

Report on the ESO workshop

Peer Review Under Review

held at ESO Headquarters, Garching, Germany and online, 6–10 February 2023

Tereza Jerabkova¹
 Francesca Primas¹
 Ferdinando Patat¹
 Annalisa De Cia¹
 Wolfgang Kerzendorf^{2,3}
 Lou Strolger⁴
 Markus Kissler-Patig⁵
 John Carpenter⁶
 Andrea Bolamperti¹
 Luca Cacciapuoti¹
 Giulia Ceci Faress¹
 Marta De Simone¹
 Elisabeth Hoppe¹
 Dominika Itrich¹
 Alzbeta Oplistolova¹
 Amanda C. Rubio¹
 Denisa Tako¹
 Claudia Toci¹



Figure 1. The workshop logo.

Inclusion; and Concrete Examples. Ample time was included for exchanges of ideas. This unique blend of participants allowed us to examine PR from various perspectives, learning from the experiences and methods of other fields, and hearing from researchers who study the PR process itself.

In the following sections we provide a short summary and highlights of the individual workshop sessions.

Peer Review at Large

The workshop session Peer Review at Large hosted an array of speakers who discussed the evolving landscape of PR in scientific research. The presentations highlighted the importance of adapting traditional systems to meet the challenges and demands of modern-day research. Several core themes emerged throughout the session, including the importance of diversity, transparency, fairness, efficiency, and evolution in the PR process.

The efficiency and evolution of the PR process were prominent aspects of this section. Tracey Weissgerber introduced an innovative project, ScreenIT, which employs automated screening tools for the initial evaluation of scientific papers, promoting efficiency and accuracy. Meanwhile, Mario Malički discussed developments in PR, including plagiarism checking, language software and semi-automated/AI checks, which all indicate a significant trend towards digitisation and automation.

Ludo Waltman addressed the mounting strain on the traditional PR system, suggesting alternatives such as preprinting and open PR, aimed at fostering genuine scientific conversations. Johanna Schnier, Christina Raasch, and Ferdinando Patat proposed re-evaluating resource allocation strategies, suggesting a two-step review procedure and dedicated funding schemes for resource-intensive proposals to avoid biases against such programmes.

This emphasis on diversity and inclusivity was echoed by Vicente Amado Olivo and Wolfgang Kerzendorf, who proposed a

Introduction to peer review

Peer review (PR) is a cornerstone of academic knowledge production and dissemination, maintaining scientific rigour and quality by scrutinising research before publication. Despite its critical role, the PR process in its current form has been under scrutiny. The system, primarily established in the 18th and 19th centuries, is perceived by many to have lagged behind the rapid expansion and specialisation of the scientific community. With a 15% increase in the number of researchers between 2014 and 2018 and a two-fold increase in publications in the field of astronomy every 14 years, the strain on PR is evident.

There is an increasing call to adapt and innovate the PR process in sync with technological advancements and the surge in publication dissemination, as the current system is susceptible to stagnation and bias. Against this backdrop, ESO hosted a workshop entitled Peer Review Under Review, aiming to create a forum in which to review the current implementations of PR and to discuss its future in a digital and interconnected science community. The workshop was attended by representatives of a wide range of organisations, including ESO, ESA, the Joint ALMA Observatory, STScI, NASA, SKAO, and NOIRLab, and a significant group of non-astronomer experts in PR. The discussions were divided into four main sessions, focusing on: Peer Review at Large; Methodologies; Diversity, Equity and

The workshop Peer Review Under Review, held at ESO Headquarters in Garching, Germany from 6 to 10 February 2023, marked a significant milestone, being the first conference focusing on peer review within the astronomical community. This unique gathering not only convened representatives from many of the major astronomical organisations but also drew experts from such diverse fields as computer science, social sciences, statistics, meta-research, and other relevant domains. This unique group of experts critically examined the current state of peer review in the scientific community. This report summarises the presentations and discussions, and the conclusions that emerged during the workshop.

¹ ESO

² Department of Physics and Astronomy, Michigan State University, USA

³ Department of Computational Mathematics, Science, and Engineering, Michigan State University, USA

⁴ Space Telescope Science Institute, Baltimore, USA

⁵ European Space Agency (ESA), European Space Astronomy Centre (ESAC), Madrid, Spain

⁶ Joint ALMA Observatory, Santiago, Chile

global registry for peer reviewers in astrophysics. Their goal is to broaden the scope of reviewers and counteract issues of author name ambiguity through innovative use of algorithms and ORCID identifiers. This initiative resonates with Cornelia Schendzielorz’s and Martin Reinhart’s idea of democratising the PR process by enlarging and diversifying reviewer pools and ensuring more qualitative deliberation in the assessments.

