# Impartial Peer Selection: An Annotated Reading List

OMER LEV
Ben-Gurion University
and
HARPER LYON
Tulane University
and
NICHOLAS MATTEI
Tulane University

The study of peer selection mechanisms presents a unique opportunity to understand and improve the practice of a group selecting its best members, despite each member of that group wanting to be selected. A prime example of such a setting is academic peer review, for which peer selection offers a variety of improvement directions. We present an annotated reading list covering the foundations of peer selection as well as recent and emerging work within the field.

Categories and Subject Descriptors: B.6.3 [**Logic Design**]: Theory of computing—*Algorithmic mechanism design*

General Terms: Algorithms, Performance, Theory

Additional Key Words and Phrases: Computational social choice, mechanism design, strategic agents

Peer selection refers to any social choice problem where agents are asked to select some subset of themselves to receive an award or benefit, which each of them would like to receive. Perhaps most notably, this serves as a (worst case) model for the way academic conferences conduct peer review, a key element of modern science.

Academic reviews and grant boards have been the subject of many empirical studies that, among other issues, have focused on the effectiveness and limits of the system, often substantiating anecdotal complaints of bias and inaccuracy [Cole et al. 1981; McNutt et al. 1990; Wenneras and Wold 1997]. These conclusions, as well as demands of practicality, suggest that broadening the base of reviewers is essential for improving the process of peer review. However, increasing the size of the reviewer pool is not without serious challenges, not the least of which is the sourcing of additional reviewers. A common solution – also proposed by the 2009 National Science Foundation's (NSF) Mechanism Design Proposal Pilot – is to require those that submit proposals to also act as reviewers, casting the problem of peer review firmly into the domain of peer selection.

Moving beyond a relatively small group of impartial experts requires robust mechanisms for soliciting and aggregating peer evaluations. Not only to support larger volumes of reviews but also because involving self-interested agents in the review process introduces new game-theoretic challenges. While we still want to select the best agents/work, we also need to worry about reviewers strategically manipulating their reviews to increase their own chances of being selected. We call a mechanism which does not incentivize strategic reviews *impartial* or *strategyproof.* The papers below highlight both the core theoretical questions underlying the tension between selecting the best work and impartiality, as well as potential mechanisms for better, more robust peer selection. In addition to the papers described below, Olckers and Walsh [2022] provides a comprehensive overview many peer review algorithms and Shah [2022] is a deeper dive into the practical considerations of conference peer review.

(1) Merrifield, M. R. and Saari, D. G. 2009. Telescope time without tears: a distributed approach to peer review. *Astronomy & Geophysics 50*, 4, 4-16

Motivated by the overwhelming number of applications to powerful telescopes, this paper presents an early application of mechanism design to the problem of peer evaluation, and uses a Borda count across submitters' rankings of a subset of proposals to determine acceptance. The mechanism also rewards reviewers according to the similarity of their reviews to the aggregate ranking, aiming to incentivize high quality reviews. This is not an impartial mechanism [Ardabili and Liu 2013], since agents are rewarded for reviews that agree with the community consensus, but the NSF using it as a basis for a pilot program, spurred research in this domain.

(2) Alon, N., Fischer, F. A., Procaccia, A. D., and Tennenholtz, M. 2011. Sum of us: strategyproof selection from the selectors. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2011), Groningen, The Netherlands, July 12-14, 2011*, K. R. Apt, Ed. ACM, 101-110

This paper explores the trade-off between impartiality and optimality in peer selection when we have approval votes. Alon et al. prove the impossibility of any deterministic impartial mechanism that finitely approximate the optimal selection, as well as to determine general bounds for randomized impartial mechanisms within the approval voting context. This work initiated a strand of research into partition mechanisms as a means of achieving good approximation guarantees, particularly when selecting a single agent [Holzman and Moulin 2013; Fischer and Klimm 2015; Bousquet et al. 2014]

