Conference Design With Strategic Authors

Yichi Zhang[1], Behzad Nabavi[2], and Grant Schoenebeck[1]

[1]University of Michigan
[2]Sharif University of Technology

Abstract

Conferences, especially computer science conferences, are seeing a surge in annual submissions. What follows is an increase in the review workload and perhaps a decrease in review quality. The tableau of different conferences and acceptance types (e.g. oral and poster presentation), which we call conference design, drives the strategic submission and production of authors, which in turn affects the review burden and author satisfaction in our scientific publication environment. We aim to understand whether a large conference should treat all accepted papers equally, adopt hierarchical acceptance (where papers can be first considered for oral acceptance and then poster acceptance), or opt for a division into two conferences (where authors should choose where to submit to in advance). Building upon prior work, our model captures several key interactions between authors and conference(s), considering authors' resubmission after rejection and their utility upon acceptance being tied to the average quality of accepted papers. We provide comprehensive discussions on which design works the best, and, more importantly, why and when it excels. Our model predicts that in most natural settings, compared with the hierarchical design, separating the conference helps preserve the reputation of the separated (sub)conferences but often experiences a higher review burden and a lower author utility.

1 Introduction

Conference peer review, the process of selecting high-quality papers for publication in a short period of time, is the key to maintaining a healthy scientific publication environment. However, the reliability of conference peer review has been greatly challenged by the record-breaking volume of submissions in recent years, especially for AI conferences.[1] Unfortunately, the growth of qualified reviewers has not kept pace with the surge in submissions [10, 17], inevitably resulting in a noisier peer review system.

This leads to a vicious cycle: with a noisier review system, top conferences have to raise their selection bars to maintain their high prestige, which results in more rejections and resubmissions of the same papers and thus a higher review burden. This further results in a counter-intuitive phenomenon, known as the *resubmission paradox* [23], where about 75% of submissions are rejected every year, yet many of them are eventually accepted at similar venues with minor or even no improvements. A recent study suggests that about 36% of the rejected papers from NeurIPS 2014 were later accepted at other prestigious CS conferences such as ICML, CVPR, and subsequent NeurIPS [2]. While some rejections are necessary to maintain the prestige of conferences, the resulting increase in resubmissions significantly adds to the review burden, diminishes author utility, and likely undermines diversity in the community[2].

One widely discussed idea is to accept every paper that is "above the bar". This idea, which we refer to as *hierarchical design*, suggests that all submissions should undergo a light review to be presented at the conference, while a subset of remarkable papers are carefully selected for distinguished recognition. For clarity, we refer to the more prestigious segment as *top conference* and the other as *secondary conference*. A significant challenge in designing the acceptance policy is managing the authors' feedback loop. For

[1]As of the registration date of this paper, the number of submissions to NeurIPS 2024 had surpassed 20,000.

[2]For example, repeated rejections may impact authors unequally. Those from more non-traditional backgrounds may become discouraged [13] and repeatedly submitting and revising the same papers may create additional barriers for authors with access to fewer resources.

example, an improperly low acceptance threshold for the secondary conference may attract a large volume of substandard submissions, which could in turn overwhelm the conference.

Another idea of mitigating the review burden is to utilize authors' information about the quality of their own papers [24, 23, 15]. The idea proposes dividing one large conference into two smaller conferences, each with its own independent review system and distinct audience. Under this *separating design*, authors reveal the quality of their work by choosing which conference to submit to. The system can leverage this self-assessment by carefully adjusting the acceptance thresholds of each conference, aiming to reject fewer papers overall.

In this paper, we study conference design, the problem of designing the tableau, and the acceptance policies of a conference while considering the incentives of authors. Building upon a prior work [23], we capture the interactions between the conference(s) and a set of authors who decide how to submit their papers as a mechanism design problem. Through both theoretical and empirical analysis, we discuss how various designs shape authors' best responses and affect the review burden of the system.

Our Model Our model considers a set of authors submitting their papers to a conference that applies various designs. The authors' papers are characterized by their qualities drawn from a commonly known (continuous) distribution. The conference aims to accept more positive-quality papers and reject more negative-quality papers. The quality of a paper is private to the author, while the conference's judgment relies on noisy reviews. For each paper of quality q, we assume the conference observes a review score of $s = q + \epsilon$ where ϵ is i.i.d. sampled from a known review noise distribution. The conference commits to a threshold acceptance policy such that only papers with review scores higher than the acceptance threshold will be accepted. Note that there are two acceptance thresholds in the hierarchical design and the separating design, corresponding to the top and secondary conference.

Authors respond to the conference's acceptance policy by deciding whether to submit (or resubmit) their papers to the conference (and if so, to which sub-conference), or to take an outside option. The utility of taking the outside option, such as putting the paper on Arxiv, always returns a utility of 1. The acceptance of a paper at the conference provides the author a utility equal to the conference value, which is modeled by the average quality of the accepted paper at that conference. The utilities of acceptance at the conference and the outside option are both time-discounted in the number of resubmissions. Based on their private information about paper qualities, authors best respond by taking the utility-maximizing action. We assume paper qualities are exogenous, meaning that qualities will not be improved within the resubmission process.

Although our model is necessarily a simplification, we believe the fundamental insights derived from our analysis (summarized below) remain robust across a reasonable range of variants of our model (see Section 5). We thus envision that our theoretical results (and simulations) can steer the discussion, uncover parameters to focus on, and inform decision-makers in practice.

Our Results We generalize the key insights about authors' best response in [23] from a one-conference setting to scenarios with two (hierarchical or separated) conferences. In spite of the acceptance bar set by the conferences, what determines the papers accepted by each conference is the authors' *de facto threshold*: a quality bar above which every paper will be submitted and resubmitted until acceptance, while below which no paper will be submitted.

- In the separating design, we show that under mild conditions, authors' best response can be captured by two de facto thresholds: a higher de facto threshold θ_1 above which authors submit and keep resubmitting only to the top conference, a lower threshold θ_2 where papers with quality between θ_2 and θ_1 are (re)submitted only to the secondary conference and below which authors immediately take the outside option.

- In the hierarchical design, we observe a similar form of authors' best response. Authors whose papers' qualities are above a higher de facto threshold θ_1 are only interested in the top conference and will withdraw their papers if they are accepted by the secondary conference; authors whose papers' qualities are between θ_2 and θ_1 will buy a lottery ticket to the top conference but are happy to embrace the secondary conference if accepted. Lastly, authors with papers' qualities lower than θ_2 will immediately take the outside option.

Because of the existence of θ_1 in the hierarchical design, we first warn the community about the idea of "accepting everything above the bar". An improperly low acceptance threshold of the secondary conference may turn it into another outside option for the authors. Therefore, authors who believe that they deserve something better will keep withdrawing and resubmitting until they get into the top conference, in which case, no review burden is saved.

We further compare various designs while conditioned on inducing the same authors' best response. Our first observation is that both the top conference and the secondary conference in the hierarchical design experience lower values than in the separating design conditioned on the same θ_1 and θ_2. Intuitively, because of the review noise, the top conference in the hierarchical design accepts a batch of papers that are sub-standard for it but are top papers for the secondary conference.

A higher conference value attracts more authors to submit, resulting in a larger gap between the acceptance threshold and the de facto threshold. Therefore, in most of the natural settings, we observe lower acceptance thresholds for both conferences in the hierarchical design. This means that conditioned on the same authors' best response, both conferences in the hierarchical design can accept submissions with fewer rounds of rejections and resubmissions, contributing to a lower review burden. However, when the review noise is large, the benefit of having a good chance of getting into the top conference outweighs the decrease in the conference value for authors with borderline papers. This leads to a lower acceptance threshold of the secondary conference in the separating design, in which case we observe a lower review burden.

1.1 Related Work

Zhang et al. [23] present a model that the current paper builds upon. In their model, the author makes a binary decision of whether to submit to a single top-tier conference or take a fixed-utility outside option. They focus on how to design the (threshold) acceptance polity to achieve the Pareto optimal tradeoff between the conference quality and the review burden. The former is modeled by the sum of the paper qualities accepted by the conference, while the latter is the total number of reviews solicited at one round of review. The key difference of our model lies in the conference values, i.e. the rewards of having an accepted paper at the conferences. The conference value is assumed to be a constant in [23], while in our model, it is tied to the average quality of the accepted papers. This allows us to investigate the performance of various designs that involve the competition between various conferences.

Considering the strategic behaviors of authors, there is a stream of works that consider eliciting authors' information about their paper qualities to assist conference peer review. Su [19] proposes the isotonic mechanism that elicits a ranking of paper qualities from an author with multiple submissions. The follow-up works generalize the idea by considering coauthorship [21] and more complicated review noise [22]. However, the isotonic mechanism is strategy-proof, meaning that truth-telling is the author's best response, only when authors' utilities are convex w.r.t. the review scores. [24] further propose the sequential review mechanism while only assuming authors' utilities are increasing w.r.t. the paper qualities conditioned on acceptance.

Despite the theoretical contributions, a substantial amount of empirical studies aim to understand and improve peer review from the reviewers' perspectives, such as review calibration [12, 14], bias [7, 5], and strategic behaviors [1, 3]. Speaking of review bias, for example, Stelmakh et al. [18] reveal a negative bias of reviewers while observing that the paper is a resubmission. Tomkins et al. [20] examine the bias of single-blind and double-blind reviews. Speaking of strategic behaviors, Stelmakh et al. [16] concern the misreporting of peer graders and reviewers in an attempt to improve their own outcome, where a test to detecting strategic behavior is proposed. Speaking of eliciting various forms of information, Shah et al. [11] conduct an experiment evaluating the effectiveness of collecting ordinal rankings from reviewers. More papers investigate the bidding process of peer review [9, 4, 6, 8].

2 Model

Building upon prior work [23], we model the interaction between authors and a conference (or multiple conferences) as a Stackelberg game. We first introduce the one-conference design where authors choose between one prestigious conference and an outside option. Then, in Section 2.1, we discuss the hierarchical design and the separating design.

Suppose an author has a paper whose quality q is sampled from a known continuous prior $\mathbf{p}$ with support $\mathbb{R}$. To simplify the analysis and avoid edge cases, we assume $\mathbf{p}$ has full support on $\mathbb{R}$, meaning that any real-valued quality has a non-zero probability.[3] Quality, in this context, signifies the level of desirability for acceptance by the conference; for instance, a quality of 0 indicates a borderline paper. We assume the quality is known to the author.[4] The author faces a binary decision: submit the paper to the conference and risk rejection, or opt for the outside option that yields a utility of 1. The outside option may capture a second-tier conference or ArXiv.

For each submitted paper, the conference assesses its quality based on a noisy review score r such that $r = q + X$, where the review noise X is i.i.d. sampled for each paper according to a known distribution. We denote f_X as the p.d.f. and F_X as the c.d.f. where we, as mentioned, assume F_X is strictly increasing on $\mathbb{R}$.

In our simulations, we frequently use a Gaussian prior model that is characterized by four hyperparameters: μ_q and σ_q are the mean and the standard deviation of the Gaussian paper quality prior, σ_r is the standard deviation of the zero-mean Gaussian review noise, and η is the author's discount factor.

The conference implements an acceptance policy that determines the probability of a paper's acceptance. We focus on the threshold acceptance policy which is shown to be optimal in terms of the review burden in this setting [23].

Definition 1. *A threshold policy is characterized by the acceptance threshold τ such that a paper is accepted if and only if its review score $r \geq \tau$.*

Under the threshold acceptance policy, the probability of a paper of quality q is $\mathrm{P}_{acc}(\tau, q) \equiv 1 - F_X(\tau - q)$.

If the paper is accepted, the author receives a utility U equal to the conference value V. The conference value is assumed to be the average quality of the papers accepted at the conference. This is the key feature that distinguishes our model from [23] and will make our model suitable for analyzing the setting of more than one conference.

If the paper is rejected, the author faces the same binary decision in the next round. However, both the author's utility of acceptance and the utility of taking the outside option are exponentially discounted by a factor of η. That's to say, after k rounds of rejections, a paper accepted by the conference yields a utility of $\eta^t V$, while taking the outside option yields a utility of η^t. We allow the submission and resubmission process to be infinitely repeated and assume the paper quality is fixed.

