Proposal:

A Political Science Peer Review and Publication Consortium

John Gerring
Professor, Department of Government
University of Texas at Austin
(jgerring@austin.utexas.edu)
(corresponding author)

Daniel Pemstein
Associate Professor, Department of Political Science
North Dakota State University

Wordcount: 31,000

Draft: 27 July 2020

Comments welcome!

ACKNOWLEDGMENTS

We are grateful for comments and feedback from Alex Debs, Zach Elkins, Colin Elman, Mike Findley, Macartan Humphreys, Nahomi Ichino, Alan Jacobs, Stephen Jessee, Sebastian Karcher, Carl Henrik Knutsen, Tim Liao, Evan Lieberman, Skip Lupia, Ellen Lust, Raul Madrid, Fred Schaffer, and Devin Stauffer. It should be clear that the views expressed in this proposal are those of the authors, and are not necessarily shared by those who have generously offered their input.

Contents

The traditional process of peer review and publication has come under intense scrutiny in recent years. To ameliorate these problems, scholars in the natural sciences have proposed or implemented a variety of overlapping initiatives. This includes pre-publication archives (e.g., arXiv), mega-journals (e.g., PLOS One, *PeerJ*, and the *Frontiers* series [Spezi et al. 2017]), open-access publication (Nosek et al. 2015**),** on-line publishing (e.g., Berkeley Press [Rodriguez et al. 2006]), a Global Peer Review Platform (Ferreira et al. 2016: 608), centralized review processes (Liao 2020), algorithm-based review scores (Riggs, Wilensky 2001; Varian 1997), portable review systems (e.g., Review Commons [www.reviewcommons.org/]), and post-publication review services (e.g., Peer Review Watch [Nosek, Bar-Anan 2010]).

The time seems ripe for a consideration of alternatives in political science. To that end, we propose a Peer Review and Publication Consortium. The Consortium retains the virtues of the traditional peer review process (governed by academic journals), while mitigating some of its vices.

In the proposed plan, journals are reconstituted as editorial teams within the rubric of the Consortium. The review process is expedited (two rounds) and geared to accept all papers that (in the judgment of editors) offer some contribution to the sum-total of human knowledge. Editorial teams manage these submissions, though they need not publish them under their masthead (some papers would be published without endorsement from an editorial team). With a consolidated system it is possible to eliminate fraudulent editorial teams (who sell space or collude with authors); detect plagiarism (or redundancy) in submissions; identify viable reviewers, exploiting the full universe of political scientists; and motivate reviewers to accept review invitations and improve their performance.

Papers are evaluated through a double- or triple-blind process along multiple dimensions (e.g., internal validity, theoretical contribution, and breadth), each of which is captured in a numeric score (0-10). These scores are adjusted by a measurement model, which reduces or eliminates some common biases, e.g., those associated with the identity of the author, the identity of the reviewer, author-reviewer relationships, insider cartels, and the like. Editorial teams could also offer their own evaluation of a manuscript. Written reviews, along with scores (raw and adjusted) across each dimension of evaluation, and the author's responses, are published on-line (ungated) along with the paper and supplementary materials. There are no page restrictions, aside from those that might be imposed by specific editorial teams. A digital space allows for post-publication commentary (with attribution) on each article.

We anticipate that this system will improve the quality and quantity of work in political science. Specifically, we anticipate that the Consortium will (a) limit the time and frustration imposed by the current review and publication process (for authors, reviewers, and editors); (b) more fully exploit the human capital of the discipline (by integrating authors and reviewers who are off the radar screen of journals); (c) offer more information, and more useful information, to end-users (researchers and policymakers in search of work on a given subject and gatekeepers such as hiring and promotion committees); (d) mitigate common biases of the review process (through measurement model adjustments); (e) ameliorate "fishing" and "file-drawer" problems and encourage replications (because a low threshold is applied to publication and pre-registration and Registered Reports are facilitated); (f) enhance the diversity of published work, including areas currently neglected by journals; (g) save money (especially for institutions that support journals or pay fees for access to journals); and (h) improve the dissemination and impact of political science research.

We begin this proposal by reviewing criticisms of the current system of peer review and publication, after which we turn to the question of the goals an ideal system ought to strive for (§I). In the following sections, we lay out the plan envisioned for the Consortium – its organization (§II); the process of peer review and publication (§III); dimensions of assessment applied to each work under review (§IV); a measurement model, which attempts to adjust raw scores so that they are equivalent and free of bias and a network analysis which identifies how central or peripheral each work

is within the discipline of political science (§V). The next sections discuss the implications of the Consortium – for improving the assessment of submitted work (§VI) and for the discipline at-large (§VII). Finally, we address problems of implementation, i.e., how the Consortium might be established and financed, if scholars think it is a good idea (§VIII).

Readers should be aware that the Consortium is a complex system, with many moving parts. Accordingly, we ask readers to be patient as they read through the following proposal. Questions that arise in the early sections may be addressed later on. (If not, please contact the authors so we can elaborate.)

I. WHERE WE ARE AND WHERE WE WANT TO GO

The traditional system of peer review and publication centered on academic journals has much to recommend it. Arguably, scientific advances realized over the past several centuries would have been impossible – or at least much slower and more halting – without it. However, the system is not without flaws.

To motivate this initiative, we begin with an inventory of shortcomings, drawn from critical work over the past several decades.[1] These shortcomings are laid out in a peremptory fashion; further discussion of each point will appear later. Of course, we cannot pretend that the Consortium will solve all of these problems. But it is plausible to claim that they will be ameliorated.

Shortcomings of the current system

The current process is *wasteful*. Countless hours are spent by authors, editors, and reviewers in the review process, especially when that process involves multiple rounds or multiple journals (as when a paper is resubmitted to a different journal after rejection at another). Binfield (2013) estimates that 15 million hours are wasted every year in redundant rounds of peer review.

Jiang et al. (2019) focus on the challenge of reformatting manuscripts to fit the stylistic requirements of various journals. They report, "There is significant variation in manuscript submission requirements for journals within the same scientific focus and only 4% of journals offered a fully format-free initial submission. Time spent on reformatting delays most publications by at least two weeks and by over three months in about 20% of manuscripts." They estimate that compliance with submission requirements costs over $1.1 billion dollars annually. On the review side, one commonly cited study estimates the cost of unpaid academic reviewing by faculty throughout the world at $3 billion a year (Research Information Network 2008; see also Copiello 2018).

