



Information Research – Vol. 30 No. iConf (2025)

Demystifying perception and decision-making in peer review: a semantic perspective

Yujie Zhang, Weikang Yuan, and Zhuoren Jiang

DOI: <https://doi.org/10.47989/ir30iConf47230>

Abstract

Introduction. Peer review is fundamental to scientific progress. However, it remains a ‘black box’ due to the ambiguous and subjective decision-making rationale, hidden within reviewers’ perceptions and judgments.

Research gaps. The scarcity of peer review data and the limited analytical scope have hindered empirical studies from thoroughly exploring these underlying factors.

Method. We collected 114,006 review records from OpenReview. Developing deep learning models, we extracted reviewers’ perceptions of key factors from the review content.

Analysis. Our analysis focuses on how these perceptual factors influence reviewers’ decisions and outcomes of the papers.

Results. Originality is the element that reviewers value and recognize the most, followed by motivation. In contrast, chairs prioritize the overall validity of the work within the field, giving more weight to the substance and soundness of papers. The role of motivation is generally supportive but has limited distinguishing power on final decisions. Perceived clarity, replicability, and meaningful comparison have a relatively minor role in the review process.

Conclusion. Our study highlights the decision-making process based on reviewers’ perceptions and uncovers perceptual differences between reviewers and chairs. This contributes to greater transparency in the peer review and helps build trust within the scientific community.

Introduction

Peer review is considered the backbone of scientific construction (Alberts et al., 2008). By filtering out substandard work and providing expert feedback, it helps ensure the quality and credibility of scientific research, contributing to the evolution and enrichment of knowledge (Bailly, 2016; Jefferson et al., 2002; Riley & Jones, 2016). However, despite its importance and decisiveness, the decision-making process in peer review mechanisms often resembles a black box (Chong & Mason, 2021; Cutler et al., 2022), which may raise concerns within the scientific community. It arises not only from the non-transparency of the procedure (Hamilton et al., 2020) but also from the ambiguous and subjective rationale underlying the reviewers' decision-making.

The review decision-making process involves complex cognitive processes (Brosch et al., 2013; Lindsay & Norman, 2013; Marietta & Barker, 2019), with perceptions and judgments formed based on the presented facts (Dror & Fraser-Mackenzie, 2008; Street & Ward, 2019). It is complex and subjective, influenced by reviewers' cognitive biases and qualifications (Benda & Engels, 2011). Their decisions—whether to accept or reject a paper—are affected by factors they have perceived such as significance, originality, methodological rigor, and presentation (Blockeel et al., 2017).

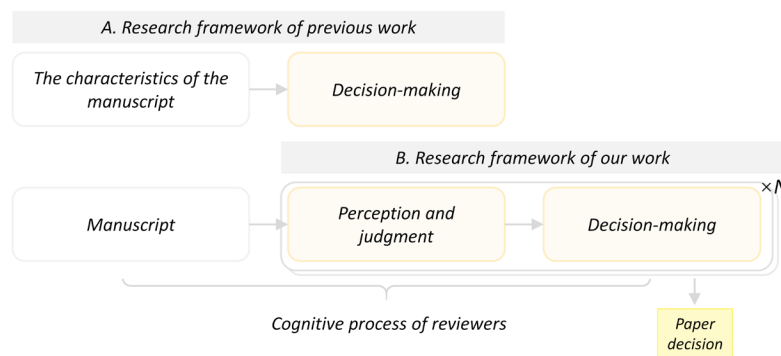


Figure 1. The research framework of previous studies compared to our work. Prior empirical studies on peer review decision-making have focused on extracting the objective characteristics of papers to reveal the preliminary process of reviewers' decisions. However, the reviewers' perceptions and judgments remain unexplored.

Existing research empirically unveils reviewers' decision-making rationale, biases, or preferences, typically employing a 'fact-decision' approach (Figure 1). For instance, they utilize external metrics, such as the disruption index, to objectively describe the characteristics of papers and then examine whether reviewers' decisions show a preference for certain work (Liang et al., 2023; Teplitskiy et al., 2022). This approach overlooks the crucial intermediate step—the 'perception and judgment' phase, closely intertwined with the cognitive processes of reviewers (Lamont, 2009; Lee et al., 2013). Exploring this can offer valuable insights into the cognitive aspects of decision-making in peer review, enhance transparency, and build greater trust in the academic community.

