November 22, 2014

Editorial:
Cosmetic Surgery in the Academic Review Process

David Hirshleifer*

Has the academic review process become excessive? I describe a model in which reviewers who seek reputations with editors for high skill recommend the repair of mere blemishes as well as significant flaws. Reviewer signal-jamming is profitable if editors have trouble distinguishing the two, leading in equilibrium to insistence upon cosmetic surgery. Indeed, if there is a chance that blemishes are unremovable, in equilibrium editor and reviewer demands sometimes cause good papers to remain unpublished. This implies a socially valuable role for active editing. Signal-jamming incentives may especially suppress innovative papers, as well as external verification by means of follow-up papers. This perspective strongly suggests that the increased burden of the review process is undesirable. I offer tentative thoughts about what to do about it.

*Merage School of Business, UC Irvine

I thank Andrew Karolyi both for very helpful comments and for giving me the opportunity to publish the *Review of Financial Studies*' (*RFS*) second editorial. It does not represent the official views of the *RFS*. I also thank Joey Engelberg, Itay Goldstein, Christopher Parsons, Matthew Spiegel, Siew Hong Teoh, Ivo Welch, and seminar participants at NUS Business School, National University of Singapore for helpful comments; and Jie Gao for helpful research assistance.

1 Introduction

There has been a strong trend in economics and other disciplines over a period of decades toward requiring more revisions of articles before publication (Ellison (2002b)). Ellison (2002b) explores a number of possible explanations based upon observable changes in the economics profession. For example, for empirical work, greater data availability and data processing technologies have reduced the costs of performing additional verifications. However, these changes reduce the cost to authors of performing such checks prior to submission, so it is not obvious that overall there should be more changes demanded in the review process, nor more rounds of review.

Another possible explanation is that with the advance of science, standards have risen. For example, there are now higher expectations in economics and finance that causality issues will be addressed with considerable care, and there is greater emphasis in theoretical work on justifying assumptions and suggesting extensions. However, authors are free to perform such work prior to initial submission, so this does not explain why more revisions would be required. In any case, since Ellison's evidence does not support such explanations, he develops a model of the review process in which arbitrary social norms about refereeing and publication standards evolve over time (Ellison (2002a)).

In an editorial for *The Review of Financial Studies*, Spiegel (2012) emphatically argues that revision demands have become excessive, and calls for "Reviewing Less - Progressing More." Spiegel advocates something close to a single round, up-or-down review process, as opposed to the current norm. As Executive Editor, he led the *Review of Financial Studies* in six years of remarkable growth, so his call to arms deserves attention.

On the other hand, there is a temptation to conclude that standards for publication are too high merely because we, as authors, feel frustrated by the high efforts now demanded for validation. The publication process should be designed to advance science, not scientists. Developing new ideas is more fun than verification, but a low verification standard causes potentially-mistaken ideas to be adopted too readily. If too many unsound new ideas are published, valid innovations get lost in the noise. So whether we consider high standards for revision to be *too* high depends on one's theory of how the review process works.

A conclusion that the review process is excessive does not come directly from the model of Ellison (2002a). Ellison is careful to avoid drawing conclusions about whether the slowdown of economic publishing is a good or bad thing.

A possible explanation, which Ellison does not emphasize, for the increasingly onerous burden of the review process derives from conflicting goals of reviewers and editors. Building

on Spiegel's insights, I describe here a model based upon reputation-building by referees which supports the view that, without any desire on the part of editors to extend the review process, a socially excessive amount of revision is required.

The basic idea is that reviewers who seek to demonstrate high skill to editors recommend the repair of mere blemishes as well as significant flaws. I illustrate this reputation-building pressure (but not the equilibrium) with a simple numerical example. Setting aside submissions that deserve flat rejection, suppose that the probability that a paper has a flaw that needs remedy is 0.5. The ex ante probability that the referee is high- rather than low-skilled is 0.5. If a paper has a flaw, a high-skilled referee detects this with probability 2/3, whereas a low-skilled referee detects this with probability 1/3. The referee never receives a signal indicating a flaw unless there actually is a flaw. The referee's only goal is to be viewed as high-skilled; no one, including the referee, knows the referee's type. Every paper has one minor blemish. The referee is free to report this blemish to the editor as if it were a serious flaw; the editor has no ability to distinguish a blemish from a flaw.