This time and money could probably put to better use. Indeed, one recent study suggests that multiple rounds of review add little to the quality of a manuscript, as judged by impact (Hadavand et al. 2020). Moreover, the burden of reviewing could certainly be more equitably distributed. According to a recent study, political scientists employed in university positions review somewhere between 5.5 (assistant prof) to 8.3 (full prof) papers a year; however, some review as many as 30 papers a year (Djupe 2015).

[1] See Daniel (1993), Ferreira et al. (2016), Godlee, Jefferson (2003), Gould (2012), Jefferson, Alderson, Wager, Davidoff (2002), Lawrence (2003), Liao (2020), Nosek, Bar-Anan (2012), Shatz (2004), Siler, Lee, Bero (2015), Smith (2006), Starbuck (2016), Suls, Martin (2009), Tennant et al. (2017), Walker, Rocha da Silva (2015). One exhaustive study of peer review in biomedical research concludes that "there is little systematic, empirical evidence to support the use of editorial peer review as a mechanism to ensure quality of reports" (Jefferson et al. 2007: 13).

The process is *slow*. For those studies that make it from submission to publication in a journal (not necessarily the journal they initially submitted to) the process is long (Azar 2006; Ellison 2002). Nosek and Bar-Anan (2012) report that a series of articles written by Nosek and various coauthors required two years from initial submission to final publication. We suspect that the analogous figure is much longer for less accomplished authors. So, new knowledge is made available to the public only after a considerable time lag.[2]

The process is *adversarial*. Authors beg to get in, and editors and reviewers fight them off. Frequently, reviews are written in a fashion that highlights shortcomings without identifying ways they might be overcome. Sometimes, reviews are terse, impolite, or even offensive (Akst 2010; Durning 2019). Sometimes, editors work hard to find ways to reject a paper, deviating from normal journal protocol and raising questions about due process (Borer 1997). Even under the best of circumstances, authors experience more rejection than acceptance (over time, the accept/reject ratio is less than one). All of this is discouraging for the author and may have a deleterious effect on her overall productivity and wellbeing (Horn 2016) – and, by extension, on productivity and wellbeing in the discipline at-large.

The process is *uninformative*. Editors are forced to make an up or down decision on every submission, reducing the complexities of research – which are inevitably multidimensional and matters of degree – to a dichotomous outcome. Although the implicit hierarchy of journals provides some information about the quality of a work it is a very noisy signal (Starbuck 2005). Journals lower down in the hierarchy are less well known, so it is difficult to judge the quality of peer review or the standards used by the editors for decisions about submission. Finally, the primary products of peer review – the reviews, and the author's responses – are not usually made public so this information is lost.

Journals are *expensive* (Bergstrom 2001) and access *restricted*. Printing and distribution of hard-copy journals is costly, and publishers take a sizeable cut in profits and administrative overhead. It is perhaps not coincidental that the publishing market has consolidated in recent years, with a handful of countries controlling the lion's share of academic publishing worldwide (Altman, Avery 2015; Gans 2017). Rich libraries must ante up enormous sums of money to purchase access to journals. Other libraries reduce their costs by reducing their subscriptions, meaning that some journals are unavailable. Few scholars can afford to purchase journals (or articles) privately. Accordingly, those without a university affiliation are generally unable to access journals, which lie behind a paywall. Pre-publication versions of articles are sometimes available but are not regarded as authoritative; they also introduce confusion since pagination is different across the two versions and content may differ. To combat this problem, some disciplines have moved from traditional journals to open access journals and some governments seem to be moving towards demanding that government-funded research be published in open-access journals.[3]

The process is *exclusionary*. Some unknown quantity of work that might make a contribution to knowledge never makes it into print, either because journals won't publish it or authors don't submit it for publication. This is often the case when results confirm accepted wisdom, e.g., the null hypothesis cannot be disproven (Franco, Malhotra, Simonovits 2014; Nosek, Spies, Motyl 2012). It may also be the case for studies that are especially innovative insofar as the review process contains a bias in favor of the status quo (Armstrong, Hubbard 1991; Grayson 2002; Horrobin 1990; Mahoney

[2] Preliminary working papers may be posted immediately. However, they are not regarded as authoritative, and often undergo changes. This generates a confusing situation where several non-identical version of a paper circulate at the same time.

[3] In Europe, the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" and "plan S" seem already in motion (www.coalition-s.org/). Meanwhile, there are rumours of a similar move in the United States (https://physicstoday.scitation.org/do/10.1063/PT.6.2.20191219a/full/#.XgMCQH0IM2c.twitter).

1977; Rennie 2003).[4] Siler and Strang (2017) suggest a third pattern: novelty is embraced by reviewers but only if it does not threaten existing theoretical frameworks, leading to a superfluity of theories. In any case, it seems clear that a great deal of work is left unpublished, and this quantity may be much greater (as a share of total work) in the social sciences than in the natural sciences. Nosek and Bar-Anan (2012: 218) note that journal "rejection rates of 70 to 90% are typical in the social sciences, whereas rejection rates of 20 to 40% are typical in the physical sciences."

The process is *idiosyncratic*. The journal review process is subject to the editorial team, or the specific editor who is in charge of a manuscript, and the judgments of 2-3 chosen reviewers. Studies have shown (and readers' experiences can probably confirm) that reviews, and review outcomes, are highly stochastic (Lindsey 1988; Peters, Ceci 1982; Pressman 2014). Papers with absurd claims are sometimes accepted for publication (Bohannan 2013). Many highly influential papers were initially rejected (Gans, Shepherd 1994; Shepherd 1995). One recent study concludes, "Peer review is a diverse method of quality control, and applied inconsistently both in theory and practice, and generally lacks any form of transparency or formal standardization. As such, it remains difficult to know precisely what a 'peer reviewed publication' means" (Tennant et al. 2017: 4; see also Mackie 1998).

The process is *biased*. Reviewers often know the identity of the author whose paper they are reviewing – even if the review process is blinded – and this may affect their judgment. In particular, studies suggest that authors with close personal or professional ties to the reviewer (Jang, Doh, Kang, Han 2016; Sandstrom, Hallsten 2008; Teplitskiy et al. 2018; Wenneras, Wold 1997), male authors (Bagues, Sylos-Labini, Zinovyeva 2017; Hengel 2017; Wenneras, Wold 1997), and authors located at top universities (Tomkins, Zhang, Heavlin 2017; Zuckerman, Merton 1971a, 1971b), and authors located at American universities (Link 1998) are granted favorable treatment.