Bias in PR was another central theme. Valentina Tartari, Hans Christian Kongsted, and Maryann Feldman highlighted biases in the allocation of scientific research funding. They suggested a more nuanced approach that takes into account factors such as career implications, visibility, and scientists’ past performance. Ferdinando Patat also tackled bias, presenting a statistical analysis of the proposal ranking process at ESO. He revealed the systematic effects that could be introduced in panel meetings. These help to minimise differences in reviewers’ opinions, leading to greater agreement among them, although this doesn’t necessarily mean the final evaluations are more accurate. Furthermore, ESO data indicate that agreement between reviewers is limited, suggesting the need for a broader statistical foundation to achieve more dependable evaluations.

Thierry Forveille provided a behind-the-scenes look at the PR process at *Astronomy & Astrophysics*, underlining the importance of selecting appropriate referees and ensuring swift responses. Similarly, Stefan Immler presented an inside perspective on NASA’s commitment to diversity, inclusion and equity in their PR processes, detailing the upcoming plans which include the introduction of Inclusion Plans for proposals and bias training for all peer reviewers.

Finally, Elena Erosheva focused on challenges associated with using numeric scores in PR settings. She proposed a combined approach of scores and rankings to enhance the accuracy and fairness of evaluations, providing a more comprehensive representation of quality estimation.

In conclusion, the presentations highlighted the need to re-evaluate and adapt

PR practices to ensure they remain fit for the purpose in the modern research landscape. They pointed towards a future in which PR is more inclusive, diversified, fair, efficient and transparent, leveraging advancements in technology to improve its efficiency and effectiveness.

Methodologies

In a riveting discussion on managing the surge in submissions, Nihar Shah showcased an automated reviewer assignment system. The system crafts a reviewer pool, gauges similarity scores for each proposal-reviewer pair, and then allocates papers based on these metrics. Nihar took the participants through the computation intricacies, from topic intersections to text matching and even bidding, culminating with the introduction of the Peer-Review4All algorithm which boasts effectiveness across varied evaluations.

Drawing attention to the potential pitfalls of the PR process, Rachel Heyard expressed concerns about the reliability of funding allocation through this method. She underscored the biases and inherent uncertainties that often mar the process, suggesting that attempts to train reviewers might fall short. Based on data from the Swiss National Science Foundation, Rachel made a compelling case for the adoption of lottery systems and continuous funding mechanisms, offering a fresh perspective that acknowledges the system’s inherent unpredictability.

On the topic of AI’s role in assessing journal articles, Mike Thelwall presented a deep dive, specifically examining the UK Research Excellence Framework (REF) 2021 and a selection of physics articles. Though AI displayed an accuracy range of 65–75% in its predictions based on a myriad of metrics, it occasionally fell short when matched against human judgment. However, the silver lining was AI’s knack for yielding unbiased results in certain spheres. While the experiment provided valuable insights, it’s important to note that it didn’t compare the deviation between AI and human evaluations with that between different human panels. Hence, one cannot conclusively determine whether the AI’s assessment deviated more significantly than what might

be observed between two human panels reviewing the same proposals. This aspect remains a crucial consideration for interpreting the results, as pointed out earlier by Ferdinando Patat.

Returning to the spotlight, Nihar Shah shared insights on the burgeoning field of Distributed Peer Review (DPR) in computer science, notably its profound impact in the machine learning sector. He delineated the clear merits of DPR but didn’t shy away from mentioning challenges like ‘commensuration bias’ and ‘miscalibration’. Nihar’s analysis, highlighting potential fraud risks in DPR, also applies to classical panels, emphasising the universal importance of implementing preventive measures to ensure the integrity, soundness and quality assurance of each review process.

Continuing this enlightening session, Fabio Sogni and Dario Dorigo described the strides they had made towards enhancing the proposal submission and review system at ESO. By the close of 2021, ESO had seamlessly integrated the ground-breaking p1Flow project, ushering in transformative components like proposalDistributor and proposalManager. Their innovative algorithm, tailored for superior proposal-reviewer congruence based on scientific relevance, marked a significant milestone in PR at ESO. Looking ahead, the successful deployment of this algorithm in DPR sets a promising precedent for its future implementation in Panel Reviews, aiming to enhance the alignment and effectiveness across all review processes.

