Stopping Submission Spam: Self-Selection via Stochastic Sequential Screening

Rigel Galgana (MIT Operations Research), Farrell Wu (MIT Operations Research)

Abstract

The past 2 years have seen the emergence of LLMs which are capable of accelerating research well beyond the point that traditional peer review can accommodate. With the lack of reviewers' monetary or disciplinary incentive to provide meaningful reviews—in conjunction with the LLM-enabled cheapness of fulfilling reciprocal reviewing duty—academic conferences are struggling to recruit the resources needed to effectively review the ever-growing deluge of submissions. Thus far, conferences have implemented a crude submission *cap* which 1) prevents genuinely productive authors and labs to submit many high quality papers and 2) fails to disincentivize submitting many low quality papers.

In this paper, we design a practical, moneyless mechanism to address both issues. Here, papers are reviewed in random order, and an author's papers will only continue to be reviewed so long as their average paper quality remains above some quality threshold θ. This lightens reviewing lessons by 1) disincentivizing authors with quality work from submitting low quality papers that could endanger their existing stronger papers, and 2) avoids having to review multiple weak papers from the same author. Using our mechanism, we simulate the 2026 ICLR conference and show that our mechanism can reduce total reviewing by 39% whilst retaining 98% of reviews of the top half of papers. We further simulate the setting where many authors submit multiple weak papers and show that our mechanism still performs well.

CCS Concepts

• **Theory of computation** → **Algorithmic game theory and mechanism design**; • **Information systems** → *World Wide Web*.

Keywords

Moneyless Mechanism Design, Peer Review, Competitive Ratio

ACM Reference Format:

Rigel Galgana (MIT Operations Research), Farrell Wu (MIT Operations Research). 2027. Stopping Submission Spam: Self-Selection via Stochastic Sequential Screening. In *Proceedings of The ACM Web Conference 2027 (WWW '27)*. ACM, New York, NY, USA, 4 pages.

1 Introduction

As AI systems make complete research papers cheaper to produce [5], the modern conference peer review system relying on unpaid expert review is unsustainable [7]. ICLR has reported suspected submissions of near-duplicate papers, including LLM-paraphrased versions [1]. In 2026, flagship computer science conferences such as ICML, NeurIPS, and ICLR received double the number of submissions compared to the previous year. The source of many of these submissions come from authors ineligible for reciprocal reviewing: ~20% of submissions had no reciprocal reviewer, and of these, ~40% had authors submitting *multiple* such papers further fueling reviewer scarcity [3]. With such volume, even expert reviewers may give low-effort reviews or delegate these duties to AI (in 2023 and 2024, an estimated ~17% of reviewers were 'heavily' LLM assisted [4])—both of which worsen conference outcomes. To combat this, conferences implemented crude *submission caps* which not only prevents productive authors from submitting good papers but also fails to stop submission slop up to said cap. We consider in this paper instead a mechanism which addresses both issues whilst abiding several natural constraints:

- **Moneyless authorship and reviewing**: conference submission is free and reviewers are typically unpaid.
- **Random order**: sequential reviewing is impractical [10].
- **Equity**: all authors are initially eligible for review.

Our mechanism. Our mechanism is simple: for each author, we maintain a running average of their paper quality (i.e., reviewer scores) and review all of their papers in random order until their average quality falls below threshold $\theta > 0$ after which, all subsequent papers are not reviewed. This mechanism is an orderless variant of the mechanism proposed in [10] wherein authors may specify their desired reviewing order. In our mechanism, adding weak papers put stronger work at risk, encouraging authors to withhold them. Furthermore, it limits the number of low quality papers reviewed as they decrease the running average; i.e., if one author's papers are all below θ, only one will be reviewed. The public ICLR 2026 data suggests that both effects will have a non-trivial effect: among ICLR 2026 submissions receiving predominantly ratings of 0/1 out of 5, 50% have an author with another accepted paper, while 11% have a coauthor whose multiple submissions are all low-scoring.

2 Model

Consider an author with a *portfolio* of *candidate* papers—each potentially with coauthors—for submission. The conference proceeds as follows.