The process is susceptible to *fraud*. Plagiarism is more common than one might think (Stitzel, Hoover, Clark 2018). Some journals are completely fraudulent, or at least highly susceptible to fraud. In other cases, fraud is episodic, but nonetheless worrisome (Butler 2013; Ferguson, Marcus, Oransky 2014). Although more common in the natural sciences (Beall 2018), "predatory" publishing is on the rise in the social sciences as well (Hasan 2018).[5]

Do we need pre-publication review?

In light of these critiques one might wonder whether pre-publication peer review should be scrapped entirely. An alternative approach is for authors to publish their work in an on-line database (e.g., SSRN in the social sciences or arXiv in the sciences), after which it might be evaluated through an automated – but possibly anonymized – system of post-publication review (Florian 2012; Nosek, Bar-Anan 2010).

We are strongly in favor of developing more sophisticated systems of post-publication assessment (§VI). However, we don't think that such systems can entirely replace the functions now performed by pre-publication review.

First, and most important, any system of post-publication review encounters the same sort of problems that a pre-publication system of review encounters, as noted above. Post-publication review is no panacea.

The only problem that post-publication review entirely solves is *speed of publication*, as a paper is published as soon as the author decides to place it on the public domain. This may be a decisive

[4] John Pencavel (2014: 85), long-time editor of the *Journal of Economic Literature*, writes: "Because the error of publishing an inadequate or poor article is much more visible and conspicuous than the error of failing to publish an excellent article, editors (and even referees) tend to play it safe. They are inclined to turn away potentially controversial papers, those that receive mixed reviews, and those that are out of the mainstream."

[5] This problem, unlike the others listed above, has become worse in the new world of online open-access publishing than it was under the traditional hard-copy journal-based system of publication.

advantage in fast-moving areas of science, where discoveries occur on a daily basis. But it is less of an advantage in slower-moving areas such as political science, which is not marked by decisive discoveries and where research is not as straightforwardly cumulative (Elman et al. 2020).

Second, some problems that are manageable in pre-publication review are impossible to manage in post-publication review. In particular, reviewers of an article post-publication know the identity of the author, and this is bound to introduce all sorts of biases, as noted. By contrast, a system of pre-publication review allows for the blinding of authors, and perhaps also of editors. Although imperfect (because sometimes knowledge of the author is impossible to disguise), the blinding process mitigates these biases.

Third, the peer review process often improves the quality of work, prior to publication. It may also improve the quality of research designs if these are reviewed prior to implementation (as part of a Registered Report). This would not be possible if peer review occurs after publication.

Fourth, if anyone can publish anything on an open platform such as SSRN a lot of drivel will be published, making it difficult—or at least time consuming—for research consumers to separate signal form noise. Social science offers a much lower barrier to participation than physics, so the experience of on-line science databases like *arXiv* is not necessarily a good indication of what we might expect in political science. Readers and reviewers in a post-publication system of review would have no way of sorting out the wheat from the chaff.

For all these reasons, we do not think that post-publication peer review can, or should, entirely replace pre-publication review.

Goals

Leaving aside the pre/post-publication debate, let us think broadly about the problem of peer review and publication. What would an ideal system look like? What objectives should it strive to achieve? Nine core objectives seem fundamental.

First, such a system should be meritocratic and inclusive, treating all manuscripts and all authors equally. It is especially important to expand the scope of the discipline so that those on the periphery are included. This is not only an issue of fairness, but also of expanding our collective human capital.

Second, such a system should seek to publish all work that adds something to the sum-total of political science knowledge. Only work that adds nothing should be discarded. This addresses the so-called "file-drawer" problem and also assures that literature reviews and meta-analyses can draw on a field of studies that is as comprehensive as possible.

Third, such a system should help to enhance the quality of published work by assisting authors to craft better research designs (if papers are submitted as Registered Reports) and – once the data collection is complete – offering suggestions for revision. This requires a collaborative, nurturing relationship between editors, reviewers, and authors.

Fourth, such a system should offer reliable and unbiased signals to authors and end-users (readers) about the contribution(s) of published work. Since contributions may be of many sorts, this requires a multi-dimensional schema of assessment.

Fifth, such a system should be resistant to fraud.

Sixth, such a system should serve as a venue for discussion and debate, beginning with the review process and continuing thereafter in post-publication formats.

Seventh, the products of such a system should be freely accessible to all. These products include not only the published paper but also appendices, data, and other explanatory material that would assist those trying to understand, replicate, or build upon the published work.

Eighth, such a system should achieve these objectives in an efficient fashion, minimizing time and expense for all involved – editors, reviewers, authors, and their employers or funders.

Finally, such a system should not sacrifice ethical considerations such as the fair treatment of human subjects and the maintenance of confidentiality agreements.

With these aims in mind, we turn to the specifics of our proposal.

II. ORGANIZATION

In this section, we lay out the envisioned organization of the Consortium. *Members* of this body include all those who submit papers (authors) or review papers (reviewers). *Editorial teams* manage the review of manuscripts. An *Oversight committee* oversees the process of peer review and publication. An *Ethics committee* sets policies concerning plagiarism, the protection of human subjects, the confidentiality of authors and reviewers, and other matters. An *Executive Council* governs the Consortium. (Depending upon workload, the Executive Council could be merged with the Oversight committee.)

We now turn to a more detailed look at each of these bodies.

Members

Anyone who submits or reviews for the Consortium becomes a member (at no charge). The following information is requested…

1. *Name*
2. *Institution*
3. *Rank (if at a university):* non-tenure track, assistant prof, associate prof, prof
4. *Year of PhD*
5. *PhD institution*
6. *PhD advisors*
7. *PhD students*
8. *Subfield*
9. *Methodology(ies):* a list of methodological specializations that the member is conversant in.
10. *Gender*
11. *Publications*
12. *Availability for review:* Yes[default] No[elective]

Biographical information might be imported from Orcid or Google Scholar. (Orcid membership could be made a condition of participation in the Consortium.) Updates would be needed only when information changes. If the person is linked to an updated Orcid account and a curated GS profile, this too could be automated.