Suppose we were to propose an equilibrium in which the referee reported truthfully (reporting a flaw only when a flaw is detected). Then if the referee observes no flaw, the referee would want to defect from the equilibrium to reporting a flaw. To see this, consider the editor's belief in the proposed equilibrium. A report claiming that there is a flaw can arise in two ways:

(1) There is a flaw, the referee is high-skilled, and the referee detects it. This has probability

$$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{2}{3}\right) = \frac{1}{6};$$

(2) There is a flaw, the referee is low-skilled, and the referee detects it. This has probability

$$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) = \frac{1}{12}.$$

So the rationally updated belief of the editor is that there is a flaw for sure, and given this fact and the fact that a flaw was reported, using Bayes' rule the probability that the referee is high-skilled is 2/3. Similarly, if no flaw is reported, a simple calculation shows that the probability that the referee is high-skilled is $4/9 < 2/3$. This breaks the proposed equilibrium, as the reputation of a referee who observes no flaw can be improved by reporting a blemish as a flaw.

In the actual equilibrium of the model I will describe, reviewers do indeed recommend the repair of mere blemishes as well as significant flaws. (This is an example of signal-jamming, not signaling; in the model the reviewer does not know his own skill level, so there

is nothing to signal.) Reviewer signal-jamming is optimal as editors cannot distinguish the two, and therefore rationally give reviewers some credit for detecting flaws or blemishes.

In the model, editors accede to demands for such cosmetic surgery, and indeed even to demands that authors remove unremovable blemishes. So in equilibrium good papers sometimes remain unpublished. This implies a socially valuable role for active editing, wherein an editor acquires a signal that allows the editor to distinguish flaws from blemishes, and instructs the author to ignore referee requests to fix blemishes.

This approach, together with insights of Ellison (2002a), may help explain the rising burden of the review process in economics. The signal-jamming perspective strongly suggests that this onerous burden is undesirable. After discussing the model and its implications, I therefore offer tentative thoughts about what to do about the problem.

Signal jamming is not the only possible source of bias in the review process. In Bayar and Chemmanur (2014), reviewers have inherent biases for or against a submission, and the editor trades off expertise versus objectivity in selecting reviewers. Also, their model concerns an editor's up-or-down publication decision, whereas my focus is on the decision of whether to require a revision.

2 The Model

There are three players: Author, Referee, and Editor. Author's submission is known to have strong promise for publication. In the basic model, what is uncertain is whether costly repair work by Author is needed before the paper will merit publication.

Submitted Papers

A paper can either have a flaw (F) or have no flaw (N), with respective prior probabilities f and $1-f$. The social value of publishing a submitted paper as-is if N is $+1$, and if F is -1. The social value of not publishing is zero. Author can repair an F paper at cost $0 < c < 1$, in which case it becomes N, implying an as-is publication value $+1$.

All papers have a blemish—an issue that looks superficially like a flaw to someone who is not an expert in the field. The blemish can also be removed by cosmetic surgery, also at cost c. Cosmetic surgery is useless, as the social value of publishing the paper is already $+1$. Since removing a blemish is costly, in a social optimum an N paper should be published as-is.

The dichotomy between fixing blemishes versus flaws is related to the dichotomy between improving the basic idea of a paper (q) versus other aspects of the paper (Ellison

(2002a)). However, in Ellison's model the review process does not improve q, whereas here an unremedied flaw is fatal to the basic contribution of the paper. Also, in Ellison's model referees are not subject to incentive problems, whereas here Editor faces the problem that Referee has bad reporting incentives.

The Players

Author

Author is passive, but Author's payoff is part of social welfare. The paper is already under submission when the game starts, and the paper's characteristics are exogenous. Author incurs cost c when required to repair a blemish or flaw before publication. (Author may prefer that any flaw be corrected, if doing so is cheap enough; that possible benefit is reflected in the positive incremental social value to repair of flaws.) Author receives -0.1 when the paper is not published; this is a simple way of allowing for a possible welfare loss when (in an extension) good papers are not published.