Lastly, to further underscore the ongoing advances in streamlining the PR process, David Harvey presented a talk on Propy. Propy is a tool that aims to modernise the scientific PR process. Moving beyond the traditional reliance on keywords for reviewer assignments, it utilises an AI-driven database of scientific articles. This approach facilitates the identification of appropriate reviewers by creating detailed profiles based on scientific concepts. While Propy seeks to improve the fairness and efficiency of PR, it reflects an ongoing evolution in the field, with its effectiveness yet to be fully assessed in the broader scientific community.

The Methodologies session was a deep dive into the current challenges and innovative solutions shaping the landscape of PR processes. With a tapestry of expertise, attendees witnessed a rich exploration spanning automated systems, funding biases, AI's burgeoning role in evaluations, and the layered world of DPR. Such revelations underscored the pressing need to refine and reshape review methodologies to better cater to the scientific community's evolving demands.

Diversity, Equity and Inclusion

The workshop session on Diversity, Equity and Inclusion provided critical insights into the PR process and the imperative need for more inclusive research evaluations. Cassidy Sugimoto highlighted the importance of such inclusive evaluations, drawing attention to the urgent need to address biases and systemic challenges present in the current PR process. She pointed out that bias in respect of who gets to become a scientist and their scientific publications and activities starts early and accumulates at each career step, hence disfavoring scientific excellence. In particular, when gender is known, women are penalised. Therefore, anonymisation is essential to both minimise discrimination and promote excellence.

Andrea Rapisarda emphasised the often underestimated role of serendipity in scientific discoveries. Using historical examples like the discovery of penicillin and the finding of the cosmic microwave background radiation, Andrea showcased how many scientific advances owe their genesis to chance events. Despite this, the present science funding landscape may sideline the role of randomness, leaning instead towards a naïve meritocracy. This structure, according to Andrea's agent-based model, tends to favour the moderately talented yet lucky individuals. To nurture genuine talent and foster true innovation, Andrea championed funding strategies that provide expansive opportunities, moving away from just rewarding past successes. The theme of randomness in the realm of scientific evaluation and funding was continued by Christophe Heger. Citing the study by Cole, Cole & Simon (1981),

Heger highlighted that a significant fraction of funding decisions are influenced by the random choice of reviewers. This brings into question the current PR system's efficacy. Tracing the origins of PR, Christophe elaborated on its aims and inherent criticisms, primarily its potential biases. To address this, reforms such as introducing lotteries or blinded reviews were debated. While these methods promised a reduction in costs and biases, a survey revealed a notable resistance from the scientific community, with a significant majority opposing the integration of lotteries into grant decisions.

The finale of the randomness discussion was the Volkswagen Foundation's Experiment! Initiative presented by Ulrike Bischler. Between 2017 and 2021, this initiative experimented with a partial randomisation method for grant selection. In this implementation, randomisation was applied to applications which were peer-reviewed in the classical way and ranked in the central, grey area, where the evaluation confusion generated by the subjectivity of the process is inherently large.

Inspired by historical precedents, such as Athenian democracy's allocation methods, the initiative aimed to challenge the prevailing PR system's limitations, including biases and conservatism. As a result of this experimental approach, there was a marked increase in representation on the part of women, early career researchers, and underrepresented disciplines. Surveys further echoed the sentiment that the grantees found the lottery system more equitable and diverse. The comparison between traditional review and this randomised method revealed no discernible difference in project outputs, reinforcing the potential benefits of such an innovative approach.

As the discussions progressed, Virginia Valian dwelt on Evaluating Merit. One of the key points she brought forward relates to the fact that a significant portion of the scientists are still convinced that they are objective in their evaluations, while they are in fact just unaware of biases. Going one step further, for example, believing that there is no gender bias actually leads to a more pronounced discrimination. Dual anonymisation definitely

helps, although it does not eliminate the differences related to gender-dependent writing styles. AI-based tools can help reduce these effects, also improving the situation for non-native speakers coming from disadvantaged backgrounds.

This was followed by Lou Strolger's discourse on "Reducing systemic biases through anonymized time-allocation peer review", and by Verne Smith's report on Planning and Deploying the NSF's NOIR-Lab Dual Anonymous Review Process. The experience of dual-anonymous PR in proposal evaluation shows that proposals from women and early-career scientists are evaluated more positively when their identity is not known.