Because authors observe the true quality of their papers and authors' utility is equally discounted between both actions (submit and quit), rejections will not change authors' decisions. That is, if an author prefers to submit in the first round, she prefers to do so no matter how many times her paper is rejected. This induces a threshold strategy of the author.

Proposition 2.1. *The author employs a threshold strategy, characterized by a threshold θ, if she submits and keeps resubmitting until the paper is accepted whenever $q \geq \theta$; and takes the outside option immediately if $q < \theta$.*

The threshold strategy is identified as one of the best responses to the conference's acceptance policy [23]. A detailed proof is provided in Appendix A.1. Intuitively, the best response is a threshold strategy because the author's utility for submission is increasing in the paper's quality.

We call θ the *de facto threshold* because it is the actual threshold that determines the set of papers that are (eventually) accepted by the conference. We further call the difference between τ and θ the *resubmission gap*, which captures how much lower the paper quality is accepted by the conference compared with the acceptance threshold set by the conference.

2.1 Designs

Here, we introduce the three designs discussed in this paper.

1. **One-conference Design.** This is the setting discussed above, where there is one large conference that attracts every author above the de facto threshold θ to submit (and keep resubmitting).

[3] We emphasize that the main takeaways from our results do not depend on this assumption.

[4] Zhang et al. [23] discuss a model where authors only observe noisy signals about their papers' quality. However, this model greatly complicates the analysis but makes little difference to the main results.

2. **Separating Design.** Separating design breaks the conference into two smaller independent conferences, each has its own review system and attracts its own set of authors. Let τ_1 and τ_2 be the acceptance thresholds of two divided conferences respectively. We call the one with acceptance threshold τ_1 the top conference and the one with τ_2 the secondary conference, where $\tau_1 > \tau_2$. At each round, the author has three options: submit to the top conference, submit to the secondary conference, or take the outside option.

3. **Hierarchical Design.** The hierarchical design sets two thresholds $\tau_1 > \tau_2$ such that papers with review score $r \geq \tau_1$ are accepted as oral presentations (top conference), while papers with review score $\tau_2 \leq r < \tau_1$ are accepted as poster presentations (secondary conference). At each round, an author with a paper that has not been accepted chooses whether to submit to the hierarchical conference or take the outside option. An author with a paper that was accepted by the secondary conference in the previous round chooses whether to embrace the secondary conference or withdraw the paper and resubmit in an attempt of getting into the top conference.

   In this design, papers accepted by the top conference include those whose authors persist in the top conference and those whose authors are open to the secondary conference but are lucky enough to be accepted by the top conference. On the other hand, papers accepted by the secondary conference consist of the rejections from the top conference whose authors are open to the secondary conference.

We are interested in investigating how various designs affect social welfare of the review system. We consider two dimensions of welfare: the review burden and the average author utility. We defer the formal definitions of both to Section 4.4. Our investigation requires us to first examine the authors' best responses.

3 Computing Author Best Response

We first emphasize that the game defined in Section 2 may encompass multiple equilibria, including mixed strategy equilibria. Nonetheless, we focus on threshold equilibria both because they can be easily characterized by one or two thresholds, making them suitable for a clear and intuitive depiction of the effects of different designs, and because they are more natural and therefore likely to arise in practice.

3.1 One Conference

In the original design, there is one conference which does not distinguish the accepted papers. In response to the conference's threshold acceptance strategy, we have shown that the threshold strategy, characterized by the de facto threshold θ, is (one of) the author's best responses. Now, we present the relationship between the acceptance threshold τ and the de facto threshold θ.

The conference value can be written as a function of θ,

$$V^o(\theta) = \frac{\int_\theta^\infty q\mathbf{p}(q)dq}{\int_\theta^\infty \mathbf{p}(q)dq},$$

where the superscript represents the one-conference design.

Let $U^o(q|\tau)$ be the author's utility of submitting a paper of quality q to the conference in the one-conference design. We sometimes omit the dependence on τ and only write the utility as a function of paper quality, i.e. $U^o(q)$. Let $\mathrm{P}_{acc}(\tau, q) = \Pr(X \geq \tau - q) = 1 - F_X(\tau - q)$. Therefore, $\mathrm{P}_{acc}(\tau, q)$ indicates the probability of a paper of quality q is accepted by the conference with acceptance threshold τ in one round of review. The expected utility of submitting for an author with a paper of quality q is thus

$$U^o(q) = \mathrm{P}_{acc}(\tau, q)V^o(\theta) + \eta(1 - \mathrm{P}_{acc}(\tau, q))\mathrm{P}_{acc}(\tau, q)V^o + \cdots = \frac{\mathrm{P}_{acc}(\tau, \theta)V^o(\theta)}{1 - \eta(1 - \mathrm{P}_{acc}(\tau, \theta))}.$$

$$
\begin{aligned}
U^o(\theta|\tau) &= \frac{\mathrm{P}_{acc}(\tau,\theta)V^o(\theta)}{1-\eta(1-\mathrm{P}_{acc}(\tau,\theta))} = 1 \\
\Rightarrow \quad P_{acc}(\tau,\theta) &= \frac{1-\eta}{V^o(\theta)-\eta} \\
\Rightarrow \quad \tau &= \theta + F_X^{-1}\left(\frac{V^o(\theta)-1}{V^o(\theta)-\eta}\right). \qquad (1)
\end{aligned}
$$

Equation (1) allows us to compute the corresponding acceptance threshold τ given a de facto threshold θ, and vice versa.

3.2 Separating Design

In the separating design, the large conference is divided into a top conference and a secondary conference, with acceptance thresholds τ_1 and τ_2 respectively. The author decides which one of these two conferences to submit her paper to or take the outside option.

Existence of the threshold best response We show that under some conditions, the following threshold strategy is the best response to the separating design.

- *Only Top*: submit (or resubmit) to the top conference until acceptance.
- *Only Secondary*: submit (or resubmit) to the secondary conference until acceptance.
- *Quit*: take the outside option and get a utility of 1.

We first write down the expected utility of each action. Let $V_1^s, V_2^s > 1$ be the values of the top and secondary conference respectively. Recall that $\mathrm{P}_{acc}(\tau, q) = 1 - F_X(\tau - q)$ is the probability of acceptance. Then, the author's utility of submitting to the top conference and the secondary conference denoted as U_A^s and U_B^s respectively, can be written as:

$$
U_A^s(q) = \frac{\mathrm{P}_{acc}(\tau_1 - q)V_1^s}{1-\eta(1-\mathrm{P}_{acc}(\tau_1-q))}, \qquad U_B^s(q) = \frac{\mathrm{P}_{acc}(\tau_2 - q)V_2^s}{1-\eta(1-\mathrm{P}_{acc}(\tau_2-q))}. \qquad (2)
$$

Proposition 3.1. *In the separating design, fixing acceptance threshold τ_1 and τ_2, if the review noise satisfies that $\frac{f_X}{(1-\eta F_X)^2}$ is increasing, there exist thresholds θ_1 and θ_2 where an author prefers to submit to the top conference if $q \geq \theta_1$, to the secondary conference if $\theta_2 \leq q < \theta_1$, and opts to quit if $q < \theta_2$.*

Proposition 3.1 provides a sufficient condition for the existence of the authors' threshold best response. The condition requires $\frac{f_X}{(1-\eta F_X)^2}$ to be increasing, which is a stronger assumption than the increasing hazard rate of the review noise. It is straightforward to see that when the review noise has an increasing hazard rate and $\eta = 1$, $\frac{f_X}{(1-\eta F_X)^2}$ is increasing. However, when η is small, the condition is violated. We note that this condition is sufficient but not necessary to induce the threshold best response. In our numerical experiments, we empirically show that the author still takes a threshold best response even when this monotonicity condition is violated in our considered settings where the review noise is smooth and has an increasing hazard rate.

Computing best response Given θ_1 and θ_2, we can easily compute the conference values.

$$
V_1^s(\theta_1) = \frac{\int_{\theta_1}^{\infty} q\mathbf{p}(q)dq}{\int_{\theta_1}^{\infty} \mathbf{p}(q)dq}, \qquad V_2^s(\theta_1,\theta_2) = \frac{\int_{\theta_2}^{\theta_1} q\mathbf{p}(q)dq}{\int_{\theta_2}^{\theta_1} \mathbf{p}(q)dq}.
$$

Then, we have two constraints that build the relationship between four thresholds. First, authors with paper quality θ_2 are indifferent between submitting to the secondary conference and the outside option, i.e. $U^s_B(\theta_2) = 1$. We use the subscript to indicate to which conference the author is submitting (1 for the top conference and 2 for the secondary conference).

$$
\begin{aligned}
U^s_B(\theta_2|\tau_2, \theta_1) = \frac{\mathrm{P}_{acc}(\tau_2 - \theta_2)V^s_2(\theta_1, \theta_2)}{1 - \eta(1 - \mathrm{P}_{acc}(\tau_2 - \theta_2))} = 1 \qquad (3)\\
\Rightarrow \quad \tau_2 = \theta_2 + F_X^{-1}\left(\frac{V^s_2(\theta_1, \theta_2) - 1}{V^s_2(\theta_1, \theta_2) - \eta}\right). \qquad (4)
\end{aligned}
$$

The second constraint requires that authors with paper quality θ_1 are indifferent between submitting to the top and secondary conferences.

$$
\begin{aligned}
& U^s_A(\theta_1|\tau_2, \tau_1, \theta_2) = U^s_B(\theta_1|\tau_2, \theta_1) \\
\Rightarrow \quad & \frac{\mathrm{P}_{acc}(\tau_1 - \theta_1)V^s_1(\theta_1)}{1 - \eta(1 - \mathrm{P}_{acc}(\tau_1 - \theta_1))} = \frac{\mathrm{P}_{acc}(\tau_2 - \theta_1)V^s_2(\theta_1, \theta_2)}{1 - \eta(1 - \mathrm{P}_{acc}(\tau_2 - \theta_1))} \qquad (5)\\
\Rightarrow \quad & \tau_1 = \theta_1 + F_x^{-1}\left(\frac{(1 - \eta + \eta\mathrm{P}_{acc}(\tau_2 - \theta_1))V^s_1(\theta_1) - \mathrm{P}_{acc}(\tau_2 - \theta_1)V^s_2(\theta_1, \theta_2)}{(1 - \eta + \eta\mathrm{P}_{acc}(\tau_2 - \theta_1))V^s_1(\theta_1) - \eta\mathrm{P}_{acc}(\tau_2 - \theta_1)V^s_2(\theta_1, \theta_2)}\right). \qquad (6)
\end{aligned}
$$

Given any two thresholds, Eq. (4) and Eq. (6) allow us to compute the other two thresholds (if feasible) in the separating design. Again, a pair of thresholds is called feasible if there exists a solution of all four thresholds to the above two equations.

3.3 Hierarchical Design

The hierarchical design allows the author to either simultaneously submit to both the top and the secondary conference or to submit only to the top conference (essentially withdrawing her paper if it does not get into the top conference). Unfortunately, this design is complex and we fail to obtain any guarantee on the threshold best response without strong assumptions. However, analogous to the previous design, we assume the author takes a threshold response and can numerically verify its robustness in our experiments.

- *Insist on Top*: submit (or resubmit) to the conference until being accepted by the top conference.
- *Open to Secondary*: submit (or resubmit) to the conference until being accepted by either the top or the secondary conference.
- *Quit*: take the outside option and get a utility of 1.

Let V^h_1 and V^h_2 be the values of the top and secondary conference respectively. We first write down the author's utility of persisting in the top conference and her utility of accepting any acceptance denoted as U^h_A and U^h_B respectively.

$$
\begin{aligned}
U^h_A(q) &= \frac{(1 - F_X(\tau_1 - q))V^h_1}{1 - \eta F_X(\tau_1 - q)} \\
U^h_B(q) &= \frac{(F_X(\tau_1 - q) - F_X(\tau_2 - q))V^h_2 + (1 - F_X(\tau_1 - q))V^h_1}{1 - \eta F_X(\tau_2 - q)}.
\end{aligned}
$$

We assume there exists de facto threshold $\theta_1 > \theta_2$ such that the author's best response is: insist on the top conference for any $q \geq \theta_1$, be open to the secondary conference when $\theta_2 \leq q < \theta_1$, take the outside option when $q < \theta_2$. The threshold best response is numerically justified in the simulation settings we consider.