(3) Kurokawa, D., Lev, O., Morgenstern, J., and Procaccia, A. D. 2015. Impartial peer review. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015*, Buenos Aires, Argentina, July 25-31, 2015, Q. Yang and M. J. Wooldridge, Eds. AAAI Press, 582-588

With the goal of expanding impartial mechanisms beyond the approval voting context and into settings where reviewers submit rankings or scores for proposals, this paper presents CredibleSubset. The CredibleSubset is an impartial peer selection mechanism which removes the potential for manipulation by building a set of all agents who might have been selected had they strategically manipulated and then randomly selecting agents from this set. This mechanism is proven to be impartial and provides a $k/_{k+m}$ approximation of their VANILLA mechanism (which is a randomized version of Borda), where $k$ is the number of desired selections and $m$ is the number of proposals each reviewer is assigned. However, to maintain impartiality, there is a probability that the mechanism returns no selection (or returns some other default option, not based on the evaluation of peers).

(4) XU, Y., ZHAO, H., SHI, X., AND SHAH, N. B. 2019. On strategyproof conference peer review. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. IJCAI'19. AAAI Press, 616-622

This paper is interested in creating an impartial reviewer assignment when we have a known set of reviewer conflicts and presents DIVIDE-AND-RANK, which assigns reviewers in a way that avoids conflicts of interests but maintains as much "expertise to review" as possible. As long as the conflict graph obeys certain size and density properties, this graph contraction-based approach is shown to be impartial and is shown to perform well empirically on ICLR submissions and reviews.

(5) AZIZ, H., LEV, O., MATTEI, N., ROSENSCHEIN, J. S., AND WALSH, T. 2019. Strategyproof peer selection using randomization, partitioning, and apportionment. *Artif. Intell. 275*, 295-309

One shortcoming of partition based mechanisms is their susceptibility to adversarial partitioning, i.e., if all the best proposals appear on one partition. To address this issue this paper presents EXACTDOLLARPARTITION, a randomized impartial mechanism in which agents are partitioned into clusters, and the number of agents selected from each cluster is apportioned by the relative scores of the agents in that cluster. While the agents selected from each cluster are the highest-ranked ones, the key insight is that the total number can vary, avoiding some problems with adversarial partitioning.

(6) JECMEN, S., ZHANG, H., LIU, R., SHAH, N. B., CONITZER, V., AND FANG, F. 2020. Mitigating manipulation in peer review via randomized reviewer assignments. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds

This paper is concerned with the more concrete problem of minimizing reviewer conflict of interest, in the interpersonal sense, when assigning conference review loads. As a result, the paper focuses on the more practical and empirical aspects of peer review, such as mismatched reviewer expertise, miscalibration between reviewer scores, and the role of group dynamics in shaping discussion. While much of this work is outside of the scope of this reading list, it is an important area of related research. Some additional examples of this style of work can be found in Ardabili and Liu [2013] and Goldberg et al. [2023].

(7) Noothigattu, R., Shah, N. B., and Procaccia, A. D. 2021. Loss functions, axioms, and peer review. *J. Artif. Intell. Res. 70*, 1481-1515

Multi-category reviews, where reviewers evaluate papers on multiple factors such as significance and readability, are common in real-life peer selection, but aggregating the scores across factors can be challenging. This paper takes a machine learning/learning to rank perspective to reproduce implicit community ranking standards. By using empirical risk minimization and extending the standard $L_p$ norms to multidimensional reviews for use as a loss function, they show that aggregation mechanism derived from the equivalent of the $L_1$ norm uniquely produce an impartial mechanism with good recall of the original rankings.

