Version: December 7, 2020

Up or Out: Resetting Norms for Peer Reviewed Publishing in the Social Sciences

Campbell R. Harvey

David Hirshleifer

The peer review process typically has multiple evaluators. Too often editors rely on the *Union Heuristic*, which requires authors to perform all requested tests and extensions suggested by referees and editors, with results confirming the paper's message. The Union Heuristic is easy, but has dysfunctional consequences— wasteful efforts on excessive empirical and theoretical robustness checks and extensions, slow publication times, bloated papers, and heavy use of referee time. We contrast this with the polar opposite (also not our recommendation), the *Intersection Heuristic*—requiring authors to perform only extensions requested by all evaluators. We propose that editors reset norms by making clear that as a matter of procedure, at editor discretion, excellent referee or coeditor suggestions will often *not* be mandated for authors. Instead, there should be a presumption of *Up or Out* (accept or reject typically one round of revisions or less) for submissions, with the extent of suggested changes held firmly in check. This will improve referee and author incentives, and streamline the publication process. It may also encourage authors to take the risks inherent in producing more innovative papers.

Campbell R. Harvey is Professor of Finance at the Fuqua School of Business, Duke University Durham, North Carolina. David Hirshleifer is Distinguished Professor of Finance and Merage Chair in Business Growth, Paul Merage School of Business, University of California-Irvine, Irvine, California. Their email addresses are: cam.harvey@duke.edu, and david.h@uci.edu.

We thank John Cochrane, Alex Edmans, Itay Goldstein, Ron Kaniel, Preston McAfee, Stefan Nagel, Bill Schwert, Rene Stulz, and Ivo Welch for helpful comments.

1. Introduction

Many authors have commented upon the growing unwieldiness of the academic review process in economics and allied fields in the social sciences. As just one illustrative datum, Hadavand et al. (2020) show for a sample of five economics journals that 10% of the papers take more than four years from submission to publication. Over time there has been an increase in numbers of required revisions, and especially in demands for robustness checks and extensions (Ellison 2002b; McAfee 2010; Spiegel 2012; Harvey 2014). By "robustness checks" we mean additional empirical tests or theoretical modelling to validate the claims of the paper. Card and DellaVigna (2013) found that recent published papers were on average three times longer than in the 1970s, a figure that excludes the rise of robustness checks in online appendices. Several authors have argued that much of the effort devoting to revisions is socially wasteful (Spiegel 2012; Hirshleifer 2015; Berk, Harvey, and Hirshleifer 2017).

Ellison (2002a) provides a model of how social norms have evolved toward increased demands for revisions. Referees have an incentive to demonstrate their skills to editors by identifying potential problems in papers—problems that may be mere blemishes rather than major flaws (see the signal-jamming model of Hirshleifer (2015)). Berk, Harvey, and Hirshleifer (2017) provide recommendations for how to improve refereeing to contribute to a fairer and more efficient review process. For example, they emphasize building a norm for a sharp and explicit separation between comments that must be dealt with to make the paper publishable, versus suggestions that are do not affect the publication decision and are therefore optional. They also emphasize the importance of referees living up to the implicit contract in invitations to revise-and-resubmit. Hamermesh (1992, 1994) and Berk, Harvey, and Hirshleifer (2016) offer advice for reviewers, which, in the latter paper, is distilled into a checklist.

In this paper, we focus on the *editorial* side of the review process and the problem of aggregated the reactions of multiple evaluators. In peer review there are typically multiple evaluators. Even when there is one referee, the editor is a second layer of evaluation. Often there are multiple referees, as well as other consulted parties (associate editors or coeditors). For convenience, in this paper we will usually (loosely) refer to the providers of recommendations as "referees" and the final decision maker as "the editor."

In our view, the role of referees is to advise the editor, and the role of the editor is to thoughtfully make use of this input for at least two purposes. First is to make a decision: to reject, invite for revision, or accept the paper. Second, if the editor invites a revision, is to formulate a set of revisions that are required for publication. The editor should not be a mere automaton that follows some general formula in processing referee comments. The editor needs to make judgments on the merits of the referee's comments. What the referee recommends has no importance per se; it is only the recommendations in conjunction with the reasons given for them that are relevant for the editor's decision.

Nevertheless, owing to the number of papers that must be evaluated, and limits to time and expertise, editors resort to cognitive short cuts, just as do all human decision makers. For example, if an editor feels uncertain about an argument for a recommendation, the editor may place greater weight upon the advice if the referee is a distinguished scholar than someone with no track record. The problem arises when the heuristics are applied too mechanically, or are not well-attuned to the decision problem.