Computing best response Again, let $p_t = \mathrm{P}_{acc}(\tau_1 - q)$, $p_s = \mathrm{P}_{acc}(\tau_2 - q) - \mathrm{P}_{acc}(\tau_1 - q)$, and $p_a = \mathrm{P}_{acc}(\tau_2 - q)$ be the probability that a paper of quality q is accepted by the top conference, secondary conference, and either conference respectively[5]. We first write down the conference values.

$$V_1^h(\theta_1,\theta_2,\tau_1,\tau_2) = \frac{\int_{\theta_1}^{\infty} q\mathbf{p}(q) + \int_{\theta_2}^{\theta_1} q\mathbf{p}(q)(p_t + (1-p_a)p_t + (1-p_a)^2 p_t + \cdots)dq}{\int_{\theta_1}^{\infty} \mathbf{p}(q) + \int_{\theta_2}^{\theta_1} \mathbf{p}(q)(p_t + (1-p_a)p_t + (1-p_a)^2 p_t + \cdots)dq}$$
$$= \frac{\int_{\theta_1}^{\infty} q\mathbf{p}(q)dq + \int_{\theta_2}^{\theta_1} q\mathbf{p}(q)\mathrm{P}_{acc}(\tau_1 - q)/\mathrm{P}_{acc}(\tau_2 - q)dq}{\int_{\theta_1}^{\infty} \mathbf{p}(q)dq + \int_{\theta_2}^{\theta_1} \mathbf{p}(q)\mathrm{P}_{acc}(\tau_1 - q)/\mathrm{P}_{acc}(\tau_2 - q)dq}.$$

$$V_2^h(\theta_1,\theta_2,\tau_1,\tau_2) = \frac{\int_{\theta_2}^{\theta_1} q\mathbf{p}(q)(p_s + (1-p_a)p_s + (1-p_a)^2 p_s \cdots)dq}{\int_{\theta_2}^{\theta_1} \mathbf{p}(q)(p_s + (1-p_a)p_s + (1-p_a)^2 p_s \cdots)dq}$$
$$= \frac{\int_{\theta_2}^{\theta_1} q\mathbf{p}(q)dq - \int_{\theta_2}^{\theta_1} q\mathbf{p}(q)\mathrm{P}_{acc}(\tau_1 - q)/\mathrm{P}_{acc}(\tau_2 - q)dq}{\int_{\theta_2}^{\theta_1} \mathbf{p}(q)dq - \int_{\theta_2}^{\theta_1} \mathbf{p}(q)\mathrm{P}_{acc}(\tau_1 - q)/\mathrm{P}_{acc}(\tau_2 - q)dq}.$$

Then, we write down two constraints. First, authors with paper quality θ_2 are indifferent between submitting to the conference while accepting any outcome and submitting to the outside option.

$$U_B^h(\theta_2|\tau_1,\tau_2,\theta_1,\theta_2) = \frac{p_t V_1^h}{1-\eta(1-p_a)} + \frac{p_s V_2^h}{1-\eta(1-p_a)} = 1$$
$$\Rightarrow \quad \mathrm{P}_{acc}(\tau_1-\theta_2)V_1^h(\theta_1,\theta_2,\tau_1,\tau_2) + (\mathrm{P}_{acc}(\tau_2-\theta_2) - \mathrm{P}_{acc}(\tau_1-\theta_2))V_2^h(\theta_1,\theta_2,\tau_1,\tau_2) = 1 - \eta(1-\mathrm{P}_{acc}(\tau_2-\theta_2))). \tag{7}$$

Second, authors with paper quality θ_1 are indifferent between sticking with submitting to the top conference and being open to the acceptance of the secondary conference.

$$U_A^h(\theta_1|\tau_1,\tau_2,\theta_1,\theta_2) = U_B^h(\theta_1|\tau_1,\tau_2,\theta_1,\theta_2)$$
$$\Rightarrow \quad \frac{\mathrm{P}_{acc}(\tau_1-\theta_1)V_1^h(\theta_1,\theta_2,\tau_1,\tau_2)}{1-\eta(1-\mathrm{P}_{acc}(\tau_1-\theta_1))} = \frac{\mathrm{P}_{acc}(\tau_1-\theta_1)V_1^h(\theta_1,\theta_2,\tau_1,\tau_2) + (\mathrm{P}_{acc}(\tau_2-\theta_1) - \mathrm{P}_{acc}(\tau_1-\theta_1))V_2^h(\theta_1,\theta_2,\tau_1,\tau_2)}{1-\eta(1-\mathrm{P}_{acc}(\tau_2-\theta_1))}$$
$$\Rightarrow \quad \frac{\eta\mathrm{P}_{acc}(\tau_1-\theta_1)}{1-\eta+\eta\mathrm{P}_{acc}(\tau_1-\theta_1)}V_1^h(\theta_1,\theta_2,\tau_1,\tau_2) = V_2^h(\theta_1,\theta_2,\tau_1,\tau_2). \tag{8}$$

Eq. (8) can be interpreted as focusing only on the case where the author's paper is rejected by the top conference but accepted by the secondary conference in the first round. This is because in any other case, the actions *Insist on Top* and *Open to Secondary* have the same utility. The right-hand side of Eq. (8) can thus be interpreted as the utility of embracing the secondary conference, which is exactly V_2^h. The left-hand side can be understood as the utility of insisting on acceptance to the top conference, and so trying again in the next round. Intuitively, when authors are less patient, i.e. when η is small, the utility of sticking with the top conference shrinks. Eq. (8) requires these two utilities to be identical for authors with paper quality of θ_1.

Given any two thresholds, Eq. (7) and Eq. (8) thus allow us to compute the other two thresholds (if feasible) in the hierarchical design.

4 Comparing Conference Design

We are primarily interested in understanding which conference design is the best at improving the social welfare of the system as defined as 1) the review burden and 2) the average author utility (Section 4.4). To better understand how various designs affect the review system, we first investigate how they impact the de facto thresholds (Section 4.1), conference values (Section 4.2), and acceptance thresholds (Section 4.3).

[5] This abuses notation as these quantities depend on q.

4.1 Feasible De Facto Thresholds

De facto thresholds characterize authors' best responses, which affect the burden of the review system, and in turn, impact the authors' own utilities. Therefore, while comparing the conference designs, we frequently conditioned on the same de facto thresholds. This section investigates what de facto thresholds can be induced as an equilibrium under each design, which is a fundamental analysis upon which our downstream comparisons are built.

As defined, a de facto threshold (or a pair of de facto thresholds) is feasible if there exists suitable acceptance thresholds such that authors with papers of quality equal to the de facto thresholds are indifferent between the two options. These constraints are characterized by the equations in Section 3. Intuitively, a necessary condition for de facto thresholds to be feasible in the separating design is that they should induce conference values that are larger than 1, which is the utility of the outside option. Otherwise, no rational author is interested in submitting to such conferences.

One Conference Design We first discuss the feasible de facto thresholds in the one-conference design as an example. In this setting, there is only one de facto threshold θ and one conference value V^o. It turns out that the necessary condition of $V^o > 1$ is also sufficient in this setting. This can be observed from Eq. (1) that any $V^o > 1$ can be mapped to a resubmission gap of $F_X^{-1}\left(\frac{V^o(\theta)-1}{V^o(\theta)-\eta}\right) \in \mathbb{R}$.[6] Furthermore, it is straightforward to show that $V^o(\theta) = \mathbb{E}_{\mathbf{p}}[Q|Q \geq \theta]$ is increasing in θ. Therefore, there exists a threshold $\bar{\theta}^o$ with $V^o(\bar{\theta}^o) = 1$ such that any $\theta > \bar{\theta}^o$ is feasible under the one-conference design.

Separating Design In the separating design, there are two de facto thresholds, corresponding to the top and the secondary conference. We have the following proposition.

Proposition 4.1. *A pair of de facto thresholds (θ_1, θ_2) is feasible under the separating design if and only if*

(a) $\theta_1 > \theta_2$;

(b) $V_2^s(\theta_1, \theta_2) > 1$;

(c) and $V_2^s(\theta_1, \theta_2) < \eta V_1^s(\theta_1) + \frac{1-\eta}{P_{acc}(\tau_2 - \theta_1)}$, *where τ_2 can be computed via Eq. (4) given θ_1 and θ_2.*

Condition (a) is required by the definition of the top conference and the secondary conference. Condition (b) and (c) are directly obtained from Eq. (4) and Eq. (6) respectively, by requiring the inverse of the c.d.f. of the review noise having real-valued solutions. Intuitively, condition (b) requires the secondary conference to be more attractive than the authors' outside option so that it can attract submissions. Condition (c) requires the secondary conference to be not too attractive so that the top conference can attract submissions.

Intuitively, fixing θ_2 and increasing θ_1, the values of both conferences increase although V_1^s increases faster than V_2^s (for natural settings considered in our simulations). Therefore, both conditions are more likely to hold when θ_1 is larger. We thus expect a convex range of feasible de facto thresholds: for any θ_2 that is not too low, any θ_1 that is larger than a threshold is feasible. We denote the minimum feasible θ_2 as $\bar{\theta}_2^s$ such that $V_2^s(\infty, \bar{\theta}_2^s) = 1$. This means that θ_2 is feasible, i.e. there exists a $\theta_1 > \theta_2$ that makes the pair of de facto thresholds feasible, if and only if $\theta_2 > \bar{\theta}_2^s$. Furthermore, by definition, $\bar{\theta}_2^s = \bar{\theta}^o$.

We present an example in Fig. 1 (the blue area). We find that the following patterns are robust across various settings in our simulations. When θ_2 is small, especially when $\theta_2 < 0$, the boundary of the feasible range is characterized by condition (b); when θ_2 is large, any $\theta_1 > \theta_2$ is feasible. Condition (c) is usually not a concern in our settings where the quality prior and the review noise are smooth.

Hierarchical Design The hierarchical design is more complicated to interpret, primarily because the conference values of both the top and the secondary conference depend not only on the de facto thresholds but also on the acceptance thresholds. We have two key findings.

First, there exists a $\underline{\theta}_2^h$ such that every $\theta_2 \leq \underline{\theta}_2^h$ is infeasible. The lower bound $\underline{\theta}_2^h$ exists because the value of the secondary conference cannot be smaller than η. Otherwise, no author is happy with the acceptance of

[6] Note that a negative resubmission gap is meaningful as it indicates the case where the conference has to greatly lower its bar to attract submissions.

the secondary conference because they can at least withdraw the paper and take the outside option in the next round which returns a utility of η. Note that in the separating design, we require $V_2^s > 1$, a stricter requirement than $V_2^h > \eta$ in the hierarchical design. This implies $\underline{\theta}_2^h < \bar{\theta}^s$. However, we do not have the guarantee that any $\theta_2 > \underline{\theta}_2^h$ is feasible under the hierarchical design. Even numerically searching for such edge cases requires high-precision computing[7] which is time-consuming but less insightful. As shown in Fig. 1 (the red area), we leave the range between $\underline{\theta}_2^h$ and -4 blank.

Second, as observed in the example, when $\theta 2 > 1$, any $\theta_1 > \theta_2$ is feasible under the separating design, while the minimum feasible θ_1 under the hierarchical design is much larger. Intuitively, when θ_2 is large, the secondary conference needs to set a high bar in both the separating and the hierarchical design. When θ_1 is not much larger than θ_2, papers with quality between θ_2 and θ_1 have about the same probability of getting into the top conference. This makes the action *Open to Secondary* especially attractive. Therefore, the top conference under the hierarchical design must lower its acceptance threshold τ_1^h so as to attract papers with quality θ_1 not taking the *Open to Secondary* option. However, this usually results in a violation of $\tau_1^h < \tau_2^h$. Nonetheless, because authors can only choose one conference to submit to in the separating design, the *Only Secondary* option is not as attractive as the *Open to Secondary* in the hierarchical design. This means that the top conference can set a high bar in the separating design, making such de facto thresholds feasible.

