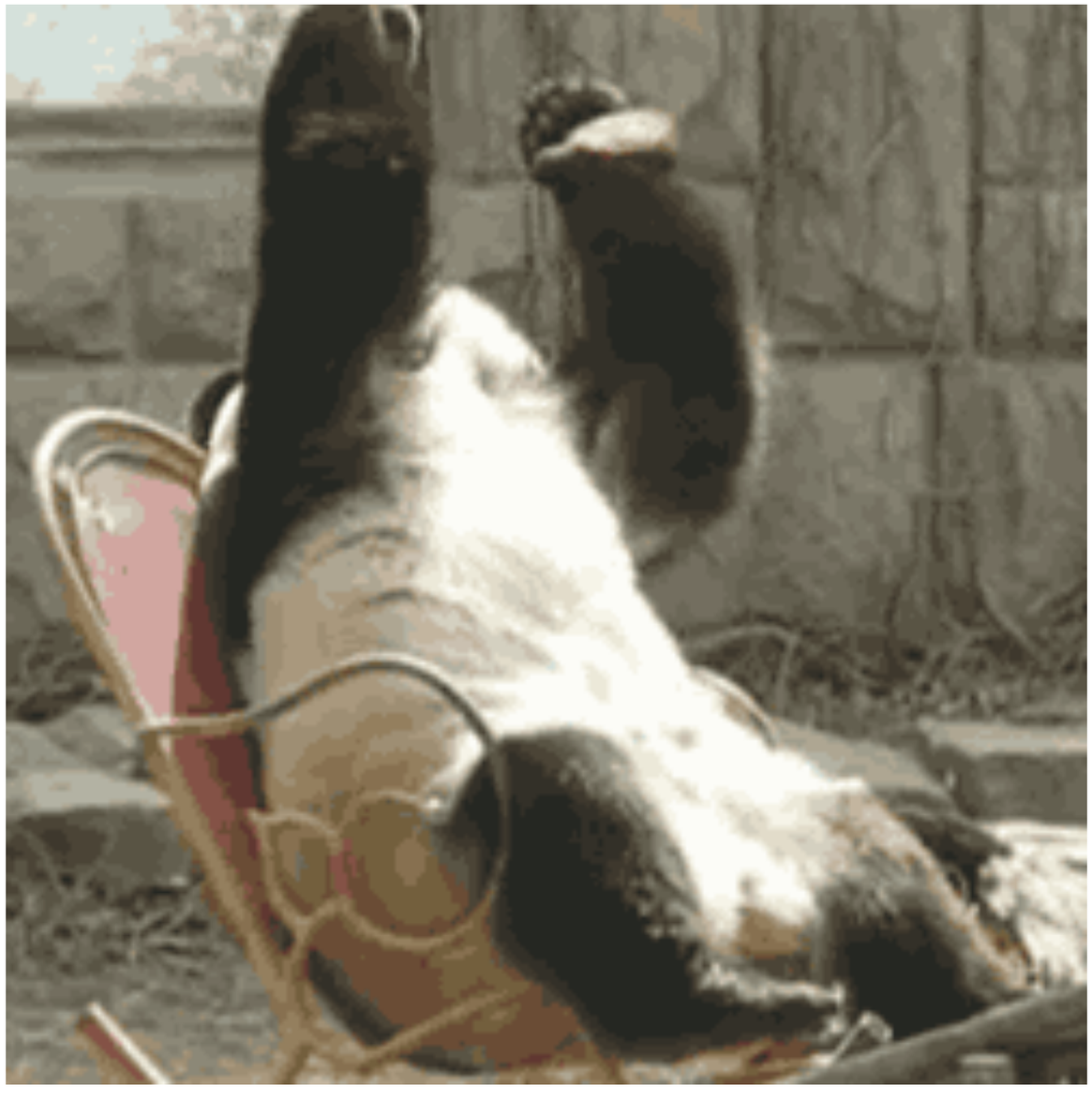


Quality over quantity

J.J. at the English language

4. Valid indicators for measuring quality (methodological rigor), impact, and quantity

In this work, we use the number of citations as a proxy for quality,



Quality over quantity

- „Quality“ is multidimensional:
  - basic aspects (methodological rigor)
  - elusive and complex aspects (innovation, creativity, ingenuity)
- Rigor - as one part of quality - can be measured (quite) objectively:
  Whether research has been **skillfully executed according to standards of good scientific practice within the field**.
- Quality cannot be *reduced* to rigor!
  - Rigor is *not a sufficient* condition for high-quality research –
    but it can be seen as a *necessary* condition for valid knowledge.