Referee

Referee can be either high or low ability (H or L). Referee's type is unknown to all the players, who have prior 0.5 that Referee's ability is H. The probabilities that the H or L types observe the flaw, if it exists, are p_H and p_L respectively, $p_H > p_L$. Referee sees a *particular* reportable flaw, not just the fact that a flaw exists.

Referee observes the blemish with probability q, with observation independent of whether there is a flaw and whether such a flaw is detected. All players rationally understand that a blemish exists; the observation here is of a particular blemish, which Author can be asked to address.

Referee must report any detected flaw to Editor. If Referee observes the blemish and no flaw, Referee can choose whether or not to report untruthfully that this blemish is a flaw. (Referee cannot report the blemish and that it is a blemish; if this were allowed, there would be no reason to do so.) If Referee observes both a flaw and a blemish, by assumption he reports only the flaw.

Referee's objective is to maximize the Editors perception of the Referees ability. So letting P denote probability, Referee solves

$$\max_R P(H|R),$$

where $R = R^F, R^N$ is the report, and where R^F is the report as a flaw of a specific flaw or blemish, and R^N is the report that there is no flaw.

Editor

Editor maximizes social welfare. If Editor publishes the paper as-is, welfare is

$$(+1)I^N + (-1)[1 - I^N],$$

where I^N is an indicator function for no flaw, with argument suppressed to minimize notation. If Editor requires revision, welfare is $1 - c$.[1]

So if Editor knew for sure that there was no flaw, Editor would accept as-is, and if Editor knew for sure about a flaw, Editor would insist that it be fixed before acceptance. I impose parameter constraints, described in the appendix, such that if $R = R^N$, suggesting no flaw, Editor publishes as-is, and such that if Editor observes a report of R^F, suggesting a flaw, Editor insists on repair. If Editor insists on repair, even if the reported flaw is actually a blemish, Author must perform a costly removal of the blemish before publication. Assume for now that Editor cannot distinguish by direct observation a blemish from a flaw.

I use strong assumptions here to focus on a few key effects. Although Editor here is ignorant and benevolent, I do not view actual editors as either fools or saints.

Conjectured equilibrium

Referee always reports R^F when possible (i.e., whenever Referee observes a flaw or blemish). Upon observing R^F, Editor updates favorably about Referee, as H Referee is more likely than L Referee to detect a flaw, if a flaw exists. (Algebraic details are in the appendix.)

If Referee reports R^N, Editor draws a negative inference, as L is more likely than H to miss a flaw when one exists. Editor requires repairs whenever a flaw is reported.

Verifying the equilibrium

Depending on parameter values, if R^F (flaw reported), generically it will either be more profitable for Editor to never require repairs, or to always require them. For realism, we focus on parameter values for which repair is insisted upon. This requires that the expected social value of repair of a flaw be sufficiently high compared to the costs of making repairs and of removing blemishes, taking into account how often, in equilibrium, blemishes are reported as flaws. Since the always-repair case involves blemish removal, resources are wasted.

To determine whether Editor will require repairs when R^F, observe that the social value when repair is insisted upon is $1 - c$, the difference between the benefit of publishing an

[1] I later discuss a more general setting in which a paper may never be published as impossible revisions are demanded. If a paper remains unpublished, there is an additional welfare component of -0.1 (author disutility).

unflawed paper and the cost of repair. Editor compares this to the expected social value of no repair,

$$P(F|R^F)(-1) + [1 - P(F|R^F)](+1) = 1 - 2P(F|R^F).$$

So Editor requires repair if

$$c < 2P(F|R^F), \tag{1}$$

which is implicitly a condition on the exogenous parameters. The explicit condition is provided in the Appendix.

This condition gives a hint about why Referee cares about having a good reputation. A stronger reputation for high ability could cause the editor to have a favorable belief about the probability that Referee writes unflawed papers. So in a dynamic setting, when Referee with a strong reputation submits a paper, (1) may fail, causing Editor to accept as-is regardless of whether a flaw is reported. An alternative reason for reputation-building would be that Referee values the possibility of being appointed in some editorial role.

To sum up, in the relevant parameter range there is a signal-jamming equilibrium; reporting a flaw makes Referee look better, so Referee reports blemishes as flaws. Since a high quality referee is better able to detect flaws, notwithstanding signal jamming it is rational for Editor to think well of a referee who reports a flaw. The report of a flaw is, in equilibrium, an indicator of the existence of a flaw, so Editor requires removal of both flaws and blemishes (which Editor cannot distinguish). So in equilibrium there is dysfunctional cosmetic surgery.