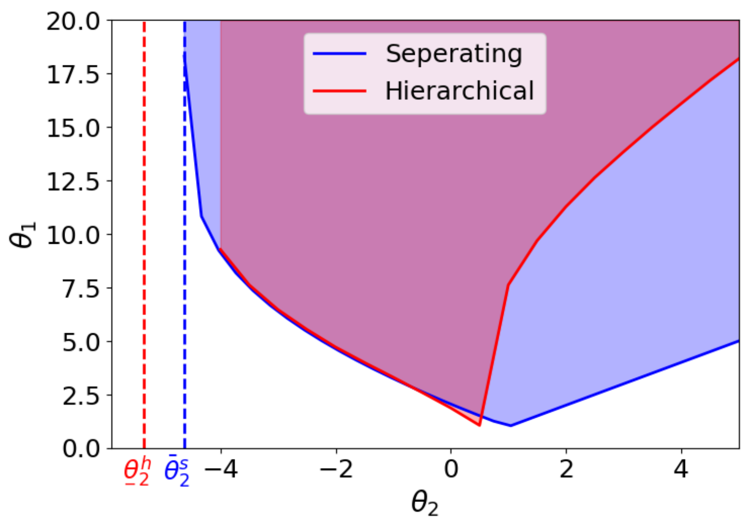


Figure 1: The range of feasible (θ_1, θ_2) under the $(\mu_q = -1, \sigma_q = 5, \sigma_r = 3, \eta = 0.7)$-Gaussian prior model.

4.2 Conference Values

The value of a conference, modeled by the average quality of its accepted papers, determines the reward of being accepted by the conference. Therefore, conference values play a key role in understanding the interactions between authors and the conferences. This section summarizes our insights regarding the impact of different designs on conference values.

Our first observation is that conditioned on the same de facto threshold, compared with the one-conference design, dividing the conference into two parts always results in a lower-value secondary conference and a higher-value top conference. This is intuitive because in every design, low-quality papers are less likely to be accepted by the top conference either because of the higher selection bar (e.g. the hierarchical design) or due to self-selection (e.g. the separating design).

Proposition 4.2. *Suppose the review noise has a strictly increasing hazard rate. Let θ be a de facto threshold of the (secondary) conference that is feasible for any design and let V be the corresponding conference value*

[7] To verify whether a pair of de facto thresholds is feasible, we have to exhaustively search a large range of acceptance thresholds. For a small θ_2, we sometimes have to accurately compute the conference values for acceptance thresholds larger than ten times the variance of the review noise.

in the one-conference design. For any other design, if the de facto threshold of the secondary conference $\theta_2 = \theta$, for any feasible θ_1, the conference values satisfy $V_2 \leq V \leq V_1$.

We defer the proof to the appendix. Proposition 4.2 presents an important property of our studied designs: dividing the conference decreases the value of the secondary conference. As we will discuss in Section 4.4, this property is the key to why dividing the conference can reduce the review burden.

Our second observation is about the separating design and the hierarchical design. Perhaps surprisingly, we find that conditioned on the same pair of de facto thresholds of the top conference and the secondary conference, the hierarchical design decreases the value of both conferences.

Proposition 4.3. *For any pair of feasible de facto thresholds $\theta_2 < \theta_1$, if the review noise distribution F_X has a strictly monotone increasing hazard rate, $V_1^h < V_1^s$ and $V_2^h < V_2^s$.*

To understand why this proposition holds, we first define some notations. Let τ_1^h and τ_2^h be the acceptance thresholds that induce the de facto thresholds θ_1 and θ_2 in the hierarchical design. Let

$$\alpha_1 = \int_{\theta_1}^{\infty} q\mathbf{p}(q)dq, \qquad \alpha_2 = \int_{\theta_2}^{\theta_1} q\mathbf{p}(q)dq, \qquad \alpha_3 = \int_{\theta_2}^{\theta_1} q\mathbf{p}(q)\mathrm{P}_{acc}(\tau_1^h, q)/\mathrm{P}_{acc}(\tau_2^h, q)dq, \tag{9}$$

$$\beta_1 = \int_{\theta_1}^{\infty} \mathbf{p}(q)dq, \qquad \beta_2 = \int_{\theta_2}^{\theta_1} \mathbf{p}(q)dq, \qquad \beta_3 = \int_{\theta_2}^{\theta_1} \mathbf{p}(q)\mathrm{P}_{acc}(\tau_1^h, q)/\mathrm{P}_{acc}(\tau_2^h, q)dq. \tag{10}$$

By definition, $V_1^s = \frac{\alpha_1}{\beta_1}$, $V_2^s = \frac{\alpha_2}{\beta_2}$, $V_1^h = \frac{\alpha_1+\alpha_3}{\beta_1+\beta_3}$, and $V_2^h = \frac{\alpha_2-\alpha_3}{\beta_2-\beta_3}$. Intuitively, compared with the separating design, the hierarchical design transfers a batch of papers whose quality is between θ_2 and θ_1 from the secondary conference to the top conference. This batch consists of β_3 papers, with a total quality of α_3.

The probability of a paper being transferred to the top conference is $\gamma(q) = \frac{\mathrm{P}_{acc}(\tau_1^h,q)}{\mathrm{P}_{acc}(\tau_2^h,q)}$, which, as we will see in the proof, is increasing in q. This indicates that the average quality of the transferred papers is higher than V_2^s but lower than V_1^s. In other words, these papers are subpar for the top conference but are likely to be top papers in the secondary conference. Therefore, separating the conference contributes to higher conference values because it prevents such "mistakes of acceptance".

Proof of Proposition 4.3. First note that because $\mathrm{P}_{acc}(\tau, q) = 1 - F_X(\tau - q)$ is decreasing in τ, $0 < \mathrm{P}_{acc}(\tau_1 - q)/\mathrm{P}_{acc}(\tau_2 - q) < 1$. This implies that $\alpha_2 > \alpha_3$ and $\beta_2 > \beta_3$. Therefore, to prove the proposition, it's sufficient to show that $\frac{\alpha_2}{\beta_2} \leq \frac{\alpha_3}{\beta_3} \leq \frac{\alpha_1}{\beta_1}$ with at least one inequality being strict.

We first show that $\frac{\alpha_3}{\beta_3} \leq \frac{\alpha_1}{\beta_1}$. Let $\boldsymbol{p}'(q) = \boldsymbol{p}(q)\mathrm{P}_{acc}(\tau_1 - q)/\mathrm{P}_{acc}(\tau_2 - q)$. Then, $\frac{\alpha_3}{\beta_3}$ is the expected quality of the paper whose quality is between θ_2 and θ_1 under the quality distribution $\boldsymbol{p}'$. This means $\frac{\alpha_3}{\beta_3} \leq \theta_1$. Furthermore, $\frac{\alpha_1}{\beta_1}$ is the expected quality of the paper whose quality is above θ_1, meaning that $\frac{\alpha_1}{\beta_1} \geq \theta_1 \geq \frac{\alpha_3}{\beta_3}$.

Next, we show that $\frac{\alpha_2}{\beta_2} < \frac{\alpha_3}{\beta_3}$. We first prove Lemma 4.4 that suggests the ratio between the acceptance probability of the top conference and the secondary conference, $\frac{\mathrm{P}_{acc}(\tau_1-q)}{\mathrm{P}_{acc}(\tau_2-q)}$, is increasing in q. With this lemma,

$$\frac{\alpha_3}{\beta_3} > \frac{\int_{\theta_2}^{\theta_1} q\mathbf{p}(q)\mathrm{P}_{acc}(\tau_1-\theta_2)/\mathrm{P}_{acc}(\tau_2-\theta_2)dq}{\int_{\theta_2}^{\theta_1} \mathbf{p}(q)\mathrm{P}_{acc}(\tau_1-\theta_2)/\mathrm{P}_{acc}(\tau_2-\theta_2)dq} = \frac{\int_{\theta_2}^{\theta_1} q\mathbf{p}(q)dq}{\int_{\theta_2}^{\theta_1} \mathbf{p}(q)dq} = \frac{\alpha_2}{\beta_2}.$$

This completes the proof.

□

Lemma 4.4. *When f_X has a strictly increasing hazard rate, $\gamma(q) = \frac{P_{acc}(\tau_1-q)}{P_{acc}(\tau_2-q)}$ is strictly increasing in q.*

Proof. The proof follows by showing the derivative of h is positive.

$$
\begin{aligned}
\frac{\partial \gamma(q)}{\partial q} &= \partial \frac{1 - f_X(\tau_1 - q)}{1 - f_X(\tau_2 - q)} / \partial q \\
&= \frac{f'_X(\tau_1 - q)(1 - f_X(\tau_2 - q)) - f'_X(\tau_2 - q)(1 - f_X(\tau_1 - q))}{(1 - f_X(\tau_2 - q))^2} \\
&= \frac{f'_X(\tau_1 - q)/(1 - f_X(\tau_1 - q)) - f'_X(\tau_2 - q)/(1 - f_X(\tau_2 - q))}{(1 - f_X(\tau_1 - q))(1 - f_X(\tau_2 - q))^3} \\
&> 0 \qquad\qquad (f_X \text{ has a strictly increasing hazard rate.})
\end{aligned}
$$

□

4.3 Acceptance Thresholds

The acceptance thresholds are key drivers of the welfare of the peer review systems. Intuitively, conditioned on the same de facto thresholds, lower acceptance thresholds imply faster acceptance and thus a lower review burden. In this section, we focus on comparing the acceptance thresholds of the separating design and the hierarchical design. In particular, conditioned on $\theta_1^s = \theta_1^h$ and $\theta_2^s = \theta_2^h$, when can we expect smaller/larger acceptance thresholds in the separating design?

We start by fixing a pair of de facto thresholds θ_1, θ_2 that are feasible in both the separating design and the hierarchical design. Let τ_1^s, τ_2^s, τ_1^h, and τ_2^h be the corresponding acceptance thresholds of the two conferences in two designs as defined before.

Top conference. Our first result suggests that under mild assumptions, the separating design always has a larger acceptance threshold for the top conference.

Proposition 4.5. *For any pair of feasible de facto thresholds $\theta_2 < \theta_1$, if the review noise distribution F_X has a strictly monotone increasing hazard rate, then $\tau_1^s > \tau_1^h$ if and only if*

$$
\frac{V_2^s}{V_1^s} < \left(1 + \frac{1 - \eta}{\eta P_{acc}(\tau_2^s, \theta_1)}\right) \frac{V_2^h}{V_1^h}.
$$

Intuitively, the top conference should set its acceptance threshold so that authors with papers of quality θ_1 are indifferent between sticking with the top conference and embracing the secondary conference. Therefore, which design has a higher τ_1 depends on how attractive is the secondary conference versus the top conference in each design. We can read two messages from Proposition 4.5.

First, when authors are impatient, the separating design tends to have a larger τ_1. This can be observed from Proposition 4.5, where conditioned on everything else, a decrease in η causes the right-hand side of the inequality to increase. The intuition is that when authors are less patient, only the first time of submission is worth the most for the authors, meaning that the benefit of sticking with the top conference greatly shrinks. Therefore, to make the top conference more attractive (so that θ_1 is a de facto threshold), the top conference has to lower its acceptance threshold in the hierarchical design, resulting in $\tau_1^h < \tau_1^s$.