Nevertheless, there are still some critical research gaps that may hinder further exploration: (a) The first challenge is about how to measure the reviewer's perception. Generally, reviewers' perceptions could be revealed in their comments like enthusiastic praise or critical criticism (Mavrogenis et al., 2020). However, most peer review processes are not publicly available and related data is extremely scarce (Baliatti et al., 2016), making it very difficult to obtain reviewer comments in real-world scenarios. Furthermore, the review contents are usually unstructured data, which makes it harder to extract meaningful insights. Prior studies examined review content through single-dimensional features like sentiment (Matsui et al., 2021), tone (Nobarany & Booth,

2015) and length (Maddi & Miotti, 2024), but have struggled to capture the multifaceted nature of the perceptions. (b) Few empirical studies assess the relationship between reviewers' perceptions and decision-making. This lack of evidence hampers theory development and highlights gaps in understanding the peer review process and perceptual factors of reviewers.

To address these research gaps, we propose the following research questions (RQs):

RQ1: how can we automate the extraction of reviewers' perceptions and judgments of a paper?

RQ2: from the perspective of reviewers' decisions, how do perceptual factors influence reviewers?

RQ3: from the perspective of papers' decisions, how do reviewers' perceptions affect the decision of a paper?

To answer the above research questions, we collected a substantial number of open review records from OpenReview and carefully filtered them to construct a valuable peer-review dataset, comprising 114,640 records from top-tier computer science conferences. Then, we develop deep learning models to automatically extract and measure reviewers' multi-dimensional perceptual factors (RQ1). Based on these factors, we conceptualize the peer review decision-making process into two primary phases: individual reviewer decisions (RQ2) and paper decisions affected by multiple reviewers (RQ3) and examine the relationship between reviewers' perceptions and decision-making in these phases.

Related work

Perception in peer review has been discussed by existing studies, particularly regarding how reviewers perceive the peer review process itself. Existing research has explored various aspects of reviewers' perceptions, such as their motivations for reviewing, preferred roles, and preferences regarding review mechanisms (Curtin et al., 2018). Other studies have examined reviewers' willingness to receive training, accept feedback, and sign their reviews (Manchikanti et al., 2015). Less work has empirically investigated reviewers' perceptions of the paper. One such study uses surveys to quantify the gap between researchers and reviewers regarding the perceived importance of specific validity and reliability issues (Street & Ward, 2019).

The process from perception to decision in peer review is inherently subjective, and this subjectivity plays a crucial role in shaping the overall review process. Existing work argues that such subjectivity is not only unavoidable but also necessary, as it may help mitigate the herding effect and encourage diverse perspectives (Park et al., 2014). This subjectivity also contributes to the phenomenon of low inter-rater reliability among reviewers. While some researchers view this inconsistency as a natural and inevitable aspect of peer review (Roediger, 1991), others consider it problematic, suggesting that it undermines the reliability of the evaluation process (Marsh et al., 2007).

The process of review is not merely about evaluating the content of a paper; it is also shaped by '*sense of self and their relative positioning*' of reviewers (Lamont, 2009). The connection between identity and evaluation raises concerns about the potential for cognitive biases, which are often cited as a threat to the progress and integrity of a field (Benda & Engels, 2011). Specifically, reviewers may show a preference for submissions from authors with similar '*schools of thought*', a phenomenon known as '*cognitive cronyism*' (Travis & Collins, 1991). Thus, while subjectivity in peer review can foster diverse viewpoints, it also presents challenges that must be addressed to ensure the fairness and accuracy of the evaluation process.

Dataset

To address the issue that the majority of journals and conferences' reviews are not publicly available, open peer review offers a promising solution to enhance the transparency and fairness of the review process (Wolfram et al., 2020). OpenReview (Soergel et al., 2013) is one of the most famous open peer review websites to promote transparency in scientific communication. It provides submitted papers alongside their corresponding reviews and comments. After a thorough examination of each venue in OpenReview, we focused on the top-tier machine learning conferences—ICLR and NeurIPS for our study (ICLR makes reviews public regardless of acceptance, while NeurIPS allows authors of rejected papers to decide whether to make their reviews public). We collected all publicly available review data for these conferences. After data pre-processing and meticulous data filtering, we obtained review information for 30,440 papers from the year 2013 to 2024, comprising 114,640 records.

Measurement of variables

Automatical extraction perceptual factors of reviewers (RQ1)

'Perceptual factors' can be defined as the reviewer's subjective perception and judgment elements, which guide their review decision-making. The reviewer's perception of one submission can be directly reflected in the review comments (Mavrogenis et al., 2020). According to previous studies (Yuan et al., 2022), we identify seven key perceptual factors (These dimensions follow the ACL review guidelines): motivation, originality, soundness, substance, replicability, meaningful comparison, and clarity.