In conclusion, this workshop session underscored the paramount importance of diversity, equity, and inclusion in the scientific PR process. The discussions provided a holistic view of the challenges present and the possible avenues for future reform. As the scientific community grapples with these issues, the overarching sentiment remains: a call for a more inclusive, unbiased, and innovative system that truly fosters talent and ground-breaking research.

Concrete Examples

The workshop session on Concrete Examples offered participants an in-depth exploration of diverse observational systems and the challenges and transitions they have faced in their PR processes.

ALMA, represented by Andrea Corvillón and John Carpenter, transitioned from a panel-based system to a DPR system in Cycle 8, as a result of an escalating number of submissions. This marked a significant development in astronomy PR, with the engagement of over 1000 individual reviewers evaluating around 1500 proposals. Feedback suggested concerns over proposal assignment and reviewer expertise, but it was noteworthy that junior researchers, including students and post-docs, provided reviews that were as constructive as those of senior counterparts. ALMA envisions enhancing this system with advanced algorithms and machine learning for better proposal assignments.

André-Nicolas Chené from NOIRLab, the US focal point for nighttime astronomy, showcased its PR system that handles over 1000 proposals annually for its semi-independent observatories. Notable in the presentation was the introduction of the Dual Anonymous Review Process and the Research Inclusion Initiative. The latter emphasises a democratic approach to scientific contributions, aiming to broaden accessibility to research opportunities.

The Australia Telescope National Facility (ATNF), a branch of Australia’s CSIRO represented by Elizabeth Mahony and Philip Edwards, has seen its Time Assignment Committee (TAC) process evolve over three decades. Today, the ATNF TAC, which includes both national and international members, reviews submissions from nearly 700 astronomers globally. A commitment to countering unconscious bias led to a transition from semi-anonymisation to complete anonymisation in 2022.

Norbert Schartel discussed the XMM-Newton mission and presented its unique PR system, stressing minimal interactions between mission staff and reviewers. With panels of experts from various countries, the system encourages diversity and inclusion while aiming for robust scientific advancements. This method serves as a bridge between the mission and the scientific community, integrating them more closely.

Rodolfo Montez illuminated the workings of the Chandra PR process. Chandra, a space-based X-ray observatory under NASA, has been operational since 1999. Montez highlighted that Chandra’s PR shows consistent success rates as between male and female PIs over the years.

Andrea Mejías spotlighted the Chilean Telescope Allocation Committee (CNTAC), responsible for allocating telescope time within Chile to a community of around 500 astronomers. The systematic structure of CNTAC ensures the optimal allocation of resources for these researchers.

Concluding the session, Vincent Larivière gave a presentation on deceptive publishers and underscored the dark side of the academic publishing world. As

publishing garners increased significance in research evaluations, predatory publishers exploiting the ‘publish or perish’ culture have emerged. Instances like the acceptance of a fake paper by most journals on a known predatory list and the dubious activities of some publishers were highlighted. Ethical publishing practices were advocated as a solution to countering these deceitful strategies.

In essence, this workshop session provided a holistic view of the various PR systems in place across diverse observatories, underscoring the continuous evolution, challenges and best practices in the field.

Remarks on the discussion session

In the discussion focusing on its fundamental principles we considered the multiple crucial purposes that PR serves in the scientific community. Firstly, it establishes trust between scientists and the general public, who fund scientific research, by ensuring that rigorous scrutiny is applied to research findings. This trust is essential for maintaining public support and credibility. Secondly, PR acts as a self-governance tool, upholding the quality and integrity of scientific work. It promotes transparency, accountability, and adherence to ethical standards. Furthermore, PR enhances the exchange of ideas and knowledge within the scientific community. It allows experts to critique and suggest improvements, ultimately elevating the quality of research. PR also facilitates the injection of fresh perspectives and novel ideas from diverse backgrounds, fostering innovation and interdisciplinarity. It aids in the allocation of limited resources like funding and telescope time, ensuring fair distribution. Despite its many shortcomings, the participants agreed that PR needs an evolution not a revolution.

Discussion of bias in PR not only covered the common topics of gender and racial diversity but also focused on the many unseen hurdles of the PR process. One major discussion point was that there remains a traditional bias in PR towards native English speakers, potentially disadvantaging non-native speaker researchers. One proposed solution included

using structured proposal documents with templates to standardise and streamline proposal writing, minimising prose and highlighting ideas. As another way to mitigate this bias, visualisations were recognised as a valuable tool that transcends language barriers. Visualisations, being inherently language agnostic, offer a universally understandable means of conveying complex ideas, promoting better comprehension, and facilitating the assessment of proposals by reviewers from diverse linguistic backgrounds. Another key issue addressed was the regional and cultural variations in argument construction, particularly relevant in more isolated astronomy communities where common proposal writing practices may not be universally taught, especially to postdocs and graduate students.

