Gatekeeper Competition

Michèle Müller-Itten*

version of January 17, 2019

Abstract

I model access to influence as a two-sided matching market between a continuum of experts and two vertically differentiated gatekeepers under sequential directed search. Real-world examples include academic publishing, venture capital, job search or political agenda setting. The equilibrium is unique and exhibits red tape in the form of wasteful fees or excessive delay. However, only the top gatekeeper artificially delays matches to increase competition, a prediction that matches observed patterns within academic publishing. This delay at the top often improves equilibrium sorting and thereby enhances aggregate match surplus.

JEL: C72, C73, C78, D47, D4, D82, L10, L86.

*University of Notre Dame (email: michele.muller@nd.edu). A previous version of this paper circulated under the title 'Gatekeeping under Asymmetric Information'. I would like to thank Haluk Ergin, William Fuchs, Shachar Kariv, David Ahn, Tymofiy Mylovanov, Hector Chade, Chris Shannon, Thomas Gresik, Ben Hermalin, Aniko Oery, Emir Kamenica, Maciej Kotowski, Marc Müller, Tadeja Gracner, Satoshi Fukuda, Joseph Root, Brett Green, Lones Smith, and the participants of various seminars for insightful comments and discussions. All errors are of course my own. Financial support from the National Science Foundation (SES 1227707) is gratefully acknowledged.

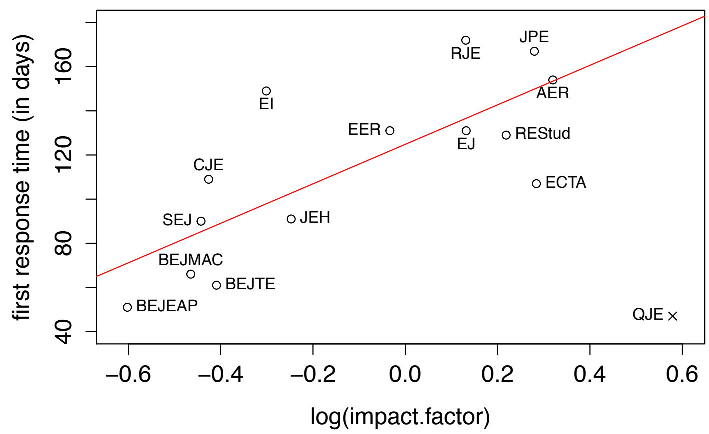
1 Introduction

Access to influence is often controlled by a small number of gatekeepers: Academics vie for publication in the top scientific journals and for funding from a handful of grant agencies, political motions must be endorsed by a member of parliament, specialists pursue a select number of jobs, and startup companies compete for financial support from a limited number of venture capitalists. There is typically no central marketplace for influence. Instead, experts approach candidate gatekeepers through directed, costly search. Understanding the implications of this market structure is crucial, since the realized matches ultimately shape public opinion, lawmaking, firm directions, technological progress and the direction of future research.

Several existing papers study one gatekeeper in isolation and offer valuable insights into the strategic considerations that shape their reactions,[1] but only a full equilibrium model is able to account for the entire effect on other market participants, explain structural features within these industries, and offer welfare comparisons. This paper aims to bridge that gap by proposing a tractable equilibrium model of decentralized matching that captures the main strategic and informational features of these markets. I am particularly interested in the emergence of 'red tape', an umbrella term that captures purely wasteful delays, tasks or fees. Complaints around gatekeeper red tape are ubiquitous, yet there is little consensus both around its structure (Are some forms of red tape more effective than others? Which gatekeepers are most likely to employ red tape in equilibrium?) and its welfare implications (Would other market participants be better off without? Would influence be attributed more efficiently?).

Since red tape plays the role of a screening device that steers expert search, I find that the most effective costs are unconditional of acceptance – e.g. submission, rather than acceptance or publication, fees (Theorem 2). Excessive delay in gatekeeper responses emerges as a dominant form of red tape, due to social or legal restrictions, equity concerns, or cost reimbursement policies. In equilibrium however, only the top gatekeeper uses unnecessary delay (Theo-

[1] A detailed discussion of the relevant literature is included at the end of the introduction.



	excluding QJE		including QJE	
$\log_{10}$(impact.factor)	89.49***	(22.11)	45.27	(28.87)
Intercept	124.84***	(7.52)	113.30***	(10.42)

Table 1: First response times at economic journals as a function of impact factor. There is a highly significant positive relationship between a journal's impact factor and editorial delays (p-value 0.16%), once the obvious outlier representing the Quarterly Journal of Economics ($\times$) is removed. The coefficient remains positive but loses its significance when the outlier is included.

rem 3). The prediction matches stylized empirical patterns among economic journals (see Table 1[2]) and echoes Ellison (2002)'s observation of a negative link between review speed and journal rank. Vaguely speaking, adding delay lowers the gatekeeper's attractiveness and thus intensifies competition. This is beneficial for the top, but detrimental for the bottom gatekeeper (Theorem 5). After all, the bottom gatekeeper attracts poor-quality experts and thus improves the quality of experts who approach the top. This positive externality may be one reason why several journals (such as Nature, AER, or PLOS)

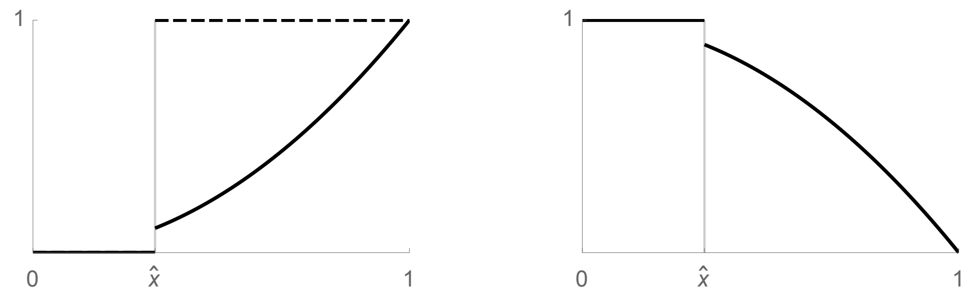
[2]Response times are as reported in Azar (2007). The sample is restricted to journals in economics, and uses the mean first response time for first submissions or the closest report thereof. 2001 impact factors are retrieved from the Journal Citation Reports, ISI Web of Knowledge, http://admin-apps.webofknowledge.com/JCR/JCR on September 19th, 2015. Missing data for Berkeley Economics Journals was replaced by 2009 ISI impact factor data compiled by the Tepper School of Business, accessed September 20th, 2015, https://server1.tepper.cmu.edu/barnett/rankings.html.

have launched lower-ranked sub-journals of lower prestige. The opposite is true for the top journal, which diverts the highest-quality experts before they approach the bottom. The desire to appear more attractive may be one reason for endogenous specialization of lower-tier journals across scientific disciplines. Indeed, by specializing on a quality-irrelevant attribute, the bottom journal signals higher acceptance odds to a subpool of authors and thus strengthens its strategic position.

Turning to the welfare questions, I find that the delay by the top gatekeeper can in fact benefit both gatekeepers through the resulting equilibrium changes in their match distribution. Indeed, as the top gatekeeper slows down, more experts first approach the bottom gatekeeper in equilibrium. In response, the top gatekeeper accepts experts' match offers more easily, the bottom less. This can even improve the payoff for certain experts. So despite not being Pareto ordered, the delay by the top gatekeeper often improves the assortativity of final matches and thus enhances overall match surplus (Theorem 4).

For concreteness, I phrase the model in terms of academic publishing, but the insights translate to the examples mentioned above. Specifically, I assume that there are two journals (or other gatekeepers such as grant agencies, elected officials, specialized firms, venture capitalists) of varying prestige (or influence) γ_j. A continuum of authors (or other experts such as interest groups, job candidates, startup companies) each try to publish a manuscript of quality x (promote an idea, land a job, obtain funding) with one of the two journals. For brevity, I refer to journals as 'she' and authors as 'he'. An author submits to one journal at a time, waits for the journal's acceptance decision, and then either matches successfully or submits elsewhere. A realized match yields instantaneous, non-transferable utility γ_j to the author and x to the journal.

The model incorporates two types of market frictions: First, information on quality is asymmetric and journals only observe a noisy signal of x. Second, search is costly since authors are impatient and submissions take time. Each author thus has to weigh potential impact against the opportunity cost of postponing other submissions. His individual acceptance odds determine the optimal search strategy. Each journal is also aware that her acceptance



(a) The top journal receives submissions from exactly those authors with high enough quality $x \geq \hat{x}$ (dashed line). Among these submissions, better authors generate better signals and thus match more frequently (solid line).

(b) The bottom journal receives submissions from authors with quality below $\hat{x}$ *as well as* from those that were previously rejected by the top journal. Since she accepts all submissions, her submission and match distribution coincide.

Figure 1: Submission and match distribution for the stylized Example 1.

strategy not only affects which manuscripts she *accepts*, but also who *submits* to her in the first place.

Example 1. To illustrate how directed search affects match outcomes, consider a stylized case. The prestige of the bottom journal is normalized to one and she *immediately* accepts *all* submissions. The top journal has twice her prestige, but takes six months before responding to submissions, accepting only those with a high enough signal. Submitting first to the top journal is optimal only for the highest-quality authors: Their signal draws are usually high, and they expect their manuscript to be accepted. Lower-quality authors instead prefer to match immediately with the bottom journal in order to avoid the delay of a rejected submission.

Consequently, the top journal here only receives submissions from the upper tail of the quality spectrum, and rejects them with positive probability. Rejected authors then submit to the bottom journal, whose submissions (and matches) thus span the entire quality range, as illustrated in Figure 1. ◇

While useful for basic intuition, blanket acceptance by the bottom journal is not realistic and fails to capture the interaction between optimal journal strategies. To study journal competition, the paper assumes that there are more manuscripts than the joint capacity of the two journals. The top journal now also receives *adversely selected* low-quality submissions, and the bottom adjusts her signal cutoff in response to policy changes at the top. The resulting equilibrium remains unique, and can be recast as a one-shot game between journals with intuitive comparative statics (Theorem 1). The competition between journals is governed by a desire to reduce low-quality submissions and thereby decrease the reliance on their own noisy assessment of manuscript quality.

Related Literature. Within the decentralized matching literature, equilibrium sorting is a central concern. I show that the equilibrium strategies are monotone when the top journal owns a small fraction of the overall capacity. This echoes a similar result by Chade et al. (2014) for costly and simultaneous one-shot search in college applications. The focus of the two papers is however quite different: The study of potential sorting failures are a central contribution of their paper, while I focus on parameters that ensure sorting and investigate the competition that results between journals.[3] There is no analogue to the competition or welfare results in the simultaneous-search environment because equilibrium multiplicity severely limits meaningful comparative statics.

Previous papers that explore competition among academic journals all focus on a partial equilibrium setup.[4] The driving factor in these models is the assumption of a *costly* refereeing process, resulting in direct savings from *any* reduction in submission load. In contrast, journals here are motivated purely by the quality of their matches, and as such the present analysis supplements,

[3]The working paper version (Müller-Itten, 2017) contrasts the sorting failures between simultaneous and sequential search in more detail.