To automatically and effectively measure these factors, we designed a pipeline that includes a discriminative model and a scoring model, each based on RoBERTa (Liu, 2019). Specifically, the discriminative model determines whether a sentence is perception related. If it does, the scoring model then assesses the multidimensional perceptual factors. To illustrate these indicators more clearly, we provide an implication and a simplified example in Figure 2. The detailed training process is outlined as follows (The hyperparameters for both models are set as follows: batch size of 16 and a learning rate of $5e-6$. We employ the best model among 8 epochs with the lowest validation loss. The model is trained on an NVIDIA V100 32GB GPU):

We use AMPERE (Hua et al., 2019) and ASAP (Yuan et al., 2022) datasets to train our models. To train the discriminative model, we leverage AMPERE, which categorizes review comments into six types. We defined the task as a classification problem to identify whether a sentence belongs to the category 'evaluation'. To train the scoring model, we use ASAP, which annotated seven dimensions of text spans with sentiment polarity. We converted it into a numeric labeling task. During training, the model takes the text as input and produces a 7-dimensional output, with each dimension representing a dimension. We use the Tanh function to scale each dimension. The loss is computed using Mean Squared Error (MSE). On the validation set (20% of total data), the model achieved a classification accuracy of 0.83 and an absolute deviation of 0.084 for each dimension.

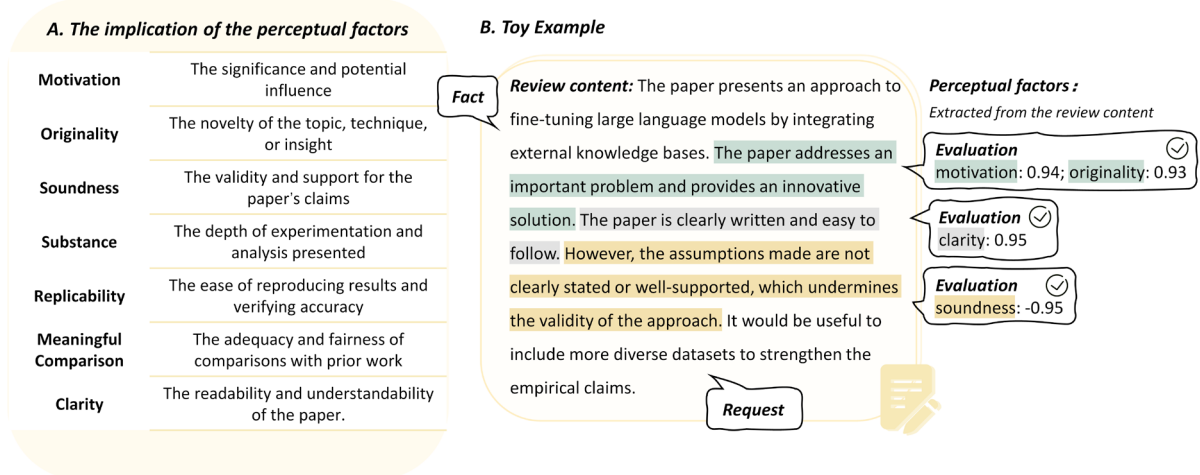


Figure 2. The implication (A) and toy example (B) of the extracted perceptual factors. Each factor’s polarity is represented by a continuous value ranging from -1 to 1, where a value closer to 1 indicates higher confidence in a positive attitude. The first and last sentences in the example are categorized as ‘fact’ and ‘request’, respectively. Unlike the ‘evaluation’ category, these sentences exhibit less subjectivity and are thus excluded from measuring the reviewer’s perceptual factors. The three sentences in between cover various perceptual factors such as ‘motivation’, ‘originality’, ‘clarity’, and ‘soundness’, showcasing the reviewer’s diverse perceptions and judgments of the paper.

After measuring the perceptual factors of each sentence, we aggregate the scores as follows:

$$PF_{pi} = \sum PF_{seg}, seg \in T_{pi}$$

$$PF_p = \frac{1}{N_p} \sum PF_{pi}$$

Where PF_{pi} represents the 7-dimensional vector of perceptual factors given by the i -th reviewer. seg represents the sentence of review content T_{pi} . PF_p denotes the average judgment of all reviewers of the paper. The more frequently reviewers mention and emphasize a particular dimension in their comments, the higher the perceptual score for that dimension.

Reviewer decision

Reviewer decisions are a key criterion for chairs in determining paper acceptance, such as ‘8: accept, good paper’ or ‘3: reject, not good enough’. Because different conferences use varying scoring ranges from year to year, we standardize all scores to a range of 0 to 10 to ensure comparability. The formula for this calculation is as follows:

$$RD_{pi} = \frac{r_{pi} - \min(R)}{\max(R) - \min(R)} \times 10, r \in R$$

$$RD_p = \frac{1}{N_p} \sum RD_{pi}$$

Where p represents a paper reviewed at a particular conference (held in a specific year). R denotes all the ratings given at this conference. r_{pi} refers to the original rating of the paper p given by the i -th reviewer, while RD_{pi} represents the processed one. N_p is the total number of reviews that the paper p received. RD_p is the average rating of all reviewers.