Research Quality Evaluation Scheme (RESQUE)

Evaluation dimension:

Research outputs (ROs):	Quality / Rigor	Impact	Quantity
Publications			
Data sets			
Research software			

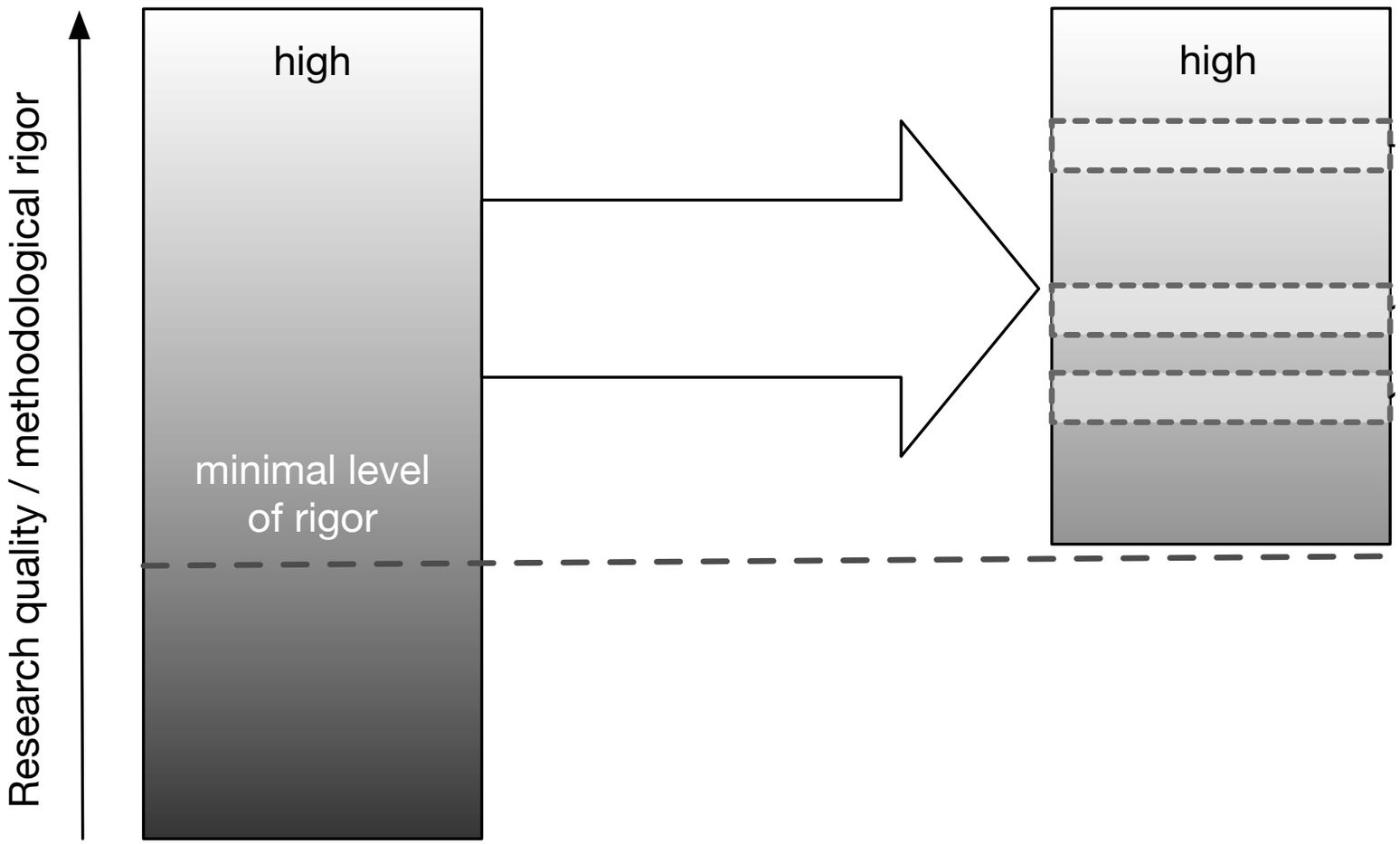
siehe Gärtner, A., Leising, D., & Schönbrodt, F. D. (2023). Empfehlungen zur Bewertung wissenschaftlicher Leistungen bei Berufungsverfahren in der Psychologie. *Psychologische Rundschau*, *74*(3), 166–174. https://doi.org/10.1026/0033-3042/a000630; sowie https://doi.org/10.31234/osf.io/rgh5b und https://doi.org/10.31234/osf.io/5yexm

5. A two-phase hiring process

Phase 1:
Negative selection with focus on *efficiency*: algorithm/indicator assisted

Longlist
(applicants who formally fit to the job description)

Shortlist
(candidates generally qualified for the job)



5. A two-phase hiring process

Phase 1:
Negative selection with focus on *efficiency*: algorithm/indicator assisted

Phase 2:
Positive selection with focus on *content*: in-depth qualitative evaluation and peer-review in committee