[4]Weitzman (1979), Oster (1980), Heintzelman and Nocetti (2009) and Salinas and Munch (2015) characterize the optimal submission strategies for academic authors. The decision problem they solve is identical to the one faced by individual authors in this paper. Baghestanian and Popov (2015) further endogenize authors' effort level and explore the impact of exogenous changes in the publishing process. Leslie (2005), Azar (2005, 2007) and Cotton (2013) discuss the possible benefits of long editorial decision times and monetary submission costs at a single journal.

rather than restates, the previous rationale for monetary fees and delay. Moreover, only the full equilibrium model generates an asymmetry across top and bottom journals, and allows for welfare statements that can anticipate the impact of technological change which drastically reduces feasible delays.

The remainder of the paper is organized as follows: Section 2 describes the formal model. Section 3 identifies the unique equilibrium in three stages: It first solves for the subgame equilibrium, then for a restricted version of the full game where response times are set exogenously, and finally for the full model with the possibility of red tape. A variant with monetary fees is included in Section 3.3.1. Welfare measures and implications are discussed in Section 4. The conclusion briefly addresses what happens under two-sided asymmetric information or endogenous journal prestige, and summarizes the main results.

2 Model Setup

Players. This paper studies competition in a decentralized, two-sided matching market between two vertically differentiated journals of prestige $\gamma_1 > \gamma_2$ and a continuum of authors. To remove reputation concerns, I assume that each author is endowed with a single manuscript. The manuscript's quality x is independently drawn according to an absolutely continuous distribution F over $X := (0, 1)$ with density f. To simplify the language, I identify authors by the quality of their manuscript and refer to journals as "she" and to authors as "he".

Mechanics. The game proceeds in three stages. In the first stage, journals simultaneously hold an editorial board meeting and commit to acceptance guidelines for referees and time targets for article turnaround. In the second stage, each author chooses a submission strategy for his manuscript. A (pure) submission strategy is a permutation $\omega \in \{(1, 2), (1, 2)\}$ that specifies the order in which the author plans to contact journals 1 and 2. The third stage is nonstrategic. It is an automated process with up to two rounds. In round 1, each author submits his manuscript to journal ω_1. Journal $j = \omega_1$ observes a quality signal $\sigma_j|x$ independently drawn from an exponential distribution G_x with rate parameter $\lambda(x) > 0$. The function $\lambda : X \to \Sigma := (0, \infty)$ is strictly decreasing to ensure a positive correlation between manuscript quality and signal values. I assume that the worst manuscript always generate the lowest signal, $\lim_{x \to 0} \lambda(x) = \infty$.[5] If this random draw meets the journal's acceptance guidelines, the manuscript is published and the author exits the game. If not, the author moves on to round two, which is structured the same way.

Strategies. A journal's strategy has two components: The acceptance guidelines $a_j : \Sigma \to [0, 1]$ denote the probability of acceptance $a_j(\sigma)$ for a manuscript with signal $\sigma \in \Sigma$. The turnaround time $T_j > 0$ determines the waiting time experienced by authors between initial manuscript submission and receipt of the journal's decision. Since all authors have the same discount rate r, I express this decision (equivalently) in terms of the resulting author discount factor $\delta_j = e^{-rT_j} \in (0, \bar{\delta}_j]$. The upper bound $\bar{\delta}_j < 1$ is exogenous, representing a feasibility constraint on the minimal turnaround time. To simplify language, I assume that the maximal discounted prestige of journal 1 exceeds that of journal 2, $\bar{\delta}_1 \gamma_1 > \bar{\delta}_2 \gamma_2$, and refer to them as the 'top' and 'bottom' journal respectively.

Conditional on journal policies $\mathcal{J} = (a_j, \delta_j)_{j \in \{1,2\}}$, each author chooses a (possibly stochastic) submission strategy in $\Delta \{(1, 2), (1, 2)\}$. I let $\psi_x^{\mathcal{J}}(\omega)$ denote the fraction of quality-x authors who submit according to permutation ω under policies $\mathcal{J}$.

Three auxiliary functions are useful for the later discussion: First, the *acceptance rate* is the likelihood of acceptance for x submissions at journal j. It is determined by j's acceptance guidelines and the conditional signal distribution, $\alpha_j(x) = \mathbb{E}_\sigma [a_j(\sigma)|x]$. Second, the *submission rate* is the ex-ante likelihood that a x manuscript is submitted to journal j during either round. This happens for sure if j is the first submission, $\pi_{\omega_1}(x|\omega) = 1$. If j is the

[5] Empirical support for the assumption of noisy but informative referee reports include Bornmann et al. (2010); Baethge et al. (2013); Welch (2014); Card and DellaVigna (2017).

second submission, this happens exactly if x's first submission is rejected, $\pi_{\omega_2}(x|\omega) = 1 - \alpha_{\omega_1}(x)$. For mixed strategies, the ex-ante likelihood of submission is given by the expectation $\pi_j(x) = \sum_\omega \psi_x^{\mathcal{J}}(\omega)\pi_j(x|\omega)$. Finally, the *match rate* is the ex-ante likelihood that a x manuscript is ultimately published in journal j. It is given by the product $\mu_j(x) = \alpha_j(x)\pi_j(x)$. When I want to stress the dependence on a specific parameter ξ, I write $\mu_j(x|\xi)$.

Payoffs. Journals care only about the quality of matches, with no regards to match timing. Published manuscripts add to the journal's *match value* $V(\mu_j) = \int_X x\mu_j(x)f(x)dx$ and *match volume* $M(\mu_j) = \int_X \mu_j(x)f(x)dx$. Journals have limited capacity κ_j; any excess matches incur a per-unit penalty $K > 1$.[6] Journal j's payoff from match μ_j is equal to her match value net of excess-capacity penalties,

$$\bar{V}(\mu_j) = V(\mu_j) - K \cdot \max\{0, M(\mu_j) - \kappa_j\}.$$

Author's care both about the final match and its timing. An acceptance by journal $j = \omega_k$ offers an instantaneous utility gain γ_j, representing the associated boost in the author's prestige. Each period k has an associated discount factor δ_{ω_k} that is controlled by journal ω_k's turnaround time. Consequently, an author with submission strategy (j,k) obtains a discounted payoff of $\delta_j\gamma_j$ if his first submission is successful, and $\delta_1\delta_2\gamma_k$ if only the second is. Accounting for the likelihood of acceptance, his expected discounted utility from submission strategy ω under journal strategies $\mathcal{J}$ equals

$$u_x^J(\omega) = \alpha_{\omega_1}(x)\delta_{\omega_1}\gamma_{\omega_1} + \big(1 - \alpha_{\omega_1}(x)\big)\alpha_{\omega_2}(x)\delta_{\omega_1}\delta_{\omega_2}\gamma_{\omega_2}.$$

Equilibrium. A (subgame perfect) equilibrium of the game specifies journal policies $\mathcal{J}^*$ and conditional author submission plans $\psi_x^{\mathcal{J}}$ such that

(i) For any journal policies $\mathcal{J}$, authors choose permutations that maximize

[6] In equilibrium, journals face no aggregate uncertainty with respect to their total match volume, and the exact size of the penalty K is therefore inconsequential.

their expected discounted match utility,

$$u_x^{\mathcal{J}}(\omega) \geq u_x^{\mathcal{J}}(\omega') \qquad \forall \omega, \omega' \in \{(1,2),(2,1)\} \text{ with } \psi_x^{\mathcal{J}}(\omega) > 0.$$

(ii) Journals maximize their own payoff by anticipating both the direct acceptance effect of a_j on $\alpha_j(x)$, as well as indirect submission effects of $\mathcal{J}_j$ on $\psi_x^{\mathcal{J}}$. In other words, $\bar{V}(\mu_j^*) \geq \bar{V}(\mu_j)$ for any $\mathcal{J}_j = (a_j, \delta_j)$, where μ_j^* is defined from journal policies $\mathcal{J}^*$ and author submission plans $\psi_x^{\mathcal{J}^*}$, and μ_j is defined from $(\mathcal{J}_j, \mathcal{J}_{-j}^*)$ and $\psi_x^{(\mathcal{J}_j, \mathcal{J}_{-j}^*)}$.

Parameter assumptions. Total capacity is insufficient to publish all manuscripts, $\kappa_1 + \kappa_2 < 1$. To capture the stickiness of journal reputation (Card and DellaVigna, 2013), journal prestige γ_j is exogenous and independent of the realized matches. Although crucial for tractability,[7] this is not the main driver behind the results. To substantiate this claim, all numerical examples are chosen so that ex-post realized mean quality $V(\mu_j)/M(\mu_j)$ is proportional to the assumed journal prestige.

3 Equilibrium analysis

3.1 Reduced Form

Using backward induction, I first show that the equilibrium features cutoff strategies for journals, which results in essentially unique subgame equilibria. This allows me to study the first stage in reduced form as a simultaneous game between journals.

Subgame Equilibrium. Authors are competing only indirectly through their impact on journal capacity. Conditional on $\mathcal{J}$, each author faces an independent, sequential search problem, where he weighs the potential prestige gain against delaying (and possibly foregoing) other submissions. Weitzman

[7]Endogenous prestige creates a coordination problem among authors (see Section 5).

(1979) and Gittins (1979) show that the optimal strategy is captured by the scores

$$z_j(\alpha_j) = \frac{\delta_j \alpha_j \gamma_j}{1 - \delta_j(1 - \alpha_j)}.$$

An author of a quality-x manuscript submits according to decreasing $z_j(\alpha_j(x))$ score.[8] Intuitively, the z_j score captures both the expected discounted benefit $\delta_j \alpha_j \gamma_j$ from an isolated submission to j as well as the externality imposed from delaying any further submissions in the case of rejection, which happens with probability $1 - \alpha_j$. Authors only vary with respect to the acceptance rate $\alpha_j(x)$, causing authors with lower-quality manuscripts to place relatively more weight on minimizing the externality. The partial derivatives of z_j indicate how submission rates vary with journal policies. It is unsurprising that high prestige, fast response times or high acceptance rates raise a journal's score, and thereby (weakly) increase the likelihood of a first submission.

Observation 1. The z_j score is increasing in prestige γ_j, discount factor δ_j and acceptance rate $\alpha_j(x)$.

Since both α_1 and α_2 grow with manuscript quality, it is not generally true that only the best manuscripts are first submitted to the top journal.[9] We will recover such monotonicity on the equilibrium path (see Theorem 1). For now, let me make a weaker observation that applies to all subgames: Each author exits the game unmatched only if both journals reject his submission. Since both the individual signal draws σ and the acceptance probabilities $a_j(\sigma)$ are history-independent, an author's strategy has no bearing on his overall match probability.

Observation 2. Author x fails to match with probability $(1 - \alpha_1(x))(1 - \alpha_2(x))$, no matter his submission strategy.

[8] The optimal strategy implies that some authors "move up the ranks" after a rejection. The intuition for this is clear: Authors may first try a safer bet, but then go for a low-odds-high-payoff journal when the remaining options offer little potential prestige. There is also empirical support for such behavior: In the Calcagno et al. (2012) dataset for instance, bout 16% of resubmissions were ultimately published in higher-impact journals.

[9] The working paper version of this article (Müller-Itten, 2017) contains a detailed discussion of possible sorting failures under a slightly more general setup.

The overall match probability is therefore trivially increasing in manuscript quality x, and rises whenever a journal accepts more signal realizations. Equilibrium uniqueness in Theorem 1 is a direct consequence of this observation.