Result

Descriptive statistics of reviewers' perceptual factors

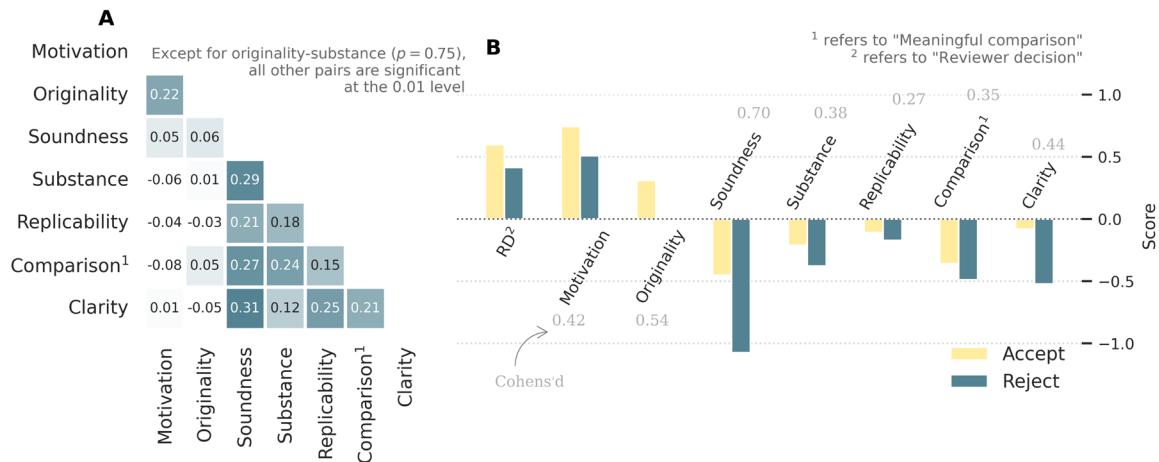


Figure 3. Description of the reviewers' perceptual factors. (A) We calculated the Pearson correlation between different factors using data samples from individual reviews. It is shown that originality and motivation were the most correlated (and vice versa); soundness and clarity were the most correlated (and vice versa); substance and soundness were the most correlated; replicability and clarity were the most correlated; and meaningful comparison and substance were the most correlated. (B) We compared the average of reviewer decisions (RD) and perceptual factors for accepted and rejected papers. Unlike Figure 3 (A), which is based on single reviews, the data in Figure 3 (B) are based on individual papers, each with multiple reviews. We found that perceptual factors, whether relating to significance (Motivation), novelty (originality), or rigor (soundness, substance, e.g.), all have the potential to influence the decisions made about the papers. It shows the differences for each factor between the accepted and rejected categories, represented by the Cohen's d .

Firstly, we observed the internal correlation of perceptual factors within reviewers' comments. As shown in Figure 3 (A), we found that the internal correlations of these metrics were generally low (≤ 0.31), which means that there is a lower likelihood of serious multicollinearity issues in modeling.

Figure 3 (B) demonstrates that accepted papers receive higher evaluations from reviewers across all criteria compared to rejected ones. Rejected papers often get positive feedback on motivation but tend to receive negative or no feedback (-0.01) on originality. This indicates reviewers are more supportive of motivation regardless of their recommendation but more critical of originality in papers they dislike. Further, reviewers often give negative evaluations on rigor-related dimensions to prompt authors to improve.

Perception-based decision-making analysis

Reviewer decision analysis (RQ2)

We analyse individual reviewers to examine how their decisions are influenced by their perceptions and judgments. The model is specified as follows:

$$RD_{pi} = \beta_0 + \beta PF_{pi} + \beta' X_{pi} + \alpha_p + \lambda_p$$

RD_{pi} represents the reviewer decision given by the i -th reviewer for paper p . The key variable PF_{pi} denotes the vector of the reviewer's perceptual factors. X_{pi} is a vector of control variables, including the request number and confidence of the reviewer. The request number represents the number of sentences related to the reviewers' requests. Confidence is the assurance of reviewers, which is scored by themselves. Since the review process is double-blind, we do not control for author-related information in this model. α_i and λ_i represent the fixed effects, capturing stable factors

affecting the reviewer's decision at the venue and time levels respectively. β_0 represents the constant term. We conducted a VIF test (Table S1) and used OLS regression for our analysis (Table 1).

