

Peer Review

Werner Nutt

Who Decides What is Published?

The peers, by peer review

What is a peer?

Merriam Webster

- a member of the British nobility, i.e., a member of one of the five ranks (as duke, marquess, earl, viscount, or baron) of the British peerage



- a person who belongs to the same age group or social group as someone else

What is a peer?

Online etymology dictionary:

- c.1300, "an equal in rank or status" (early 13c. in Anglo-Latin), from Anglo-French *peir*, Old French *per* (10c.), from Latin *par* "equal" (see *par* (n.)).
- Sense of "a noble" (late 14c.) is from Charlemagne's Twelve Peers in the old romances, who, like the Arthurian knights of the Round Table, originally were so called because all were equal.
- Sociological sense of "one of the same age group or social set" is from 1944.
- **Peer review** attested by 1970.
- **Peer pressure** is first recorded 1971.

Peer review in academia

Peers decide about

- which **papers** are published (editorial peer review)
- which **project proposals** are funded (funding review)
- (*hiring as a researcher, professor, ...
the higher the rank, the more peerish the decision*)

How does editorial peer review work?

Two models

- Conference Reviewing
- Journal Reviewing

Conference Reviewing

Call for papers

PC is assembled [60 if 10/reviewer, 150 if 4/reviewer, 300 if 10/rev, 750 if 4/r]

Papers are submitted [200 - 1000]

Reviewers are assigned or choose papers [4 to 10]

Papers are evaluated by reviewers [during 6-8 weeks]

i.e., reviewers write reviews/referee reports + give a grade

Reviews are consolidated [3]

during a discussion, moderated by a senior PC member

Reviewers reach final judgment, if not, senior PC member decides whether to accept or not (and writes a meta review)

Variants:

- **Rebuttal phase:** authors are given the chance to respond to reviews
- **Possibility of revision:** acceptance conditional on modifications of the paper

Conference Reviewing: Characteristics

- Program committee (PC)
 - program **chairs** (appointed by organizing committee of the conference series)
 - PC chairs invite **area chairs** and other **PC members**
 - PC members may introduce **additional reviewers**
- Strict deadlines, **short time** for reviewing (4-6 weeks)
- Papers are **divided among PC**
- **5-10 papers** per PC member (batch reviewing)
- Decision (in most cases): **accept/reject**

Conference Reviewing: Timeline

Before the reviewing proper

- Program committee assembled
- Call for papers, includes topics, scope of conference
- Paper bidding
- Chairs assign paper

Reviewing proper

- Write individual review
- Discussion among reviewers

Potentially

- Rebuttal
- Revision

Final Decision

Journal Reviewing

- Editorial board
 - editor(s) in chief and editorial board members
 - one editor takes care of a submitted paper
 - editor asks ~3 experts for reviews
- Reviewer has 4-6 weeks time for a review (can be extended)
- Decision:
 - accept as is (rare)
 - accept with minor revision (also rare)
 - revision and resubmission
 - reject

Blind vs. Double-blind Reviewing

Blind vs. double-blind reviewing

- reviewers are unknown: *blind* or *single-blind*
- authors are unknown, too: double blind

Double-blind reviewing means: authors must make any reasonable effort to anonymize the paper. i.e.,

- no author names
- no reference to own work

instead of, “we have shown in [RN11]”,
say “Razniewski and Nutt have shown in [RN11]”

R. Snodgrass has summarized research on single vs. double-blind reviewing. “Single-Versus Double-Blind Reviewing: An Analysis of the Literature”, SIGMOD Record, 2006

Paper bidding

PC members bid for paper assignment, based on

- title
- abstract
- *(keywords, areas)*
- *(authors and their affiliation)*

PC members express degrees of preference for paper:

- e.g., high, medium, low in EasyChair
- assignment based on preferences and possibly area match