First-stage equilibrium. Journals try to limit matches with low-quality authors in favor of those with high-quality. Obvious profitable deviations for a journal are those that replace bad matches one-for-one with good ones. Formally, $\tilde{\mu}_j$ is *steeper* than μ_j around anchor x_0 if $(\tilde{\mu}_j(x) - \mu_j(x))(x - x_0) \geq 0$ with strict inequality on a set of qualities with positive F-measure. In other words, $\tilde{\mu}_j$ reduces match rates below quality x_0 and increases them above. This obviously raises journal payoff as long as the match volume is also preserved, $M(\tilde{\mu}_j) = M(\mu_j)$. We can replace the volume restriction with a range and continuity assumption on the set of feasible deviations by invoking the Intermediate Value theorem.

Lemma 1. *Let M denote the metric space of all measurable functions from X to $[0, 1]$. Assume that there exist deviations $\mathcal{M} : X \to M$ that each steepen the match rate relative to μ_j, parametrized by their anchors $x_0 \in X$. If $M(\mathcal{M}(x_0))$ is continuous in x_0, then there exists $\hat{x}_0 \in \mathit{int}(X)$ with $\bar{V}(\mathcal{M}(\hat{x})) > \bar{V}(\mu_j)$.*

Proof. See Appendix A. □

In equilibrium, there are no profitable deviations, and so Lemma 1 strongly reduces the set of feasible equilibrium strategies. One consequence is that all acceptance policies take the form of cutoff rules.

Lemma 2. *In equilibrium, all acceptance strategies take the form of cutoff rules.*

Proof. See Appendix A. □

Cutoff rules favor high signals, which skews acceptance towards high-quality manuscripts. However, since journals move first, the proof has to account for the change in author submissions. Luckily, the monotonicity of z_j implies that authors respond to j's increased acceptance by moving their journal-j submission to a (weakly) earlier round. The submission rate π_j thus *amplifies* any changes in the acceptance rate α_j. By using cutoff rules, a journal both accepts and attracts more high-quality manuscripts.

An important assumption of the model is that no manuscripts are of negative quality,[10] nor do any dominate the excess-capacity penalty $K > 1$. This rules out profitable capacity manipulation as in Sönmez (1997): Since the match rate is decreasing in the cutoff $\underline{\sigma}_j$ at *every* quality level x, any $\underline{\sigma}_j$ above the capacity-clearing cutoff forgoes valuable matches with *all* authors x, while the excess-capacity penalties for any $\underline{\sigma}_j$ below negate the gains from any additional matches. In equilibrium, each journal thus either accepts all submissions, $\underline{\sigma}_j = 0$, or exhausts her capacity, $M_j = \int_X \mu_j(x) f(x) dx = \kappa_j$. However, since joint capacity is insufficient to publish all manuscripts, and overall match probability is independent of submission order (Observation 2), the former does not occur in equilibrium.

Observation 3. In any equilibrium, $(\underline{\sigma}_1, \underline{\sigma}_2) > 0$, and both journals are at full capacity.

The exponential signal structure also simplifies the equilibrium analysis: It ensures that in any subgame where journals use cutoff strategies, all but a F-measure zero set of authors have a unique dominant strategy. The purely technical proof is based on counting roots of a generalized Dirichlet polynomial, and has been relegated to the appendix (Lemma 4). The resulting uniqueness[11] of subgame equilibria allows us to study the game in reduced form as a simultaneous game between journals.

3.2 Restricted Equilibrium

I first solve the model for exogenously fixed turnaround times or, equivalently, fixed discount factors δ_j. In this restricted strategy space, the model admits

[10]This assumption can be replaced with distributional assumptions that make sure that total capacity is far smaller than the inflow of worthwhile submissions. The crucial feature is captured in Observation 3: All capacities are met in equilibrium.

[11]I consider two strategy profiles identical if they only differ by the behavior of a F-measure-zero set of authors.

a succinct proof of equilibrium existence and uniqueness.[12] It is also possible to show that when the top journal is small enough relative to total capacity, equilibrium strategies are monotone in the sense that the top journal is more selective, $\underline{\sigma}_1 > \underline{\sigma}_2$, and receives first submissions only from the highest-quality authors, $u_x^{(\boldsymbol{\sigma},\boldsymbol{\delta})}\big((1,2)\big) \geq u_x^{(\boldsymbol{\sigma},\boldsymbol{\delta})}\big((2,1)\big)$ if and only if $x \geq \hat{x}$.

Theorem 1. *For any fixed discount factors* $(\delta_1, \delta_2) \in (0, \bar{\delta}_1] \times (0, \bar{\delta}_2]$*, a unique equilibrium exists. If* κ_1 *is small enough relative to total capacity* $\kappa_1 + \kappa_2$*, equilibrium strategies are monotone for any discount factors. Finally, journal cutoffs vary continuously in all external parameters.*

Proof. See Appendix A. □

The proof relies on the strategic complementarities among journal cutoffs. The intuition for this is twofold: As an author's acceptance rate at the competing journal k decreases, so does his z_k score and he might now propose to journal j *earlier*. In addition, even *conditional* on submission order, the author proposes to j more often since any previous submissions are more likely rejected. To counter her surge in submissions, journal j therefore also accepts fewer signals. The complementarities ensure that the best response function Φ is monotonic, and hence equilibrium existence follows from Tarski's Fixed Point Theorem. Uniqueness is a consequence of Observations 2 and 3: Two distinct ordered fixed points would differ in the total match volume irrespective of authors' submission orders, and at most one meets total capacity $\kappa_1 + \kappa_2$. This uniqueness also implies continuity by Berge's Theorem.

Finally, a sufficient condition for monotone strategies is obtained by computing the hypothetical match volumes (M_1, M_2) that result if both journals employ the same cutoff s and all authors first submit to the bottom journal. If $\kappa_1 \leq M_1$, yet $\kappa_1 + \kappa_2 = M_1 + M_2$, the monotonicity of total matches requires $\underline{\sigma}_2 \leq s \leq \underline{\sigma}_1$. Indeed, the cutoffs have to lay on either side of s in order to

[12] A significantly longer proof in the working paper version (Müller-Itten, 2017) establishes equilibrium existence for any finite number of journals, and any signal structure that satisfies the monotone likelihood ratio property. Journal cutoffs remain unique in the more general setting.

maintain total match volume (which is independent of author strategies by Observation 2), and the top journal would exceed capacity if $\underline{\sigma}_1 \leq s \leq \underline{\sigma}_2$.[13] To prove that authors behave monotonically in any subgame with monotone cutoffs, I show algebraically that the score differential $z_1(\alpha_1(x)) - z_2(\alpha_2(x))$ satisfies a single crossing property.

It is worth highlighting that the distributional assumption $\lim_{x \to 0} \lambda(x) = \infty$ implies a non-zero mass of authors who first submit to the bottom journal, since

$$\lim_{x \to 0} \frac{z_2(x)}{z_1(x)} = \lim_{x \to 0} \frac{(1-\delta_1)\delta_2 \alpha_2(x) \gamma_2}{\delta_1(1-\delta_2)\alpha_1(x)\gamma_1} = \frac{(1-\delta_1)\delta_2\gamma_2}{\delta_1(1-\delta_2)\gamma_1} \lim_{x \to 0} e^{\lambda(x)(\underline{\sigma}_1 - \underline{\sigma}_2)} = \infty$$

under monotone cutoffs $\underline{\sigma}_1 > \underline{\sigma}_2$. For the remainder of the paper, I focus exclusively on monotone equilibria by assuming that the top journal κ_1 is small relative to total capacity $\kappa_1 + \kappa_2$. Any equilibrium can then be parsimoniously characterized by the journal cutoffs $\underline{\boldsymbol{\sigma}}$, discount factors $\boldsymbol{\delta}$, and the quality of the unique indifferent author $\hat{x} \in X = (0, 1)$.

We have so far treated the discount factors $\boldsymbol{\delta}$ as exogenous. Allowing journals to modify turnaround times endogenously is the focus of the next section. To do so, it is useful to formally state the comparative statics of the equilibrium, which are entirely intuitive in the case of only two journals.

Lemma 3. *A rise in γ_j or δ_j causes a rise in journal j's cutoff and a drop in her opponent's cutoff.*

Proof. See Appendix A. □

An increase in prestige or turnaround speed increases the attractiveness of a journal, which increases the mass of first submissions. The journal responds by being more selective. Meanwhile, the opponent is less selective in order to maintain capacity despite a slump in submissions.

[13] Match volume for the top journal increases through three channels: The drop in its own cutoff $\underline{\sigma}_1$, the increase in the opponent cutoff $\underline{\sigma}_2$, and any positive mass of first submissions.

3.3 Red Tape

The coupling of author directed search and information frictions generates subtle strategic tradeoffs for a journal: She wants to reject bad submissions, but relying on her noisy signal will inadvertently discourage submissions across the quality spectrum. Many other features of real-world gatekeepers also factor into the optimal author search: Ease of access, regulatory hurdles, customization requirements, explicit monetary costs or response times to name a few. To the extent that a gatekeeper can control these, she may exploit them as additional screening mechanisms.

In this section, I want to ask whether a gatekeeper ever implements purely wasteful delays, tasks or fees, which I refer to jointly as *red tape*. Since red tape is profitable when it discourages mostly low-quality submissions, important follow-up questions also emerge: Are some forms of red tape more effective than others? Which gatekeepers use red tape in equilibrium? And what are the social ramifications from increasing red tape?

3.3.1 Monetary fees

The main model does not include monetary fees as a way to deter authors with low acceptance rates. Instead, it focuses solely on spurious delay as a way to screen authors. In some applications (e.g. political agenda-setting), it would in fact be illegal for gatekeepers to levy monetary access fees. In scientific publishing, the fee structure itself suggests a very limited effect on author payoffs, for reasons that I'll explore here. I slightly expand the model to allow for two types of monetary fees. The first charge $c_j \geq 0$ is due upon manuscript submission irrespective of its ultimate acceptance. I refer to this as a 'submission fee', it is also known as a handling charge. The second charge $p_j \geq 0$ is due conditional upon acceptance of the manuscript. I refer to this as a 'publication fee', it is also known as a manuscript processing charge.

Revenue from either type of fee obviously increases a journal's financial profitability, but that is besides the point of this paper. Instead, I want to focus on firm competition when its customers are engaged in directed search.

To this end, I assume that any financial returns are 'burned' without creating residual benefit to the journal, and ask: Can monetary fees serve as a screening device for a journal that is motivated purely by manuscript quality? – To answer this question, I show that the different forms of red tape are imperfect substitutes: All three lower journal attractiveness, but their relative salience varies with the acceptance rate α_j. Intuitively, submission fees are offset by an expected prestige boost only if acceptance is likely enough, and thus tend to dissuade primarily low-α_j authors. Publication fees are incurred only by successful authors, which makes them more salient for high-α_j authors. Finally, slow turnaround times are particularly costly to authors who delay other promising submissions. These tend to be intermediate-quality authors, for whom a rejection by the current and later acceptance by another journal is likely. In other words, in terms of screening authors for quality, submission fees are more effective than delay, which in turn is more effective than publication fees.