Remarks

What does the model tell us about the increase in revisions required for publication in economics, and in particular financial economics, over time? Compared to a half century ago, when modern finance was pursued by a very small number of scholars, the number of submissions at top journals has vastly increased, as has the range of topics to be evaluated and the specialized skill needed to evaluate each topic. It is now much harder for editors to distinguish a flaw from a blemish, and with more submissions, much less time is available to thoroughly assess each alleged flaw. Furthermore, with more finance scholars, there is much less direct observation by editors of the quality of others; as such, there are stronger incentives for referees to manipulate editor perceptions of quality through their reports.

In this spirit, the signal-jamming approach therefore suggests that newborn specialty journals will have weaker signal-jamming incentives than general interest journals, resulting in less insistence upon cosmetic surgery. Such incentives should be weaker when there are fewer submissions, a narrower area within which editors can distinguish flaws from

blemishes more readily, and referees whose qualities are better known to the editor by direct personal experience. Of course, specialty journals can grow to be quite large (as with finance journals within the economics field), intensifying signal-jamming incentives.

The expanding size of the finance profession seems an appealing explanation for the increased burden of revisions in finance, but expansion not seem to be enough to explain the slowdown in the publication process in economics as a whole. According to Ellison (2002b), the economics profession has not grown substantially over recent decades, and Ellison also argues that in economics generally, it has not become harder to evaluate papers.

This suggests that a better explanation, for economics as a whole, may be Ellison's hypothesis that there are multiple equilibria in which it can be self-confirming for reviewers to have norms for either higher or lower demands for revisions, and in which the norm migrates upward over time. In Ellison (2002a), the upward migration is driven by authors thinking that the quality of their own papers is higher than it really is, causing them to infer from their experience as authors and reviewers that the prevailing standard for the r component of paper quality (the kind that can be improved through the revision process rather than the unimprovable quality of the main idea) is high.

This is plausible, but the approach here suggests a different possible driver of evolving norms. Reviewers are highly demanding because of the signal-jamming incentives for cosmetic surgery described here. Authors observe that editors tend to go along with the recommendations of highly-demanding reviewers. Authors therefore absorb the social norm for high standards of polish, and raise their own subjective standards in their roles as authors and referees. Authors therefore perfect papers more before submission. This reduces the number of severe flaws and minor blemishes for reviewers to catch. But reviewers still have a reputational incentive to identify imperfections, and scour papers looking for blemishes that can be presented as flaws. As referees and editors absorb the rising social norm, ever-smaller imperfections can credibly be represented as significant flaws. So it seems plausible, in analogy with Ellison (2002a), that this self-reinforcing process can result in standards that evolve upward over time.

Extensions

So far there is little that editors, in their invincible ignorance, can do about reviewer signal-jamming. But suppose now that with probability r, Editor observes whether the reported flaw is a blemish. When r is arbitrarily close to zero, the equilibrium is essentially the same as before, but we can now consider histories in which Editor learns that the reported flaw is actually a blemish. In that circumstance, Editor ignores Referee's report

and publishes the paper as-is. This raises the possibility of a socially valuable role for active editing. If Editor receives a noisy signal about whether the reported flaw is a blemish, and if the noise is sufficiently small, Editor will sometimes overrule Referee in favor of as-is publication. This raises welfare by reducing cosmetic surgery.

When Editor catches Referee reporting a blemish as a flaw, the reputational penalty is limited, as in equilibrium both H and L types report blemishes. There is some stigma; H is less likely than L to be caught reporting a blemish as H is more likely than L to catch and report a flaw instead. But this difference in probabilities could be quite small.

So even if the model were varied to allow Editor to punish Referee for misreporting, misreporting is not necessarily deterred. (Bayar and Chemmanur (2014) explicitly model editor sanctions of referee misbehavior.) In the absence of boiling in oil, if the skill difference between referees and/or detection probability r is too small, such punishments will not deter Referee. Also, Editor may not have any *ex post* incentive to impose punishments, in which case commitment is needed.