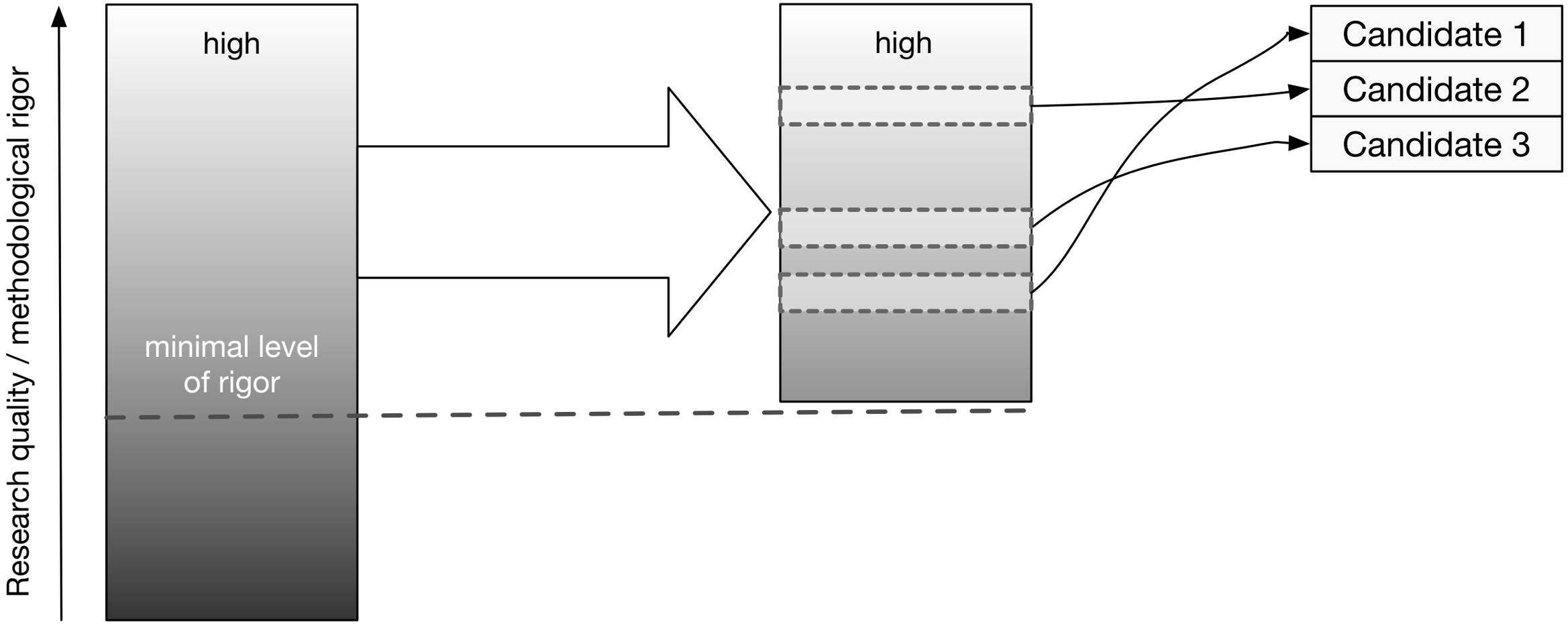
Longlist
(applicants who formally fit to the job description)

→

Shortlist
(candidates generally qualified for the job)

→

Final list with ranked candidates



Scoring-Principles in Phase 1

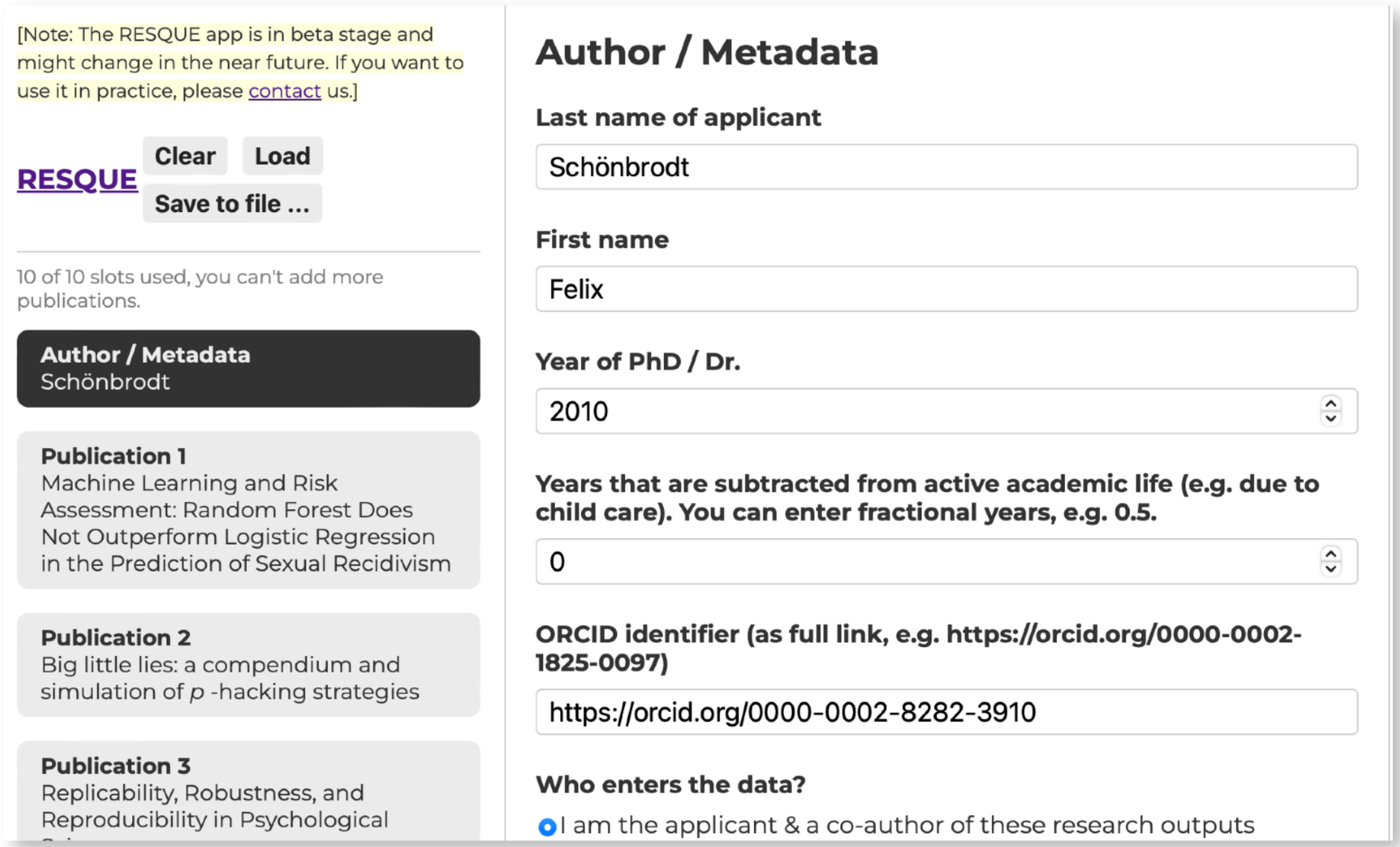
- Basic principles:
  - There are many ways to do science. But what you do, you should do well.
  - Example: You don't have to do theory-based work - but if you do it, it should be good.
  - In phase 1, we measure “fast and frugal indicators” that can be assessed as objectively as possible.
  - The focus is on basic *hygiene factors*. (We don't even try to measure innovation etc.).
  - You shouldn't be penalized if you can't get points in principle
- Relative Rigor Score
  - “POMP": Percentage of maximum points
  - If an indicator is justifiably “not applicable”, then the maximum score is reduced accordingly and you can still achieve 100%.