Theorem 2. *If monetary fees affect author utility, journals screen authors through submission fees, rather than delay. In the absence of submission fees, they use excessive delay rather than publication fees.*

Proof. See Appendix A. □

Formally, I show that a journal can steepen its match rate by replacing unnecessary delay with submission fees. By Lemma 1, any journal would thus charge maximal submission fees[14] before resorting to delay. Previous papers have noted a similar tradeoff (Azar, 2006; Heintzelman and Nocetti, 2009), which they then broadly interpreted as evidence in favor of monetary fees, especially as a way of speeding up the publication process.

However, the second part of Theorem 2 casts some doubt on that line of logic: It shows that in the absence of submission fees, a journal can steepen its match rate by replacing publication fees with unnecessary delay. This clear ranking of fees seems at odds with empirical observations. Indeed, submission

[14] An upper bound on these fees could arise from equity concerns (Cotton, 2013) or social norms.

fees are relatively rare, and virtually unheard of in many disciplines other than business, economics, finance, and experimental biology (Ware, 2010). Especially among open-access journals, *publication* fees (if any) are the norm (Solomon and Björk, 2012; West et al., 2014; Van Noorden, 2013). The second part of Theorem 2 casts some doubt on that line of logic: It shows that in the absence of submission fees, a journal can steepen its match rate by replacing publication fees with unnecessary delay. This clear ranking of fees seems at odds with empirical observations. Indeed, submission fees are relatively rare, and virtually unheard of in many disciplines other than business, economics, finance, and experimental biology (Ware, 2010). Especially among open-access journals, publication fees (if any) are the norm (Solomon and Björk, 2012; West et al., 2014). A frequent concern regarding submission fees is that they may discourage promising submissions. Theorem 2 shows that this criticism applies even more to publication fees. Are the many journals who charge publication fees missing this basic intuition? Or is the model wrong when it assumes that authors account for monetary fees in their cost-benefit analysis? – After all, these fees are usually covered by grants or adsorbed by the author's research department. Only delay is borne privately by the author. So another way of reconciling the dominant fee structures with the model is to interpret it as evidence that delay is the only cost affecting the author's submission strategy. This is the approach I take in this paper.

3.3.2 Spurious delay

A capacity-preserving change that lowers both turnaround speed and signal cutoffs affects authors' optimal submission strategy. However, such a change does not typically result in a steepening of the match rate. Instead, a marginal reduction of $\underline{\sigma}_j$ increases acceptance rates everywhere, whereas the marginal reduction of δ_j lowers submission rates only for the indifferent author $\hat{x}$. Thus, spurious delay enhances journal payoffs if and only if $\hat{x}$ is low enough. As the next result shows, the equilibrium implications are two-fold: The top journal employs spurious delay when the prestige ratio γ_1/γ_2 is big enough. However; the bottom always minimizes turnaround times to strengthen its competitiveness.

Theorem 3. *If κ_1 is small enough relative to total capacity $\kappa_1 + \kappa_2$, a unique equilibrium exists. Delay is minimal at the bottom, $\delta_2 = \bar{\delta}_2$. If the prestige ratio γ_1/γ_2 is large enough, the top journal employs spurious delay, $\delta_1 < \bar{\delta}_1$.*

Proof. See Appendix A. □

Spurious delay implies that journal j replaces matches with $\hat{x}$ in favor of matches with other authors $x \neq \hat{x}$. Because total capacity is preserved, the rate of substitution $d\underline{\sigma}_j \frac{\partial \alpha_j(x)}{\partial \underline{\sigma}_j} \pi_j(x) f(x)$ is dictated by the change in her signal cutoff. The net impact on her payoff is thus proportional to

$$S_j := \int_X (x - \hat{x}) g(\underline{\sigma}_j | x) \pi_j(x) f(x) dx \tag{1}$$

where $x - \hat{x}$ is the payoff difference between matches with x and $\hat{x}$. The deviation is profitable if and only if $S_j > 0$ or, equivalently, if the marginal first submission $\hat{x}$ is low enough relative to the expected quality of a marginal acceptance. In a monotone equilibrium, the top journal has both a higher acceptance threshold $\underline{\sigma}_1 > \underline{\sigma}_2$ and receives the best manuscripts first. Fixing submissions, the quality of a marginal acceptance at the top thus first order stochastically dominates that at the bottom. This generates an asymmetry for equilibrium play: Since S_1 is nonpositive in equilibrium, S_2 is always negative and the bottom journal always competes on time ($\delta_2 = \bar{\delta}_2$) in order to attract better submissions. Meanwhile, the top journal is "too attractive for its own good" when the prestige ratio γ_1/γ_2 is large and too many authors flood submit to it first. Delay lowers its attractiveness, and thus raises the quality of the marginal first submission $\hat{x}$.

4 Welfare Considerations

4.1 Red Tape

There are winners and losers to the introduction of red tape, and individual payoffs are not Pareto ranked. Yet, despite what casual reflection may suggest, the top journal is not the only winner here: Her additional delay can actually improve the payoff for *both* journals and also make some intermediate-quality authors better off. Indeed, the surge in $\hat{x}$-quality matches may be attractive to a high-capacity bottom journal with low-quality marginal acceptances. Similarly, the reduction in $\underline{\sigma}_1$ increases acceptance rates at the top, which intermediate quality authors often value over longer delay.[15]

To estimate the social welfare impact of red tape, one therefore needs to resort to an aggregate measure of surplus. As is standard, suppose therefore that a match between author x and journal j creates surplus $\phi(x, \gamma_j) > 0$. I assume that ϕ is increasing in both parameters and supermodular, making assortative matching the socially desirable outcome.[16] A social discount rate $\rho > 0$ accounts for costly match delay, $(\tau_1, \tau_2)(x) = (T_1, T_1 + T_2)$ for manuscripts first submitted to the top ($x > \hat{x}$) and $(\tau_1, \tau_2)(x) = (T_1 + T_2, T_2)$ otherwise. Taken together, match rates μ_1, μ_2 yield social surplus

$$W = \sum_{j \in J} \int_X \phi(x, \gamma_j) e^{-\rho \tau_j(x)} \mu_j(x) f(x) dx. \tag{2}$$

Despite the negative connotation of the term 'red tape',[17] the next result

[15] For a numerical example, consider parameters $(\gamma_1, \bar{\delta}_1, \kappa_1) = (0.694, 0.95, 0.036)$, $(\gamma_2, \bar{\delta}_2, \kappa_2) = (0.446, 0.95, 0.424)$, $\lambda(x) = 1/x$ and distribution $f(x) = 2(1-x)$. With red tape ($\delta_1 = 0.907$), γ_1 and γ_2 correspond to the average match quality at each journal. Without red tape ($\delta_1^{nRT} = 0.95$), the average quality reduces to $\gamma_1^{nRT} = 0.682$ and $\gamma_2^{nRT} = 0.445$. Since journal equilibrium payoffs equal $\kappa_j \gamma_j$ in this notation, the bottom journal is better off in the equilibrium with delay. So are authors with quality $x \in (0.31, 0.56)$.

[16] One candidate surplus function is $\phi(x, \gamma_j) = x \cdot \gamma_j$, in which case authors and the planner care equally about readership. If authors do not internalize the full gains from an extended audience, the match surplus function might be more strongly supermodular.

[17] Seminal papers in development economics identify red tape as a symptom of corruption: In Shleifer and Vishny (1993), a central bureaucrat with monopoly power raises the official costs for services through red tape, and then charges bribes in exchange for lower access

shows that such spurious delay often improves match assortativity, and thus may enhance overall surplus.

Theorem 4. *If κ_1 is small enough relative to total capacity $\kappa_1 + \kappa_2$ and the prestige ratio is large enough, some red tape is welfare enhancing under supermodular match values and a sufficiently low social discount rate $\rho > 0$.*

Proof. See Appendix A. $\square$

The proof is similar in structure to that of Theorem 3, except that it accounts for equilibrium cutoff adjustments by both journals. In essence, I show that delay lowers $\underline{\sigma}_1$ and raises $\underline{\sigma}_2$, thus bringing the cutoffs closer together. This improves the quality distribution of papers at the top journal, and lowers the quality of unpublished papers.[18] Authors ultimately fall into one of three bins of given capacity: 'unmatched', or 'matched with journal j' for $j \in \{2, 1\}$. If the quality of authors in the first bin decreases, and that in the last bin increases, the assignment becomes more assortative. This is socially valuable as long as the brunt of the delay is privately borne by authors and of minimal concern for welfare, i.e. $r \gg \rho$.

From a market design perspective, Theorem 4 identifies red tape as beneficial in situations where the indifferent author $\hat{x}$ is of low quality. This result is relevant to a market designer who controls gatekeeper prestige γ_j. By way of example, committee membership in the senate is an important determinant of a politician's influence (McCubbins et al., 1994, p.18). Ideally, the designer would like to appoint the gatekeeper ('journal') with the highest-quality matches to a key position. However, shifting prestige increases the ratio γ_1/γ_2 and thus negatively affects the equilibrium match distribution – unless red tape negates the relative prestige gain in the eyes of the petitioners ('authors'). Low values of $\hat{x}$ also occur if response times are generally short

hurdles. In Banerjee (1997), a corrupt official is tasked with the allocation of goods to cash-constrained buyers. The official has to allocate goods efficiently for fear of detection by the government. However, rather than identifying buyers' valuation through price discrimination, the official charges maximal fees and implements red tape as a sorting mechanism.

[18] After all, the best way to 'screen out' poor papers with two signal draws is to set $\underline{\sigma}_1 = \underline{\sigma}_2$.

since $z_j(x) \to \gamma_j$ as $\delta_j \to 1$. This implies that even if technological innovations drastically lower *all* minimal response times, it is socially preferable that the top journal *do not* reduce delay as much as technically feasible. By keeping some red tape $\delta_1 < \bar{\delta}_1$, she nudges the lowest quality authors towards the bottom journal first, and improves social surplus.

The role of editorial delays has been studied before in a partial equilibrium framework (Leslie, 2005; Azar, 2007; Heintzelman and Nocetti, 2009; Cotton, 2013), mostly with a focus on contrasting them to monetary submission costs. These papers associate an explicit cost to refereeing and articulate why journals want to raise submission hurdles. However, the arguments presented here do not rely on reductions in refereeing load, and as such are *in addition* to the previously identified channels. Also, none of these previous models incorporates adjustments made by other journals, and thus they do not allow comparable welfare statements.

4.2 Externality

The competition between journals is not a zero-sum game. Indeed, while journals compete for the best manuscripts, they also benefit when a competitor diverts low-quality submissions. A journal can then fill its limited capacity with higher-quality manuscripts and improve her equilibrium payoff. Formally, I say that journal j exerts a positive (negative) externality on journal k if the equilibrium payoff V_k is increasing (decreasing) in κ_j. In a monotone equilibrium, it is the bottom journal that attracts and diverts the most low-quality submissions. In line with this intuition, I find that the bottom journal generally exerts a positive externality on the top journal, while the top journal exerts a negative externality on the bottom journal.

Theorem 5. *Assume κ_1 is small relative to $\kappa_1 + \kappa_2$. The top journal exerts a negative externality on the bottom journal, while the bottom journal exerts a positive externality on the top journal as long as the prestige ratio γ_1/γ_2 is large enough.*

Proof. See Appendix A. □

The formal proof is a pure exercise in algebra. The caveat regarding the prestige ratio ensures that the equilibrium features endogenous delay, and hence the marginal acceptance of the top journal is equal in quality to $\hat{x}$. A marginal increase in κ_2 reduces submissions at and below $\hat{x}$ in favor of additional acceptances with mean quality $\hat{x}$.