The session also explored strategies for instigating change within the PR process, debating whether it’s more effective to encourage proposers to drive change from within the system or to push for systemic changes externally. The consensus leaned towards the necessity for both approaches to ensure a fair and effective PR system.

Participants exchanged views on the effect of the panel discussions in PR. Evaluation agreement among different people generally increases after panel discussions, but not necessarily towards accuracy. On the one hand, panel discussions allow reviewers to share expertise and correct potential misunderstandings. Being exposed to the evaluations of other reviewers can also be beneficial in improving the way reviews evaluate. On the other hand, people change their minds during panel discussions, feeling a need to conform, or when being challenged, or in confrontations with prestigious names. Biases and psychological aspects of panel discussions are hard to quantify.

Another topic discussed was the effect of dual-anonymous PR and its perception in the community. One strong advantage of dual anonymisation is that it reduces the biases in the evaluation, for example against less established researchers and those from underrepresented minorities. Biases are not completely removed though. Dual anonymisation allows reviewers to focus purely on science,

rather than on personal information. On the other hand, it is sometimes hard to reach full anonymity, especially in small communities. Anonymisation is at odds with the concept of open science, where both reviewers and applicants are openly known. A two-step approach, where science only is evaluated in the first place and information on the applicants is disclosed later, could be feasible and beneficial in different fields.

Workshop demographics

The workshop saw an impressive turnout, with a total of 173 registered attendees. Among them, 46 were present in person, while 127 opted for remote participation. However, from the online group, on average 30 participants were consistently active throughout the workshop sessions. The gender distribution was relatively balanced, with 73 females, 88 males, and 12 who preferred not to specify. Breaking it down by career level, there were 111 staff members, 24 postdocs, 17 students, and 21 in the Other category. The workshop featured 34 presentations. These statistics showcase a broad spectrum of attendees, underscoring the widespread interest in the workshop's theme and content.

Concluding remarks

The Peer Review Under Review workshop was a landmark event in the

assessment of PR systems within the astronomical community. The assembly of diverse experts cast a critical eye over current practices, noting challenges in managing the burgeoning volume of research, inherent biases, and the adaptability of systems in today's digital age. Throughout the sessions, there was a clear call for innovation, increased efficiency and inclusivity. Discussions ranged from the utilisation of AI and machine learning in paper evaluations, the role of chance in scientific discovery and funding, to case studies of various observatories and their unique PR systems. A particular emphasis was placed on the importance of diversity, equity, and inclusion in the PR process. Moreover, the emergence of predatory publishing underscored the need for ethical practices.

Other points concerning possible future improvements in the PR process were discussed. These included: reviewer training, motivation and awards; a more structured PR to minimise subjectivity; the possibility of making the reviews available to the readers, particularly in the context of open science; a more open and collaborative approach to PR (for example starting from preprints). Finally, the challenge posed by the rapidly growing scientific production was considered, in the context of PR sustainability and costs for society. AI tools may help to cope with this important problem, possibly providing a viable way of running a pre-screening.

The workshop¹ with all its contributions (available online via Zenodo²) was a clarion call for the evolution of PR, ensuring the quality, efficiency and fairness of future scientific research.

It is now the task of all those directly involved in PR to continue the discussions and to rise to the challenges that were identified, all this for more equitable, fair and effective review systems in the near future.

Acknowledgements

We extend our heartfelt gratitude to ESO for their generous financial support and essential assistance, which were pivotal in the successful execution of this workshop. Additionally, we would like to acknowledge the invaluable support provided by ESO's IT department, whose expertise and dedication played a crucial role in facilitating the smooth operation of the event. Finally, we are grateful for the valuable assistance offered by OpenAI's language model, ChatGPT.

References

Cole, S., Cole, J. R. & Simon, G. A. 1981, *Science*, 214, 881

Links

¹ Workshop webpage: <https://www.eso.org/sci/meetings/2023/PRUR.html>

² Workshop contributions: <https://zenodo.org/communities/peerreview23>

P. Horálek/ESO



All of ESO's observatories are based in Chile, but the organisation's headquarters, and the newest exciting addition, sits in Garching, a small city near Munich in Germany. This image shows a striking and unusual view of this new member of the ESO family – the ESO Supernova Planetarium & Visitor Centre.