The RESearch QUality Evaluation Framework (RESQUE)

https://www.resque.info

Demo Collector-App:

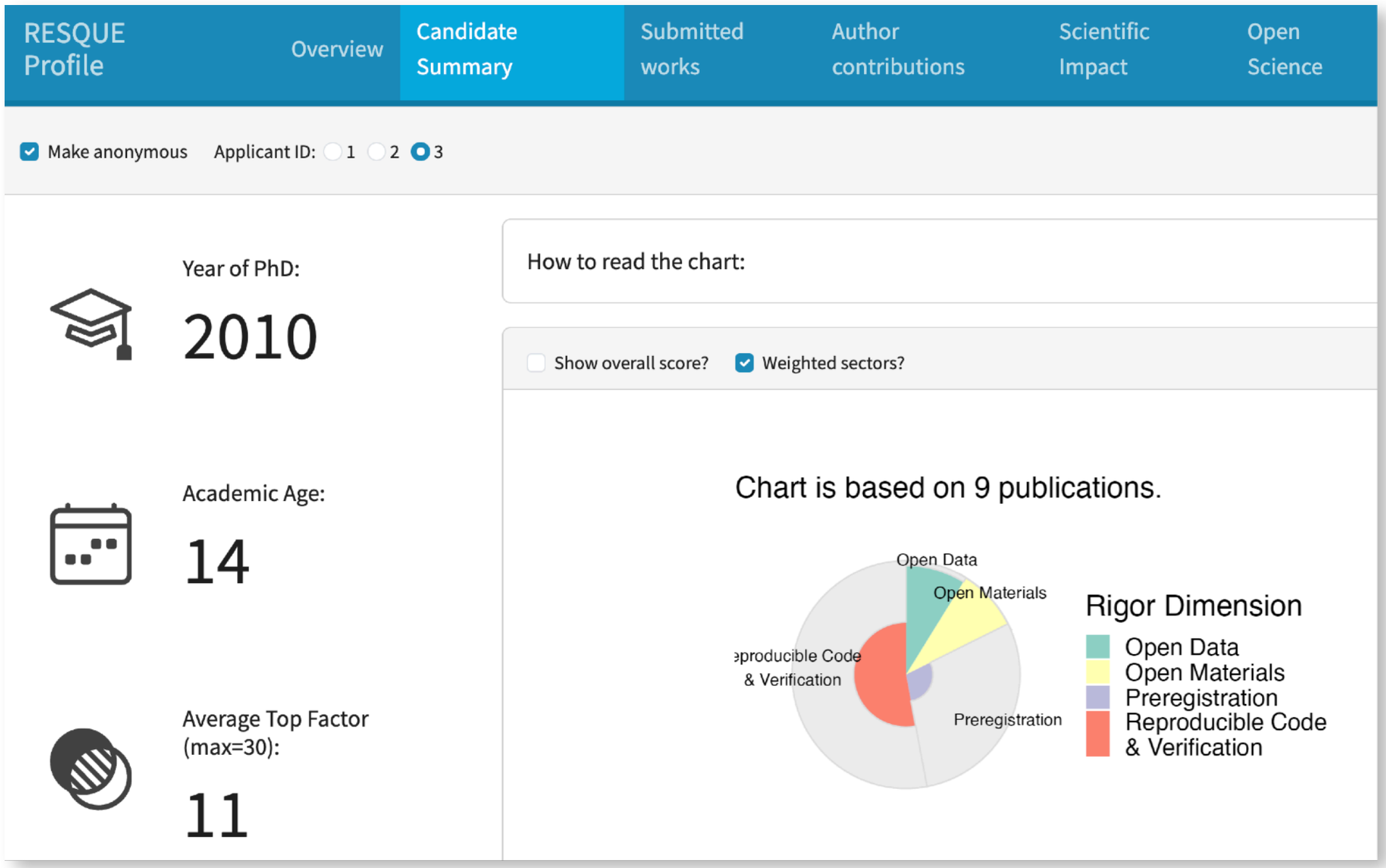
https://resque-framework.github.io/collector-app/



Demo Profile-App:

R-package **RESQUER**: https://github.com/RESQUE-Framework/RESQUER

Get your profile online (work in progress): https://shiny.psy.lmu.de/felix/RESQUE_profile/



CoARA Action Plans

Organisation	Organisation type	Country	Status of Submission	Date of joining CoARA	Due date of first action plan
Karlstad University	Universities and their associations	Sweden	10.5281/zenodo.11191147	30/05/2022	30/05/2023
Eurodoc	Academies, learned societies, and their associations, and associations of researchers	Europe	10.5281/zenodo.8407034	03/10/2022	03/10/2023
University of Rijeka	Universities and their associations	Croatia	10.5281/zenodo.10634416	03/10/2022	03/10/2023
Vilnius University	Universities and their associations	Lithuania	10.5281/zenodo.11612471	03/10/2022	03/10/2023
Autonomous University of Barcelona	Universities and their associations	Spain	10.5281/zenodo.11657234	03/10/2022	03/10/2023

Discussion: Implications for ECRs?

- Is a change in hiring criteria unfair?
- What are potential negative side-effects?
- Goodhart's law: How could you hack the new system?
- Strategy: Don't be the 0% open science person - show at least some experience in the key practices (FAIR open data, pre-registration, open code)

Speicher

Subtitle

Discussion

- What are potential negative side-effects?
- Goodhart's law: How could you hack the new system?
- Barriers for implementation: What would the chair of your next hiring committee say when you propose to switch to the new system?

Why is the uptake so slow?

1. No idea how to do it better
2. Too much effort
3. Restricting the academic freedom of committees?
4. Social dilemma: First movers have a disadvantage
5. Committee members maybe excelled on the old metrics (but not necessarily on the new ones?)
6. A sudden change in assessment criteria is unfair (after all, we spent years optimizing the old ones)