As a consequence of this dynamic, the bottom journal tries to avoid head-on competition while the top welcomes it.[19] This may be one factor that has pushed several flagship journals (such as Nature, AER, or PLOS) to create associated, but lower-ranked research outlets. An attractive set of alternatives reduces unwanted submissions at the top, which often improves, rather than lowers, the quality of the remaining publications.

Another way in which these externalities manifest themselves is in regards to specialization. Suppose that each manuscript also has a payoff-irrelevant field attribute θ, chosen with equal odds from $\{A, B\}$. In the initial board meeting, journals have to decide what percentage of their capacity they want to allocate to each of these fields. (Cutoffs $\underline{\sigma}_j^\theta$ are then determined from the capacity constraint.) The top journal benefits from a high-capacity bottom competitor, and thus tends to allocate capacity to mimic the bottom journal. The opposite is true for the bottom journal. In fact, even if the top journal allocates equal capacity across fields ($\kappa_1^A = \kappa_1^B$), the bottom journal may be able to improve its payoff through 'specialization' on either field, e.g. $\kappa_2^A > \kappa_2^B$. Allocating excessive capacity for field-A submissions signals higher acceptance rates to those authors, and thus yields more field-A first submissions above $\hat{x}$ and fewer field-B submissions below $\hat{x}$. Informally, the bottom journal is trying to 'drown' the top journal by condensing her matches among a sub-pool of authors.[20]

The combination of general-interest top journals and specialized second-tier journals is the norm in many disciplines. While there are certainly several competing explanations for this phenomenon (e.g. referee recruitment or en-

[19]Indeed, in some sense red tape 'dampens' the prestige advantage of the top journal and thus increases competition.

[20]For a more formal treatment of this off-equilibrium extension, see the working paper version of this article (Müller-Itten, 2017).

dogenous prestige), the journal's signaling incentive to attract good target submissions may be one factor among many. Similarly, the venture capital literature also discusses that specialization can be enhancing firm performance through improved access to potential deals (Gompers et al., 2009; Norton and Tenenbaum, 1993).

5 Discussion and Conclusion

This paper studies competition for agents engaged in costly directed search under one-sided asymmetric information. A tractable equilibrium analysis offers novel intuition regarding the strategic trade-offs and equilibrium presence of various screening mechanisms; among them monetary fees, excess delay, or specialization. Relevant market design problems are addressed with a welfare analysis that highlights how red tape can improve match assortativity and thus help ensure that the best ideas garner the most influence.

The model's main intuition extends beyond the environment studied here: For instance, I assume that authors have perfect knowledge of the manuscript quality x. Reality may be better captured by a model with two-sided imperfect information or, via the equivalence result of Chade et al. (2014), a situation where signals are positively affiliated. This generates an 'acceptance remorse' among authors matched with the bottom journal, since the revelation of a high signal draw $\sigma \geq \underline{\sigma}_2$ renders authors more optimistic about their chances at the top journal. To avoid this remorse, more authors submit first to the top, lowering the quality of the indifferent author $\hat{x}$ in each subgame. However, since red tape essentially allows journals to select the location of $\hat{x}$, the qualitative features of the equilibrium remain intact and the main results of the paper still hold.

I also assume that journal prestige is exogenously given. Over the relatively short time horizon of a tenure track researcher, this may be an appropriate assumption, as reputations are indeed rather sticky (Card and DellaVigna, 2014). In the long term however, and in so far as publications serve primarily as signals for manuscript quality, a more sensible assumption may be to set journal prestige equal to the expected quality among its matches. The most appropriate extension would consider a setup with multiple generations of authors and long-lived journals, where prestige is determined by past matches only.[21] For myopic journals, this would be nothing more than a sequence of equilibria studied in this paper. Forward looking journals may accept suboptimal payoffs in one generation in favor of a more favorable competitive position in the future. Note however that the top journal's use of red tape already implies that she actively (and successfully) lowers her perceived attractiveness. If she can expect no gain from higher future prestige, and no cost from lowering her future prestige through additional red tape, there is no reason for her to deviate from myopic utility maximization. Things are different for the bottom journal, which benefits from higher prestige: She may now choose to forego available capacity in order to improve mean match quality, and raise future prestige. One-sided capacity manipulation as in Sönmez (1997) is thus a likely feature of such an extension.

A Additional Proofs

Proof of Lemma 1. By the very definition of steepness, $M(\mathcal{M}(0)) > M(\mu_j) > M(\mathcal{M}(1))$. The Intermediate Value Theorem and continuity of $M(\mathcal{M})$ then guarantee the existence of $\hat{x} \in (0,1)$ with $M(\mathcal{M}(\hat{x})) = M(\mu_j)$. Match value increases from μ_j to $\mathcal{M}(\hat{x})$ since

$$V(\mathcal{M}(\hat{x})) - V(\mu_j) - \hat{x}\underbrace{\left(M(\mathcal{M}(\hat{x})) - M(\mu_j)\right)}_{=\,0} = \int_X (x-\hat{x})\left(\mathcal{M}(\hat{x})(x) - \mu_j(x)\right) f(x)dx > 0,$$

and so the conclusion follows. □

Proof of Lemma 2. By contradiction, assume $a_j : \Sigma \to [0,1]$ is not almost everywhere equal to a cutoff function. Let $S : X \to \Sigma$ be defined through the

[21] This is to avoid that journal prestige becomes a self-fulfilling prophecy due to author coordination.

quantile function $S(x_0) = G_{x_0}^{-1}(1 - \alpha_j(x_0))$ to ensure that the acceptance rate of author x_0 remains unchanged.

By monotonicity of λ, I first show that the acceptance rate $1 - G_x(S(x_0))$ is steeper than $\alpha_j(x)$ around anchor x_0. Specifically, for all $x \neq x_0$,

$$\begin{aligned}
\big(1 - G_x(S(x_0)) - \alpha_j(x)\big)(x - x_0) &= \left(e^{-\lambda(x)S(x_0)} - \int_0^\infty a_j(s)\lambda(x)e^{-\lambda(x)s}ds\right)(x - x_0) \\
&= \lambda(x)(x - x_0)\left(\int_0^{S(x_0)} -a_j(s)e^{-\lambda(x)s}ds + \int_{S(x_0)}^\infty (1 - a_j(s))e^{-\lambda(x)s}ds\right) \\
&> \lambda(x)(x - x_0)e^{(\lambda(x_0)-\lambda(x))S(x_0)}\left(\int_0^{S(x_0)} -a_j(s)e^{-\lambda(x_0)s}ds + \int_{S(x_0)}^\infty (1 - a_j(s))e^{-\lambda(x_0)s}ds\right) \\
&= \frac{\lambda(x)}{\lambda(x_0)}(x - x_0)e^{(\lambda(x_0)-\lambda(x))S(x_0)}\underbrace{\big((1 - G_{x_0}(S(x_0)) - \alpha_j(x_0)\big)}_{=0 \text{ by definition of } S(x_0)}(x - x_0) = 0.
\end{aligned}$$

The inequality is obtained by inserting a factor $\Gamma(s) = e^{(\lambda(x_0)-\lambda(x))(S(x_0)-s)}$ into each integral. For $x > x_0$, this factor removes weight ($\Gamma(s) > 1$) where the integrand is negative ($s < S(x_0)$) and adds weight ($\Gamma(s) < 1$) where it is positive ($s > S(x_0)$). The direction is reversed for $x < x_0$, but when multiplied by the negative factor $(x - x_0)$, this operation still reduces the overall value.

To translate to match rates, note that π_j is weakly increasing in α_j, and hence the cutoff rule also steepens match rates $\mu_j(x) = \alpha_j(x)\pi_j(x)$ around the same anchor x_0. Lemma 1 shows that journal j can improve her payoff by unilaterally switching to the cutoff rule, which contradicts the equilibrium assumption. □

Proof of Theorem 1. Existence. By differentiability of z, F and G, each journal's match volume M_j is differentiable in the cutoff vector $\boldsymbol{\sigma} = (\underline{\sigma}_1, \underline{\sigma}_2)$. Moreover, an increase in $\underline{\sigma}_j$ strictly lowers acceptance rates $\alpha_j(x)$ for all authors x. Since $\partial z_j/\partial\alpha_j > 0$, the submission rate $\pi_j(x)$ is pointwise weakly decreasing in the journal's own cutoff $\underline{\sigma}_j$ and weakly increasing in the opponent's cutoff $\underline{\sigma}_{-j}$. As in the proof of Lemma 2, these features translate to the match rate $\mu_j^{\boldsymbol{\sigma}}(x) := \mu_j(x|\underline{\boldsymbol{\sigma}})$, ensuring that $\partial M(\mu_j^{\boldsymbol{\sigma}})/\partial\underline{\sigma}_j < 0$ and

$\partial M(\mu_j^{\boldsymbol{\sigma}})/\partial\underline{\sigma}_{-j} \geq 0$. This monotonicity extends to the journal best response function $\Phi : \Sigma^2 \to \Sigma^2$, given by the capacity-clearing condition $\Phi_j(\underline{\sigma}) = \min\left\{\tilde{\underline{\sigma}}_j \in \Sigma \mid M\left(\mu_j^{(\tilde{\underline{\sigma}}_j,\underline{\sigma}_{-j})}\right) \leq \kappa_j\right\}$ according to Observation 3. The image of Φ is bounded above by the cutoff $\underline{\sigma}_j^0$ that a journal would set under full submissions $\pi(x) = 1$. It is defined implicitly as the solution to the capacity constraint $\int_0^1 \alpha_j(x|\underline{\sigma}_j^0)f(x)dx = \kappa_j$. Restricting thus Φ to the complete lattice $([0, \underline{\sigma}_1^0] \times [0, \underline{\sigma}_2^0], \leq)$, Tarski's fixed point theorem (Tarski, 1955) states that the fixed points of Φ also form a nonempty and complete lattice along with the partial order $\leq$. Any such fixed point represents an equilibrium of the game when accompanied by any submission orders that follow the $z_j(x)$ scores.

Uniqueness.By Observation 2, any cutoff vector $\underline{\boldsymbol{\sigma}}$ yields strictly higher total odds of matching for each author than $\underline{\boldsymbol{\sigma}}' > \underline{\boldsymbol{\sigma}}$. Any two distinct ordered fixed points would therefore differ in total match volume $M(\mu_1) + M(\mu_2)$. Since that sum equals total capacity $\kappa_1 + \kappa_2$ by Observation 3, equilibrium is therefore unique.