(1) Paper j has quality $q_j \in [0, 1]$. Authors privately know q_j, which is revealed to the conference only upon review. A paper is *weak* if $q_j < \theta \in [0, 1)$; *strong* otherwise.
(2) Authors value $\sum_j q_j r_j$, where r_j is paper j's review probability and the sum is over their candidate papers. This serves as a proxy for their expected number of acceptances.
(3) Starting from no submissions, authors consider candidates once in decreasing quality and submit a paper whenever every coauthor's objective weakly increases, accounting for its effect on their other submissions.
(4) All authors start eligible. Papers are reviewed in uniformly random order and skipped if any of the coauthor's averages have already fallen below θ.

We say that a paper is *withheld* when it is unsubmitted and *skipped* when submitted but not reviewed. Only reviewed papers affect averages, so skipping does not affect the authors' averages. The

conference sets the threshold θ and measures *welfare* as

$$W_\theta = \sum_j \big[(1-\theta)q_j r_j + \theta(1-q_j)(1-r_j)\big], \tag{1}$$

$$W_\theta^\star = \sum_j \max\{(1-\theta)q_j, \theta(1-q_j)\}. \tag{2}$$

The optimum $W_\theta^\star$ reviews exclusively strong papers, so a weak review loses $\theta - q_j$ relative to this benchmark, while leaving a strong paper unreviewed loses $q_j - \theta$. We also consider a second objective which is more amenable to our third practical constraint by only penalizing *repeated* weak reviews; i.e., at least one author has an previous weak review. The adjusted objective

$$W_\theta^{(1)} = W_\theta + \mathbb{E} \sum_{\text{initial weak review } j} (\theta - q_j) \tag{3}$$

does not penalize authors' initial weak reviews whilst penalizing repeat weak reviews and missed strong papers. Here, the objectives satisfy $W_\theta \le W_\theta^{(1)} \le W_\theta^\star$, where these only sum only over an author's candidate papers, measuring individual contribution to conference welfare rather than the entire conference.

3 Results

Our first four results consider a single author with $n \ge 1$ candidates whose reviews cannot be blocked by collaborators. Section 3.1 allows overlapping portfolios; all proofs are in Appendix A.

Self-screening. Consider adding a weak paper of quality $q < \theta$. Let $L(q)$ be the expected loss from existing submissions in the author's objective, conditional on the new paper being reviewed. WLOG, indexing submitted papers by decreasing quality, we have:

Theorem 1. *A weak paper is submitted exactly when $q \ge L(q)$, and submitted qualities must satisfy for every $k \ge 2$*

$$q_k \ge \min\left\{\theta, \frac{1}{k}\sum_{\ell<k} q_\ell\right\}. \tag{4}$$

In other words, candidates must offer enough improvement to one's expected acceptances to outweigh risking stronger work.

Early review stopping. When authors do not withhold weak papers, early stopping can still reduce the total expected number of reviews required. Of the n candidate papers, assume n_+ are strong and let R (resp. R_-) be the total of (resp. weak) papers reviewed:

Theorem 2. *The sharp bounds on the number of reviews are*

$$\mathbb{E}[R] \le n_+ + \frac{n-n_+}{n}, \quad \mathbb{E}[R_-] \le \frac{(n-n_+)(n_+ + 1)}{n}. \tag{5}$$

The expected number of reviews is at most one more than the number of strong papers, regardless of however many weak candidates the author may have.

Welfare guarantees. We now establish worst-case bounds on the objectives W_θ and $W_\theta^{(1)}$ as a function of W_θ^*:

Corollary 3. *For $n \ge 2$ and $\theta \le 1/2$ respectively, we have:*

$$\frac{W_\theta}{W_\theta^\star} \ge \min\left\{\frac{1}{2}, \frac{2(1-\theta)}{2-\theta}\right\}, \quad \frac{W_\theta^{(1)}}{W_\theta^\star} \ge 1 - \frac{1}{6(1-\theta)}. \tag{6}$$

These bounds are tight. Moreover, $W_\theta^\star - W_\theta \le \theta(n-n_+)/n \le \theta$.

At $\theta = 1/2$, the guarantees are one-half of optimal welfare, or two-thirds when first weak reviews are allowed. We also have an additive gap of at most θ between the optimum W_θ^* and W_θ, which trivially implies the same guarantee for $W_\theta^{(1)}$.