2. The Union Heuristic and the Intersection Heuristic

Too often, editors apply something akin to what we call the *Union Heuristic*. The Union Heuristic requires authors to perform all requested tests and extensions suggested by referees and editors. It further requires that for publication all (or almost all) of these extensions confirm the paper's message.

Every editor is different; we certainly do not mean to make sweeping claims about all editors. Most editors of good journals are extremely industrious, few use a pure union heuristic, and many make a point of digesting referee reports and giving authors clear guidance about which comments to focus on and which to ignore. Nevertheless, it is our experience as authors that many editors do have at least some tilt toward the union heuristic.

The Union Heuristic has several attractions for editors. First, it economizes on their time the editor does not need to evaluate the suggestions.

Second, for papers in areas outside the editor's specialty, even an industrious and wellintentioned editor may not be able to distinguish valuable from inconsequential suggested revisions. So, the Union Heuristic economizes on expertise. Furthermore, some editors may fear the embarrassment of publishing a paper which experts in the field will instantly recognize as suffering from a blatant, fatal flaw. This is usually more widely visible to large sets of readers than the opposite mistake of rejecting a good paper. Overruling a referee on a substantive point creates a risk of such embarrassment.

Omission bias may create a further tilt toward rejection. This is the psychological phenomenon that actions that lead to bad outcomes tend to be viewed more negatively than when inaction leads to a bad outcome—even when the odds are worse under inaction (Ritov and Baron 1990). For example, people tend to find more dreadful the idea that they chose to have their child vaccinated, and that this caused their child's death, than that they refrained from vaccination, and that this resulted in their child's death. For an editor, choosing to publish a paper is the active choice; papers are, by default, not published.

Third, the Union Heuristic avoids offending referees and members of the editorial team. Editors sometimes find it uncomfortable ignoring the suggestions of referees or other members of the editorial team. Even if the editor does not specifically notify team members about detailed outcomes, such outcomes may eventually be noticed.

Although attractive to editors, the Union Heuristic generates major social costs, to the detriment of scholarship and science. It contributes to bloat and delay in the editorial process. It also shifts the focus of the review process, for editors, reviewers, and authors, from importance and novelty of a paper to specific flaws, some of which are merely minor. It shifts the incentives of authors from achieving excellence to avoiding any topic or approach that might be perceived to have blemishes.

This is the most damaging feature of the Union Heuristic; it acts as a barrier to the most innovative papers. Papers that are breaking new ground typically lack the smoothness of incremental research. Incremental papers have the benefit that past literature has gradually worked out answers to various objections. So, the authors only need to defend a small further twist. A new general approach typically raises a correspondingly rich set of plausible objections. Indeed, most proposed new approaches do have serious flaws. But if new approaches are strangled in the cradle, innovation stagnates.

Of course, closed-mindedness of reviewers to innovation often takes the form of recommending rejection rather than demanding a long list of feasible robustness checks. The Union Heuristic does not exacerbate this problem. However, referee close-mindedness can also tip the recommendation from demanding just a few (or relatively easy) robustness checks to insisting on many (or very difficult) robustness checks. In such cases, the Union Heuristic is damaging.

Several observers have commented that referees and editors tend to be overly risk averse about new approaches. As Arrow (1995) put it: "I think the publication selection procedure at the major journals has become methodologically more conservative, more given to preferring small wrinkles in existing analysis to genuinely new ideas."

It is at first glance not obvious why this would be the case. Research scholars have chosen their profession because they enjoy ideas, and are curious about how the world really works. Why would they have an undue, knee-jerk preference for the tried and true?

A plausible answer is that on the whole they do not. Individually, scholars find novel ideas fun and exciting. The problem is with process rather than individuals. Use of the Union Heuristic by editors amplifies criticisms of creative new approaches even more than criticisms of small incremental ideas. It induces a conservative bias in the editorial process.

For example, consider an idea which has a 50% chance of ultimately being proven wrong. If the idea is an incremental twist, the paper may not be worth publishing. The potential upside for changing how we understand the world is limited. Why waste reader time absorbing the idea if the validity is unclear? At the opposite extreme, suppose that the idea, if true, will shift the paradigm for understanding the field. Then this is an excellent bet. To move science forward, scholars need to absorb the idea and address it critically to determine, through further research, whether it really works. In other words, high variance ideas have greater option value.

The Union Heuristic biases decisions in exactly the wrong direction. A paper with a 50% chance of being right will probably generate more objections and requested changes if it blazes a new trail. So, the Union Heuristic may limit the dissemination of more innovative ideas.