Membership might lapse if the individual is inactive for a period of time. Individuals may also be removed for malfeasance.

Editorial teams

Editorial teams are formed in an open market with minimal barriers to entry. Any group of political scientists could form a new editorial team at any time, subject to vetting by the Consortium, who would implement background checks on each proposal so as to avoid editorial teams that might be fraudulent.

If one is worried about excessive fragmentation one might impose some sort of requirement, e.g., (a) the number of people who submit to, or review for, the editorial team, (b) the number of published articles processed by the editorial team.

Even with this sort of requirement, barriers to entry would be much lower under the Consortium than under the traditional system of journal publication. Starting an editorial team requires only a set of people willing to take on the tasks of editing and another set of people (authors) willing to submit their work. The infrastructure is already in place. This, in turn, should institute greater competition among teams.

Editorial teams could set out distinctive missions, specifying the sort of work they are interested in publishing. Authors who wish to publish with that editorial team would work to achieve those objectives, or find another editorial team whose desiderata are more closely aligned with their own. Some teams might specialize in particular methods or epistemologies; others might specialize in particular substantive topics or areas of the world. Some teams might offer a venue for exploratory work, or for work that lays out new theoretical frameworks. Others might focus on work with strong internal validity, perhaps with the requirement that studies be pre-registered and reviewed prior to data collection (as Registered Reports). Some might specialize in disciplinary issues, book reviews, reviews of the literature, topics of public interest, or public policy.

There is no limit, in principle, to the number or type of editorial teams that might develop, and the size and composition of each editorial team would presumably vary with the volume of submissions and the number of discrete areas of expertise those submissions require. Likewise, we do not feel the need to define tight boundaries around what "political science" is, or should be. Indeed, the Consortium could easily be extended to include adjacent fields in the social sciences.

Teams would institute their own rules for self-selection. Some teams might by chosen by, and responsible to, organized sections of APSA (Comparative Politics, Politics and History,…) or associations (Western, Southern,…). The only caveat is that rules for selection be publicly stated and adhered to. An editorial team that does not follow its own rules is subject to suspension.

Each editorial team would select a title, just as journals do. The title would be noted prominently on the front page of every article published under their auspices. We suspect this will provide an identity for the team, even though – in contrast to a journal – teams would not have distinctive fonts, formatting, or hardbound copies to distribute.

The number of teams in operation at any given point in time would be subject to demand. If authors do not like the options available they are free to start a new team. If an editorial team loses traction (i.e., submissions) it may fold or merge with another team.

Presumably, competition among editorial teams will have positive effects on the peer review process. If an editorial team is slow to review papers, chooses idiosyncratic or uninformed reviewers, or is negligent in some other respect, authors can defect (submitting future papers to a different team) or start their own team. At the same time, teams can build brands that add structure, and information, to the system.

Editorial teams would choose how narrowly or broadly to craft their brand. This includes methodological as well as substantive matters. A narrow brand presumably encompasses a smaller set of potential authors. However, a large brand may be so diffuse that it is perceived as meaningless; it doesn't stand for anything and hence is unattractive to authors.

It is difficult to predict how the competition among editorial teams might sort out. Conceivably, the array of teams in the Consortium might look very much like the current system of journals, though we suspect it will be a somewhat smaller field.

Ethics

Any venture of this magnitude has myriad ethical ramifications. To grapple with these issues in a coherent and comprehensive fashion we envision an Ethics committee. Though we cannot predict what might result from that committee's deliberations, we want to highlight several issues.

Ethical issues associated with peer review are not very different under the Consortium than they are under the current journal-centered regime. Authors retain control of what they make public in their article and background materials, so it is their judgments about issues of transparency and replication that count. (It is not clear that this is the case for all journals.)

One issue is a little different. If the Consortium becomes the primary system of publication in political science it will contain an enormous database including all reviews, initial drafts, communications between author and editor and between author and reviewers. It seems inappropriate to make this information available to the editors of every editorial team. However, it is important to make the previous history of an editorial team available to *that* editorial team. This follows current journal practice. (Every new editor of the APSR is privy to the APSR archives.)

Another issue concerns the measurement model (§V). Since this draws information from the biographies of reviewers, publication of the entire database would make it possible to identify who reviewed a given paper. However, it is vital that the measurement be open to public scrutiny. To achieve this without compromising reviewer identity we would need to carefully vet, and monitor, anyone who seeks to work with the entire database.

Governance

The Consortium is a system of multi-level governance in which each unit enjoys some degree of autonomy but must also coordinate with the others. As with any system of multilevel governance there are questions about what degree of centralization is optimal. Our intention is to offer editorial teams as much freedom as possible to run their ship as they see fit, while maintaining the economies of scale offered by the Consortium and preserving a set of core features that all teams follow (and can presumably agree upon).

Even so, there will be plenty of issues to resolve in this complex and evolving system. A proper governance structure must be in place to deal with disagreements in the present and to meet new challenges in the future.

In our view, the Consortium should be primarily accountable to its members – in particular, the editorial teams who perform most of the work of the Consortium and, presumably, represent the various intellectual currents of the discipline. Accordingly, each editorial team might select a member to join the Executive Council, with votes distributed according to the number of papers reviewed by each team.

Special committees such as the Oversight committee and Ethics committee could also be selected by the editorial teams. If there are a small number of teams, committees might be constituted by allowing each team to select one member. If there are a large number, a system of elective representation will need to be instituted in which teams nominate and vote for members, with votes weighted by the size (number of papers reviewed) of each team.

Committees would presumably conduct most of their work through online interfaces. Face-to-face meetings could be planned once a year at APSA, and virtual meetings (via Skype) could occur as needed.

Above the Consortium, we can envision a relationship with the American Political Science Association (APSA). In this capacity, the APSA Council might serve a general oversight function, while decisions about internal governance would be left to the Consortium. Another option would be

to associate with an existing body (e.g., CQRM) or to develop a stand-alone body with its own tax-exempt ID, so as not to be beholden to any organization.