Monotonicity. Let $M_j(s)$ denote the match volume $M(\mu_j)$ resulting if both journals employ the signal cutoff s and all authors first submit to the bottom journal, i.e. $\pi_2(x) = 1$ and $\pi_1(x) = 1 - \alpha_2(x)$. By Observation 2, the total match volume $M_{\mathsf{tot}}(s) = M_1(s) + M_2(s) = 1 - \int_X G(s|x)^2 f(x)dx$ is independent of author strategies, strictly decreasing and continuous. Whenever $M_{\mathsf{tot}}(s) = \kappa_1+\kappa_2$, the actual equilibrium cutoffs need to satisfy $\min\{\underline{\sigma}_1, \underline{\sigma}_2\} \leq s \leq \min\{\underline{\sigma}_1, \underline{\sigma}_2\}$ due to the monotonicity of total match volume. By the argument outlined in the first paragraph of the proof, any cutoffs $\underline{\sigma}_1 \leq s \leq \underline{\sigma}_2$ raise the match volume for the top journal, $M(\mu_1) \geq M_1(s)$ (with strict inequality if at least one of the cutoffs changes). In other words, if $k_1(\kappa_1+\kappa_2) := M_1(M_{\mathsf{tot}}^{-1}(\kappa_1 + \kappa_2))$, the only possible equilibrium cutoffs satisfy $\underline{\sigma}_2 \leq s \leq \underline{\sigma}_1$. It remains to show that in response to such monotone cutoffs, only the best authors (if any) first submit to the top journal. Formally, note that $z_1(e^{-\lambda(x)\underline{\sigma}_1}) \leq z_2(e^{-\lambda(x)\underline{\sigma}_2})$ if and only if

$$\Gamma(\lambda) := -\underbrace{\delta_2(1-\delta_1)\gamma_2}_{\Gamma_1>0} e^{\lambda\underline{\sigma}_1} + \underbrace{\delta_1(1-\delta_2)\gamma_1}_{\Gamma_2>0} e^{\lambda\underline{\sigma}_2} + \underbrace{\delta_1\delta_2(\gamma_1-\gamma_2)}_{\Gamma_0>0} \leq 0$$

for $\lambda = \lambda(x)$. Since $\Gamma' = \underline{\sigma}_2\Gamma - (\underline{\sigma}_1 - \underline{\sigma}_2)\Gamma_1 - \underline{\sigma}_2\Gamma_0$, this equation admits a single crossing condition: Whenever $\Gamma(\lambda(x)) \leq 0$, then both the slope $\Gamma'(\lambda(x))$ and hence $\Gamma(\lambda(x'))$ are negative for any $x' \leq x$.

CONTINUITY. The fixed point minimizes the continuous function $\|\boldsymbol{\Phi}(\underline{\boldsymbol{\sigma}}) - \boldsymbol{\sigma}\|^2$. By Berge's Theorem of the maximum, the unique optimum $\underline{\boldsymbol{\sigma}}$ varies continuously in all external parameters. □

Proof of Lemma 3. The change in parameters strictly increases $z_j(\alpha_j(x))$ for all authors x, and thus moves the quality of the indifferent author $\hat{x}$. The additional submissions increase total matches μ_j, and hence the best response function $\Phi_j(\underline{\sigma})$.

Yet, overall matches need to remain equal to total capacity $\kappa_1 + \kappa_2$ at the fixed point $\underline{\boldsymbol{\sigma}}$, so the two cutoffs need to move in opposite directions. Because of the above, the fixed point condition $\underline{\boldsymbol{\sigma}} = \boldsymbol{\Phi}(\underline{\boldsymbol{\sigma}})$ requires $\underline{\sigma}_j$ to go up and $\underline{\sigma}_{-j}$ to go down. □

Lemma 4. *If $\gamma_1 > \gamma_2$, the set of indifferent authors, $\{x | z_1(\alpha_1(x)) = z_2(\alpha_2(x))\}$, has F-measure zero for any journal cutoffs $(\underline{\sigma}_1, \underline{\sigma}_2) \in \Sigma^2$.*

Proof. The indifference condition $z_2(e^{-\lambda(x)\underline{\sigma}_2}) = z_1(e^{-\lambda(x)\underline{\sigma}_1})$ simplifies to

$$-\Gamma_1 e^{\lambda(x)\underline{\sigma}_1} + \Gamma_2 e^{\lambda(x)\underline{\sigma}_2} + \Gamma_0 = 0 \tag{3}$$

for $\Gamma_1 = \delta_2(1-\delta_1)\gamma_2$, $\Gamma_2 = \delta_1(1-\delta_2)\gamma_1$ and $\Gamma_0 = \delta_1\delta_2(\gamma_1 - \gamma_2)$. As a function of $\lambda(x)$, the generalized Dirichlet polynomial (3) is of length three. By Jameson (2006, Corollary 3.2), it has at most two positive zeros. By monotonicity of λ, this therefore also bounds the cardinality of the set of indifferent authors. As any finite set, it thus has F-measure zero by absolute continuity. □

Proof of Theorem 2. To establish this result, I show that submission fees steepen the match rate relative to red tape, and that (in the absence of submission fees) red tape steepens the match rate relative to publication fees. It then follows by Lemma 1 that journals would first maximize any feasible submission fees before employing red tape, and would always opt for submission over publication fees.

In the presence of monetary fees, the z-scores from Weitzman (1979) expands to

$$\tilde{z}_j(\alpha_j) = \frac{\delta_j \alpha_j(\gamma_j - p_j) - c_j}{1 - \delta_j(1 - \alpha_j)}.$$

Again, the denominator captures the expected discounted benefit of an isolated submission to j. The return from a successful submission is reduced by the size of the publication fee p_j, and any submission incurs the submission fee c_j irrespective of its success. If $\tilde{z}_j(\alpha_j) < 0$, the author exits the game voluntarily before submitting to journal j. Since all matches add to journal value, journals limit their fees so that at least the best authors eventually submit, $\tilde{z}(\alpha_j(1)) > 0$. It is also without loss of generality to assume the existence of an indifferent author $\hat{x}$ with $\tilde{z}_j(\alpha_j(\hat{x})) = \tilde{z}_k(\alpha_k(\hat{x}))$ or $\tilde{z}_j(\alpha_j(\hat{x})) = 0$, for otherwise the journal can lower δ_j without affecting its matches.

The first step is to show is that there always exists a compensated marginal change $\delta_j \mapsto \delta_j + d\delta_j$ and $c_j \mapsto c_j + dc_j(x_0)$ with $d\delta_j > 0$, $dc_j(x_0) \geq 0$ that steepens the match rate around any anchor $x_0 \in X$. Indeed, let α_j^0 denote the acceptance rate at x_0 if $\tilde{z}_j(\alpha_j(x_0)) \geq 0$; otherwise set it to the break-even acceptance rate satisfying $\tilde{z}_j(\alpha_j^0) = 0$. Furthermore, set

$$dc_j(x_0) = (\tilde{z}_j(\alpha_j^0) + c_j)\frac{d\delta_j}{\delta_j} \geq 0.$$

Since $z_j(\alpha_j|\delta_j + d\delta_j, p_j, c_j + dc_j(x_0)) = \frac{\partial z_j}{\partial \delta_j} d\delta_j + \frac{\partial z_j}{\partial c_j} dc_j(x_0)$, it is a pure exercise in algebra to verify that the change in journal policy leaves the score $\tilde{z}_j$ unchanged at α_j^0, while strictly increasing (decreasing) $\tilde{z}_j$ at higher (lower) acceptance rates. Consequently, submission rates $\pi_j(x)$ remain at zero whenever $\tilde{z}_j(\alpha_j(x)) < 0$, and weakly increase (decrease) for all other $x \geq x_0$ ($x \leq x_0$). The change is strict over a non-zero set around the indifferent author $\hat{x}$. The same is obviously true for match rates $\mu_j(x) = \pi_j(x)\alpha_j(x)$. By Lemma 1, equilibrium play therefore implies that each journals either charges maximal submission fees or employs no red tape.

The second step is to show that without submission fees ($c_j = 0$), there always exists a compensated marginal change $p_j \mapsto p_j - dp_j(x_0)$ and $\delta_j \mapsto \delta_j - d\delta_j$ with $d\delta_j, dp_j(x_0) > 0$ that steepens the match rate around any anchor $x_0 \in X$. Indeed, set

$$dp_j(x_0) = \frac{\gamma_j - p_j}{\delta_j(1 - \delta_j(1 - \alpha_j(x_0)))} d\delta_j > 0.$$

Since $z_j(\alpha_j | \delta_j - d\delta_j, p_j - dp_j(x_0), 0) = -\frac{\partial z_j}{\partial \delta_j} d\delta_j - \frac{\partial z_j}{\partial p_j} dp_j(x_0)$, it is a pure exercise in algebra to verify that the change in journal policy leaves the score $\tilde{z}_j$ unchanged at quality x_0, while strictly increasing (decreasing) $\tilde{z}_j$ at higher (lower) quality levels. Submission and match rates weakly increase (decrease) for all $x \geq x_0$ ($x \leq x_0$). As before, the increase is strict over a non-zero set around the indifferent author $\hat{x}$. Since red tape can always be increased further (the set of feasible δ_j does not contain the zero lower bound), Lemma 1 implies that no journal would ever levy publication fees only. Instead, publication fees would always be accompanied by submission fees in equilibrium. □

Proof of Theorem 3. Theorem 1 establishes a capacity bound $\kappa_1 \leq \underline{k}_1(\kappa_2)$ that is independent of the discount factors (δ_1, δ_2) and ensures monotone equilibrium behavior, $\underline{\sigma}_1 \geq \underline{\sigma}_2$ and $\pi_1(x) = 1$ if and only if $x \geq \hat{x}$.

The proof proceeds in four steps: First, I show that adding delay is profitable whenever S_j (Equation (1)) is positive. Second, I show that this rules out artificial delay by the bottom journal in equilibrium. Among the remaining strategy profiles, I then show that there exists an unique equilibrium. Finally, I show that red tape is increasing in the prestige ratio.

Delay is payoff-enhancing if $S_j > 0$. Consider a capacity-preserving compensated marginal change $(d\delta_j, d\underline{\sigma}_j) < 0$ with

$$\begin{aligned} 0 &= d\delta_j \frac{\partial M_j}{\partial \delta_j} + d\underline{\sigma}_j \frac{\partial M_j}{\partial \underline{\sigma}_j} \\ &= \left(d\delta_j \left| \frac{\partial \hat{x}}{\partial \delta_j} \right| - d\underline{\sigma}_j \left| \frac{\partial \hat{x}}{\partial \underline{\sigma}_j} \right| \right) \hat{\alpha}_1 \hat{\alpha}_2 f(\hat{x}) - d\underline{\sigma}_j \int_X g(\underline{\sigma}_j | x) \pi_j(x) f(x) dx, \end{aligned} \quad (4)$$

where $\hat{\alpha}_j = \alpha_j(\hat{x})$. The absolute values capture the fact that $\hat{x}$ moves in the opposite direction for $j = 1$ and $j = 2$, but with the same qualitative effect on the deviator's match volume. The effect on j's match value can similarly be written as

$$\begin{aligned}
d\delta_j \frac{\partial V_j}{\partial \delta_j} &+ d\underline{\sigma}_j \frac{\partial V_j}{\partial \underline{\sigma}_j} \\
&= \left(d\delta_j \left| \frac{\partial \hat{x}}{\partial \delta_j} \right| - d\underline{\sigma}_j \left| \frac{\partial \hat{x}}{\partial \underline{\sigma}_j} \right| \right) \hat{x}\hat{\alpha}_1\hat{\alpha}_2 f(\hat{x}) - d\underline{\sigma}_j \int_X x g(\underline{\sigma}_j|x)\pi_j(x)f(x)dx \\
&\overset{(4)}{=} d\underline{\sigma}_j \int_X (\hat{x} - x) g(\underline{\sigma}_j|x)\pi_j(x)f(x)dx. \qquad (5)
\end{aligned}$$

Since $d\underline{\sigma}_j < 0$, this deviation is profitable if and only if Equation (1) is positive.