Second, we investigate when we can expect $\frac{V_2^s}{V_1^s} < \frac{V_2^h}{V_2^h}$, in which case τ_1^s is always larger than τ_1^h. Let $\alpha_3, \beta_1, \beta_2, \beta_3$ be the same notations as defined in Eq. (10). Let $V_3 = \frac{\alpha_3}{\beta_3}$ be the expected quality of the papers transferred from the secondary conference to the top conference in the hierarchical design. Note that $V_2^s \le V_3 \le V_1^s$.

$$
\begin{aligned}
\frac{V_2^s}{V_1^s} - \frac{V_2^h}{V_1^h} < 0 \qquad &\Leftrightarrow \qquad \frac{V_2^s}{V_1^s} - \frac{\frac{V_2^s\beta_2 - V_3\beta_3}{\beta_2 - \beta_3}}{\frac{V_1^s\beta_1 + V_3\beta_3}{\beta_1 + \beta_3}} < 0 \\
&\Leftrightarrow \qquad V_2^s(\beta_2 - \beta_3)\left(V_1^s\beta_1 + V_3\beta_3\right) - V_1^s(\beta_1 + \beta_3)\left(V_2^s\beta_2 - V_3\beta_3\right) < 0 \\
&\Leftrightarrow \qquad V_2^s\left(V_1^s - V_3\right)\beta_2 - V_1^s\left(V_3 - V_2^s\right)\beta_1 - V_3\left(V_1^s - V_2^s\right)\beta_3 > 0. \qquad (11)
\end{aligned}
$$

The three terms on the left-hand side of Eq. (11) correspond to the volume of the papers that are accepted by the secondary conference in the separating design β_2, accepted by the top conference in the separating design β_1, and transferred from the secondary conference to the top conference in the hierarchical design β_3. Without further assumptions, it is possible for Eq. (11) to be either positive or negative. For example, suppose we have a very selective top conference so that both β_1 and β_3 are small. This may correspond to the scenario where some AI conferences only select less than 5% of the submissions as "oral presentations" (top conference). In this case, the majority of the submissions are absorbed by the secondary conference, i.e. $\beta_2 \gg \beta_1$ and $\beta_2 \gg \beta_3$. Consequently, the left-hand side of the equation is dominated by the first term, which is positive, meaning that we should expect a larger τ_1 if we separate the conferences.

However, suppose we have two conferences such that the more prestigious one (the top conference) has a larger volume of submissions, i.e. $\beta_1 \geq \beta_2$. Suppose the top conference has a good review system such that only papers with quality close to θ_1 have a chance of getting in. This implies that β_3 is small and V_3 is close to θ_1. In this case, the first two terms dominate the left-hand side of Eq. (11). Therefore, we can expect the left-hand side to be negative if $V_3 - V_2^s \geq V_1^s - V_3$, in which case the separating design has a smaller τ_1.

Secondary conference. Our second result compares the acceptance thresholds of the secondary conference. Let $\gamma(\theta_2) = \frac{\mathrm{P}_{acc}(\tau_1^h - \theta_2)}{\mathrm{P}_{acc}(\tau_2^h - \theta_2)}$.

Proposition 4.6. *For any pair of feasible de facto thresholds $\theta_2 < \theta_1$, if the review noise distribution F_X has a strictly monotone increasing hazard rate, then there always exists a threshold $0 < t < 1$ such that $\tau_2^s > \tau_2^h$ if and only if $\gamma(\theta_2) < t$.*

Proposition 4.6 suggests whether the separating design induces a larger acceptance threshold of the secondary conference depends on $\gamma(\theta_2)$. This ratio between $\mathrm{P}_{acc}(\tau_1^h, \theta_2)$ and $\mathrm{P}_{acc}(\tau_2^h, \theta_2)$ indicates the relative likelihood of a paper of quality θ_2 being accepted at the top conference compared to the secondary conference in a hierarchical design.

We illustrate the intuition of this result using the case where $\gamma(\theta_2) < t$, while the reasoning of the opposite case is analogous. When $\gamma(\theta_2)$ is small, which is usually the case when the review noise is small, the borderline papers (with quality $q = \theta_2$)) are unlikely to get into the top conference under the hierarchical design. Consequently, both in the separating design and the hierarchical design, the utility for authors of borderline papers (primarily) comes from the reward of being accepted at the secondary conference. As indicated by Proposition 4.3, the secondary conference has a larger value in the separating design than in the hierarchical design. Therefore, when $\gamma(\theta_2)$ is small, the secondary conference must raise its acceptance threshold in the separating design so that the reduced acceptance probability can offset the increased value of acceptance, thereby maintaining θ_2 as the de facto threshold in both designs.

4.4 Social Welfare

We assess different designs based on their efficacy in improving the social welfare of the review system. We primarily focus on two concepts: the review burden and the average author utility.

4.4.1 Review Burden and Average Author Utility

We use the one-conference design as an example to illustrate these two measures. For any round t, the submissions to the conference are composed of the new papers generated in round t that are decided to submit, resubmissions that are generated and submitted in round $t-1$ but got rejected, resubmissions that are generated in round $t-2$ and are rejected for two rounds, and so on. Suppose t is infinity and the quality distribution is fixed across all rounds. This implies that the review burden has converged to a fixed point. Let τ and θ be the conference's acceptance threshold and the corresponding de facto threshold, respectively. Let $p_{rej}(q) = 1 - \mathrm{P}_{acc}(\tau, q)$ be the probability that a paper of quality q is rejected in one round under acceptance threshold τ. Then, we write the review burden as

$$R(\theta, \tau) = \frac{1}{n} \int_{\theta}^{\infty} n\mathbf{p}(q)(1 + p_{rej}(q) + p_{rej}^2(q) + \cdots)dq = \int_{\theta}^{\infty} \mathbf{p}(q)/\mathrm{P}_{acc}(\tau, q)dq.$$

With the same recipe, we can write down the review burden of the other designs.

$$R^s(\theta_1,\theta_2,\tau_1,\tau_2) = R^h(\theta_1,\theta_2,\tau_1,\tau_2) = \int_{\theta_2}^{\theta_1} \mathbf{p}(q)/\mathrm{P}_{acc}(q,\tau_2)dq + \int_{\theta_1}^{\infty} \mathbf{p}(q)/\mathrm{P}_{acc}(q,\tau_1)dq.$$

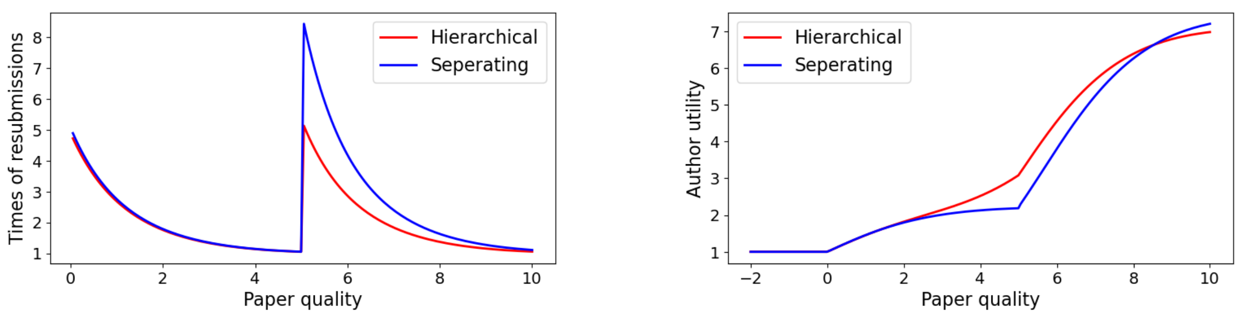
We note that the review burden in the separating design and the hierarchical design looks identical (conditioned on the same thresholds). This is because, for an author with a paper of quality q, the condition for her to consider resubmitting is the same under these two designs. Figure 2 (a) presents the number of times a paper of quality q is (re)submitted under each design. The review burden of the system is the integral of this curve weighted by the quality prior.

The average author utility is calculated by integrating the author's utility under her best response. Figure 2 (b) presents the utility of an author with a paper of quality q under each design. The average author utility is the integral of this curve weighted by the quality prior.

$$\bar{U}(\theta,\tau) = \int_{-\infty}^{\theta} \mathbf{p}(q)dq + \int_{\theta}^{\infty} \mathbf{p}(q)\frac{\mathrm{P}_{acc}(\tau,q)V}{1-\eta(1-\mathrm{P}_{acc}(\tau,q))}dq,$$

$$\bar{U}^s(\theta_1,\theta_2,\tau_1,\tau_2) = \int_{-\infty}^{\theta_2} \mathbf{p}(q)dq + \int_{\theta_2}^{\theta_1} \mathbf{p}(q)\frac{\mathrm{P}_{acc}(\tau_2-q)V_2}{(1-\eta(1-\mathrm{P}_{acc}(\tau_2-q)))}dq + \int_{\theta_1}^{\infty} \mathbf{p}(q)\frac{\mathrm{P}_{acc}(\tau_1-q)V_1}{1-\eta(1-\mathrm{P}_{acc}(\tau_1-q))}dq.$$

$$\bar{U}^h(\theta_1,\theta_2,\tau_1,\tau_2) = \int_{-\infty}^{\theta_2} \mathbf{p}(q)dq + \int_{\theta_2}^{\theta_1} \mathbf{p}(q)\frac{(\mathrm{P}_{acc}(\tau_2-q)-\mathrm{P}_{acc}(\tau_1-q))V_2+\mathrm{P}_{acc}(\tau_1-q)V_1}{1-\eta(1-\mathrm{P}_{acc}(\tau_2-q))}dq + \int_{\theta_1}^{\infty} \mathbf{p}(q)\frac{\mathrm{P}_{acc}(\tau_1-q)V_1}{1-\eta(1-\mathrm{P}_{acc}(\tau_1-q))}dq.$$



(a) Times of (re)submissions

(b) Author utility

Figure 2: (a) The times of resubmissions and (b) the author's utility as a function of paper quality q. The example is in the ($\mu_q=-1, \sigma_q=5, \sigma_r=2, \eta=0.7$)-Gaussian prior model with $\theta_1=5$ and $\theta_2=0$.

Aligning Review Burden and Average Author Utility. We provide an intuitive connection between the review burden and the a

4.4.2 Comparing Social Welfare Using Theoretical Insights

Figure 3 presents how different model variables vary with the de facto threshold of the top conference conditioned on $\theta_2=0$. Note that $\theta_2=0$ implies that all papers with positive quality are submitted and eventually accepted by either of the conferences. Additional examples with different hyperparameters are deferred to the appendix. We first summarize our theoretical insights.

- Proposition 4.2: Compared with the conference values in the one-conference design, the separating design and the hierarchical design both have a lower-value secondary conference and a higher-value top conference, i.e. $V_2 < V < V_1$.

- Proposition 4.3: Conditioned on the same de facto thresholds, the separating design has higher values for both the top and the secondary conference compared with the hierarchical design, i.e. $V_1^s > V_1^h$ and $V_2^s > V_2^h$.

- Proposition 4.5: Because conferences are more attractive to authors in the separating design, the top conference has to raise its acceptance threshold, $\tau_1^s > \tau_1^h$ (in most natural settings we tested).

- Proposition 4.6: Whether the separating design or the hierarchical design has a lower τ_2 depends on the ratio between the likelihood of a borderline paper being accepted by the top conference and the secondary conference under the hierarchical design, i.e. $\gamma(\theta_2) = \frac{\mathrm{P}_{acc}(\tau_1^h - \theta_2)}{\mathrm{P}_{acc}(\tau_2^h - \theta_2)}$. The proposition predicts that the event of $\tau_2^s < \tau_2^h$ occurs when the ratio is large, which happens when θ_1 is small. This pattern is illustrated in Fig. 3.