Reviewing proper

PC members

- produce a report
 - write report themselves, or
 - give the task to an auxiliary reviewer (colleague, PhD student, postdoc)
- summarize the report in a (provisional) judgment with grades (e.g., strong accept, accept, weak accept, borderline, weak reject, reject)
- view other reports after having submitted their own
- discuss their reviews and possibly adjust their judgment and review
 - if asked by PC chairs or area chairs (= meta-reviewers) to do so in case the judgments vary (significantly)
- PC/area chairs finally decide about acceptance or rejection

Rebuttal phase

May be inserted between

- first judgment of reviewers and
- final judgment of reviewers
- Authors respond to reviews
 - in short period, e.g., 4 days
 - with limited space, e.g., 500 words
- Reviewers may revise their reviews (and grades) based on other reviews and author rebuttal
- Meta-reviewers report to PC chairs
- PC chairs decide

Revision Phase

Recently introduced in some conferences (VLDB, SIGMOD, ICDE)
(not together with rebuttal)

- Reviewers not only decide about acceptance/rejection, but may also request a revision
- PC chairs/meta-reviewers decide about acceptance, rejection, or revision
- Authors have limited period (approx. 4 weeks) to revise the paper according to comments of reviewers
- Reviewers comment on the outcome of the revision and discuss
- PC chairs/meta-reviewers come up with final decision

Structure of a review

- Summary of the paper
- Character of the work, relevance for outlet
- Contribution of the paper, relationship to state of the art
 - originality
 - depth
- Strengths
- Weaknesses
- Writing
- Explanation of judgment
- Detailed comments

See also Allen S. Lee “Reviewing a Manuscript for Publication”
(<http://www.people.vcu.edu/~aslee/referee.htm>)

Review Structure CIKM, DB Track

- Overall Rating
 - Strong Accept, Accept, Neutral, Reject, Strong Reject
- Top 3 Strengths
- Top 3 Weaknesses
- Detailed Comments
- Author feedback needed?
- What specific feedback do you like the authors to provide
- Comments to Program Chair

Review Structure IJCAI

- **Summary:** Describe the paper in 2-3 sentences
- **Relevance:** Is the work relevant to AI?
- **Originality:** Does the paper clearly point out differences from related research? Are the problems or approaches new?
- **Significance:** Is the work important? Does the paper make a valuable contribution to knowledge and understanding in the AI area? Does it advance the state of the art? Does the paper add to our understanding of some aspect of agent systems? Does the paper stimulate discussion of important issues or alternative points of view? Does the paper carefully evaluate the strengths and limitations of its contributions, and draw lessons for future work?
- **Technical Quality:** Is there a careful evaluation of the proposed method and the results? Is the paper technically sound, with compelling arguments?
- **Readability and Organization:** Is the paper clearly written? Does the paper motivate the research? Are results clearly described and evaluated? Is the paper well organized?

Review Structure VLDB

- Overall Recommendation: Accept, Revise, Reject
- Summary: what is proposed? in what context? brief justification of recommendation
- 3 strong points
- 3 weak points
- Relevance
- Novelty
- Significance
- Technical Depth, quality of content
- Presentation
- Would you champion acceptance?
- Detailed comments

Let's look at examples

- CAiSE 2017 (Int'l. Conf. on Advanced Inf. Systems Engineering)
 - 3 reviews, rejection
- BPM 2018 (Int'l. Conf. on Business Process Management)
 - 3 reviews, acceptance
- ISWC 2016
 - 3 reviews
 - rebuttal
 - rejection
- SIGMOD 2015
 - 3 reviews
 - revision
 - acceptance

We organize BZMC 2019

... the 6th UNIBZ PhD Student Mock Conference 2019

with EasyChair

Review Structure for BZMC 19

Summary

Comment

Grade

Deadline for reviews: Thu, 8am

Review structure for BZMC 18

- Summary
- 3 Strong Points
- 3 Weak Points
- Presentation and Clarity
- Comments (Significance, Thoroughness, Originality)