Average reviewed paper quality. Let Q be the total quality of reviewed papers; the minimum below is over submitted papers. We have the following average quality guarantee:

Corollary 4. *If $n_+ \ge 1$,*

$$\frac{\mathbb{E}[Q]}{\mathbb{E}[R]} \ge \theta \frac{n_+(n_+ + 1)}{n_+(n_+ + 1) + 1} \ge \frac{2\theta}{3}. \tag{7}$$

The final constant $2/3$ is sharp. For every reviewing order:

$$\frac{Q}{R} \ge \theta - \frac{\max\{0, \theta - \min_j q_j\}}{R} \ge \theta\left(1 - \frac{1}{R}\right). \tag{8}$$

With strong work to protect, mean quality per review across reviewing orders is at least $2\theta/3$, which approaches θ as the number of reviewed strong papers grows larger.

3.1 Extensions

Thus far, we have considered the mechanism's guarantees for a single author in isolation, rather than the entire conference. We extend our results to this broader setting where 1) all authors must agree to submit a paper and 2) an author with weak submissions can endanger the probability that the papers they coauthored are reviewed. For one author, let R_δ count reviewed papers with $q_j \le \theta - \delta$, where $\delta > 0$; the sum below is over their candidate papers.

Theorem 5. *For any submission choices and review order,*

$$R_\delta \le 1 + \left\lfloor \sum_j \max\{q_j - \theta, 0\}/\delta \right\rfloor. \tag{9}$$

Thus, an author with several strong papers can only sustain a small number of weak reviews. One immediate consequence of this is that an author whose papers are all weak receives at most one review, regardless of their number of submissions or coauthors—a useful guarantee to prevent authors from mass producing and submitting low quality papers (e.g., using AI [8]).

4 Experiments

We use 18,788 ICLR 2026 main-track papers representing 62,077 authors [2, 9] (including scored withdrawals but excluding desk rejections). We focus on ICLR as *all* submissions and reviews are posted to OpenReview, which avoids the issue of selection bias in opt-in large conferences such as NeurIPS or ICML. For each paper, we sample ratings according to a mixture of Gaussians with mean equal to the review score and weight proportional to reviewer confidence. We then standardize, and add unit-variance Gaussian noise. A paper's percentile among the samples gives q_j. We randomly sample 20 sets of quality assignments and vary $\theta = 0, 0.05, \ldots, 0.75$. In Table 1, we have three separate experiments.

(1) *Self-select* uses our model's decision procedure (all authors must benefit from submission) and implements early stopping.
(2) *Original* assumes that every paper is submitted, and only considers the savings from the stopping rule.

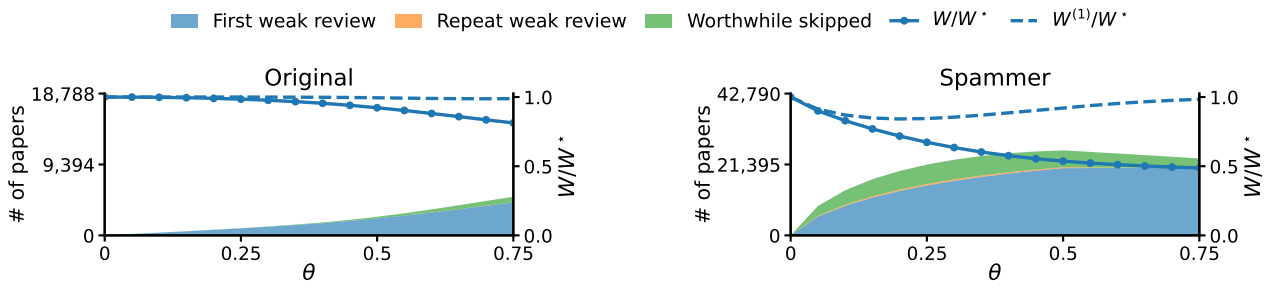


Figure 1: Original and Spammer performance on ICLR 2026 both with self-selection and early stopping.

Table 1: Reviewing outcomes at $\theta = 1/2$ for ICLR 2026.