This is exacerbated by the fact that the Union Heuristic magnifies conflict of interest problems on the part of referees. An opportunistic referee can delay a paper extensively by recommending a revision with a long and demanding list of requirements. This tends to occur when the idea is exciting enough that there are competing teams racing to win the priority tournament.

Of course, robustness checks and extensions often do create crucial social value. Up to a point, they are essential. The problem is when they are allowed to hijack the review process excessively. There will never be enough robustness checks to prove that the authors' interpretation of an empirical result is valid. There are always more and more possibilities to be ruled out. Furthermore, in some ways the authors of a paper are poorly suited to perform robustness checks, as authors have strong incentives to search until they find a specification that

"works." So, there are diminishing social returns to pushing authors to generate more and more extensions.¹

To the extent that referees have undue incentives to insist upon changes, the Union Heuristic exacerbates the problem by accepting such recommendations uncritically and without constraint. Furthermore, there tends to be very high disagreement in the outcome recommendations of different referees (Gilliland and Cortina (1997), Welch (2014)). In our experience as editors, high disagreement is also present in recommendations by referees for additional tests and extensions. Consequently, the Union Heuristic expands the requested workload on authors especially heavily.

Also, although we argued that the Union Heuristic economizes on editor time, an exception may occur if the editor ignores (or misses) the disagreement among referees in the initial found. This could lead to an increase in the editor workload in later rounds as the editor will have to mediate the disagreement.

Before turning to our recommended solution, it is instructive to consider the opposite extreme to the Union Heuristic, which we call the Intersection Heuristic. This is to require only those robustness checks and extensions that all evaluators have suggested. We hasten to add that neither the Union Heuristic nor the Intersection Heuristic is optimal. For one thing, just as the Union Heuristic is too inclusive, the Intersection Heuristic is often too restrictive. If two evaluators see the same problem, that likely indicates an issue to be addressed, even if a third evaluator does not. Even more importantly, both heuristics are too mechanical.

However, if one had to use a mechanical approach, there is no clear superiority of the Union Heuristic over the Intersection Heuristic. So, there is reason to shift from the current informal status quo more in the direction of the Intersection Heuristic.

In particular, when at least two evaluators share a similar concern, that concern tends to be far more likely to be valid than if it is limited to a single evaluator. The Union Heuristic puts heavy weight on evaluator idiosyncrasies. The Intersection Heuristic usefully emphasizes that evaluator commonalities are much more likely to be important. We therefore believe that the Intersection Heuristic is a useful reference point. It is a valuable reminder that the Union Heuristic is a poor makeshift, and that authors should not be required to perform all suggested changes.

We recognize that even shared concerns are not necessarily important ones, or even valid ones. Referees often recommend linkage to familiar and obvious directions. So requests that are common to multiple referees should not be automatically required by the editor. (Sometimes the issues raised by such requests may, however, deserve some succinct verbal mention in the paper.)

A further caveat is that sometimes there are bad ideas that are generally tempting to referees, leading to bad overlapping recommendations. This can arise, for example, when there are fads in the application of a new technique that multiple referees are excited about. As another example, in empirical corporate finance papers, referees will almost always complain about endogeneity (very often with good reason). So it would not be unusual for multiple referees to request robustness checks for endogeneity, even

¹ Of course, the external replication process is an imperfect human endeavor as well. For example, replicators have a "reverse" p-hacking incentive to find a specification that invalidates a paper's conclusion (see Harvey 2020).

in cases where the conjectured endogeneity problem is amorphous or contrived. A related point is illustrated by the running joke that at every finance seminar someone will always ask, "What about tax effects?"

3. So what should editors actually do?

We detail a series of recommendations.

3.1 Resetting norms, and instructing authors which comments to ignore

One commentator on this paper provided a four-word summary: "Editors, do your job." Another commentator said, less succinctly but more colorfully, "Editors (get off your butts and) make the decisions." Specifically, for papers that are invited for revision, an editor should give guidance to the authors, based on careful thought, about what referee suggestions are required for publication, and crucially, which are *not* required. (And in some cases, which suggestions must not be performed at all).

We strongly endorse this call for active editorship. Editors can contribute crucially by engaging intellectually with submissions and reviewer comments. This refines decisions and improves referee incentives.

For example, suppose that the editor receives one report that requests five rather boring extensions, and another report that identifies a potentially fatal flaw that invalidates the paper unless it can be addressed. Then if the editor chooses to invite a revision, clearly the editor should only request that the potentially fatal flaw be addressed. This is neither the union nor the intersection of the referee suggestions.

Comments that are common to multiple referees will often deserve to be addressed. But as we have discussed, active editor discretion is still essential. The editor plays a key role in identifying which of the comments that are idiosyncratic to a single referee are also important enough to be insisted upon.