III. PROCESS

Under the Consortium, the peer review and publication process would work something like this…

1. An author submits a blinded version of her article to the Oversight committee. She signs a consent form to publish the paper if it is not desk-rejected. She is asked to specify if she thinks any criteria of assessment (see §IV) do *not* apply to her paper. She is also asked to submit any background materials (appendices, data, code, et al.) that will appear along with the article and a data storage and transparency plan that explains her choices about what to make public and what to keep private – including a discussion of ethical issues and IRB agreements. If the author intends some of these background materials to be available only to reviewers, this must be clarified. Finally, the author identifies five editorial teams, in order of preference, as possible destinations for her article.
2. The Oversight committee checks to make sure the materials are anonymized. To check for redundancy, the manuscript is compared to all published papers, books, and papers previously submitted to the Consortium using plagiarism detection software (e.g., CrossCheck). Any potential problems that arise are dealt with by the Oversight committee.
3. The article is sent to the five editorial teams identified by the author, in order of preference. Each team may decide to send the paper out for review or decline. If the latter, they indicate whether they believe the paper should be published at all. If all five teams refuse to review, the paper is sent back to the Consortium, which makes a final decision on whether or not to desk-reject the paper, taking into account feedback from the five editorial teams. If the decision is to desk-reject the author may be given an opportunity to resubmit, and some guidelines on what would be expected in that resubmission. If the decision is to publish, the Oversight committee designates an editorial team to supervise the review process. That team must accept the responsibility of overseeing the review process, though it need not publish the manuscript under its masthead.
4. The chosen editorial team identifies six reviewers (and back-ups if any of these decline), optionally drawing on recommendations provided by the Consortium database and measurement model. Reviewers fill out a survey, assigning scores to the manuscript across different dimensions (§IV). For each question, they specify their level of confidence. They also offer written comments, which should explain their scores and offer suggestions for how the manuscript might be improved.
5. Invitations to reviewers are issued from the main Consortium office, without identification of the editorial team that is in charge of the manuscript. (This ensures that reviewers do not tailor their comments and scores to norms and standards that are perceived to be specific to that team.) The review process is double-blind or triple-blind, according to the policies adopted by the editorial team.
6. Reviewers and authors may contact each other at any point in the process in order to clarify points about the manuscript or the reviews. Communications occur through an anonymized protocol (a la *Craig's List*) so that anonymity is preserved. (This should streamline the review

process, which is constrained to two rounds of review and therefore must come to a conclusion expeditiously.)

7. The manuscript may be withdrawn if the author and a majority of reviewers agree. (We expect this will occur only in a few very rare cases, e.g., where the reviewers notice problems that were not identified at the desk-reject phase.)
8. If not withdrawn, the author is expected to revise in light of comments and scores received from the reviewers, and to resubmit. After this point, no further modifications can be made to the article or the background materials. If background materials are housed separately these must be submitted with a time-stamp. (This eliminates the possibility of slippage between the version that is reviewed and the version that is published.)
9. In the second round, the six reviewers are asked to review the author's responses and the revised manuscript, along with each other's scores and comments, and to revise their own comments and scores. They are discouraged from raising new issues (unless they are a response to changes the author made in the revised manuscript) as the author has no further opportunity revise.
10. The author is given one last opportunity to respond to the reviewers' comments and scores – though she is not permitted to alter the text of the article or the background materials.
11. Three *breadth* reviewers are chosen by algorithm to answer just one question – the paper's breadth of appeal to the discipline. (This question is also addressed by the six main reviewers.) The algorithm chooses three people whose work falls in different subfields, none of which is the paper's (declared) subfield. They view the final (revised) version of the paper and are not privy to the reviews. Presumably, this reviewing task can be accomplished by perusing the paper's abstract, introduction, and conclusion.
12. The editorial team now makes a final decision about whether to publish the paper. If rejected, the paper goes to the next editorial team on the author's list (if that team has not already desk-rejected the paper) and so on until a team accepts the paper. If no team on the author's list of five claims the paper it is published without an editorial masthead, as a free-standing paper in the Consortium. (This is analogous to a book published with a university press outside of any book series.)
13. The article is copy-edited, typeset, and published on-line under a Creative Commons license. The title of the editorial team (if any) appears prominently on the front page of the published article, along with the editor and the team who oversaw the review. Following the article are the products of the peer review process – the reviewers' final written comments (anonymous), the author's final responses to those comments, and the reviewers' final scores – including the raw scores, adjusted scores from the measurement model, and confidence intervals around those adjusted scores. If the editorial team has chosen to write a review of the manuscript, this also appears.
14. Post-publication, a digital community space allows discussion of a published paper to continue. Here, any member of the Consortium may comment on the article and the author may respond. Authors may also submit corrected versions of the manuscript or background materials. Each entry would be permanent and attached to a unique DOI. Each comment would include the author's name, institutional affiliation (if any), and email, so as to discourage scurrilous comments. No editing would be required except (if necessary) to redact inappropriate comments.

15. Where serious mistakes in a published work are discovered the editorial team that supervised the review process makes a decision about retraction or correction. Reviews and scores accompanying the original article may be withdrawn or revised.

This is the process, in a nutshell. We turn now to a more detailed consideration of various aspects.

Choosing editorial teams

Upon submitting an article the author chooses a set of five editorial teams, in order of preference. To help with this decision, basic information about each team would be readily available. This includes information provided by each editorial team, e.g., their title, names of editors, and mission statement. In addition, statistical information would be generated by the Consortium database, e.g., number of submissions, mean submission ranking (1-5), number of desk-rejects, number of articles reviewed, number of published articles, mean article length, minimum article length, maximum article length, mean article score (raw and adjusted, across each dimension of assessment laid out in §IV), acceptance rate, mean time under review (counting only periods when an article is in the hands of the editorial team), mean length of written comments, mean reviewer score (a rating provided by each author at the end of the review process that rates the helpfulness of the reviews received for her article), mean Google Scholar citation count of published articles and reviewed articles.

Statistics would be aggregated at one- and five-year intervals and could be updated on a weekly basis. Tables containing this information could be organized by the user so that editorial teams are rank-ordered according to a chosen criterion, or set of criteria.

Desk rejection

If five editorial teams decline to review an article the decision to desk-reject becomes the responsibility of the Oversight committee. This committee may delegate decisions to individual members or to staff; but it must oversee the process and ensure consistency.

In making a determination, the committee should not try to estimate the potential of the paper *after* revision, as they have little control over that process. If they have specific suggestions, and especially if they have an idea for revision that might bump a paper above the threshold, they should communicate this to the author as part of their decision to "desk reject with possibility of resubmission."