	Original	Original (no self-select)	Spammer	Spammer (no self-select)
Total candidates	18,788	18,788	42,790	42,790
Total submitted	11,649	18,788	35,651	42,790
Total reviewed	**11,425**	9,348	24,821	22,966
Weak withheld	7,139	0	7,139	0
Weak skipped	6	4,568	5,712	13,172
Weak reviewed	2,248	4,826	20,545	20,224
Repeat weak	3	402	266	459
Strong skipped	217	4,872	5,118	6,653
Strong kept (%)	**97.69**	48.14	45.52	29.18
Mean reviewed q	0.66	0.49	0.28	0.26
Mean q non-repeat	0.75	0.71	0.72	0.67
$W_\theta/W_\theta^\star$ (%)	92.27	65.20	53.64	53.37
$W_\theta^{(1)}/W_\theta^\star$ (%)	99.45	81.24	92.04	89.41

(3) The *Spammer* setting assumes ~24000 additional, low-quality solo paper candidates (which roughly matches the 42790 submissions of ICLR 2027).[1]

Figure 1 and Table 1 show our headline empirical result which is that we retain most (97.69%) of strong papers whilst reducing the total reviewing workload by ~39% (see bolded entries in first column). We break this down further by separating the effects of self-selection versus avoiding repeated weak reviews.

- In our mechanism (column 1), we retain 98% of strong papers rather than 48% when authors do *not* self-select (column 2).
- For avoiding repeated weak reviews, we consider the no-self select or spammer settings, which save several thousand reviews.[2]
- We also note the impact of our equity constraint as W_θ and average quality deteriorate significantly in the spammer/no-self select settings, though $W_\theta^{(1)}$ and the mean q of non-repeat weak papers remain very strong.

[1]Each paper assigned to a uniform randomly chosen author, with quality uniform between zero and the smaller of 0.5 and that author's lowest original quality. Every addition is considered for submission alongside the original.

[2]Note that these savings could be much larger depending on how many AI paper candidates an author considers, which in our experiments is approximately 0.38 (62077 unique authors and 24000 added AI candidate papers).

5 Conclusion

Our proposed mechanism saves scarce reviewing effort by encouraging authors to withhold weak work and limiting repeated weak reviews which we show theoretically and empirically whilst maintaining that most strong papers are reviewed. Our mechanism respects three practical constraints, such as not requiring monetary transfers from authors or to reviewers, not enforcing a specific reviewing order, and requiring that every author is given a fair chance at review. We acknowledge the limitations of our analysis that 1) authors know q_j (often overestimated [6]), 2) the reviewers can gauge q_j accurately, and 3) papers receive one review. As for *practical* drawbacks of our mechanism, we note that the early stopping mechanism may incentivize reviewers to delay reviewing. However, we may weaken our mechanism (e.g., reducing θ) or combine with others to achieve the desired reduction in reviewing.

References

[1] ICLR 2026 Program Chairs. 2025. ICLR 2026 Response to LLM-Generated Papers and Reviews. https://blog.iclr.cc/2025/11/19/iclr-2026-response-to-llm-generated-papers-and-reviews/

[2] ICLR 2026 Program Chairs. 2026. ICLR 2026 Author Guide. https://iclr.cc/Conferences/2026/AuthorGuide

[3] ICLR 2027 Program Chairs. 2026. Submission policies for ICLR 2027. https://blog.iclr.cc/2026/09/02/submission-policies-for-iclr-2027/

[4] Weixin Liang, Zachary Izzo, Yaohui Zhang, Haley Lepp, Hancheng Cao, Xuandong Zhao, Lingjiao Chen, Haotian Ye, Sheng Liu, Zhi Huang, Daniel McFarland, and James Y. Zou. 2024. Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews. In *Proceedings of the 41st International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 235)*. PMLR, 29575–29620.

[5] Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. 2024. The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. arXiv:2408.06292.

[6] Charvi Rastogi, Ivan Stelmakh, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, Jennifer Wortman Vaughan, Zhenyu Xue, Hal Daumé III, Emma Pierson, and Nihar B. Shah. 2022. How do Authors' Perceptions of their Papers Compare with Co-authors' Perceptions and Peer-review Decisions? arXiv:2211.12966 [cs.LG] https://arxiv.org/abs/2211.12966

[7] Sakana AI. 2025. The AI Scientist Generates its First Peer-Reviewed Scientific Publication. https://sakana.ai/ai-scientist-first-publication/

[8] Rong Shan, Te Gao, Hang Zheng, Yunjia Xi, Jiachen Zhu, Zeyu Zheng, Yong Yu, Weinan Zhang, and Jianghao Lin. 2026. Position: Academic Conferences are Potentially Facing Denominator Gaming Caused by Fully Automated Scientific Agents. In *Proceedings of the 43rd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 306)*. PMLR, 172030–172042.