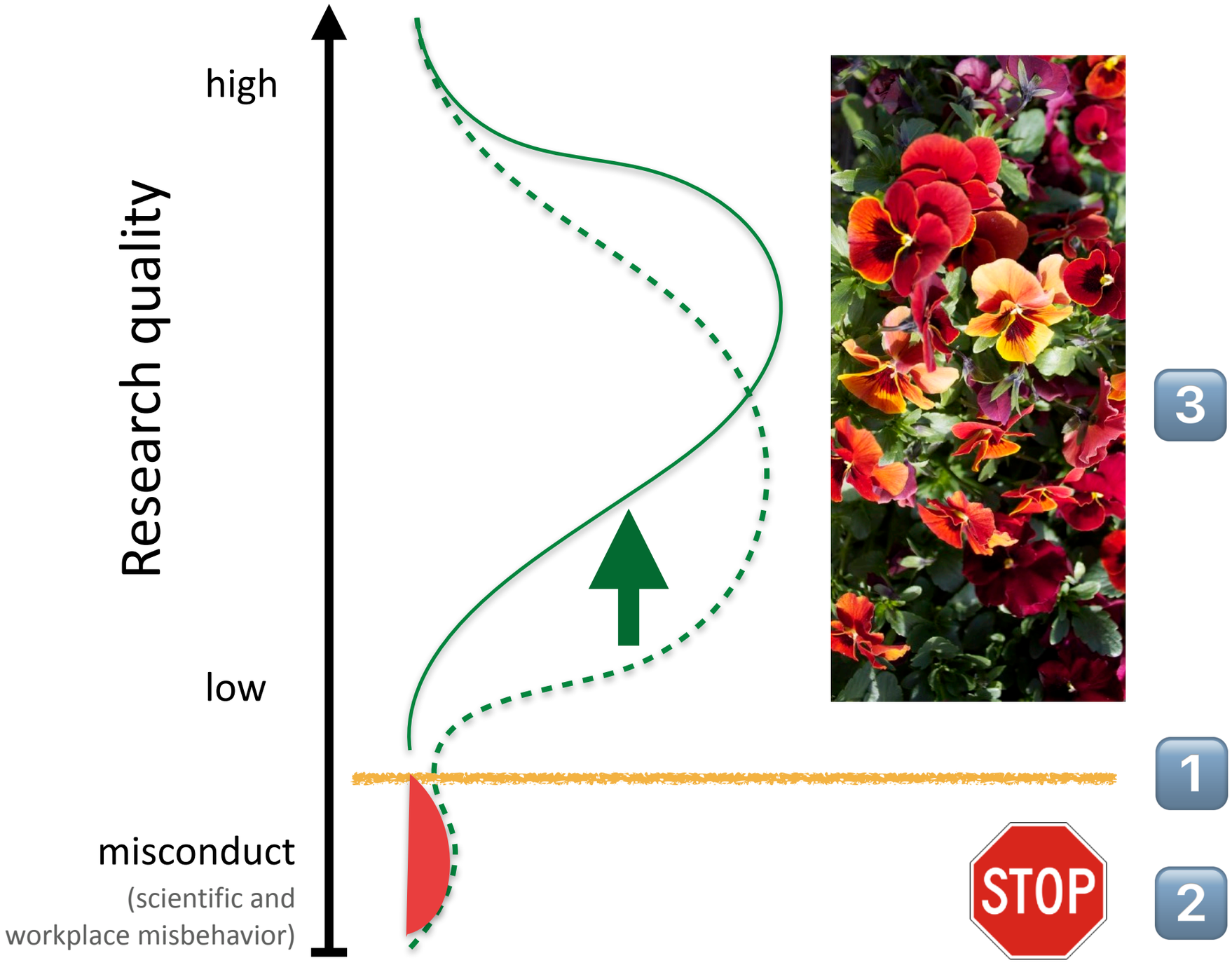
Scientific progress I

- When the goal is scientific progress, defined as achieving valid and credible knowledge, it is important to differentiate *progress* and *quality*:
- "**Quality** is primarily an activity-oriented concept, concerning the skill and competence in the performance of some task.
- **Progress** is a result-oriented concept, concerning the success of a product relative to some goal.
- All acceptable work in science has to fulfill certain standards of quality. But it seems that there are no necessary connections between quality and progress in science. Sometimes very well-qualified research projects fail to produce important new results, while less competent but more lucky works lead to success.
- *Nevertheless, the skillful use of the methods of science will make progress highly probable. Hence, the best practical strategy in promoting scientific progress is to support high-quality research.*" (Niiniluoto, 2019, p. 6).

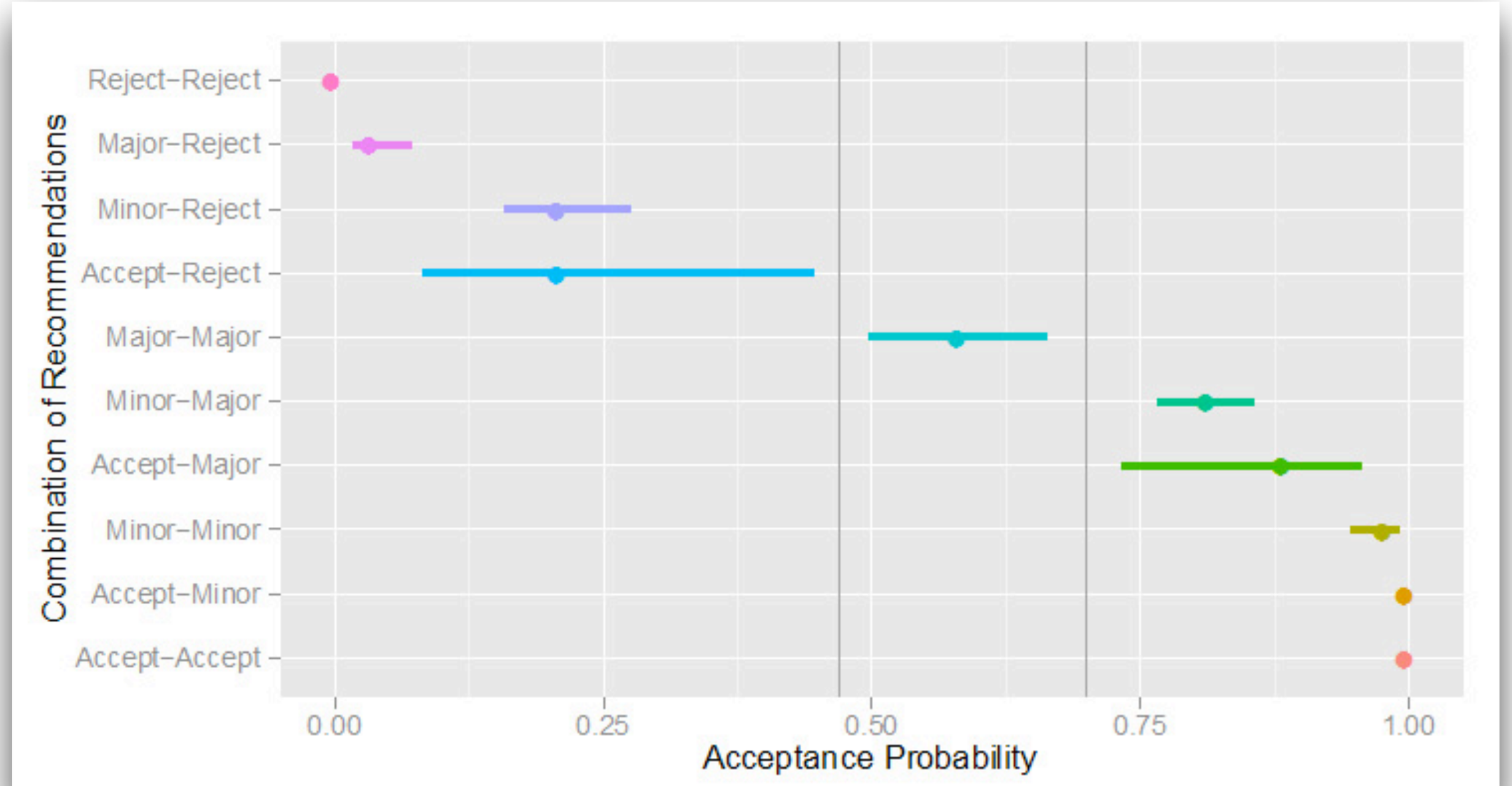
Scientific progress II

- Assumption 1: We will never be able to predict what research will be excellent, useful, or impactful (in the real world).
- Assumption 2: We know quite well what bad science is.
- Solution to foster scientific progress:
  - Weed out bad science
  - Support researchers to achieve high standards of methodological rigor
  - (See next slide)

1. Scientific fields should debate and find a consensus about the basic level of necessary good research practices („craftsmanship“).
2. These should be required, controlled and enforced by universities, funders, journals, supervisors.
3. Those who comply to this minimal standard should be free to thrive, with as few regulations and bureaucratic compliance as possible.