All this said, the sad reality is that often, for reasons of time and expertise, the editor will not be sure which suggestions should or should not be required. This is where mechanical rules come into play--not as sacred text that editors blindly follow, but as general signposts that tend to point decisions one way or another.

It is in this gray zone where we think that the influence of the Union Heuristic is excessive and pernicious. Recognizing that the optimum for many submissions is closer to Intersection Heuristic than the Union Heuristic can be an efficient guidepost for improving editorial thinking.

To reiterate, mechanical adherence to the Intersection Heuristic would be suboptimal, and that is not our recommendation. To give just one example, suppose that an excellent referee offers several key suggestions, and a careless referee does not provide any useful suggestions. Then it may make perfect sense for the editor to suggest following the suggestions of the excellent referee, not the empty intersection of the two reports. Again, our overall recommendation is to criticize the Union Heuristic as an implicit heuristic signpost, and that a more reasonable heuristic guideline lies somewhere between the Union Heuristic and the Intersection Heuristic. Again, this is a mere signpost, not a rule to be followed in every instance.

3.2 The Up or Out norm and triage

The use of the Union Heuristic in aggregating requests across referees is not the only reason that the review process has become so unwieldy. A reinforcing problem is that until very recently, editors have tended to be too passive in deferring to referee requests for further revisions over multiple rounds. This could be viewed as a different kind of Union Heuristic: taking the Union of referee requests across different rounds.

Many editors try to control this, at least in principle, via the norm that referees must not introduce completely new requests after round one. However, it is very often the case that referees are not satisfied with how the authors have addressed previous-round requests, and recommend giving the authors another chance to revise.

At this point, it is extremely tempting for editors to go along with such recommendations. Rejecting the paper is unattractive when it has already been invited for revision, and has a significant chance of still succeeding. Accepting the paper immediately, even though a referee still views it as unacceptable, is also unattractive.

However, a general practice of going along with such recommendations creates damaging incentives. Knowing that several rounds are common, referees do not feel compelled to think as hard about making the full scope of their concerns crystal clear. Authors feel more free to submit raw work, viewing the review process as one where there will likely be multiple opportunities to refine the research. Overall, there is a too much wasteful iteration. To avoid this, there is a need for editorial commitment to keep the process streamlined.

We therefore suggest that editors reset norms by creating a general presumption of up (clear-cut pathway to acceptance) or *out* (reject) for submissions, with the extent of suggested changes held firmly in check. Specifically, we define the Up or Out norm as one in which accept or reject is typically made with one round of revisions or less, where invited revisions are typically accepted (if the authors execute the requested changes effectively). In other words, many papers are rejected without invitation to revise, and among those that are invited for revision, we suggest that 90% are accepted on the first revision. Only occasionally there will be circumstances in which two or three rounds are needed for resolution.²

The Up or Out norm comes close to ruling out referee suggestions to write what is effectively a different paper by obtaining new data, developing a new model, or starting a major new empirical enterprise. Instead, referees should typically be asked what improvements to the *existing* paper are called for. Discipline is needed on the part of editors as well. It is hard to reject a paper which has real promise, but which requires fundamental reworking. But under the Up or Out norm, such papers are insufficiently developed for publication. There can be cases in which a suggestion is so promising that the editor wishes in effect to invite a new paper. But under the Up or Out norm, such cases are very much the exception.

 $^{^{2}}$ We define Up or Out to be a soft norm. A hard Up or Out norm requires the editor to make an accept/reject decision without requiring revision. Such a policy has advantages, and was instituted at *Economic Inquiry* by Preston McAfee.

For editors, implementing Up or Out requires triage. When a paper is being rejected, there is no need to devote time to signaling meticulousness to authors, e.g., by demonstrating attention even to inessential details. When a paper is publishable, it should be accepted, rather than attempting to make it perfect, which is impossible anyways. The part of the triage where editors need to devote heaviest attention is for papers that need work to become publishable and influential. For these, editors need to devote substantial time to studying the paper, referee reports and author responses.

One basic step that facilitates establishing the Up or Out norm is for editors set a clear and explicit expectation that the Union Heuristic will not be a standard default or benchmark procedure. There are many excellent ideas for extension that do not need to be executed in a given submission. A well-functioning system requires that authors often be directed fairly to perform only a subset of the changes viewed as essential by the reviewers.

Importantly, we are not referring just to suggestions viewed by reviewers as optional. As emphasized by Berk, Harvey and Hirshleifer (2017), referees should be asked to sharply distinguish between optional suggestions and ones that the referee views as required for publication.³ This is helpful, but is not enough. Even the smaller set of required changes tends to become bloated when journals have multiple evaluators each of whom produces more requirements.