[9] Jing Yang, Qiyao Wei, and Jiaxin Pei. 2026. Paper Copilot: Tracking the Evolution of Peer Review in AI Conferences. In *International Conference on Learning Representations*.

[10] Yichi Zhang, Grant Schoenebeck, and Weijie Su. 2024. Eliciting Honest Information from Authors Using Sequential Review. *Proceedings of the AAAI Conference on Artificial Intelligence* 38, 9 (2024), 9977–9984.

A Proofs

A.1 Submission screening (Theorem 1)

Only candidates that reach review affect the author's objective. Conditional on review, the expected change is $q - L(q)$. Since the candidate can be first, it is submitted exactly when $q \geq L(q)$.

For (4), insert $q = q_k < \theta$ at each of the k positions in a review order of the $k-1$ existing submissions. The inserted paper can only lower the running average, and inserting it first prevents every existing submission from being reviewed.

(1) **The original running average stays at or above θ.** The sum of changes in the author's objective across the k positions is at most $kq - \sum_{\ell<k} q_\ell$.
(2) **The original running average first falls below θ at review t.** The candidate is reviewed only if inserted before one of those t papers. The sum of changes is at most tq minus their total quality, so it is nonpositive.

If $kq < \sum_{\ell<k} q_\ell$, both bounds are nonpositive, with strict loss when the order starts with a paper of quality above q. Averaging yields (4).

A.2 Reviewing limits (Theorem 2)

With all n candidates submitted, condition on the first paper.

(1) **Weak first paper.** This occurs with probability $(n - n_+)/n$ and produces exactly one review.
(2) **Strong first paper.** At most n papers, including $n - n_+$ weak papers, are reviewed.

Taking expectations yields (5). Both bounds are nondecreasing in the number of submitted papers of either type, so they hold for all possible subsets of papers. For tightness, consider strong papers with quality 1 and weak papers with quality $\theta - \varepsilon$, with $0 < \varepsilon < \theta$. If $(n - n_+)\varepsilon \leq 1 - \theta$, every paper is reviewed after a strong first paper because the running average stays at least θ.

A.3 Welfare (Corollary 3)

Write $Q_+ = \sum_{q_j \geq \theta} q_j$. All strong papers are submitted first, and each further submission weakly increases $\mathbb{E}[Q]$. Thus $\mathbb{E}[Q] \geq Q_+$, and (1)–(2) and Theorem 2 imply

$$W_\theta^\star - W_\theta = \theta(\mathbb{E}[R] - n_+) - \mathbb{E}[Q] + Q_+ \leq \theta(n - n_+)/n. \qquad (10)$$

When both strong and weak papers are submitted, omit withheld papers, which contribute the same amount to W_θ, $W_\theta^{(1)}$, and $W_\theta^\star$. Let s be the number of submitted weak papers and $q_{\min}$ the lowest submitted quality. By (4), we have $Q_+ < (n_+ + 1)\theta$ as

$$q_{\min} \geq \frac{Q_+ + (s-1)q_{\min}}{n_+ + s} \quad \Longrightarrow \quad q_{\min} \geq \frac{Q_+}{n_+ + 1}\,.$$

(1) **No weak submissions.** All strong papers are guaranteed to be reviewed, so we trivially have $W_\theta = W_\theta^{(1)} = W_\theta^\star$.
(2) **No strong papers.** Only high quality papers $q = \max_j q_j$ are submitted, and exactly one is reviewed. Thus $W_\theta^{(1)} = W_\theta^\star$ and

$$W_\theta^\star - W_\theta = \theta - q \leq W_\theta^\star/n, \quad \text{where } W_\theta^\star \geq n\theta(1-q)\,.$$

(3) **At least two strong papers and some weak submissions.** By (8), $Q - \theta R \geq q_{\min} - \theta$. Substituting this in (10) gives

$$W_\theta^\star - W_\theta \leq Q_+ - (n_+ - 1)\theta - q_{\min} \leq \frac{Q_+}{n_+ + 1}.$$

Since $W_\theta \geq (1-\theta)Q_+$, we have $W_\theta/W_\theta^\star \geq 3(1-\theta)/(4-3\theta)$ giving the first bound in (6). Also, (10) and $W_\theta^\star \geq (n_+ + s)\theta(1-\theta)$ yield

$$1 - \frac{W_\theta^{(1)}}{W_\theta^\star} \leq \frac{s}{(n_+ + s)^2(1-\theta)} \leq \frac{1}{8(1-\theta)}.$$