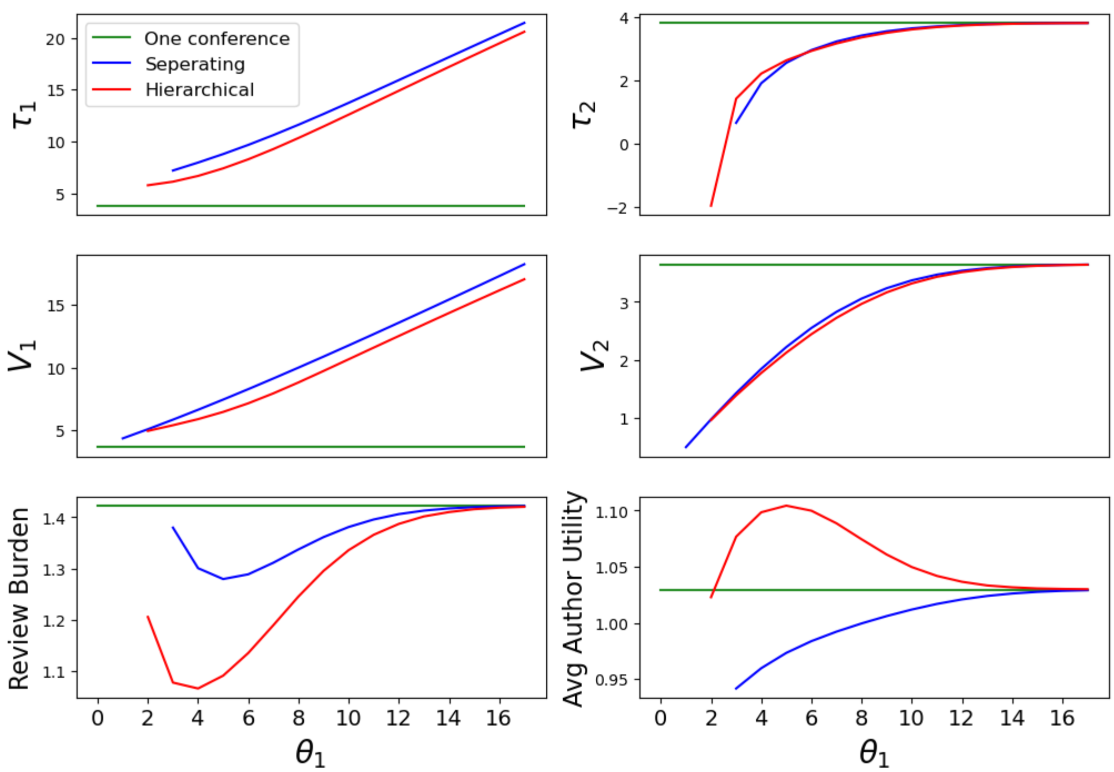
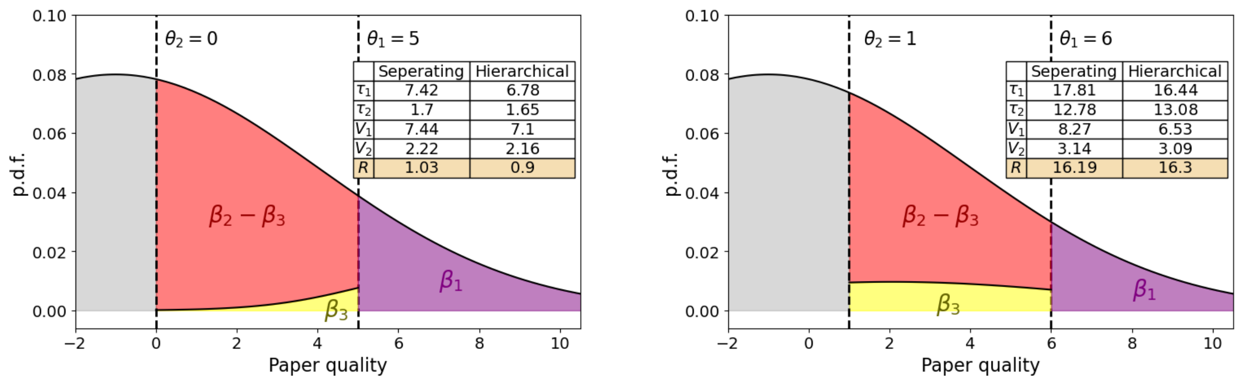


Figure 3: Variation of key model variables with θ_1 while fixing $\theta_2 = 0$. We use the ($\mu_q = -1, \sigma_q = 5, \sigma_r = 3, \eta = 0.7$)-Gaussian prior model.

Our theoretical insights allow a better comparison of the social welfare of different designs. We observe that in most of the natural settings, the hierarchical design experiences a smaller review burden and a larger average author utility. This is because conditioned on the same de facto thresholds, conferences under the separating design have larger values, which results in $\tau_1^s > \tau_1^h$ in almost all of the natural settings and $\tau_2^s > \tau_2^h$ whenever θ_1 is large (meaning that the secondary conference attracts most of the submission). Note that larger acceptance thresholds result in more rejections and resubmissions, which lead to a higher review burden and a smaller author utility. This implies that the separating design tends to be in a disadvantaged position due to its larger conference values.

However, this does not rule out the possibility that the separating design can reduce the review burden. The separating design reduces the review burden when $\tau_2^h < \tau_2^s$ and the secondary conference attracts most of the submissions. Nonetheless, we only find such examples in extreme cases where reviews are very noisy and authors are very patient. Figure 4 visualize the compositions of the conferences in two parameter settings each corresponding to the case that the separating design has (a) a larger review burden and (b) a smaller review burden. A key difference between the two examples is that β_3, which is the portion of papers that are "transferred" from the secondary conference to the top conference under the hierarchical design (the yellow area), is much larger for $q = \theta_2$ in (b) compared with (a). Therefore, the secondary conference in the hierarchical design has to raise its acceptance threshold in example (b), resulting in $\tau_2^s < \tau_2^h$ and thus a lower review burden under the separating design.



(a) Hierarchical design has a smaller review burden

(b) Seperating design has a smaller review burden

Figure 4: Two examples of the comparison between the separating design and the hierarchical design conditioned on the same de facto thresholds. β_1 represents the papers submitted to the top conference in the separating design; β_2 represents the papers submitted to the secondary conference in the separating design; β_3 represents the papers transferred from the secondary conference to the top conference in the hierarchical design. Example (a) is in the $(\mu_q = -1, \sigma_q = 5, \sigma_r = 2, \eta = 0.7)$-Gaussian prior model, and example (b) is in the $(\mu_q = -1, \sigma_q = 5, \sigma_r = 5, \eta = 0.98)$-Gaussian prior model.

Two feasible solutions for the hierarchical design. We further note an interesting observation that the same pairs of de facto thresholds may result in two solutions of acceptance thresholds under the hierarchical design when θ_2 is large (see Fig. 6 in the appendix). In particular, one solution results in a $\tau_1^h > \theta_1$ and another solution results in a $\tau_1^h < \theta_1$. This is due to a unique feature of the hierarchical design that the conference values are affected by not only the de facto thresholds but also the acceptance thresholds. As a result, by lowering τ_1^h, the top conference in the hierarchical design can increase its acceptance probability at the cost of decreasing its value while maintaining the same level of attractiveness to the authors. This is the key to two feasible solutions. However, as the top conference is taking more papers from the secondary conference, lowering τ_1^h also reduces V_2^h. Therefore, the second solution is only feasible when V_2^h is large enough, which happens when θ_2 is large enough. In practice, the second solution of $\tau_1^h < \theta_1$ corresponds to a less prestigious top conference but is more generous in accepting papers.

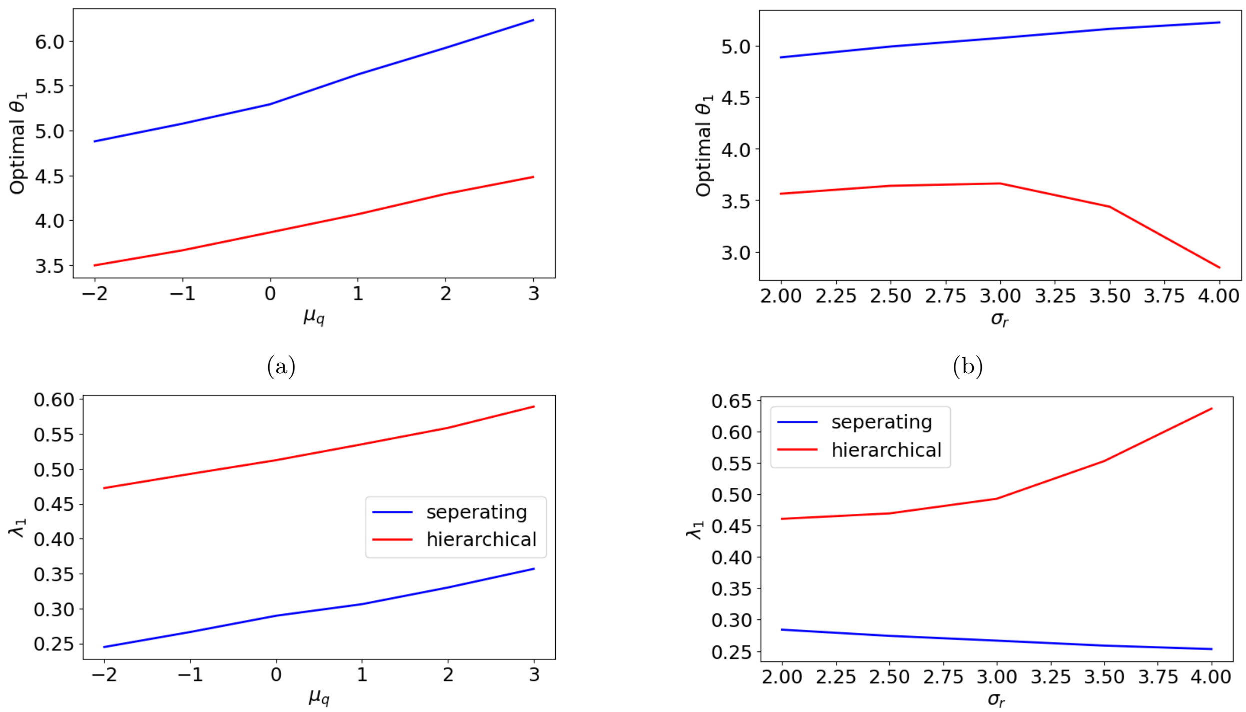
4.4.3 Optimizing Review Burden

In Fig. 3, we observe that the review burden first decreases and then increases as the top conference raises its selection bar. This pattern is driven by two factors of the review system. On one hand, as θ_1 increases, more authors opt for the option of *Only Secondary* or *Open to Secondary*. Because $\tau_2 < \tau_1$, this results in a faster acceptance for more papers, which decreases the review burden. On the other hand, increasing θ_1 also increases both τ_1 and τ_2. Consequently, for both the top and the secondary conference, papers experience more rejections before acceptance, which increases the review burden. When θ_1 is close to θ_2, the first effect dominates, while for large θ_1, the second effect is more significant. This interplay causes the non-monotonic relationship between the review burden and θ_1.

The optimal de facto threshold $\theta_1^h < \theta_1^s$. We then ask when different designs will reach the optimal de facto threshold θ_1 that minimizes the review burden. As shown in Fig. 5 (a) and (b), we observe that the optimal θ_1 is always smaller in the hierarchical design compared with the separating design. This implies that the hierarchical design always leads to a larger top conference while optimizing the review burden, shown in Fig. 5 (c) and (d) where λ_1 is the ratio between the number of papers accepted by the top conference and the number of papers accepted by both conferences.

How do the quality prior and review noise affect the optimal thresholds of conferences? From Fig. 5 (a) and (c), we observe that as the mean of paper quality μ_q increases, both the optimal θ_1 and the relative size of the top conference rise under both designs. At a high level, the optimal θ_1 is reached when the marginal benefit of increasing θ_1 — having more papers being submitted to the secondary conference with a smaller acceptance threshold ($\tau_2 < \tau_1$) — offsets the marginal loss — where both τ_1 and τ_2 increase with θ_1. Since a higher μ_q amplifies the former effect, the critical point is achieved at a larger θ_1. Therefore, we see that the optimal θ_1 increases with μ_q.

Fig. 5 (b) indicates that the optimal θ_1 is increasing with the review noise in the separating design, while the relationship between them is non-monotonic in the hierarchical design. Fixing the quality prior and the de facto threshold of the secondary conference, this further suggests that as review noise increases, the size of the top conference is expected to increase under the hierarchical design but shrink under the separating design (see Fig. 5 (d)).



(a) (b) (c) (d)

Figure 5: (a) and (b) are the de facto threshold θ_1 that minimizes the review burden while (c) and (d) are the relative size of the top conference λ_1 under the optimal de facto threshold. Unless otherwise specified in the figures, we consider the $(\mu_q = -1, \sigma_q = 5, \sigma_r = 3, \eta = 0.7)$-Gaussian prior model as the default parameter setting.