- Scale of 1-5 for marks for Relevance, Thoroughness, Originality

Conference Management Systems

- Conference management has typical workflows
- Conference management systems allow organizers
 - to define the specific workflow of for their conference
 - invite a PC
 - receive submissions
 - organize paper bidding and allocation
 - collect of reviews
 - organize discussions
 - let participants of the reviewing process communicate
 - etc.
- Frequently used
 - EasyChair (<http://www.easychair.org>)
 - Microsoft CMT (<https://cmt.research.microsoft.com/cmt/default.html>)
(see also <https://cmt.research.microsoft.com/cmt/userroles.html>)

Dynamics of reviewer discussions

- Who is likely to start?
 - The reviewer with the positive review?
 - Or the one with a negative review?

*Hint: Which reviews are usually longer?
The positive or the negative ones?*
- What is the most probable outcome of a discussion and an adjustment?

The NIPS Experiment

- The Conference on Neural Information Processing Systems (NIPS) is a major Machine Learning Conference
- NIPS 2014 ran an experiment
 - The program committee was split into two equal halves
 - 10% of the submissions (= 166 papers) were reviewed by both halves
 - The two committees disagreed on 43 papers
- Analysis
 - NIPS 2014 had an acceptance rate of 22.5%, which corresponds to 37 papers to accept
 - Disagreement on 43 papers means:
 - Half 1 accepted 21 papers rejected by Half 2
 - Half 2 accepted 22 papers rejected by Half 1
 - This is a disagreement in 22/37 papers, i.e, in 59% of cases

(see <http://blog.mrtz.org/2014/12/15/the-nips-experiment.html>)

Journal Reviewing

In my experience, typically, single-blind

Differences with conference reviewing

- one paper, no batch
- possibility to have a look at the paper before agreeing to review
- possibility to decline a request
- wider range of judgments: accept, minor revision, revision, reject
- interaction with authors: authors explain how they took the reviewers' comments into account for the revision
- iteration: up to 2 revisions possible
- no discussion among reviewers,
co-reviewers do not reveal their identity

Journal Reviewing: Example

- VLDB Journal 2015
 - 3 reviews
 - revision
 - 3 reviews
 - revision
 - acceptance

Blind vs. double-blind

- Historically, **completely open** review:
 - author submits to journal editor (handwritten manuscript)
 - editor decides.
- Then: blind review

Ongoing debate: blind vs. double-blind

- Pros:
- Cons:

1991: Blank's Study

- Landmark Paper
 - *American Economic Review* (AER) is a top-tier economics journal
 - *American Economic Association's Committee on the Status of Women in the Economics Profession* expressed concern that women may suffer negative effects from blind reviewing
- Editors asked M. Blank to devise a study (1987-89)
 - 50% of papers were reviewed blind, 50% double blind
- Results: Under a double-blind system
 - acceptance rates are lower and referee reports are more critical
 - no change for authors from top-tier and bottom-tier institutions, but those from almost-top suffer
 - women fare slightly better, but this is statistically insignificant
 - female referees gave lower ratings to nonblind papers than do men and tend to give higher ratings to blind papers, while male referees show the opposite pattern.

Rebecca M. Blank

The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from
The American Economic Review

Source: The American Economic Review, Vol. 81, No. 5 (Dec., 1991), pp. 1041-1067

Other Experiences

MLA (Modern Language Association) had conference reviewing

- single-blind until 1973
- starting from 1974, double-blind:
 - #papers of women and newcomers doubled
 - in 1975, the same
 - by 1978, women, newcomers and established male researchers had equal acceptance rates
- Consequence: since 1979, double blind

WSDM'17 (Conf. Web Search and Data Mining) split PC into 2 halves

- one half could see author names, the author half not (2 reviewers from each half per paper)
- Outcomes: bias in favour of famous companies, universities, and researchers, statistically insignificant bias against female authors

Andrew Tomkinsa, Min Zhang, and William D. Heavlin
Reviewer bias in single- versus double-blind peer review
PNAS | November 28, 2017 | vol. 114 | no. 48

Consequences at AER

- In 1991, move to double-blind reviewing.
- In 2011, moved back to blind reviewing:
 - author identity can often be found out by search engines
 - if authors are known, conflicts of interest can more easily be identified
- “In the age of Google, double-blind has become a fiction”
(Jonathan Katz, Caltech)
- Researchers publicise their work by
 - preprints
 - talks
 - websites

Can one Recognize a Good Journal ?