Of course, this does not solve the core problem of establishing where the threshold of acceptance should lie and how it should be applied across submissions of very different sorts. Not everything deserves to be published. But everything that adds something to the sum-total of human knowledge should be published. This is our motto.

One way to ensure that this motto is achieved is to adopt a rule that a paper is published if any member of the Oversight committee recommends publication. Accordingly, once a paper reaches the committee it might send it direct to the member(s) who is most likely to be positively disposed. (Since committee members are chosen by editorial teams, and in some sense represent those teams, there is an expectation that if a committee member from a particular team recommends publication her team will be willing to review that paper.)

The rationale for this inclusive protocol may be briefly stated. First, if the Consortium comes to serve as the principal vehicle for publication in political science then papers rejected by the Consortium have no other publication options. Second, the long-term import of a publication may not be immediately apparent. Third, publishing a bad paper is probably less harmful – especially if

accompanied by poor scores and critical written comments, indicating its low quality – than failing to publish a good paper.

Of course, different standards must be applied to different sorts of contributions. A diverse field such as political science does not facilitate a single threshold of "soundness," on the model of PLOS and other mega-journals (Bjork, Catani 2016; Spezi et al. 2017). For a replication, the key issue is research design and an honest portrayal of results (no hatchet jobs). For a paper that purports to offer a new theory, the standard must focus on the theory. And so forth.

We need not speculate about what portion of submissions might be rejected by the committee. This obviously depends upon the volume and quality of submissions. It is worth noting that some science mega-journals reject nearly half of their submissions (Bjork, Catani 2016: 10). They are not come-one-come-all publication mills, as sometimes portrayed.

In any case, if a large number of papers on a given subject, or with a given methodology, are desk-rejected it may behoove the authors of those papers to form their own editorial team under the auspices of the Consortium. They would then be free to accept papers for review and publish their own brand of political science. Of course, they would also be responsible for finding reviewers for those papers.

As a practical matter, the decision to review a paper implies that trusted reviewers can be found to conduct that review. If this becomes a problem for low-quality submissions it may be necessary to desk-reject those submissions. This issue will need to be sorted out over time, and may vary across subfields or methodologies.

In any case, worries about an avalanche of bad submissions to the Consortium may be overblown. Desk-rejecting a very bad paper is not very time-consuming; the worse it is, the easier the decision. If a paper manages to beat the threshold and goes out for review but comes back with poor scores the author is obliged to publish that paper, whose scores remain in the public domain in perpetuity. (The rare exception is if the author and a majority of reviewers agree it should be withdrawn.) There are substantial reputational costs for submitting a poorly constructed paper, even if – one might say, *especially* if – it is published.

Editorial responsibilities

The editorial team in charge of a submission has six mandatory responsibilities and one optional responsibility.

Accept or reject. They must decide whether to publish the paper under their rubric, or not.

Dimensions of evaluation. They must decide which dimensions of evaluation (§IV) are relevant for a paper. For example, if a paper is purely theoretical, dimensions of evaluation dealing with validity are not applicable. These irrelevant dimensions should not appear before reviewers, lest they cause confusion. To be clear, this is not an expression of the editors' preferences (what they care about) but about the dimensions of evaluation relevant to the paper under review – relevant, that is, to at least some audiences (even if not to the editors).

Recruitment. They must recruit a full (or nearly full) complement of reviewers and make sure that these reviewers follow through with their responsibilities in a timely manner. A system of reminders and nudges may be programmed through the Consortium system, but sometimes the editorial team will need to send notes to motivate wayward reviewers or to judge complex situations where a review may not be forthcoming.

Redaction. If there are scurrilous comments in the reviewers' reports or the author's responses the editors must redact those comments – ideally, before it reaches the other party but in any case before it is published.

Reviewer rating. To give feedback to reviewers, and to motivate them, the editorial board must rate each full reviewer on a five-point scale. To what extent have they been helpful in providing direction to the author for improvement? Have they given enough clarity and detail in their written comments? Have they refrained from scurrilous comments? Are their written comments in sync with their scores? (Do they explain low scores?) Have they revised (or at least explained) their comments in the second round, in response to the author's responses? Have they approached their task in a conscientious and timely manner? If there are extenuating circumstances (e.g., an emergency contingency in the reviewer's life), this should be taken into account when assigning a reviewer's rating. (This is a fair bit of work so if it becomes onerous it might be reconsidered. Some of these issues would be picked up in the MM, e.g., reliability, length of comments, timeliness.)

Misconduct or serious error. Cases of plagiarism and other varieties of fraud must be judged by the editorial team, in consultation with the Ethics committee. If, after publication, it is realized – by the author or by other researchers – that a paper is marred by serious error, the editorial team will need to make a decision about retraction or correction. Reviews and scores accompanying the original paper might be withdrawn or revised. Although the editorial team handles day-to-day issues, the Ethics committee should oversee their decisions and provide a court of appeals when there are disagreements (either within the editorial team, with the author, with reviewers, or other parties).

Editorial reviews (optional). In addition to the foregoing mandatory responsibilities, the editorial team may choose to write its own "editorial" review of the paper, assigning scores (across the standard dimensions) and written comments. These will not be included in the measurement model but they will be made public, under the signature of the managing editor or the entire editorial team. Editorial reviews conducted during the first round would presumably reflect on the reviews, perhaps offering an opinion about which issues raised in the reviews are most crucial and perhaps taking issue with some of the reviewers. If revisions are recommended by the editors they should clearly state whether those revisions are necessary, or necessary and sufficient, for publication under their masthead. We would not want editorial teams to demand revisions and then reject a manuscript that has successfully implemented those revisions. (Note that this might make the paper *less* suitable for publication with a different editorial team.)

An editorial review offered in the second round can comment on a paper, and on the reviews, but is not in a position to request revisions – for the obvious reason that the second round is the final round.

In any case, editorial teams who wish to play a more active role in the process may do so. Editorial teams who feel that the reviewers know best, or do not have time to devote to this task, may avoid it.

Choosing reviewers

The choice of reviewers is perhaps the most consequential aspect of peer review. If the reviewers are qualified and conscientious the review is likely to be timely and helpful. If not, then not.