(8) Cembrano, J., Fischer, F. A., Hannon, D., and Klimm, M. 2022. Impartial selection with additive guarantees via iterated deletion. In *EC '22: The 23rd ACM Conference on Economics and Computation, Boulder, CO, USA, July 11 – 15, 2022*, D. M. Pennock, I. Segal, and S. Seuken, Eds. ACM, 1104-1105

Taking inspiration from elimination style voting mechansims like IRV and seeking mechanisms with good approximations bounds for optimal peer selection, this paper describes a "twin threshold" mechanism for approval voting peer selection, where approval votes are removed from agents below a threshold of approval and then selects the agents receiving the maximum non-deleted votes, provided that the agent received more votes than the second, higher, threshold. This mechanism is shown to be impartial and achieves an additive approximation guarantee of $O(n^{1+k}2)$, for $n$ the number of agents and $k \in [0, 1]$, but only when agents' number of approval votes can be bounded by $O(n^k)$.

(9) Aziz, H., Micha, E., and Shah, N. 2023. Group fairness in peer review. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 – 16, 2023*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds

Uniquely for this list, this paper is concerned with a problem other than impartial reviews, as Aziz et al. consider the problem of balancing review assignments for large conferences between research communities to avoid requiring reviewers to evaluate papers far outside of their specialties. Their approach is to extend the game theoretic notion of core fairness to the review assignment setting, and present CoBRA, an algorithm which determines a core-respecting assignment on any set of separable and consistent preferences over the assignments.

(10) Lev, O., Mattei, N., Turrini, P., and Zhydkov, S. 2023. Peernomina-

tion: A novel peer selection algorithm to handle strategic and noisy assessments. *Artif. Intell. 316*, 103843

It is an unfortunate reality that reviewers may be imperfectly accurate when it comes to selecting the best papers, but how do we deal with this noise? To address this issue this paper presents PeerNomination, an impartial peer selection mechanism that is able to account for the presence of noisy and/or unreliable reviewers. This mechanism dynamically weighs reviewers' scores, comparing them to other reviewers of the same agents, thus reducing their impact on the outcome. In order to maintain the mechanism's impartiality despite this new manipulation direction, the reviewer assignments have to follow a particular structure, for which an algorithm is also given.

## Acknowledgements

Harper Lyon and Nicholas Mattei was supported in part by NSF Awards IIS-RI-2007955, IIS-III-2107505, and IIS-RI-2134857 as well as the Tulane University Jurist Center for Artificial Intelligence and the Tulane University Center for Community-Engaged Artificial Intelligence. Omer Lev was supported, in part, by NSF-BSF grant #2021659, and by Israel Science Fund (ISF) grants #1965/20 and #3152/20.

## REFERENCES

Ardabili, P. N. and Liu, M. 2013. Incentives, quality, and risks: A look into the NSF proposal review pilot. *CoRR abs/1307.6528.*

Bousquet, N., Norin, S., and Vetta, A. 2014. A near-optimal mechanism for impartial selection. In *Proceedings of the 10th International Conference on Web and Internet Economics (WINE).* Beijing, China, 133–146.

Cole, S., Cole, J., and Simon, G. 1981. Chance and consensus in peer review. *Science 214,* 4523, 881–886.

Fischer, F. and Klimm, M. 2015. Optimal impartial selection. *SIAM Journal on Computing 44,* 5, 1263–1285.

Goldberg, A., Stelmakh, I., Cho, K., Oh, A. H., Agarwal, A., Belgrave, D., and Shah, N. B. 2023. Peer reviews of peer reviews: A randomized controlled trial and other experiments. *CoRR abs/2311.09497.*

Holzman, R. and Moulin, H. 2013. Impartial nominations for a prize. *Econometrica 81,* 1 (January), 173–196.

McNutt, R. A., Evans, A. T., Fletcher, R. H., and Fletcher, S. W. 1990. The effects of blinding on the quality of peer review: A randomized trial. *The Journal of the American Medical Association 263,* 10, 1371–1376.

Olckers, M. and Walsh, T. 2022. Manipulation and peer mechanisms: A survey. *CoRR abs/2210.01984.*

Shah, N. B. 2022. Challenges, experiments, and computational solutions in peer review. *Commun. ACM 65,* 6, 76–87.

Wenneras, C. and Wold, A. 1997. Nepotism and sexism in peer-review. *Nature 387*, 341–343.