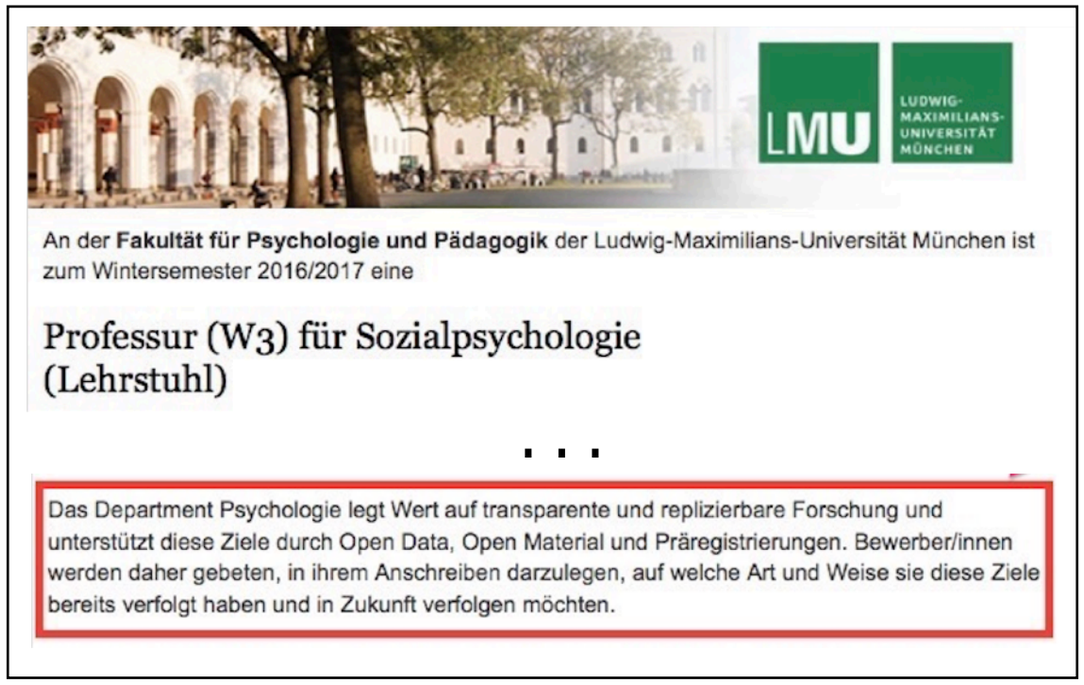
„But the reviewers do not decide about rejection and acceptance - the editor weighs several sources of information to reach an informed decision.“
➞ in the ideal case, yes. But in general, the decision is closely related to the reviewers assessment:



https://www.elsevier.com/connect/is-peer-review-just-a-crapshoot

Changing the incentive structure: Professorship hiring committees

Hiring committees: Make „open science“ a desirable or essential job characteristic



LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

An der **Fakultät für Psychologie und Pädagogik** der Ludwig-Maximilians-Universität München ist zum Wintersemester 2016/2017 eine

Professur (W3) für Sozialpsychologie
(Lehrstuhl)

. . .

Das Department Psychologie legt Wert auf transparente und replizierbare Forschung und unterstützt diese Ziele durch Open Data, Open Material und Präregistrierungen. Bewerber/innen werden daher gebeten, in ihrem Anschreiben darzulegen, auf welche Art und Weise sie diese Ziele bereits verfolgt haben und in Zukunft verfolgen möchten.