5 Discussions and Limitations

We discuss several limitations of our model and explore room for potential improvement.

First, we assume paper qualities are exogenous which will improve through resubmissions. This assumption implies that our definition of the review burden is an upper bound, as authors may experience fewer rounds of resubmissions with improved qualities. Although the absolute values of review burden are not calibrated, further analysis concerns the improvements of papers are unlikely to alter our main insights, which are primarily based on the relative relationship between review burden under various designs. However, in the endogenous quality setting, other dimensions of the review system should be considered. For example, we hypothesize that the separating design is better at encouraging higher effort from authors to improve their paper qualities. Intuitively, the separating design improves the utility of authors with high-quality papers (see Fig. 2 (b)).

Second, we assume authors' self-evaluations of their papers are accurate. As discussed in [23], noisy evaluations greatly complicate authors' response strategies. For example, authors' beliefs about their papers can be changed by historical reviews so that they may stop submitting after several rounds of rejections. However, numerical analysis in the prior work suggests that this assumption is not a first-order concern in the dynamics between authors and conferences.

Third, we only consider separating the conference into two segments. This simplification enables a clean and intuitive analysis of our conference design problem and we believe is the most natural implementation in practice. However, future work may discuss how much the review burden can be reduced while the number of hierarchies increases. What will be the limit of various designs?

6 Conclusion

We investigate conference design with strategic authors from the mechanism design aspect. We generalize a prior model [23] to capture the interactions between two (hierarchical or separated) conferences and a set of strategic authors. Our analysis of the authors' best responses suggests that the design of the tableau and the acceptance policies of conferences must consider the strategic nature of the authors. Therefore, we should be careful about implementing some ideas that are plausible but may be problematic in practice, such as the advice of accepting more (or even all) submissions. We furthermore compare two designs proposed in the paper from the aspects of review burden and author utility. Our results suggest that, compared with the hierarchical design, although separating the conference results in a better selection of papers, it often experiences a higher review burden and a lower author utility.

References

[1] Adam Cohen, Smita Pattanaik, Praveen Kumar, Robert R Bies, Anthonius De Boer, Albert Ferro, Annette Gilchrist, Geoffrey K Isbister, Sarah Ross, and Andrew J Webb. Organised crime against the academic peer review system. *British Journal of Clinical Pharmacology*, 81(6):1012, 2016.

[2] Corinna Cortes and Neil D. Lawrence. Inconsistency in conference peer review: Revisiting the 2014 neurips experiment, 2021.

[3] Daniele Fanelli. How many scientists fabricate and falsify research? a systematic review and meta-analysis of survey data. *PloS one*, 4(5):e5738, 2009.

[4] Tanner Fiez, Nihar Shah, and Lillian Ratliff. A super* algorithm to optimize paper bidding in peer review. In *Conference on uncertainty in artificial intelligence*, pages 580–589. PMLR, 2020.

[5] Samir Haffar, Fateh Bazerbachi, and M Hassan Murad. Peer review bias: a critical review. In *Mayo Clinic Proceedings*, volume 94, pages 670–676. Elsevier, 2019.

[6] Ari Kobren, Barna Saha, and Andrew McCallum. Paper matching with local fairness constraints. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1247–1257, 2019.

[7] Carole J Lee, Cassidy R Sugimoto, Guo Zhang, and Blaise Cronin. Bias in peer review. *Journal of the American Society for information Science and Technology*, 64(1):2–17, 2013.

[8] Cheng Long, Raymond Chi-Wing Wong, Yu Peng, and Liangliang Ye. On good and fair paper-reviewer assignment. In *2013 IEEE 13th international conference on data mining*, pages 1145–1150. IEEE, 2013.

[9] Reshef Meir, Jérôme Lang, Julien Lesca, Nicholas Mattei, and Natan Kaminsky. A market-inspired bidding scheme for peer review paper assignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4776–4784, 2021.

[10] D Sculley, Jasper Snoek, and Alex Wiltschko. Avoiding a tragedy of the commons in the peer review process, 2018.

[11] Nihar B Shah, Behzad Tabibian, Krikamol Muandet, Isabelle Guyon, and Ulrike Von Luxburg. Design and analysis of the nips 2016 review process. *Journal of machine learning research*, 19(49):1–34, 2018.

[12] S S Siegelman. Assassins and zealots: variations in peer review. special report. *Radiology*, 178(3): 637–642, 1991. doi: 10.1148/radiology.178.3.1994394. URL `https://doi.org/10.1148/radiology.178.3.1994394`. PMID: 1994394.

[13] Jennifer Sills, George P. Chrousos, and Alexios-Fotios A. Mentis. Imposter syndrome threatens diversity. *Science*, 367(6479):749–750, 2020. doi: 10.1126/science.aba8039. URL `https://www.science.org/doi/abs/10.1126/science.aba8039`.

[14] Arnaldo Spalvieri, Silvio Mandelli, Maurizio Magarini, and Giuseppe Bianchi. Weighting peer reviewers. In *2014 Twelfth Annual International Conference on Privacy, Security and Trust*, pages 414–419. IEEE, 2014.

[15] Siddarth Srinivasan and Jamie Morgenstern. Auctions and prediction markets for scientific peer review. *arXiv preprint arXiv:2109.00923*, 8, 2021.

[16] Ivan Stelmakh, Nihar B Shah, and Aarti Singh. Catch me if i can: Detecting strategic behaviour in peer assessment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4794–4802, 2021.

[17] Ivan Stelmakh, Nihar B Shah, Aarti Singh, and Hal Daumé III. A novice-reviewer experiment to address scarcity of qualified reviewers in large conferences. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4785–4793, 2021.

[18] Ivan Stelmakh, Nihar B Shah, Aarti Singh, and Hal Daumé III. Prior and prejudice: The novice reviewers' bias against resubmissions in conference peer review. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–17, 2021.

[19] Weijie J Su. You are the best reviewer of your own papers: An owner-assisted scoring mechanism. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021. URL `https://openreview.net/forum?id=xmx5rE9QP7R`.

[20] Andrew Tomkins, Min Zhang, and William D Heavlin. Reviewer bias in single-versus double-blind peer review. *Proceedings of the National Academy of Sciences*, 114(48):12708–12713, 2017.

[21] Jibang Wu, Haifeng Xu, Yifan Guo, and Weijie Su. An isotonic mechanism for overlapping ownership, 2023.

[22] Yuling Yan, Weijie J. Su, and Jianqing Fan. The isotonic mechanism for exponential family estimation, 2023.

[23] Yichi Zhang, Fang-Yi Yu, Grant Schoenebeck, and David Kempe. A system-level analysis of conference peer review. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 1041–1080, 2022.

[24] Yichi Zhang, Grant Schoenebeck, and Weijie Su. Eliciting honest information from authors using sequential review. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 9977–9984, 2024.

A Additional Proofs

A.1 Threshold Best Responses

This section justifies the threshold best response of authors in the one-conference design and the separating design.

Proof of Proposition 2.1. The acceptance probability for a submission with quality $Q = q$, represented as $\mathrm{P}_{acc}(\tau, q)$, is non-decreasing with respect to q due to the monotone acceptance policy. For policies with a non-trivial threshold, this probability strictly increases with q. An author is motivated to submit their paper when the expected utility of doing so is greater than the value of the outside option, which is 1. Conversely, the author will refrain from submission if this expected utility is less than 1. Where the expected utility is precisely 1, the author is indifferent between submitting to the conference and choosing the outside option. This decision-making process is repeated in subsequent rounds, implying a persistent submission strategy by the author until their work is accepted. The expected utility can be quantified as the time-discounted sum of the utility from acceptance:

$$U^o(q|\tau) = \sum_{t=1}^{\infty} V^o \cdot \eta^{t-1} \mathrm{P}_{acc}(\tau, q) \cdot (1 - \mathrm{P}_{acc}(\tau, q))^{t-1} = \frac{V \cdot \mathrm{P}_{acc}(\tau, q)}{1 - \eta \cdot (1 - \mathrm{P}_{acc}(\tau, q))}$$

Let $\rho = \frac{V-\eta}{1-\eta}$. By analyzing the conditions $U^o(q|\tau) > 1$, $U^o(q|\tau) < 1$, and $U^o(q|\tau) = 1$ in relation to q, an author will opt to submit when $\mathrm{P}_{acc}(\tau, q) > 1/\rho$, choose the outside option when $\mathrm{P}_{acc}(\tau, q) < 1/\rho$, and will be indifferent when $\mathrm{P}_{acc}(\tau, q) = 1/\rho$.

Let $\bar{\theta} = \sup\{q \mid \mathrm{P}_{acc}(\tau, q) < 1/\rho\}$ and $\underline{\theta} = \inf\{q \mid \mathrm{P}_{acc}(\tau, q) > 1/\rho\}$. By the monotone acceptance policy, $\bar{\theta} \leq \underline{\theta}$. Therefore, an author wants to submit (and persist in resubmitting until acceptance) if $q > \underline{\theta}$, and to withhold submission if $q < \bar{\theta}$. Should $\bar{\theta}$ and $\underline{\theta}$ converge, then any θ equating $\bar{\theta}$ and $\underline{\theta}$ satisfies the claim. Conversely, if $\bar{\theta} \neq \underline{\theta}$, then for all q within $(\bar{\theta}, \underline{\theta})$, $\mathrm{P}_{acc}(\tau, q) = 1/\rho$, so the author is indifferent between submitting and not submitting for all such q. Thus, any $\theta \in (\bar{\theta}, \underline{\theta})$ satisfies the claim. □

Proof of Proposition 3.1. We first show that $U_B^s(q)$ is increasing in q by showing the first-order derivative w.r.t. q is positive.

$$\frac{\partial U_B^s}{\partial q} = \frac{(1-\eta) f_X(\tau_2 - q) V_2^s}{(1 - \eta F_X(\tau_2 - q))^2} > 0.$$

Because $U_B^s(-\infty) = 0$ and $U_B^s(\infty) = V_2^s > 1$, the monotonicity of U_B^s thus implies a threshold θ_2 such that the author prefers submitting to the secondary conference if $q \geq \theta_2$ and taking the outside option otherwise.

Next, we show that $U_A^s(q) - U_B^s(q)$ is increasing in q. Again, we write down the first-order derivative.

$$\begin{aligned}
\frac{\partial U_A^s - U_B^s}{\partial q} &= \frac{(1-\eta) f_X(\tau_1 - q) V_1^s}{(1 - \eta F_X(\tau_1 - q))^2} - \frac{(1-\eta) f_X(\tau_2 - q) V_2^s}{(1 - \eta F_X(\tau_2 - q))^2} \\
&\geq (1-\eta) V_2^s \left(\frac{f_X(\tau_1 - q)}{(1 - \eta F_X(\tau_1 - q))^2} - \frac{f_X(\tau_2 - q)}{(1 - \eta F_X(\tau_2 - q))^2} \right) \\
&\geq 0.
\end{aligned}$$

Therefore, there exists a θ_1 such that the author prefers submitting to the top conference if $q \geq \theta_1$ and submitting to the secondary conference otherwise.