Can one Recognize a Good Journal ?

Among the possible answers

- impact factor (related to the number of citations, computed and published by Thomson Reuters)
- editors (scientists responsible for a journal)
- publisher
- CORE ranking (<http://www.core.edu.au>)
- Scimago ranking (<https://www.scimagojr.com>)

... and What About Bad Journals?

... and What About Bad Journals?

Possible answer:

- Bealls List (<https://beallslist.weebly.com>)

Jeffrey Beall was librarian at Denver, Colorado

He wrote a blog about predatory open access journals, i.e.,

- journals that cost nothing to the reader (open access)
- but where authors pay for publishing their work
- where reviewing is sloppy to non-existent
- publishers earn money by publishing shoddy research for authors who need their work be published

He took his web page down after being put under pressure by OMICS Publishing Group (whom Beall accused of being predatory)

Stories about Predatory Journals

- *“Fake Science”*

German TV broadcast on effects of faulty science distributed via predatory journals (English subtitles)

<https://www.daserste.de/information/reportage-dokumentation/dokus/videos/exklusiv-im-ersten-fake-science-die-luegenmacher-englische-version-video-100.html>

- *“I fooled millions into thinking chocolate helps weight loss. Here’s how.”*

Report on the press response to a journal paper with fake science

<http://wisconsinailyindependent.com/i-fooled-millions-into-thinking-chocolate-helps-weight-loss-heres-how/>

- *“Betrug statt Spitzenforschung - Wenn Wissenschaftler schummeln”* (Fraud in Place of Top Level Research)

German investigative TV report publishes papers in OMICS journals, showing that eating Chia increases intelligence

<https://www1.wdr.de/mediathek/video/sendungen/quarks-und-co/video-betrug-statt-spitzenforschung---wenn-wissenschaftler-schummeln-100.html>

Open reviewing models

- Submissions and reviews are visible to the world, reviewers are anonymous
 - e.g., peerj.com (<https://peerj.com>) (in life sciences and psychology)
Example paper by Daniel Graziotin et al.
(<https://peerj.com/articles/289/>)
- Submissions and reviews are visible to the world, reviewers identify themselves (but may remain anonymous), other researchers may contribute reviews, all contributing reviewers are acknowledged in the end
 - e.g., [Semantic Web Journal](#)

Peer Review for Research Funding

Funding bodies/agencies base decisions on peer reviews, e.g.,

- South Tyrol:
 - internal projects of UNIBZ, evaluated by Univ. Research Committee 1mio/yr
 - calls by province for research and innovation projects ~2-3 mio/yr
- Italy:
 - PRIN ~390 mio EUR in 2018, no funding in 2016/17, 100 mio/yr before
 - FIRB (no funding in 2016/17)
- Austria:
 - FWF (= Fonds zur Förderung der wissenschaftlichen Forschung) 180 mio/yr
 - also FFG, AWS, ...
- Germany:
 - DFG (= Deutsche Forschungsgemeinschaft) 3.2 billion EUR/year
32,000 projects funded in 2017, 22,500 reviews/yr
 - BMBF (= Bundesministerium für Bildung und Forschung)