If the model were varied so that Referee sought a reputation for honesty as well as competence, a new pressure would be introduced for good Referee behavior. If Editor can sometimes observe whether R^F was a blemish or a flaw, Editor would be able to assess veracity, which would therefore improve referee incentives.

It might be argued that signal-jamming incentives could be banished by having editors always passively following referee recommendations. Under editorial passivity, when a referee later submits a paper, reputation with the editor for high ability as a referee and author won't affect the decision. However, in a broader setting in which referees need to exert costly effort to generate signals, completely eliminating reputational incentives could destroy the informativeness of the reviewing process.

The situation described here is far from first-best. There are several reasons to suspect that in practice even this second-best outcome will not be achieved.

- **Editor time constraints.** As the submission load rises, editors are less able to generate precise independent signals about submission quality. The organizational structure of the *Review of Financial Studies* divides the submission load among a (small) set of decision-making editors, which is designed to promote active editorship.

- **Moralistic training.** In academia, error and careless analysis are considered serious evils. This extends moralistic training in favor of carefulness that starts from childhood, as with the saying, "For the want of a nail, the kingdom was lost." Editors and

referees who have absorbed a moralistic anti-error and anti-carelessness ethic may insist on an unreasonable degree of care. The moral pressure for high standards by an editor are especially intense when an imperfection indisputably exists and has been highlighted by a reviewer. Tolerating such an imperfection may feel inappropriate, even if the cost of remedying it exceeds the benefit.

- **Editor reputational incentives.** An editor may fear seeming foolish to a reviewer in allowing a valid but minor criticism to remain unaddressed. Or, an editor may feel that addressing every detected concern protects the editor's reputation with readers.

- **Other agency problems and biases.** Unethical or undedicated referees have other reasons to be too tough: to preserve journal space for their own work, to suppress competing papers, or to economize on their own time. However, these possibilities do not clearly imply too many revisions; they can just as easily result in rejection of a paper that would otherwise be invited for revision. Similarly, the fact that the reviewer-specific component of recommendations is large (Welch (2014)) suggests that it is hard to come to correct decisions, but does not have an obvious implication for whether an excessive number of revisions will be required.

Further extensions: Innovation and empirical testing

Consider now a variation on the model in which, with probability w close to zero, a given blemish or a given flaw cannot be fixed. With w small, the equilibrium is essentially identical to what we have already analyzed. But now, when a blemish is not fixable, if Editor insists upon repair, the paper is never resubmitted and never published. So good papers sometimes die. This shows that signal-jamming incentives can suppress new good ideas.

Going slightly beyond the model, innovative papers are especially likely to have imperfections according to the conventions of existing literature. This gives authors an incentive to work on relatively narrow topics which do not advance thinking much, but which can be executed blemish-free. This suggests that signal-jamming incentives tend to suppress innovative work especially.

As a final model variation, suppose that Referee may have a preference for telling the truth. Specifically, with probability x close to zero, Referee's utility function gives Referee utility $1 - \lambda$ for reporting truthfully (never report a flaw when there is no flaw), where λ is known to all. There is also a utility component with weight λ on Referee's reputation

for skill. With $x \approx 0$, the equilibrium is essentially identical to what we have already analyzed. However, now, in the event that Referee cares about truth, if λ is below some easily-calculated cutoff, Referee reports truthfully, because of high weight on the direct utility component derived from doing so. When λ is above the cutoff, despite caring about truth to some extent, Referee still reports blemishes as flaws.

So as λ increases, the probability that a blemish is reported as a flaw increases. This yields the empirical implication that when Referee cares more about reputation, more cosmetic surgery is recommended and performed. (As described here, since $x \approx 0$ the effect is very small, but a similar point applies with non-negligible x.)

A possible means of testing this is based upon comparison of junior with senior reviewers. For standard reasons, we expect more junior or untenured individuals to be subject to heavier reputational pressures. If so, they will report more flaws to be corrected; empirical proxies for this would be longer referee reports, and (conditioning on a recommendation to revise) more critical reports. We would also expect that for such reviewers, editors more frequently ask authors to ignore portions of reports, even after controlling for reviewer quality proxies (i.e., the professional stature achieved by reviewers at comparable ages and seasoning). To control (partially) for selection effects, it would be interesting to see whether there is a difference in length of report and favorability of report for papers that have one senior and one junior reviewer.