No red tape at the bottom. In a monotone equilibrium, $\pi_2(x) < \pi_1(x)$ if and only if $x > \hat{x}$. Similarly, $\underline{\sigma}_2 \leq \underline{\sigma}_1$ implies that

$$e^{(\lambda(\hat{x})-\lambda(x))\underline{\sigma}_2} \leq e^{(\lambda(\hat{x})-\lambda(x))\underline{\sigma}_1} \qquad \text{if and only if } x \geq \hat{x}.$$

Together, it follows that

$$\begin{aligned}
S_2 < \int_X (x - \hat{x}) g(\underline{\sigma}_2|x)\pi_1(x)f(x)dx &= e^{-\lambda(\hat{x})\underline{\sigma}_2} \int_X \lambda(x)(x - \hat{x}) e^{(\lambda(\hat{x})-\lambda(x))\underline{\sigma}_2}\pi_1(x)f(x)dx \\
\leq e^{-\lambda(\hat{x})\underline{\sigma}_2} \int_X \lambda(x)(x - \hat{x}) e^{(\lambda(\hat{x})-\lambda(x))\underline{\sigma}_1}\pi_1(x)f(x)dx &= e^{\lambda(\hat{x})(\underline{\sigma}_1 - \underline{\sigma}_2)} S_1. \qquad (6)
\end{aligned}$$

This implies that either the bottom journal seeks to *reduce* delay ($S_2 < 0$) or the top journal has a strict incentive to *increase* delay ($S_1 > 0$). Since delay is unbounded above, the latter cannot occur in equilibrium, and hence $\delta_2 = \bar{\delta}_2$.

A unique equilibrium exists. By the previous part, an equilibrium has to satisfy the conditions of Theorem 1 with $\delta_2 = \bar{\delta}_2$ and either $S_1 = 0$ or *both* $S_1 < 0$ and $\delta_1 = \bar{\delta}_2$. Consider therefore what happens to S_1 in the restricted equilibrium as $\delta_1 \in (0, \bar{\delta}_1]$ and $\delta_2 = \bar{\delta}_2$. The comparative statics in Lemma 3 imply that $\hat{x}$ and $\underline{\sigma}_2$ are continuously decreasing, and $\underline{\sigma}_1$ is continuously increasing in δ_1. In turn, S_1 is decreasing in both $\hat{x}$ and $\underline{\sigma}_2$ and (for the same reason as in Equation (6)) satisfies

$$S_1(\hat{x}, \underline{\sigma}_1, \underline{\sigma}_2) < e^{\lambda(\hat{x})(\tilde{\underline{\sigma}}_1 - \underline{\sigma}_1)} S_1(\hat{x}, \tilde{\underline{\sigma}}_1, \underline{\sigma}_2) \qquad \text{if } \tilde{\underline{\sigma}}_1 \geq \underline{\sigma}_1.$$

Together, the two imply a single crossing property: When $S_1 \geq 0$ for any δ_1, then $S_1 > 0$ for any larger δ_1. Also, note that as $\delta_1 \to 0$, $\hat{x} \to 1$ and hence S_1 starts out negative for small δ_1. Both equilibrium existence and uniqueness follow from this observation, as S_1 equals zero for at most one $\delta_1 \in (0, \bar{\delta}_1]$, and if it remains negative throughout, the unique equilibrium has minimal delay $\delta_1 = \bar{\delta}_1$.

Monotonicity in γ_1. Both prestige γ_1 and discount factor δ_1 affect journal payoffs only through author submissions. The optimal discount rate for $\gamma_1' > \gamma_1$ thus maintains the indifferent author by solving $z_1(\hat{\alpha}_1|\gamma_1) = z_1(\hat{\alpha}_1|\gamma_1')$ for $\hat{\alpha}_1 = \alpha_1(\hat{x})$, implying

$$\tilde{\delta}_1 = \frac{\delta_1 \gamma_1}{\gamma_1' - (1 - \hat{\alpha}_1)\delta_1(\gamma_1' - \gamma_1)}, \tag{7}$$

which is decreasing in γ_1' and $\hat{\alpha}_1$. In other words, the z_1 score for authors with higher (lower) acceptance rate than $\hat{\alpha}_1$ increases (decreases), and so all author maintain their submission order. Since submission and acceptance rates are thus unchanged, the new discount factor $\delta_1' < \delta_1$ describes the unique equilibrium under γ_1'. □

Proof of Theorem 4. According to Lemma 3, a marginal increase in red tape $dT_1 > 0$ changes the submission order of the indifferent author, $d\hat{x} > 0$, and the equilibrium cutoffs, $d\underline{\sigma}_1 < 0$ and $d\underline{\sigma}_2 > 0$. In what follows, I will show that for $\hat{x}$ low enough (caused by a large prestige ratio γ_1/γ_2), this change improves match assortativity. Welfare is affected by the delay, $0 > \frac{\partial W}{\partial T_1} > -\rho W$, and through the changes in match rates,[22] here expressed as changes in

[22]The derivatives are obtained by differentiating $\mu_1(x) = \alpha_1(x)(1 - \alpha_2(x)\mathbf{1}_{x \leq \hat{x}})$, and $\mu_1(x) + \mu_2(x) = 1 - G(x|\underline{\sigma}_1)G(x|\underline{\sigma}_2)$ by Observation 2.

μ_1,

$$d\mu_1(x) = \begin{cases} -g(x|\underline{\sigma}_1)d\underline{\sigma}_1 & \text{if } x > \hat{x}, \\ -\alpha_1(\hat{x})\alpha_2(\hat{x})d\hat{x} & \text{if } x = \hat{x}, \\ g(\underline{\sigma}_2|x)\alpha_1(x)d\underline{\sigma}_2 - G(\underline{\sigma}_2|x)g(\underline{\sigma}_1|x)d\underline{\sigma}_1 & \text{if } x < \hat{x}, \end{cases}$$

and changes in the total match rate $\mu_1 + \mu_2$,

$$\begin{aligned} d(\mu_1(x) + \mu_2(x)) &= -g(\underline{\sigma}_1|x)G(\underline{\sigma}_2|x)d\underline{\sigma}_1 - g(\underline{\sigma}_2|x)G(\underline{\sigma}_1|x)d\underline{\sigma}_2 \\ &= \lambda(x)\left(e^{-\lambda(x)(\underline{\sigma}_1+\underline{\sigma}_2)}(d\underline{\sigma}_1 + d\underline{\sigma}_2) - e^{-\lambda(x)\underline{\sigma}_1}d\underline{\sigma}_1 - e^{-\lambda(x)\underline{\sigma}_2}d\underline{\sigma}_2\right). \end{aligned}$$

I will first study the limit case with $\rho = 0$. The surplus generated by authors of quality x can then be decomposed into

$$\underbrace{\big(\phi(x,\gamma_1) - \phi(x,\gamma_2)\big)}_{\Delta\phi(x)>0}\mu_1(x) + \phi(x,\gamma_2)\big(\mu_1(x) + \mu_2(x)\big).$$

Note that $d\mu_1(x)$ is positive at all $x \neq \hat{x}$. As $\hat{x} \to 0$, red tape maintains capacity κ_1 by replacing matches of vanishing surplus $\Delta\phi(\hat{x}) \to 0$ one-for-one with others of positive surplus $\Delta\phi(x) > 0$. At the same time, red tape steepens the total match rate due to a single crossing property: Indeed, $d(\mu_1(x) + \mu_2(x)) > 0$ is equivalent to $d\underline{\sigma}_1 + d\underline{\sigma}_2 > e^{\lambda(x)\underline{\sigma}_2}d\underline{\sigma}_1 + e^{\lambda(x)\underline{\sigma}_1}d\underline{\sigma}_2 = H(\lambda(x))$. The right side of that equation is decreasing ($H'(\lambda) \leq 0$) exactly when $e^{\lambda(\underline{\sigma}_1 - \underline{\sigma}_2)} \leq -\frac{\underline{\sigma}_2}{\underline{\sigma}_1}\frac{d\underline{\sigma}_1}{d\underline{\sigma}_2}$ which occurs, if ever, for λ small enough. Since $\lim_{\lambda\to 0} H(\lambda) = d\underline{\sigma}_1 + d\underline{\sigma}_2$ and total capacity $\kappa_1 + \kappa_2$ is maintained, red tape increases the total match rate for small λ (high-surplus x) and decreases for large λ (low-surplus x). At $\rho = 0$, the marginal introduction of red-tape is thus strictly welfare enhancing.

By continuity, this remains true for $\rho > 0$ small enough. □

Lemma 5. *Consider any two equilibria with cutoffs $(\underline{\sigma}_1, \underline{\sigma}_2)$ and $(\underline{\sigma}'_1, \underline{\sigma}'_2)$. If $\underline{\sigma}_1 \geq \underline{\sigma}'_1$, $\underline{\sigma}_2 \leq \underline{\sigma}'_2$ and at least one of the inequalities is strict, then $\hat{x} > \hat{x}'$. Consequently, the match volume increases for journal 1 and decreases for journal 2.*

Proof. Equation (1) states that the top journal does not want to add further red tape if and only if

$$S_1(\underline{\sigma}_1, \underline{\sigma}_2, \hat{x}) = \int_X (x - \hat{x})\lambda(x)e^{-\lambda(x)\underline{\sigma}_1}(1 - e^{-\lambda(x)\underline{\sigma}_2}\mathbf{1}_{x\le\hat{x}})f(x)dx \le 0,$$

where I've added the relevant parameters in parentheses for clarity. Since $(\underline{\sigma}_1, \underline{\sigma}_2)$ constitutes an equilibrium with indifferent author $\hat{x}$, $S_1(\underline{\sigma}_1, \underline{\sigma}_2, \hat{x}) \le 0$. Substituting $\underline{\sigma}'_1$ maintains the sign because $e^{(\lambda(\hat{x})-\lambda(x))\underline{\sigma}'_1} \le e^{(\lambda(\hat{x})-\lambda(x))\underline{\sigma}_1}$ if and only if $(x - \hat{x})$ is positive,

$$\begin{aligned}
S_1(\underline{\sigma}'_1, \underline{\sigma}_2, \hat{x}) &= e^{-\lambda(\hat{x})\underline{\sigma}'_1}\int_X (x - \hat{x})\lambda(x)e^{(\lambda(\hat{x})-\lambda(x))\underline{\sigma}'_1}(1 - e^{-\lambda(x)\underline{\sigma}_2}\mathbf{1}_{x\le\hat{x}})f(x)dx \\
&\le e^{-\lambda(\hat{x})\underline{\sigma}'_1}\int_X (x - \hat{x})\lambda(x)e^{(\lambda(\hat{x})-\lambda(x))\underline{\sigma}_1}(1 - e^{-\lambda(x)\underline{\sigma}_2}\mathbf{1}_{x\le\hat{x}})f(x)dx \\
&= e^{\lambda(\hat{x})(\underline{\sigma}_1-\underline{\sigma}'_1)}S_1(\underline{\sigma}_1, \underline{\sigma}_2, \hat{x}) \le 0.
\end{aligned}$$

Substituting $\underline{\sigma}'_2$ further lowers the value of S_1 by adding additional matches below $\hat{x}$, implying $S_1(\underline{\sigma}'_1, \underline{\sigma}'_2, \hat{x}) \le S_1(\underline{\sigma}'_1, \underline{\sigma}_2, \hat{x}) \le 0$. Any nonzero change in a cutoff makes the corresponding inequality strict.