One benefit of explicitly resetting norms is that it reduces pressure on editors in individual cases to use the Union Heuristic. Once this expectation is established, a referee or coeditor has no reason to feel disparaged when an editor tells an author to only execute a subset of the required changes made by several parties. Instead, it should be well understood by all that this is the typical situation. Editors can remind referees that even excellent suggestions for extensions and robustness checks are sometimes best addressed by subsequent literature rather than as a condition for publication. (And if an individual is overly sensitive about the use of editor discretion, that person is probably not an appropriate for the referee or coeditor role.)

An instruction to the authors to perform only a subset of the referee-required tests would be mere lip service if authors have reason to think that a negative recommendation from any referee is likely to doom the submission. In that circumstance, the editor would in effect still be using the Union Heuristic in the next round. Authors would still feel compelled to perform all suggested tests (except in extreme cases such as when a test is completely infeasible) in order to appease referees.

So, to make instructions to authors to use discretion and address suggestions selectively, the editor needs to make clear that the editor will not mechanically obey negative referee recommendations in the next round. There must be an explicit policy and practice of overruling negative referees if the author has made a good faith effort and has made reasonable choices as to what robustness checks to actually perform.

This will require that editors make hard choices. On occasion, referees will be dissatisfied, and owing to lack of expertise, the editor cannot be sure whether the author or referees are correct.

³ Editors can reinforce this distinction by emphasizing to authors and referees that the optional suggestions are indeed optional, and perhaps even specifically request that authors not respond to optional suggestions in the "Response to the Referee" document.

There cannot be any blanket rule of what to do in such situations. The editor must balance the evidence as effectively as possible.

What is clear is that when there are multiple referees, a practice of always rejecting whenever at least one referee dislikes the revision creates harmful incentives for authors. That would destroy the ex ante incentive for them to perform robustness checks selectively, and the even earlier ex ante incentive for them to pursue risky research ideas. It would restore an equilibrium in which authors must always execute all referee suggestions. That would restore the dysfunctional outcomes discussed at the start of this paper).

Another approach to avoiding the problems associated with aggregating the suggestions of multiple evaluators is to use only one referee and to not solicit input from other members of the editorial team. With only one producer of request extensions (other than the decision-making editor), the Union Heuristic and the Intersection Heuristic are equivalent. This saves the editor time, but eliminates the editor's flexibility to choose the best among a wider set of referee suggestions. Furthermore, some editors may feel that they have greater discretion as to whether to accept a paper when there are two sets of referee suggestions. Each approach to avoiding the costs associated with bloated requests for revision has its own costs and benefits.

We have focused our discussion on what set of referee proposed changes for editors to insist upon in revisions. A related issue is whether to follow referee advice as to whether to accept, invite a revision, or reject. In the same spirit as our recommendation that editors do not mechanically take the union of all referee suggestions, we do not favor a mechanical system of rejecting whenever one or more referees recommends rejecting. This could be phrased as rejecting whenever the union of all referee recommendations contains a reject. This can be called the Union Heuristic*, where the asterisk indicates that this is for recommendations rather than proposed changes. The Union Heuristic* amounts to veto power on the part of any referee.

There are several obvious reasons for editors to use discretion, and quite likely almost all editors do so. However, the Union Heuristic* is tempting as a guidepost, and may have an undue influence on decisions on the margin. An editor may think that following this approach in most cases is maintaining high standards. The same considerations that make the Union Heuristic tempting—work reduction, deferral to the criteria and expectations of each referee—make the Union Heuristic* tempting as well.

But the Union Heuristic^{*} is also subject to many of the same drawbacks as the Union Heuristic. It can pressure authors to accede to unreasonable demands for robustness checks. It can encourage gaming by referees who seek to delay or block a paper. It tends to block innovative papers especially. Of course few editors would endorse the Union Heuristic^{*} in pure form, as it is so mechanical. But we further suggest that it should not be an implicit signpost or a typical case. Disagreement on the part of the referees is a symptom of a need for an editor to actively engage with the issues without undue regard for reviewers' bottom line recommendations.

In recent years many journals and editors have at least aspirationally adopted something akin to the Up or Out norm. This is a crucial advance for the profession. However, even with the best of intentions, the Union Heuristic makes it very hard to stick to Up or Out. When the attention of authors, referees and editors is spread across numerous requests for changes, the risk compounds that at least one of these issues is viewed as being not addressed acceptably by at least one referee or editor. At the same time, editors and referees typically recognize that submitters have done a great deal of responsive hard work, so there is reluctance to reject the revision based on just one or a few remaining issues. So too often an extra-round with significant effort or risk (rather than relatively routine clean-up) is required.