3 Policy and Concluding Remarks

The signal-jamming effects described here impose direct social costs associated with wasteful revisions and the occasional suppression of good papers. The long-run possible effect is that standards evolve to become even more dysfunctional.

A fairly obvious policy implication of the signal-jamming approach is that editors should impose sanctions when a reviewer reports a blemish as a flaw, or at least express disapproval. A character in a novel by Jack Vance described how bitterly society would condemn a marauding claque that clashed cymbals loudly whenever a musical ensemble was scheduled to perform. This is what reviewer signal-jamming does to research performances.

A review process requiring extensive revision affects *who* verifies, not just *whether* someone verifies. Authors are well tooled up for the task, but have an incentive to seek out specifications that support the papers conclusions. Furthermore, the requirement of extensive cosmetic surgery slows down the publication of corrective papers. This results in a self-reinforcing process. The absence of external checks gives reviewers and editors a

further reason for insisting on more author-performed robustness checks.

Some observers are troubled at the lack of external correction in finance and economics as compared, e.g., with physical sciences. For the reasons just described, a general streamlining of the revision process would probably help.

Taking a broader view (and setting aside signal jamming for the moment), in trying to determine the optimum amount of verification before publication, a key consideration is limited reader attention. Authors can always self-publish. This would be just as good as journal publication if readers had time to read everything carefully. Readers do not, so editor curation matters.

At first glance, this suggests that submissions should be held to very high standards, as readers currently do not have time to read even the papers published in top journals (even after accounting for reader specialization and interest in only a subset of what is published).

However, a key benefit of top journal filtering is to supply readers with validated sound bites. If a paper concludes, for example, that the evidence opposes the signaling theory of IPO underpricing, then readers can absorb that conclusion, at least tentatively, and incorporate it into their thinking without studying the paper in detail. This take-on-faith heuristic is dangerous, of course, but inevitable. Readers can acquire a broader view of how the world works by provisionally accepting more ideas and evidence than can be processed through direct personal validation. (Of course, a good scholar avoids overcommitting to the conventional wisdom as encapsulated in sound-bite summary of past research.)

If readers tend to take published conclusions on faith, it becomes all the more important that published papers be accurate. However, the sound-bite validation role of the publication process also implies that there is no need to restrict the number of acceptances to the number of papers that scholars would have time to read in detail.

How much author verification should be required also depends on the relative costs of accepting bad ideas without verification, versus of rejecting good ideas. Suppression of scientific innovation is recurrent, as with the famous example of Galileo. This point is raised emphatically in Matthew Spiegel's editorial: "Presumably, academic journals exist and publish articles to disseminate new ideas. Somehow that simple goal has been lost." I have suggested here that signal-jamming incentives and the current culture in financial economics tend to push authors away from innovative research toward routine extensions that can be executed with a minimum of minor blemishes.

Since it is very hard to realistically quantify the costs and benefits of revision, whether we are on the wrong track right now is a judgment call—ideally guided by stylized facts

and modeling. Spiegel's view that there is vastly too much author verification relies in part on the trend at top journals toward longer papers and greater numbers of revisions. But this is inconclusive, as there may have been too little author verification before, so that now we're doing better. Or maybe the costs of verification have gone down (computing and datasets have improved, for example), so that the optimum has shifted toward greater verification.

The signal-jamming approach gives what I view as more persuasive support for the idea that the review process tends to be tilted too heavily toward over-perfecting papers, and toward suppression of innovation.

I don't know whether, as Spiegel suggests, editors, on receiving a first submission, should simply make an up or down call, plus, perhaps a few editorial suggestions. However, I emphatically endorse one of his final suggestions: "Editors can help by routinely paring down referee demands. Let authors know that they are free to ignore various comments." Active editing of this sort was certainly a key aspiration for the editors during my term at the *RFS*. Given the equilibrium bias toward cosmetic surgery described in the model, there is a clear benefit to editors routinely recognizing and dismissing unreasonable reviewer demands.

How can such changes be implemented? First, consider journal contractual and organizational structures that make use of editor time effectively, to help make active editing feasible. Some possibilities include paying editors to substitute time away from other activities, the founding of specialty journals, the subdivision of submissions among multiple non-anonymous decision-making editors, and the separation of administrative from evaluative editorial roles. Second, and more importantly, we need to change the culture. This means publicly recognizing the problem, and devoting attention to changing our behaviors as referees and editors.