We identified the impact of various perceptual factors on reviewer decisions. Among all perceptual factors (Figure 4), reviewers prioritize originality, followed by motivation, as the primary basis for their decisions. Next in importance are soundness and substance, underscoring the paper's overall persuasiveness and rigor. Lastly, reviewers focus on meaningful comparison, clarity, and replicability. They determine whether a paper is easy to follow and disseminate and they also form the basis for the paper's credibility.

In Figure 5, we conducted a heterogeneity analysis of reviewers grouped by high and low confidence levels (the boundary is set at 5). The results showed that decisions made by high-confidence reviewers were more strongly associated with perceptual factors (except for replicability and clarity).

Dependent Variable: Reviewer decision	Coefficient	SE	P value	Confidence interval
Motivation	0.244	0.005	0.000	[0.235, 0.253]
Originality	0.372	0.005	0.000	[0.363, 0.381]
Soundness	0.214	0.003	0.000	[0.208, 0.220]
Substance	0.213	0.006	0.000	[0.201, 0.225]
Replicability	0.098	0.011	0.000	[0.076, 0.120]
Meaningful comparison	0.172	0.007	0.000	[0.158, 0.187]
Clarity	0.159	0.003	0.000	[0.154, 0.165]
Request number	0.020	0.001	0.000	[0.018, 0.021]
Confidence	-0.060	0.002	0.000	[-0.065, -0.056]

Table 1. The regression result of reviewer decision analysis. The model has an R-squared value of 0.27, and the observation number is 107,272.

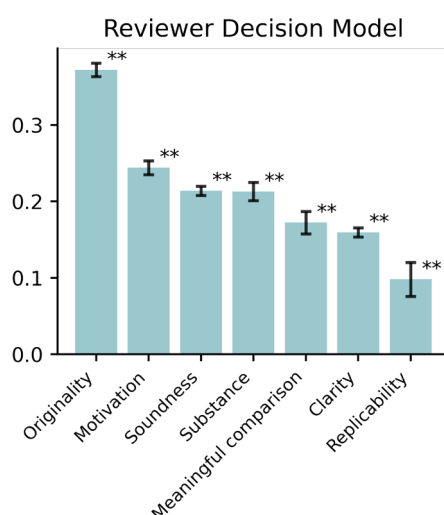


Figure 4. The coefficient comparison of reviewer decision analysis. Among all the perceptual factors, originality has the greatest impact on the reviewer's decision, followed by the perception and judgment of motivation. ** represents $p < 0.01$, and * represents $p < 0.05$.

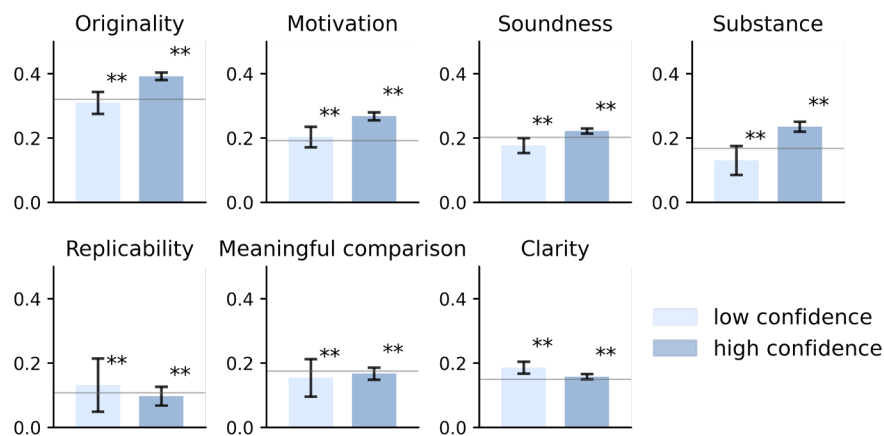


Figure 5. The impact of perceptions reflected in review content between high-confidence and low-confidence reviewers on their decisions (using the standardized median value, i.e., 5, as a cut-off, with the standardization process similar to *RD*). The middle line represents the coefficient for confidence at the median value of 5. The study found that, except for clarity and replicability, the perceptions of high-confidence reviewers are most closely associated with their decisions across various aspects. This reflects that decisions made by high-confidence reviewers are more interpretable and follow a logical progression from perception to decision-making.