Establishing a professional norm of Up or Out, and that the Union Heuristic is inappropriate, will also improve referee incentives. Referees will have reduced incentives to engage in signal-jamming by proposing numerous cosmetic changes to prove to the editor how smart they are. Referees will realize that editors are looking for a few really important suggestions, not a laundry list of numerous ones.

An Up or Out norm will provide at least two important social benefits. First, it will reduce wasteful time spent by referees coming up with minor extensions, and will reduce the cost to editors and authors of deciding which proposed extensions should actually be performed. Second, by protecting authors from rejection based on miscellaneous blemishes, it will especially streamline the publication process for more innovative papers, since such papers are almost sure to raise a concern on the part of at least one reviewer. So the Up or Out norm is also likely to encourage authors to devote efforts to producing more innovative papers.

A countervailing concern is that a strict or Up or Out norm might push editors to conservative evaluation of papers, since more innovative papers are often harder to evaluate based upon the first round of reviews. However, we are not endorsing one major round as an inflexible rule. We are instead suggesting Up or Out as sign post and heuristic. Editors should optimize, not blindly obey mechanical heuristics. But in most cases one-round will be optimal in about 90% of cases, so the principle of Up or Out acts as a good reminder to editors to avoid inviting revision to marginal papers that are fairly ordinary and have a chance---but not a great one---of clearing the hurdle in the next round.

3.3 Editor time-management

A benefit to editors of replacing the Union Heuristic with the Up or Out norm is that this entails greater editor engagement with the intellectual substance of submissions. However, an ideal implementation of this is not without cost. It requires editors, in some cases, to spend more time assessing which reviewer comments should be implemented, and deciding when to overrule an insistent reviewer. So implementation of the Up or Out norm would ideally be accompanied by measures to leverage editorial time.

Although not the main focus of this paper, several possibilities for leveraging editors' time come to mind. One is to increase the number of autonomous decision-making editors on editorial teams. (We refer to individual editor decision-making; editorial committee interactions would be time-consuming.) This of course has costs as well as benefits. It may, for example, create incentive problem. There is greater prestige to being the decisive managing editor of a top journal to being one among many coeditors, especially if editors cycle in and out of their positions. Members of a large team of coeditors have less incentive to promote journal success than a managing editor whose reputation rises and falls with the journal.

Performing ruthless triage on papers (as discussed earlier) is a crucial time-saver for editors. Some of the profession has the misconception that it is the editor's job to provide specific commentary on every submission, or to write in a way that demonstrates that the editor has

mastered all nuances of the paper and the referee reports. Instead, it is sufficient for an editor to simply express agreement with the bottom line of a negative referee. Indeed, in some cases a paper is far from an editor's expertise, and even a generally-active editor may primarily be deferring to the referee's judgement rather than adding noise. Editors can communicate the norm that editors do not provide specific commentary on all submissions, and the reasons for it. Editors can also and explain that the role of the review process is mainly to decide what to publish, not to improve papers that are being rejected.

A relative mundane way to open greater editor time is to increase staff support to offload journal administrative issues as much as possible. Another is to increase release from teaching or other types of administrative duties.

4. Wasteful cross-journal allocation of evaluator effort

Our focus has been on strategies of individual journals and editors. However, there are also problems that derive from the process of sequential submission of papers across journals. For example, there is extensive duplication of referee time as rejected papers pass from journal to journal. One problem is that authors may fail to appropriately take into account useful comments, perhaps owing to a perception that editors are unlikely to be aware of the comments of reviewers for other journals (Hirshleifer, Schwert and Singleton 2013). But even if authors address comments industriously, a paper may just not be right for a top journal. In the finance field, as an example, such papers may potentially be evaluated by three top finance journals and five top economics journals. The potential burden on referees and editors from this process is massive.

To address this problem, a commentator on an early draft of this paper offers the bold proposal that referee reports should be uploaded to a publicly available archive. Authors could upload their responses. Conference discussants and others would also be allowed to upload commentary. It would also be important to post the various versions of the paper that referees and discussants are commenting on. (A related idea is to have public replication repositories; see Harvey (2014).)

There are pros and cons to this approach. Some curation is needed to prevent unproductive or abusive posts. In our view, analysis of the pros and cons can only go so far. It is valuable to have *experimentation*. Much will be learned about the effectiveness of this approach when some journal decides to give it a try.