S_1 is also strictly decreasing in $\hat{x}$, so any $\hat{x}' \ge \hat{x}$ would result in $S_1(\underline{\sigma}'_1, \underline{\sigma}'_2, \hat{x}') < 0$, which occurs in equilibrium only if the top journal uses minimal delay $\delta'_1 = \bar{\delta}_1$. That in turn would increase the z_1 score and decrease the z_2 score for *all* authors, since the second equilibrium increases both turnaround speed and acceptance rates at the top journal, while reducing the acceptance rate at the bottom journal. In particular, any author of quality $x \ge \hat{x}$ would now first submit to the top journal, contradicting the existence of an indifferent author $\hat{x}' \ge \hat{x}$.

Finally, given the direction of the changes in cutoffs and indifferent author quality, the match rate μ_1 (μ_2) increases (decreases) pointwise, causing a raise (fall) in the match volume. □

Proof of Theorem 5. By Lemma 5, a marginal increase in either cutoff moves both equilibrium cutoffs down, $d\underline{\sigma}_1, d\underline{\sigma}_2 < 0$.[23]

Match rates for each journal are affected by changes in the cutoff and changes in author strategies (as captured by $\hat{x}$). If journal 2 maintains capacity, simplification yields

$$\begin{aligned} 0 &= d\underline{\sigma}_1 \frac{\partial M_2}{\partial \underline{\sigma}_1} + d\underline{\sigma}_2 \frac{\partial M_2}{\partial \underline{\sigma}_2} + d\hat{x} \frac{\partial M_2}{\partial \hat{x}} \\ &= \int_X \lambda(x) e^{-\lambda(x)\underline{\sigma}_2} \big(- d\underline{\sigma}_2 + (d\underline{\sigma}_1 + d\underline{\sigma}_2) e^{-\lambda(x)\underline{\sigma}_1} \mathbf{1}_{x \geq \hat{x}} \big) f(x) dx + d\hat{x} \alpha_1(\hat{x}) \alpha_2(\hat{x}) f(\hat{x}). \end{aligned} \tag{8}$$

The effect on the bottom journal's equilibrium payoff is negative,

$$\begin{aligned} dV_2 &= d\underline{\sigma}_1 \frac{\partial V_2}{\partial \underline{\sigma}_1} + d\underline{\sigma}_2 \frac{\partial V_2}{\partial \underline{\sigma}_2} + d\hat{x} \frac{\partial V_2}{\partial \hat{x}} \\ &= \int_X x \lambda(x) e^{-\lambda(x)\underline{\sigma}_2} \big(- d\underline{\sigma}_2 + (d\underline{\sigma}_1 + d\underline{\sigma}_2) e^{-\lambda(x)\underline{\sigma}_1} \mathbf{1}_{x \geq \hat{x}} \big) f(x) dx + \hat{x} d\hat{x} \alpha_1(\hat{x}) \alpha_2(\hat{x}) f(\hat{x}) \\ &\overset{(8)}{=} \underbrace{-d\underline{\sigma}_2}_{>0} \underbrace{S_2}_{\leq 0} + \underbrace{d\underline{\sigma}_1}_{<0} \int_{\hat{x}}^1 \underbrace{(x - \hat{x})}_{>0} g(\underline{\sigma}_2 | x) e^{-\lambda(x)\underline{\sigma}_1} f(x) dx < 0, \end{aligned}$$

since $S_2 \leq 0$ in any equilibrium.

The analysis for the top journal is analogous, except that an increase in $\hat{x}$ now *decreases* matches locally,

$$\begin{aligned} 0 &= d\underline{\sigma}_1 \frac{\partial M_1}{\partial \underline{\sigma}_1} + d\underline{\sigma}_2 \frac{\partial M_1}{\partial \underline{\sigma}_2} + d\hat{x} \frac{\partial M_1}{\partial \hat{x}} \\ &= \int_X \lambda(x) e^{-\lambda(x)\underline{\sigma}_1} \big(- d\underline{\sigma}_1 + (d\underline{\sigma}_2 + d\underline{\sigma}_1) e^{-\lambda(x)\underline{\sigma}_2} \mathbf{1}_{x \leq \hat{x}} \big) f(x) dx - d\hat{x} \alpha_1(\hat{x}) \alpha_2(\hat{x}) f(\hat{x}). \end{aligned} \tag{9}$$

[23] The increase in total capacity along with Observation 2 implies that at least one cutoff decreases, and if they move in opposite directions, Lemma 5 implies slack capacity for at least one journal.

The effect on the top journal's equilibrium payoff is therefore positive,

$$
\begin{aligned}
dV_2 &= d\underline{\sigma}_1 \frac{\partial V_1}{\partial \underline{\sigma}_1} + d\underline{\sigma}_2 \frac{\partial V_1}{\partial \underline{\sigma}_2} + d\hat{x} \frac{\partial V_1}{\partial \hat{x}} \\
&= \int_X x\lambda(x) e^{-\lambda(x)\underline{\sigma}_1} \big(- d\underline{\sigma}_1 + (d\underline{\sigma}_2 + d\underline{\sigma}_1) e^{-\lambda(x)\underline{\sigma}_2} \mathbf{1}_{x \leq \hat{x}} \big) f(x) dx - \hat{x} d\hat{x} \alpha_1(\hat{x}) \alpha_2(\hat{x}) f(\hat{x}). \\
&\overset{(9)}{=} \underbrace{-d\underline{\sigma}_1}_{>0} \underbrace{S_1}_{=0} + \underbrace{d\underline{\sigma}_2}_{<0} \int_0^{\hat{x}} \underbrace{(x - \hat{x})}_{<0} g(\underline{\sigma}_1 | x) e^{-\lambda(x)\underline{\sigma}_2} f(x) dx > 0,
\end{aligned}
$$

as long as $S_1 = 0$ in equilibrium. By the proof of Theorem 3, this happens precisely when the prestige ratio is large and the top journal employs red tape.

□

References

Azar, Ofer H., "The Review Process in Economics: Is It Too Fast?," *Southern Economic Journal*, 2005, *72* (2), pp. 482–491.

_ , "The Academic Review Process: How can we make it more efficient?," *The American Economist*, 2006, *50* (1), pp. 37–50.

_ , "The Slowdown in first-response times of economic journals: Can it be beneficial?," *Economic Inquiry*, 2007, *45* (1), 179–187.

Baethge, Christopher, Jeremy Franklin, and Stephan Mertens, "Substantial agreement of referee recommendations at a general medical journal–a peer review evaluation at Deutsches Ärzteblatt International," *PLoS One*, 2013, *8* (5), e61401.

Baghestanian, Sascha and Sergey V Popov, "On Publication, Refereeing, and Working Hard," September 2015. Available at http://dx.doi.org/10.2139/ssrn.2424868.

Banerjee, Abhijit V, "A theory of misgovernance," *The Quarterly Journal of Economics*, 1997, *112* (4), 1289–1332.

Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel, "A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants," *PloS one*, 2010, *5* (12), e14331.

Calcagno, V., E. Demoinet, K. Gollner, L. Guidi, D. Ruths, and C. de Mazancourt, "Flows of Research Manuscripts Among Scientific Journals Reveal Hidden Submission Patterns," *Science*, 2012, *338* (6110), 1065–1069.

Card, David and Stefano DellaVigna, "Nine Facts about Top Journals in Economics," *Journal of Economic Literature*, 2013, *51* (1), 144–61.

_ **and** _ , "Page Limits on Economics Articles: Evidence from Two Journals," *The Journal of Economic Perspectives*, 2014, pp. 149–167.

_ **and** _ , "What do editors maximize? Evidence from four leading economics journals," Technical Report, National Bureau of Economic Research 2017.

Chade, Hector, Gregory Lewis, and Lones Smith, "Student Portfolios and the College Admissions Problem," *The Review of Economic Studies*, 2014, *81* (3), 971–1002.

Cotton, Christopher, "Submission Fees and Response Times in Academic Publishing," *American Economic Review*, 2013, *103* (1), 501–09.

Ellison, Glenn, "The Slowdown of the Economics Publishing Process," *Journal of Political Economy*, 2002, *110* (5), pp. 947–993.

Gittins, John C, "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1979, pp. 148–177.

Gompers, Paul, Anna Kovner, and Josh Lerner, "Specialization and Success: Evidence from Venture Capital," *Journal of Economics & Management Strategy*, 2009, *18* (3), 817–844.

Heintzelman, Martin and Diego Nocetti, "Where should we submit our manuscript? An analysis of journal submission strategies," *The BE Journal of Economic Analysis & Policy*, 2009, *9* (1).

Jameson, Graham J.O, "Counting Zeros of Generalised Polynomials: Descartes' Rule of Signs and Laguerre's Extensions," *The Mathematical Gazette*, 2006, pp. 223–234.

Leslie, Derek, "Are Delays in Academic Publishing Necessary?," *The American Economic Review*, 2005, *95* (1), pp. 407–413.

McCubbins, Mathew D., Roger G. Noll, and Barry R. Weingast, "Legislative Intent: The Use of Positive Political Theory in Statutory Interpretation," *Law and Contemporary Problems*, 1994, *57* (1), pp. 3–37.

Müller-Itten, Michèle, "Gatekeeping Under Asymmetric Information," *Unpublished Working Paper*, June 2017. Available at SSRN: https://ssrn.com/abstract=3186803.

Noorden, Richard Van, "The true cost of science publishing," *Nature*, 2013, *495* (7442), 426–429.

Norton, Edgar and Bernard H. Tenenbaum, "Specialization versus diversification as a venture capital investment strategy," *Journal of Business Venturing*, 1993, *8* (5), 431 – 442.

Oster, Sharon, "The Optimal Order for Submitting Manuscripts," *The American Economic Review*, 1980, *70* (3), pp. 444–448.

Salinas, Santiago and Stephan B. Munch, "Where Should I Send It? Optimizing the Submission Decision Process," *PLoS ONE*, 01 2015, *10* (1), e0115451.

Shleifer, A. and R.W. Vishny, "Corruption," *The Quarterly Journal of Economics*, 1993, *108* (3), 599–617.

Solomon, David J. and Bo-Christer Björk, "A study of open access journals using article processing charges," *Journal of the American Society for Information Science and Technology*, 2012, *63* (8), 1485–1495.

Sönmez, Tayfun, "Manipulation via Capacities in Two-Sided Matching Markets," *Journal of Economic Theory*, 1997, *77* (1), 197 – 204.

Tarski, Alfred, "A lattice-theoretical fixpoint theorem and its applications," *Pacific Journal of Mathematics*, 1955, *5* (2), 285–309.

Ware, Mark, "Submission Fees – A tool in the transition to open access?," Technical Report March 2010. Accessed June 8, 2018. http://www.knowledge-exchange.info/event/submission-fees.

Weitzman, Martin L, "Optimal search for the best alternative," *Econometrica: Journal of the Econometric Society*, 1979, pp. 641–654.

Welch, Ivo, "Referee Recommendations," *Review of Financial Studies*, 2014.

West, Jevin D., Theodore Bergstrom, and Carl T. Bergstrom, "Cost effectiveness of Open Access Publications," *Economic Inquiry*, 2014, *52* (4), 1315–1321.