Paper decision analysis (RQ3)

When deciding whether to accept a paper, the conference chairs take into account the reviewers' scores and feedback to make the final decision on acceptance or rejection. We further analyse how the chairs' decision is made to gain insights into the process of decision-making in peer review. X_p is the vector of control variables, including average *RD*, request number and the prestige of paper, where prestige refers to the average h-index of all authors, matched with S2AG (Version: May 21, 2024) (Wade, 2022). $PaperDecision_p$ represents the final decision for paper p . It is a binary variable representing whether the paper is accepted. The model is specified as follows:

$$PaperDecision_p = \beta_0 + \beta_1 RD_p + \beta PF_p + \beta' X_p \alpha_p + \lambda_p$$

We conducted a VIF test (Table S2) and used logistic regression for our analysis (Table 2). We uncovered an interesting distinction (Figure 6): from the perspective of the chairs, the most critical perceptual factors are not originality or motivation, but rather substance and soundness. This contrasts with the focus of reviewers. The findings suggest that chairs place greater emphasis on ensuring the overall validity in the field. Moreover, factors such as motivation, replicability, clarity, and meaningful comparison do not have a significant impact on paper decisions. This could be due to the supportive nature (as shown in Figure 3) of factors like motivation, which may lack strong differentiating power from the perspective of papers' decisions. It is shown in Figure 7 that in the high-*RD* group (*RD* below the median), positive judgments on validity-related dimensions, are more influential on papers' decisions.

Dependent Variable: Paper decision	Coefficient	SE	P value	Odds multiplier	Confidence interval
Motivation	0.078	0.052	0.132	1.08	[-0.024, 0.180]
Originality	0.110	0.047	0.020	1.12	[0.017, 0.203]
Soundness	0.184	0.034	0.000	1.20	[0.118, 0.250]
Substance	0.192	0.064	0.003	1.21	[0.066, 0.318]
Replicability	-0.172	0.123	0.160	0.84	[-0.413, 0.068]
Meaningful comparison	-0.067	0.079	0.397	0.94	[-0.222, 0.088]
Clarity	0.059	0.032	0.070	1.06	[-0.005, 0.122]
Reviewer Decision	2.810	0.056	0.000	16.60	[2.700, 2.919]
Request number	-0.005	0.009	0.603	1.00	[-0.023, 0.013]
Prestige	0.005	0.002	0.020	1.01	[0.001, 0.010]

Table 2. The regression results for paper decision analysis show that the pseudo-R-squared is 0.55, with a sample size of 17,620 observations. Since chairs' decisions are heavily influenced by the average scores provided by reviewers, we excluded papers rated below the 1st percentile of accepted submissions (a total of 5,464 papers, with a 97% probability of rejection) and those rated above the 99th percentile of rejected submissions (a total of 2,268 papers, with a 97% probability of acceptance).

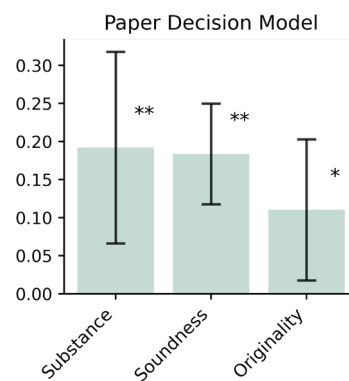


Figure 6. The coefficient comparison of paper decision analysis. We only present the main variables that are significant in the regression analysis. Among all the perceptual factors, substance and soundness have a significant positive correlation with the chair's decision, followed by originality. The effects of other main variables on the paper's decision are not significant. ** represents $p < 0.01$, and * represents $p < 0.05$.

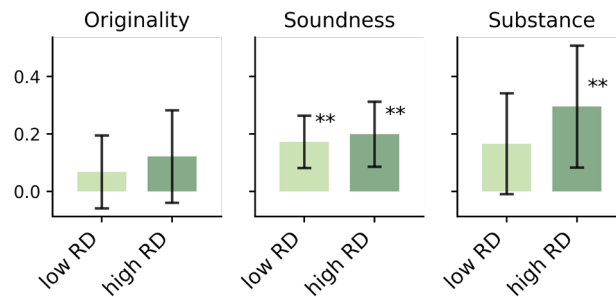


Figure 7. We compared the high and low *RD* groups (divided by the median, with extremely high and low scores excluded in the same way as in Table 2) regarding reviewers' perceptions of the papers. We found that decisions on papers in the high-*RD* group were more likely to be influenced by the reviewers' perceptions. High scores already indicate the overall superiority of the paper's quality. If reviewers' evaluations of the validity-related factors are positive, it indicates that the chairs' concerns have been addressed, thus increasing the likelihood that the chairs will accept the paper.

Conclusion and discussion

Due to limited review content and narrow analytical dimensions, there's little empirical research on the perception-to-decision process in peer review. Our findings show that originality is the most critical factor in reviewers' decisions, followed by motivation. Chairs rely on reviewers' decisions but prioritize assessments of soundness and substance, focusing on the research's overall validity. Reviewers' feedback on motivation is reassuring but has limited distinguishing ability for chairs. The Overall impact of perceived meaningful comparison, clarity, and replicability on decision-making in peer review is weak.