One of the advantages of this approach, at the system level, is that editors of the second or third journal will be able to form an initial assessment of the paper without requiring replication of referee effort. The editor may decide that, despite the posted author response, the editor agrees with the rejection. The editor may, alternatively, decide to accept without further review! The editor would still have the option to seek a new report to resolve remaining uncertainties.

Another advantage of this approach is that it improves referee incentives. It is not very rewarding to identify a key weakness of a paper, only to see that the paper accepted at another comparable journal where another referee was not working as hard. With comments and reports publicly archived, a referee can feel sure that the referee's insights have entered into the evaluation process for other journals (though of course, other editors and referees may have different perspectives).

An objection to this approach is that it is fairer for authors to receive a fresh chance of acceptance at each journal. The implicit concern is presumably that one or a few early negative referee recommendations would be disproportionately influential in inducing reject at many journals. Opposing this concern is the fact that authors can respond, that and there are other natural commentators (such as conference discussants) who might post divergent (positive) views, and that active editors can evaluate referee recommendations critically.

A different concern is that workload might increase instead of decreasing, to the extent that individuals who are writing for public posting (either referees or authors) may feel a greater need to polish their reports or response documents. However, since referees are anonymous, their pressure to polish is reduced. Furthermore, even under the current system authors have a very strong incentive to polish, as authors are seeking to persuade editors and referees to accept the paper.

5. Journal experimentation

We further suggest that journals seriously consider more radical departures from the current system in economics-related fields. A look at some historical innovations suggests why a willingness to experiment is valuable. We have mentioned the strict Up or Out system of *Economic Inquiry*. The system of fast processing of submissions, with high payments to referees, was an innovation of the *Journal of Financial Economics* which, by forcing other journals to compete, reduced turn-around time for the entire finance field. The *Review of Finance* has introduced an expedited option for papers (a higher fee, very fast Up or Out option). The *Quarterly Journal of Economics* has achieved fast average turnaround (by the standards of acquired competitive advantage) by desk rejecting a large proportion of papers, and devoting greater attention to the few papers that are likely to be published. Engineering conference publications use a very different system from economics and finance journals. Experimentation is needed for the system to improve.

Improving norms is largely the job of editors and referees, but authors and readers can help too. Authors can help by submitting their papers to journals that follow the Up or Out system, avoid the Union Heuristic in the use of multiple evaluators, and have policies of active editing. Readers can help by according high prestige to excellent journals, and working to ensure that academic promotion processes reflect the quality of a journal's published papers. Journal reputations are sticky, but for outcomes to improve, it is important for the users of journals to use independent judgement, and to reward excellence.

6. Conclusion

We argue that the current system for peer-reviewed publication is increasingly dysfunctional. Economics papers have tripled in length over the last 50 years as referees pile on endless robustness checks of doubtful utility. It is common for authors to resubmit papers where the response to the referee report had more pages than the original manuscript. Furthermore, with Internet appendices becoming increasingly routine, the full research product (paper plus all appendices) often runs over 100 pages. The social cost is significant. Every day spent on robustness checks is a day subtracted from other innovative research. The costs is not just borne

by authors. Referees waste time writing multipage referee reports that, in many cases, could be a single page. Editors are burdened to sort through recommendations, many of which are inessential.

We offer a perspective about the editorial process and recommend changes. A submission has typically multiple evaluators, and editors often seem to rely on what we call the *Union Heuristic*, which requires authors to perform all requested tests and extensions suggested by referees and editors. We argue that the Union Heuristic is severely dysfunctional. At the other extreme is the *Intersection Heuristic*—requiring authors to perform only extensions requested by all evaluators. The Intersection Heuristic avoids some of the worst problems of the Union Heuristic, but is still clearly suboptimal – especially when referees have diverse perspectives.

Although neither heuristic is optimal, they are useful to think about as signposts—hints about what should or should not typically occur. A useful target for editors lies somewhere between the Union heuristic and the Intersection Heuristic—acceding to all requests made by only a subset of referees should not regularly occur. However, editors should not treat these mechanical heuristics as inflexible rules. Editors must invest the time to actively flag the specific referee comments must be addressed for publication and the ones that can be ignored. This requires a great deal of time from the editor. To free up time, editors need to ruthlessly minimize time devoted to less important activities, such as justifying rejections or giving comments on papers that are clear rejections. The more socially valuable use of editorial time is to shepherd innovative papers through the publication process.

Our other key recommendation is Up or Out. Conditional on a revise and resubmit, we believe that 90% of papers should be accepted on the first revision. There should be a clear roadmap given to the author in the first decision letter. In almost all cases, if the authors have addressed the crucial issues detailed in the roadmap, the paper is accepted.