Peer Review for Research Funding/2

Funding bodies/agencies base decisions on peer reviews, e.g.,

- UK: Research Councils
 - EPSRC (= Engineering and Physical Sciences Res. C.) 900 mio GBP/yr
2000 proposals/yr
 - BBSRC (= Biotechnology and Biological Sciences Res. C.) 450 mio GBP/yr
 - all councils: combined budget 3.5 bio GBP/yr, out which 1 bio for grants
- France:
 - ANR (Agence Nationale de la Recherche) 1 billion EUR
- Belgium:
 - FSO (= Fonds Wetenschappelijk Onderzoek) in Flanders, 200 mio EUR/yr
 - FNRS (= Fonds de la Recherche Scientifique) in Wallonie
- USA:
 - NSF, DARPA, NIH, ...
- EU:

Structure of a Grant Proposal

- Problem to be solved, background, related work
- Significance
- Research outcomes, e.g., theorems, prototypes, emp. study
- Impact on research community, economy, society
- Approach, work programme
 - Work packages: tasks, deliverable, milestones responsibilities, how measure success
- Resources: how much money for what?
- Investigators (PI, col) team: competence, track record
- Management: coordination, risk management

Workflow

- Applicant submits project proposal
- Agency nominates reviewers
 - proposed by staff
 - approved by academic on boards
- Panel of researchers ranks proposals
 - 1 rapporteur for proposal reports on reviews
 - top ranked proposals are funded

Funding Rates

Percentage of proposals accepted

- UNIBZ internal projects: ~30%
- Province: ~30%
- DFG: ~30% of proposals, ~38% of funds requested
- FWF: ~30%
- ANR: ~25%
- EPSRC: ~34% (highest among all research councils)

Review Criteria EPSRC

- Quality: Comment on
 - The novelty, relationship to the context, and timeliness
 - The ambition, adventure, and transformative aspects identified
 - The appropriateness of the proposed methodology.
- National importance
- Impact: Comment on
 - the pathway to impact identified for this work
 - Applicant
- Applicant
 - Comment on the applicant's ability to deliver the proposed project
- Resources and Management
- Overall Assessment

EPSRC Assessment Grades

- 1) This proposal is scientifically or technically flawed
- 2) This proposal does not meet one or more of the assessment criteria
- 3) This proposal meets all assessment criteria but with clear weaknesses
- 4) This is a good proposal that meets all assessment criteria but with minor weaknesses
- 5) This is a strong proposal that broadly meets all assessment criteria
- 6) This is a very strong proposal that fully meets all assessment criteria

ANR Criteria

- Technical and scientific quality
- Methodology, quality of project construction and coordination
- Overall impact of the project
- Quality of the consortium or of the team association
- Appropriateness of project resources/Project feasibility
- General opinion
 - Strong points
 - Weak points
 - Recommendations

Horizon 2020 (EU) Criteria

- Excellence
 - Clarity and pertinence of the objectives;
 - Credibility of the proposed approach;
 - Soundness of the concept, including trans-disciplinary considerations, where relevant;
 - Extent that proposed work is ambitious, has innovation potential, and is beyond the state of the art (e.g. ground-breaking objectives, novel concepts and approaches).
- Impact
 - The expected impacts listed in the work programme under the relevant topic;
 - ...
- Quality and efficiency of the implementation
 - Coherence and effectiveness of the work plan, including appropriateness of the allocation of tasks and resources;

Horizon 2020 Scores (for each criterion)

- **0** — The **proposal fails to address the criterion** or cannot be assessed due to missing or incomplete information.
- **1** — **Poor**. The criterion is inadequately addressed, or there are serious inherent weaknesses.
- **2** — **Fair**. The proposal broadly addresses the criterion, but there are significant weaknesses.
- **3** — **Good**. The proposal addresses the criterion well, but a number of shortcomings are present.
- **4** — **Very Good**. The proposal addresses the criterion very well, but a small number of shortcomings are present.
- **5** — **Excellent**. The proposal successfully addresses all relevant aspects of the criterion. Any shortcomings are minor.

How would you proceed when writing a review?

- Discussion

How can one's writing influence the decision about acceptance/rejection?

- Discussion