While many believe that novelty is always hampered in the publication process (Liang et al., 2023; Wang et al., 2017), our study empirically uncovers that reviewers place a substantial emphasis on their perception of originality and are more inclined to accept papers they deem novel. This highlights that originality is the most valued and recognized element by reviewers. However, the transition from perception to decision is complex and still largely unexplored. Specifically, reviewers' perspectives can be influenced by preconceptions or past experiences, and there is often a gap between perception and its expression. Additionally, cognitive limitations may obscure underlying social biases. The intricate mechanisms and overall effectiveness of this system demand ongoing exploration and refinement. Within this context, reviewers' cognition, perception, and decision-making are indispensable and critical components. While the peer review system benefits science, it also necessitates a continuous cycle of feedback to evolve and improve (Ward et al., 2015).

Limitation and future work

Due to challenges in accessing review content, our data has some constraints. We have also lightly addressed chairs' perceptions, which we will explore in more depth. The requests from reviewers also contain important information, and there is potential for further analysis.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (72104212).

About the authors

Yujie Zhang is a master student from the Department of Information Resources Management at Zhejiang University, China. They can be contacted at yj.zhang@zju.edu.cn.

Weikang Yuan is a doctoral student from the Department of Information Resources Management at Zhejiang University, China. They can be contacted at yuanwk@zju.edu.cn.

Zhuoren Jiang is an Assistant Professor the *corresponding author* of this paper from the Department of Information Resources Management at Zhejiang University, China. They can be contacted at jiangzhuoren@zju.edu.cn.

References

- Alberts, B., Hanson, B., & Kelner, K. L. (2008). Reviewing Peer Review. *Science*, 321(5885), 15–15. <https://doi.org/10.1126/science.1162115>
- Bailly, B. A. F. L. (2016). Learning from peer review. *Nature Nanotechnology*, 11 2, 204. <https://doi.org/10.1038/nnano.2016.4>
- Balietti, S., Goldstone, R. L., & Helbing, D. (2016). Peer review and competition in the Art Exhibition Game. *Proceedings of the National Academy of Sciences*, 113(30), 8414–8419.
- Benda, W. G., & Engels, T. C. (2011). The predictive validity of peer review: A selective review of the judgmental forecasting qualities of peers, and implications for innovation in science. *International Journal of Forecasting*, 27(1), 166–182.
- Blockeel, C., Drakopoulos, P., Polyzos, N., Tournaye, H., & García-Velasco, J. (2017). Review the ‘peer review’. *Reproductive Biomedicine Online*, 35 6, 747–749. <https://doi.org/10.1016/j.rbmo.2017.08.017>
- Brosch, T., Scherer, K., Grandjean, D., & Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision-making. *Swiss Medical Weekly*, 143(1920), w13786–w13786.
- Chong, S. W., & Mason, S. (2021). Demystifying the process of scholarly peer-review: An autoethnographic investigation of feedback literacy of two award-winning peer reviewers. *Humanities and Social Sciences Communications*, 8(1), 1–11.
- Curtin, P. A., Russial, J., & Tefertiller, A. C. (2018). Reviewers’ Perceptions of the Peer Review Process in Journalism and Mass Communication. *Journalism & Mass Communication Quarterly*, 95, 278–299. <https://doi.org/10.1177/1077699017736031>
- Cutler, S., Xia, Y., & Beddoes, K. (2022). Peering into the Black Box of Peer Review: From the Perspective of Editors. 9th Research in Engineering Education Symposium (REES 2021) and 32nd Australasian Association for Engineering Education Conference (REES AAEE 2021). <https://doi.org/10.52202/066488-0095>
- Dror, I. E., & Fraser-Mackenzie, P. A. (2008). Cognitive biases in human perception, judgment, and decision making: Bridging theory and the real world. In *Criminal investigative failures* (pp. 79–94). Routledge.
- García, J. A., Rodríguez-Sánchez, R., & Fdez-Valdivia, J. (2015). Bias and effort in peer review. *Journal of the Association for Information Science and Technology*, 66(10), Article 10.
- Hamilton, D., Fraser, H., Hoekstra, R., & Fidler, F. (2020). Journal policies and editors’ opinions on peer review. *eLife*, 9. <https://doi.org/10.7554/eLife.62529>
- Howard, L., & Wilkinson, G. (1998). Peer review and editorial decision-making. *The British Journal of Psychiatry*, 173(2), 110–113.