If $\theta_1 > \theta_2$, the author will submit to the top conference when $q \geq \theta_1$, submit to the secondary conference when $\theta_2 \leq q < \theta_1$, and take the outside option if $q < \theta_2$. If $\theta_1 \leq \theta_2$, the model reduces to the one-conference design where we have $\theta_2 = \theta_1$. □

A.2 Conference Values

Proof of Proposition 4.2. We first define some notations. Let

$$\alpha_1 = \int_{\theta_1}^{\infty} q\mathbf{p}(q)dq, \qquad \alpha_2 = \int_{\theta}^{\theta_1} q\mathbf{p}(q)dq, \qquad \alpha_3 = \int_{\theta}^{\theta_1} q\mathbf{p}(q)\mathrm{P}_{acc}(\tau_1 - q)/\mathrm{P}_{acc}(\tau_2 - q)dq,$$
$$\beta_1 = \int_{\theta_1}^{\infty} \mathbf{p}(q)dq, \qquad \beta_2 = \int_{\theta_2}^{\theta_1} \mathbf{p}(q)dq, \qquad \beta_3 = \int_{\theta_2}^{\theta_1} \mathbf{p}(q)\mathrm{P}_{acc}(\tau_1 - q)/\mathrm{P}_{acc}(\tau_2 - q)dq.$$

Therefore,

$$V^o = \frac{\alpha_1 + \alpha_2}{\beta_1 + \beta_2}, \qquad V_1^s = \frac{\alpha_1}{\beta_1}, \qquad V_1^h = \frac{\alpha_1 + \alpha_3}{\beta_1 + \beta_3},$$
$$V_2^s = \frac{\alpha_2}{\beta_2}, \qquad V_2^h = \frac{\alpha_2 - \alpha_3}{\beta_2 - \beta_3}.$$

Separating Design We want to show that $V_2^s < V^o < V_1^s$ for any $\theta_1 > \theta_2$. It is sufficient to show that $\frac{\alpha_1}{\beta_1} > \frac{\alpha_2}{\beta_2}$, i.e. $V_1^s > V_2^s$. This is straightforward because when the paper quality distribution has full support on $\mathbb{R}$, $V_1^s = \mathbb{E}_{\mathbf{p}}[Q|Q \geq \theta_1] > \theta_1$. As V_2^s is defined as the average quality of papers between θ_2 and θ_1, $V_2^s < \theta_1$.

Hierarchical Design We first show that the value of the secondary conference is lower than the one-conference design, i.e. we want to show that $\frac{\alpha_2-\alpha_3}{\beta_2-\beta_3} < \frac{\alpha_2+\alpha_1}{\beta_2+\beta_1}$. It is sufficient to show that $\frac{\alpha_1}{\beta_1} > \frac{\alpha_3}{\beta_3}$. Let $\boldsymbol{p}'(q) = \boldsymbol{p}(q)\mathrm{P}_{acc}(\tau_1 - q)/\mathrm{P}_{acc}(\tau_2 - q)$. Then, $\frac{\alpha_3}{\beta_3}$ is the expected quality of the paper whose quality is between θ_2 and θ_1 under the quality distribution $\boldsymbol{p}'$. This means $\frac{\alpha_3}{\beta_3} \leq \theta_1$. Furthermore, we have shown that $\frac{\alpha_1}{\beta_1}$ is strictly above θ_1, meaning that $\frac{\alpha_1}{\beta_1} > \theta_1 \geq \frac{\alpha_3}{\beta_3}$.

We now prove that when the review noise has a strictly increasing hazard rate, the top conference under the hierarchical design has a larger value than the one-conference design, i.e. we want to show that $\frac{\alpha_1+\alpha_3}{\beta_1+\beta_3} < \frac{\alpha_1+\alpha_2}{\beta_1+\beta_2}$. It is sufficient to show that $\frac{\alpha_3}{\beta_3} > \frac{\alpha_2}{\beta_2}$. We first prove Lemma 4.4 that suggests the ratio between the acceptance probability of the top conference and the secondary conference, $\frac{\mathrm{P}_{acc}(\tau_1-q)}{\mathrm{P}_{acc}(\tau_2-q)}$, is increasing in q. With this lemma,

$$\frac{\alpha_3}{\beta_3} > \frac{\int_{\theta_2}^{\theta_1} q\mathbf{p}(q)\mathrm{P}_{acc}(\tau_1 - \theta_2)/\mathrm{P}_{acc}(\tau_2 - \theta_2)dq}{\int_{\theta_2}^{\theta_1} \mathbf{p}(q)\mathrm{P}_{acc}(\tau_1 - \theta_2)/\mathrm{P}_{acc}(\tau_2 - \theta_2)dq} = \frac{\int_{\theta_2}^{\theta_1} q\mathbf{p}(q)dq}{\int_{\theta_2}^{\theta_1} \mathbf{p}(q)dq} = \frac{\alpha_2}{\beta_2}.$$

This completes the proof.

A.3 Acceptance Thresholds

Proof of Proposition 4.5. The proof follows by comparing Eq. (5) and Eq. (8), the conditions that induce θ_1 as a de facto threshold of the top conference in the separating design and the hierarchical design respectively.

Equation (5) can be rewritten as

$$\frac{\mathrm{P}_{acc}(\tau_1^s, \theta_1)}{1 - \eta + \eta\mathrm{P}_{acc}(\tau_1^s, \theta_1))} = \frac{\mathrm{P}_{acc}(\tau_2^s, \theta_1)}{1 - \eta(1 - \mathrm{P}_{acc}(\tau_2^s, \theta_1))} \frac{V_2^s(\theta_1, \theta_2)}{V_1^s(\theta_1)}.$$

Equation (8) can be rewritten as

$$\frac{\mathrm{P}_{acc}(\tau_1^h, \theta_1)}{1 - \eta + \eta\mathrm{P}_{acc}(\tau_1^h, \theta_1))} = \frac{1}{\eta} \frac{V_2^h(\theta_1, \theta_2, \tau_1^h, \tau_2^h)}{V_1^h(\theta_1, \theta_2, \tau_1^h, \tau_2^h)}.$$

In the remaining part of the proof, we omit the dependence of conference values on the thresholds. Note that because F_X has a strictly increasing hazard rate, $\mathrm{P}_{acc}(\tau, q) = 1 - F_X(\tau - q)$ is monotone decreasing in the gap $\tau - q$. Therefore, the left-hand side of the above two equations is decreasing in τ_1^s and τ_1^h respectively.

This implies that $\tau_1^s > \tau_1^h$ if and only if

$$
\begin{aligned}
& \frac{1}{\eta}\frac{V_2^h}{V_1^h} > \frac{\mathrm{P}_{acc}(\tau_2^s,\theta_1)}{1-\eta(1-\mathrm{P}_{acc}(\tau_2^s,\theta_1))}\frac{V_2^s}{V_1^s} \\
\Rightarrow \qquad & \frac{V_2^s}{V_1^s} < \left(1+\frac{1-\eta}{\eta\mathrm{P}_{acc}(\tau_2^s,\theta_1)}\right)\frac{V_2^h}{V_1^h}
\end{aligned}
$$

□

Proof of Proposition 4.6. The proof follows by comparing the Eq. (3) and Eq. (7), the conditions that induce θ_2 as a de facto threshold of the secondary conference in the separating design and the hierarchical design respectively.

These conditions imply that the expected utility of an author with a paper of quality θ_2 must be equal to 1 in both designs. This means,

$$
\begin{aligned}
& U_2^s(\theta_2|\tau_2^s,\theta_1) = U_2^h(\theta_2|\tau_1^h,\tau_2^h,\theta_1,\theta_2) \\
\Rightarrow \qquad & \frac{\mathrm{P}_{acc}(\tau_2^s,\theta_2)V_2^s}{1-\eta+\eta\mathrm{P}_{acc}(\tau_2^s,\theta_2))} = \frac{\mathrm{P}_{acc}(\tau_1^h,\theta_2)V_1^h + \left(\mathrm{P}_{acc}(\tau_2^h,\theta_2)-\mathrm{P}_{acc}(\tau_1^h,\theta_2)\right)V_2^h}{1-\eta+\eta\mathrm{P}_{acc}(\tau_2^h,\theta_2))}.
\end{aligned}
$$

Let $\hat{\mathrm{P}}_{acc}(\tau,q) \coloneqq \frac{\mathrm{P}_{acc}(\tau,q)}{1-\eta+\eta\mathrm{P}_{acc}(\tau,q)}$, which is decreasing in $\tau - q$. The above equation can be rewritten as

$$
\begin{aligned}
\Rightarrow \qquad & \hat{\mathrm{P}}_{acc}(\tau_2^s,\theta_2)V_2^s = \hat{\mathrm{P}}_{acc}(\tau_2^h,\theta_2)\left(V_2^h+\gamma(\theta_2)V_1^h\right) \\
\Rightarrow \qquad & \frac{\hat{\mathrm{P}}_{acc}(\tau_2^s-\theta_2)}{\hat{\mathrm{P}}_{acc}(\tau_2^h,\theta_2)} = \frac{V_2^h+\gamma(\theta_2)V_1^h}{V_2^s}.
\end{aligned}
$$

To compare τ_2^s with τ_2^h, we only have to focus on the right-hand side of the above equation. First, note that by Proposition 4.3, $V_2^s > V_2^h$. Furthermore, we have $V_1^h > \theta_1 > V_2^s$, where we ignore the edge case that the quality prior $\mathbf{p}$ has zero density everywhere when $q > \theta_1$. This means the numerator of the right-hand side is a weighted average between a value that is smaller than the denominator (V_2^h) and a value that is larger than the denominator (V_1^h) where the weight is $\gamma(\theta_2)$. Therefore, if $\gamma(\theta_2) = 0$, the right-hand side is smaller than 1, meaning that $\tau_2^s > \tau_2^h$; if $\gamma(\theta_2) = 1$, the right-hand side is greater than 1, meaning that $\tau_2^s < \tau_2^h$. Because the right-hand side is linear in $\gamma(\theta_2)$, there must exist a threshold $0 < t < 1$ such that the proposition holds. □

□

B Additional Figures

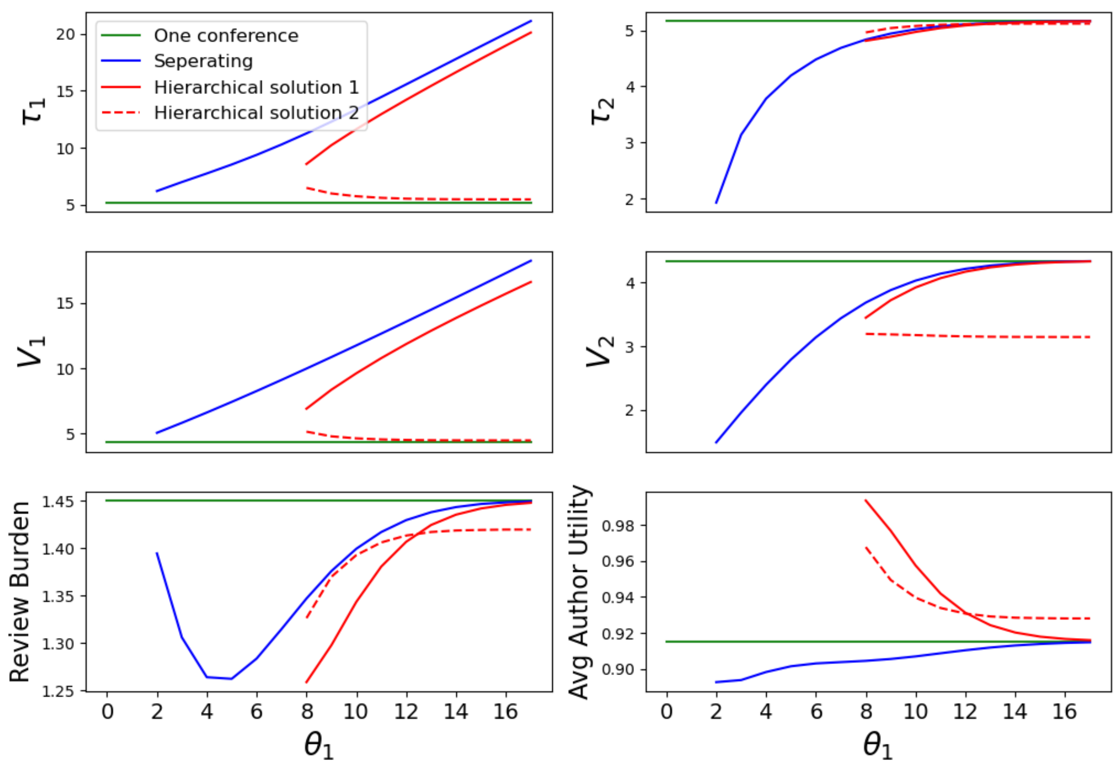


Figure 6: Variation of key model variables with θ_1 while fixing $\theta_2 = 1$. We use the ($\mu_q = -1, \sigma_q = 5, \sigma_r = 3, \eta = 0.7$)-Gaussian prior model.