(4) **One strong paper and some weak submissions.** Write q_+ for the strong quality and $\bar{q}_-$ for the weak submissions' mean. The strong paper is reviewed only when first. A weak paper is reviewed only if it or the strong paper is first, so its review probability is at most $2/(s+1)$. Hence $\mathbb{E}[Q] \geq q_+$ implies $\bar{q}_- \geq q_+/2$, and $W_\theta^\star = (1-\theta)q_+ + \theta s(1-\bar{q}_-)$, as well as

$$W_\theta^\star - W_\theta \leq \frac{s(q_+ + \theta - 2\bar{q}_-)}{s+1}, \quad W_\theta^\star - W_\theta^{(1)} \leq \frac{sq_+ - \theta - (s-1)\bar{q}_-}{s+1}.$$

The last inequality subtracts $\theta - \bar{q}_-$ because the first weak review is equally likely to be any weak submission. Dividing the bound on $W_\theta^\star - W_\theta$ by $W_\theta^\star$ gives a fraction that decreases in $\bar{q}_-$. Substituting $\bar{q}_- = q_+/2$ yields

$$1 - \frac{W_\theta}{W_\theta^\star} \leq \frac{\theta}{(s+1)[(1-\theta)q_+/s + \theta(1 - q_+/2)]}.$$

The denominator is minimized at $s = 1$ since $\theta(1 - q_+/2) \geq (1-\theta)q_+/s$ for $s \geq 2$, using $q_+ \leq 2\theta$. The resulting bound $\theta/[2\theta + (2-3\theta)q_+]$ is at most $\max\{1/2, \theta/(2-\theta)\}$ because $0 \leq q_+ \leq 1$.
For $W_\theta^{(1)}$, use $\bar{q}_- \geq q_+/2$ in its loss bound and $W_\theta^\star \geq (1-\theta)(q_+ + s\theta)$:

$$(1-\theta)\left(1 - \frac{W_\theta^{(1)}}{W_\theta^\star}\right) \leq \frac{q_+/2 - \theta/(s+1)}{q_+ + s\theta} \leq \frac{s}{(s+1)(s+2)} \leq \frac{1}{6}.$$

The 2nd term increases with q_+, so the 2nd inequality uses $q_+ \leq 2\theta$.

To show tightness: the qualities $(0, 0)$ yield $W_\theta/W_\theta^\star = 1/2$ for $\theta > 0$. The qualities $(1, 1/2)$ give $W_\theta/W_\theta^\star = 2(1-\theta)/(2-\theta)$ for $\theta \geq 2/3$. For $0 < \theta \leq 1/2$, qualities $(2\theta - 2\varepsilon, \theta - \varepsilon)$ yield $W_\theta^{(1)}/W_\theta^\star \to 1 - 1/[6(1-\theta)]$ as $\varepsilon \downarrow 0$.

A.4 Average quality (Corollary 4)

The first $R-1$ reviews have total quality at least $\theta(R-1)$, so $Q - \theta R \geq q_{\min} - \theta$ for the lowest submitted quality $q_{\min}$.

(1) **No weak submissions.** Every reviewed quality is at least θ, which gives both bounds.
(2) **Some weak submissions.** The inequality above yields (8). The bound $q_{\min} \geq Q_+/(n_+ + 1)$ implies

$$\frac{\mathbb{E}[Q]}{\mathbb{E}[R]} \geq \frac{\theta\mathbb{E}[Q]}{\mathbb{E}[Q] + \theta - Q_+/(n_+ + 1)}.$$

Using $\mathbb{E}[Q] \geq Q_+ \geq n_+\theta$ gives (7).

To show tightness, papers with qualities $(\theta, \theta/2)$ attain $\mathbb{E}[Q]/\mathbb{E}[R] = 2\theta/3$, establishing sharpness of the final constant.

A.5 General authorship (Theorem 5)

Before the last review counted by R_δ, the sum of $q_j - \theta$ over the author's reviewed papers is nonnegative. Each of the preceding $R_\delta - 1$ such reviews has $q_j - \theta \leq -\delta$. Thus

$$(R_\delta - 1)\delta \leq \sum_j \max\{q_j - \theta, 0\}\,.$$