- Hua, X., Nikolov, M., Badugu, N., & Wang, L. (2019). Argument Mining for Understanding Peer Reviews. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2131–2137.
- Jefferson, T., Wager, E., & Davidoff, F. (2002). Measuring the quality of editorial peer review. *Jama*, 287(21), 2786–2790.
- Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Harvard University Press.
- Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2–17.
- Liang, Z., Mao, J., & Li, G. (2023). Bias against scientific novelty: A prepublication perspective. *Journal of the Association for Information Science and Technology*, 74(1), 99–114.
- Lindsay, P. H., & Norman, D. A. (2013). *Human information processing: An introduction to psychology*. Academic press.
- Liu, Y. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv Preprint arXiv:1907.11692*.
- Maddi, A., & Miotti, L. (2024). On the peer review reports: Does size matter? *Scientometrics*, 1–21.
- Manchikanti, L., Kaye, A., Boswell, M., & Hirsch, J. (2015). Medical journal peer review: Process and bias. *Pain Physician*, 18 1, E1–E14.
- Marietta, M., & Barker, D. (2019). The Psychology of Fact Perceptions II. One Nation, Two Realities. <https://doi.org/10.1093/OSO/9780190677176.003.0006>
- Marsh, H. W., Bond, N. W., & Jayasinghe, U. W. (2007). Peer review process: Assessments by applicant-nominated referees are biased, inflated, unreliable and invalid. *Australian Psychologist*, 42(1), 33–38.
- Matsui, A., Chen, E., Wang, Y., & Ferrara, E. (2021). The impact of peer review on the contribution potential of scientific papers. *PeerJ*, 9, e11999.
- Mavrogenis, A. F., Quaile, A., & Scarlat, M. M. (2020). The good, the bad and the rude peer-review. In *International Orthopaedics* (Vol. 44, pp. 413–415). Springer.
- Nobarany, S., & Booth, K. S. (2015). Use of politeness strategies in signed open peer review. *Journal of the Association for Information Science and Technology*, 66(5), 1048–1064.
- Park, I.-U., Peacey, M. W., & Munafo, M. R. (2014). Modelling the effects of subjective and objective decision making in scientific peer review. *Nature*, 506(7486), 93–96.
- Peters, D. P., & Ceci, S. J. (1982). Peer-review research: Objections and obligations. *Behavioral and Brain Sciences*, 5(2), 246–255.
- Phillips, J. (2011). Expert bias in peer review. *Current Medical Research and Opinion*, 27, 2229–2233. <https://doi.org/10.1185/03007995.2011.624090>
- Riley, B., & Jones, R. B. (2016). Peer review: Acknowledging its value and recognising the reviewers. *The British Journal of General Practice : The Journal of the Royal College of General Practitioners*, 66 653, 629–630. <https://doi.org/10.3399/BJGP16X688285>

- Roediger, H. L. (1991). Is unreliability in peer review harmful? *Behavioral and Brain Sciences*, 14(1), 159–160.
- Soergel, D., Saunders, A., & McCallum, A. (2013). Open Scholarship and Peer Review: A Time for Experimentation.
- Street, C. T., & Ward, K. (2019). Cognitive Bias in the Peer Review Process: Understanding a Source of Friction between Reviewers and Researchers. *Data Base*, 50, 52–70.
<https://doi.org/10.1145/3371041.3371046>
- Teplitskiy, M., Peng, H., Blasco, A., & Lakhani, K. R. (2022). Is novel research worth doing? Evidence from peer review at 49 journals. *Proceedings of the National Academy of Sciences*, 119(47), e2118046119.
- Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Reviewer bias in single-versus double-blind peer review. *Proceedings of the National Academy of Sciences*, 114(48), 12708–12713.
- Travis, G. D. L., & Collins, H. M. (1991). New light on old boys: Cognitive and institutional particularism in the peer review system. *Science, Technology, & Human Values*, 16(3), 322–341.
- Tvina, A., Spellecy, R., & Palatnik, A. (2019). Bias in the Peer Review Process: Can We Do Better? *Obstetrics & Gynecology*. <https://doi.org/10.1097/AOG.0000000000003260>
- Wade, A. D. (2022). The semantic scholar academic graph (s2ag). *Companion Proceedings of the Web Conference 2022*, 739–739.
- Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8), 1416–1436.
- Ward, P., Graber, K. C., & Van Der Mars, H. (2015). Writing quality peer reviews of research manuscripts. *Journal of Teaching in Physical Education*, 34(4), 700–715.
- Wenneras, C., & Wold, A. (2001). Nepotism and sexism in peer-review.
- Wolfram, D., Wang, P., Hembree, A., & Park, H. (2020). Open peer review: Promoting transparency in open science. *Scientometrics*, 125(2), 1033–1051.
- Yuan, W., Liu, P., & Neubig, G. (2022). Can we automate scientific reviewing? *Journal of Artificial Intelligence Research*, 75, 171–212.

© [CC-BY-NC 4.0](#) The Author(s). For more information, see our [Open Access Policy](#).

Appendix

Supplementary materials

See [supplementary materials](#).