Another message of this essay is that there is a need for greater experimentation by journals with different review procedures. Complementary with this, there is value to quantitative research on the functioning and performance of the academic review process. For example, how often is it that papers that receive demands for robustness checks actually fail to pass the test? If this is rare, this may indicate that such demands are indeed excessive. With respect to replication of effort of referees across journals, how many reports does a typical paper receive from good journals before ultimate placement?

Our proposals do not address all problems with the academic paper review process. A major source of wastage is replication of referee effort. As a paper bounces from journal to journal, there is often a very large number of referees expending redundant effort. This problem also applies to a paper presented at multiple conferences, wherein discussants engage in redundant evaluation. While we do propose some ideas here, we do not claim to know whether there is a workable system that would address this wastage.

Overall, we call for editors to reset the expectations of authors and referees by making clear that editors will routinely perform actively evaluation of papers that are to be invited for revision, and that streamlining the publication process will be an important consideration. As a consequence, crucially, even excellent referee or coeditor suggestions will often *not* be mandated for authors---the Union Heuristic will not be employed. Instead, editors will balance the costs as well as the benefits of additional prepublication extensions and robustness checks. We further recommend that editors establish a strong presumption of Up or Out. In addition to addressing the bloated publication process, we argue this will improve referee and author incentives. An

important example of this derives from the fact that papers that deviate from conventional approaches tend to stimulate a larger number of demands for robustness checks. So these norm changes promise to encourage authors to undertake risky innovative papers.

References

Arrow, Kenneth, "Forward," in *Rejected: Leading Economists Ponder the Publication Process*, George B. Shepherd, Ed., Thomas Horton and Daughters, Sun Lakes, Arizona, 1995, p. vii.

Berk, Jonathan, **Campbell R. Harvey** and **David A. Hirshleifer**, "A Checklist for Reviewing a Paper," 2016, Available at SSRN: *https://ssrn.com/abstract=2887708*.

Berk, Jonathan and **Campbell R. Harvey** and **David A. Hirshleifer**, "<u>How to Write an Effective</u> <u>Referee Report and Improve the Scientific Review Process</u>," January 2017, *Journal of Economic Perspectives*, 31(1):231-244.

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

Ellison, **Glenn.** "Evolving standards for academic publishing: A *q*-*r* theory", *Journal of Political Economy*, 2002a, 110, 5, 994-1034, October.

Ellison, **Glenn.** "The Slowdown of the Economics Publishing Process", *Journal of Political Economy*, 2002b, 110, 5, 947-993, October.

Gilliland, Stephen W. and **Jose M. Cortina.** "Reviewer and Editor Decision Making in the Journal Review Process". *Personnel Psychology*, 1997, 50:2, 427-452.

Hadavand, Aboozar, Daniel S. Hamermesh and Wesley W. Wilson, "Is Scholarly Refereeing Productive (at the Margin)?" NBER Working Paper 26614 <u>http://www.nber.org/papers/w26614</u>, January 2020.

Hamermesh, Daniel S., "The Young Economist's Guide to Professional Etiquette", *Journal of Economic Perspectives*, 1992, 6:1, Winter, 162-179.

Hamermesh, Daniel S., "Facts and Myths about Refereeing", *Journal of Economic Perspectives*, 1994, 8:1, Winter, 153-163.

Harvey, Campbell R. "Reflections on Editing the Journal of Finance: 2006-2012", in *Secrets of Economics Editors*, Szenbert, Michael and Lall Ramrattan, eds., Cambridge: MIT Press, 2014. 67-81.

Harvey, Campbell R. "Editorial: Replication in Financial Economics", *Critical Finance Review*, 8, No.1-2, 2020, 1-9.

Hirshleifer, David. <u>Cosmetic Surgery in the Academic Review Process</u>, *Review of Financial Studies* (2015) 28(3):637-649.

Hirshleifer, David, G. William Schwert, Kenneth J. Singleton, Joint Editorial, *Review of Financial Studies*, Volume 26, Issue 11, November 2013, Pages 2685–2686

McAfee, R. Preston. Edifying Editing, *The American Economist*, 55:1, Spring 2010, 1-8.

Ritov, Ilana and Baron, Jonathan. Reluctance to vaccinate: Omission bias and ambiguity, *Journal of Behavioral Decision Making*, 1990, 3:4, October/December, 263-277.

Spiegel, Matthew. <u>Reviewing Less—Progressing More</u>, *Review of Financial Studies* (2012) 25(5):1331-1338.

Welch, Ivo. <u>Referee Recommendations</u>, *Review of Financial Studies* (2014) 27(9):2773-2804.

15