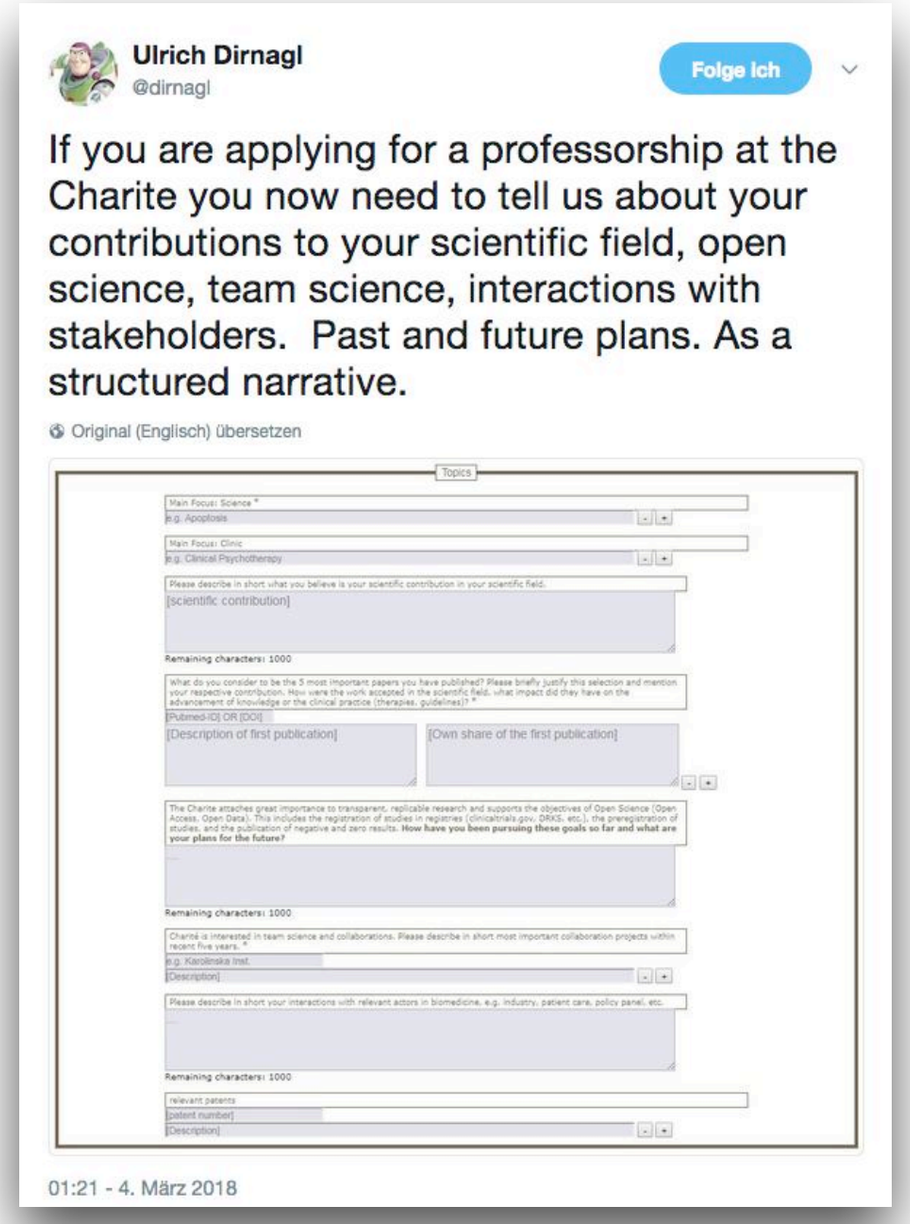
Since 2015: All professorship job descriptions use this requirement

Full Professor (W3)
of Social Psychology

to be filled as soon as possible.

. . .

The Department of Psychology aims for transparent and reproducible research (including Open Data, Open Materials, and Preregistrations). Applicants are asked to illustrate how they have pursued these goals in the past and/or how they plan to do so in the future.



Ulrich Dirnagl
@dirnagl

Folge ich

If you are applying for a professorship at the Charite you now need to tell us about your contributions to your scientific field, open science, team science, interactions with stakeholders. Past and future plans. As a structured narrative.

Original (Englisch) übersetzen

Topics

Main Focus: Science *
e.g. Apoptosis

Main Focus: Clinic
e.g. Clinical Psychotherapy

Please describe in short what you believe is your scientific contribution in your scientific field.
[scientific contribution]
Remaining characters: 1000

What do you consider to be the 5 most important papers you have published? Please briefly justify this selection and mention your respective contribution. How were the work accepted in the scientific field, what impact did they have on the advancement of knowledge or the clinical practice (therapies, guidelines)? *
[Pubmed-ID] OR [DOI]
[Description of first publication]
[Own share of the first publication]

The Charite attaches great importance to transparent, replicable research and supports the objectives of Open Science (Open Access, Open Data). This includes the registration of studies in registries (clinicaltrials.gov, DRKS, etc.), the preregistration of studies, and the publication of negative and zero results. **How have you been pursuing these goals so far and what are your plans for the future?**
Remaining characters: 1000

Charité is interested in team science and collaborations. Please describe in short most important collaboration projects within recent five years. *
e.g. Karolinska Inst.
[Description]

Please describe in short your interactions with relevant actors in biomedicine, e.g. industry, patient care, policy panel, etc.
Remaining characters: 1000

relevant patents
[patent number]
[Description]

01:21 - 4. März 2018

See more such prof job ads at:

https://docs.google.com/document/d/1ty43Syw0Flkh8ncjW8MZArIkvYe8hLwwhLIlwbtSk_Y/edit?usp=drive_web&ouid=108982640291853577145

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR PSYCHOLOGIE UND PÄDAGOGIK

Google™ benutzerdefinierte Su | www.lmu.de | Sitemap

FAKULTÄT
STUDIUM
FORSCHUNG
PSYCHOLOGIE
Studium und Lehre
Forschung
Open-Science-Committee
About our OSC
Recognizing Open Research Practices in Our Hiring Policy
Workshops and Talks
Lehr- und Forschungseinheiten
Ambulanzen und Testlab

drucken

Recognizing Open Research Practices in Our Hiring Policy

In December 2015, the Department Psychology of the LMU Munich added a paragraph to a professorship announcement which emphasized the department's commitment to responsible research and asked applicants to write a short statement about their open science practices:

"Our department embraces the values of open science and strives for replicable and reproducible research. For this goal we support transparent research with open data, open materials, and study pre-registration. Candidates are asked to describe in what way they already pursued and plan to pursue these goals."

Since then, all further professorship job advertisements of our department had this requirement.

In May 2018, the department's steering committee unanimously voted for an explicit policy to always include this (or a similar) statement to all future professorship job advertisements. It is the task of the appointment committee to value the existing open science activities as well as future commitments of applicants appropriately. By including this statement, our department aims to communicate core values of good scientific practice and to attract excellent researchers who aim for transparent and credible research.

http://www.fak11.lmu.de/dep_psychologie/osc/open-science-hiring-policy/index.html