Finding those qualified, conscientious, and available reviewers is no easy task (Chetty, Emmanuel, Laszlo 2014; Stafford 2018). Journal editors must first identify potential reviewers, wait for a reply, and then find new reviewers if the first group refuses or accepts but does not fulfill their duties in a timely fashion.

The Consortium database would assist this arduous process by suggesting eligible reviewers. In principle, the entire system of reviewer choice could be automated. However, we are not sure how feasible it will be. Consequently, we assume that editors will play a role. How often they choose to deviate from algorithmically derived suggestions is up to them.

In any case, we suppose that the algorithm and the editors would aim for a balanced team of six reviewers, where individual reviewers satisfy somewhat different desiderata.

Subject knowledge could be judged by the overlap in work between (a) the paper under review and (b) all the papers published by the potential reviewer. Assuming the latter are in the Consortium database or are accessible through a library portal, one could match (a) keywords, (b) words in the abstract or full text, and (c) references. More holistically, the MM (see §V) could match reviewers to papers both by comparing the topic compositions of potential reviewers' published work to the topics in the paper under review, and by matching paper topics to scholarly communities, and recommending reviewers hailing from those communities. We believe this will be superior to relying on vague descriptions of the reviewer's area(s) of expertise.

Methodological knowledge could be judged in a topic model by overlap in the use of key methodological terms ("panel design," "IRT," et al.). Again, the scholarly network and topic modeling aspects of the MM that we describe in §V can do a lot of the work here, improving on simple keyword-matching approaches.

Conscientiousness could be gauged by a reviewer's track record. How promptly are reviews completed? Has she ever failed an assignment (gone AWOL)? Assuming editorial teams rate each reviewer (discussed above), this would offer a strong basis for judging conscientiousness.

Personal links to the author could be gauged by a network analysis that takes into account the place and year of PhD, departmental affiliations, coauthorships, and so forth. We presume that members of the author's department, her PhD students, and her former PhD advisor, would be excluded. Others who are less closely linked to the author might be chosen, but it is an issue editors should be aware of. Presumably, one would want a mix of "close" and "distant" reviewers. The "scholarly community" membership estimates produced by the MM may prove especially useful here, helping editors to find appropriate reviewers, while balancing community match with breadth of perspectives. As we mention above, it would be possible to provide editors with tools that would allow them to generate selections of recommended reviewers that would allow them to best estimate quality from the perspective of their intended audiences.

Availability could be gauged by looking at how many review assignments the potential reviewer has been given over the past year, and how many she has outstanding (not yet completed). And, of course, by checking whether the reviewer has listed herself as available in her Member profile.

Motivating reviewers

The task of finding qualified reviewers who will carry out their task diligently has become more difficult in recent years. There are many time-demands on academics and little payoff for conducting careful reviews.

The Consortium offers several possible mechanisms for incentivizing reviewers. (We leave open which of these, or which combination, might be implemented.) Each draws on the unique feature of the Consortium – its comprehensiveness. If most political scientists publish in the Consortium this gives the Consortium a variety of instruments that are not available to the editor of a traditional journal – particularly if that journal occupies a lower rung in the journal hierarchy.

First, one might require that for every paper published by the Consortium a (solo) author be willing (if asked) to conduct six reviews. For coauthored papers, the requirement would be a fraction of that number (3 papers if there are 2 coauthors, 2 papers if there are 3 coauthors, and so forth). This means that, after an initial post-sign-up grace period, if an author has not met her quota and says No to a request for review she is prevented from submitting new papers until the quota is reached – or, if she is not asked to review, until a specified period of time has elapsed.

Second, one might issue an annual public account of how many reviews each person in the Consortium database has conducted and how many invitations they have declined. In this fashion, researchers can claim credit for what they have done, which in some cases may be regarded as an important aspect of professional service, perhaps even a requirement for promotion. It also serves as a shaming device for those who have reviewed little and declined a lot.

Note that many journals list reviewers at the end of each year. However, they do not say how many manuscripts they reviewed, or how many they declined to review. And because a journal is a small space it is possible that reviewer confidentiality will be compromised. By contrast, in the extremely large pool of Consortium reviewers it is unlikely that listing the reviewer, and the number of reviews she has conducted, will reveal her identity.

Third, one might take advantage of the "availability" option on each member's profile (see above). If turned off, the reviewer is not pestered and will not accrue "declines." However, this member is also prevented from submitting new manuscripts. When a manuscript is submitted and sent out for review (not desk-rejected), the author is expected to be available for the subsequent six months. (The availability switch cannot be turned off.) Of course, the author may still decline invitations, but those will go on record (as declines).

With these various points of leverage, motivating members to review should be easy. Motivating reviewers to conduct thorough and insightful reviews is harder.

Publishing these reviews, as we propose to do, could help. Under the current system, some reviewer comments are ignored or over-ridden by the editors; they disappear from view and the reviewer may feel that her efforts are for naught. When reviews are published, the reviewer knows that her objections will become part of the permanent record, leaving future readers to consider the matter and come to their own conclusion. The reviewer's voice is heard.

More generally, there are reasons to suggest that publishing reviews enhances their quality. Although the reviewer's name would not appear, she is likely to be conscious of the fact that her review will be in the public domain in perpetuity. Tennant et al. (2017: 15) summarize several studies on the subject, concluding that

> when referees' comments were made public, significantly more cooperative interactions were formed, while the risk of incorrect comments decreased, suggesting that prior knowledge of publication encourages referees to be more constructive and careful with their reviews. Moreover, referees and authors who participated in cooperative interactions had a reviewing accuracy rate that was 11% higher. On the other hand, the possibility of publishing the reviews online has also been associated with a high decline rate among potential peer reviewers, and an increase in the amount of time taken to write a review.

As a final measure to incentivize reviewers (and to assist in identifying future reviewers), we ask editorial teams to rate the quality of reviews (both the first and second rounds) received from each reviewer on a five-point scale in which 0 = did not complete the assignment, 1 = barely adequate, 3 = adequate, 5 = very helpful. This might be made public, becoming part of that reviewer's profile. Surely, scholars would work hard to avoid a low rating on this public assessment of their performance.

Blinding

We assume a review process that is double-blind (authors and reviewers are blinded) or triple-blind (editors are also blinded).

We realize that reviewers are often aware of the identity of the author, or at least have a guess about whom it might be. Accordingly, we ask reviewers several questions about their knowledge of the reviewer and include this as a feature in the measurement model (§V).