Appendix

Verification that Editor Requires Revision If and Only If Referee Reports a Flaw

To rule out the uninteresting case in which papers with no reported flaws are rejected, we consider parameter values such that in equilibrium $P(N|R^N)(+1) + P(F|R^N)(-1) > 0$. This is an implicit constraint on the exogenous parameters.

We now express condition (1), which implies that Editor requires repair when a flaw is reported, in terms of the exogenous parameters. By Bayes rule,

$$\begin{aligned} P(F|R^F) &= \frac{P(R^F|F)P(F)}{P(R^F|F)P(F) + P(R^F|N)P(N)} \\ &= \frac{fP(R^F|F)}{P(R^F|F)f + P(R^F|N)(1-f)}. \end{aligned}$$

In turn,

$$\begin{aligned} P(R^F|F) &= 0.5p_H + 0.5p_L \\ P(R^F|N) &= q. \end{aligned}$$

So

$$P(F|R^F) = \frac{f(0.5p_H + 0.5p_L)}{f(0.5p_H + 0.5p_L) + q(1-f)}.$$

So Editor requires repair if

$$c < \frac{2f(0.5p_H + 0.5p_L)}{f(0.5p_H + 0.5p_L) + q(1-f)},$$

a further constraint on the exogenous parameters.

Verification that Referee Maximizes Reputation by Reporting Blemishes as Flaws

We calculate the probability on the equilibrium path that the Referee is H conditional on R^F vs. R^N.

$$P(R^F \mid H) = f[p_H + (1-p_H)q] + (1-f)q$$
$$P(R^F \mid L) = f[p_L + (1-p_L)q] + (1-f)q$$

$$\begin{aligned} P(H \mid R^F) &= \frac{\{f[p_H + (1-p_H)q] + (1-f)q\}0.5}{\{f[p_H + (1-p_H)q] + (1-f)q\}0.5 + \{f[p_L + (1-p_L)q] + (1-f)q\}0.5} \\ &= \frac{fp_H + q + f(1-p_H)q}{fp_H + q + f(1-p_H)q + fp_L + q + f(1-p_L)q} \\ &= \frac{p_H f(1-q) + q}{p_H f(1-q) + q + p_L f(1-q) + q} \\ &= \frac{p_H f(1-q) + q}{(p_H + p_L)f(1-q) + 2q} > \frac{1}{2} \qquad \text{since} \quad p_H > p_L. \end{aligned}$$

$P(H \mid R^N):$
$P(R^N \mid H) = 1 - f[p_H + (1-p_H)q] - (1-f)q$
$P(R^N \mid L) = 1 - f[p_L + (1-p_L)q] - (1-f)q$

$$\begin{aligned}
P(H \mid R^N) &= \frac{0.5\{1 - f[p_H + (1-p_H)q] - (1-f)q\}}{0.5\{1 - f[p_H + (1-p_H)q] - (1-f)q\} + 0.5\{1 - f[p_L + (1-p_L)q] - (1-f)q\}} \\
&= \frac{1 - f[p_H + (1-p_H)q] - (1-f)q}{2 - f[(p_H + p_L) + (2 - p_H - p_L)q] - 2(1-f)q} \\
&= \frac{1 - q - fp_H(1-q)}{2 - 2q - f(1-q)(p_H + p_L)} < \frac{1}{2} \quad \text{since} \quad p_H > p_L.
\end{aligned}$$

So $P(H \mid R^N) < P(H \mid R^F)$.

References

Bayar, O. and T. J. Chemmanur, 2014, A model of the editorial process in scientific journals, Working paper, University fo Texas at San Antonio.

Ellison, G., 2002, Evolving standards for academic publishing: A q-r theory, *Journal of Political Economy* **110**, 994–1034.

Ellison, G., 2002, The slowdown of the economics publishing process, *Journal of Political Economy* **110**, 947–993.

Spiegel, M., 2012, Reviewing less - Progressing more, *Review of Financial Studies* **25**, 1331–1338.

Welch, I., 2014, Referee recommendations, *Review of Financial Studies* **27**, 2773–2804.