Even so, we prefer a blinded process as this keeps the focus where it should be – on the manuscript and not on the author. Note that those authors who probably benefit most from anonymity are those who are lesser known, perhaps because they are young, unemployed, or employed in obscure places. For these authors, maintaining anonymity is feasible, especially if they have refrained from posting their paper online or listing it on their CV.

A triple-blind process is fairly easy to implement in the Consortium. To be sure, editors might mistakenly suggest the author as a reviewer. In this event, the system would not allow it. While this provides a clue to the identity of the author it is not definitive, as other potential reviewers – e.g., members of an author's department, former students, and former PhD advisors – are also peremptorily rejected.

Background materials

For most manuscripts, there are additional background materials that explain in greater depth how the author reached her conclusions and how others might do the same, if they wished. Where did the evidence come from? What are the possible sources of bias in this evidence? What additional analyses were conducted (other than those in the published article)? For qualitative work, it is often helpful to know what research sites the author visited, how she identified those sites, interacted with participants, and so forth. For quantitative work, it is often helpful to know how the data was gathered, what data transformations were performed prior to analysis, what additional tests (not reported or not fully presented in the article) were conducted, and so forth. For any empirical research there is data – of a qualitative or quantitative sort – that might be made available to future researchers, if properly curated and stored on a permanent site (e.g., Dataverse or the Qualitative Data Repository).

The goal of this background information is to make the steps involved in research as transparent as possible so that a manuscript can be properly evaluated by reviewers as well as by subsequent readers. Its second purpose is to allow future researchers to replicate the findings of a paper and to build on those findings, perhaps re-purposing the data for another project.

The tricky question is *how much* of this background information should be revealed, given ethical concerns (and formal IRB agreements) about protecting human subjects and the time-demands on authors, who must carefully curate their data.

We do not want to impose an editor's or reviewer's judgment on the author – who, after all, bears the primary moral and legal responsibility, not to mention the workload. Note that once a paper is sent out for review it is assured of publication, so there is no danger of editors or reviewers forcing authors to do things they do not wish to do. Authors remain in control of their article and their background materials.

At the same time, we want to encourage authors to be as forthcoming as possible, and we want to give signals to end-users about how conscientiously the author has pursued this goal. Accordingly, we ask authors to submit a data storage and transparency plan, specifying how they are approaching the task and detailing any ethical or practical issues that constrain their ability to be fully transparent. We ask reviewers of the manuscript to read the author's plan, review the background materials that have been submitted, and offer a score reflecting their judgment about how conscientiously the author has pursued this goal.

To be clear, the highest possible score may be given – and ought to be given – to an author who is unable to divulge details about her sources because of IRB concerns. This is not a hidden mandate for absolute transparency, and we trust that reviewers – chosen for their knowledge of the subject matter and methodologies employed by the author – will be sensitive to these considerations.

Where quantitative analysis is employed in a paper, we encourage authors to employ *Code Ocean*. This system allows authors to embed their code in a capsule, which runs on software at Code

Ocean. Anyone reading an article (on-line) can reproduce all the tables simply by pressing a button. In this fashion, the process of cross-checking tables and figures in a manuscript with the code and data used to generate those tables and figures is a simple matter of comparing what appears on the page with what appears from Code Ocean.[6]

Pre-registration and Registered Reports

Several protocols have been developed in recent years to separate the theory and research design stage of research from the testing and reporting stage. For purposes of theory testing, it is important to tie authors' hands, so they cannot fish for exciting results among a large number of potential hypothesis or manipulate particular configurations of data, specifications, and estimators that offer statistically significant results. For purposes of avoiding publication bias, it is important to motivate reviewers and editors to publish work that has strong validity, even if it offers no novel findings.

Pre-registration occurs where an author specifies the hypothesis(es), research design, and pre-analysis plan for a study on a public web site (e.g., E-Gap), which is preserved with a time-stamp and suitably anonymized.

A *Registered Report* (Nosek, Lakens 2014) is pre-registered and then submitted for review. Assuming it is not desk-rejected, an editorial team would send the paper out for review in the form of a proposal for work to be conducted later. (This would form the first round of the Consortium review.) Suggestions from the reviewers might be incorporated (or not) into the final research design. After the research is carried out it is resubmitted to the same reviewers, who offer a final judgment (the second round of review).

It should be obvious that these protocols apply only to work that is "confirmatory," where a hypothesis and research design can be specified in advance, prior to seeing the data, and does not in any way impugn the importance of exploratory research – which may occur ex post (if a study has been pre-registered) or as part of a study whose research design is exploratory from the outset.

At first glance, there is less need for these special protocols in the Consortium. Since all work that passes a low bar is published, authors have less incentive to fish for interesting results. However, there may be competition to publish with certain editorial teams, and those teams may employ "interesting" as a criterion for acceptance. Moreover, authors are likely to be cognizant of the reception of their work post-publication. Since sexy work generally obtains more citations, there may be an incentive to fish for flashy results even if publication is assured. Another argument for these protocols is that they enhance authors' consideration of important research design issues. (And if the pre-registered research design goes out for review, as it does in a Registered Report, the author also benefits from input on the research design from reviewers.)

For all these reasons, we believe that the case for pre-registration and results-blind review is nearly as strong under the Consortium as it is under the current journal-centered review process. We do not see the need for an official policy with respect to when authors should follow these protocols. But we do want to guarantee that these options are available, when warranted. And, we want to publicize these features clearly so that reviewers and end-users know which protocols have been followed, and with respect to what – since some aspects of a paper may follow pre-registered protocols while others deviate.

[6] The system is currently in use at *BJPS* and *PSRM* (see https://www.cambridge.org/core/journals/political-science-research-and-methods/article/who-gets-the-credit-legislative-responsiveness-and-evaluations-of-members-parties-and-the-us-congress/FED4A70A65D3458C1AC1AD7756967D5B/core-reader), under a collaboration with Cambridge UP (http://admin.cambridge.org/about-us/news/cambridge-university-press-and-code-ocean-announce-partnership/).

Digital-only publication

Under the current system, older, high-status journals tend to be published in hard-copy format and on-line (in a digital format) while newer or lower-status journals tend to be published only in digital format. Because publication formats seem to signal quality (though not very reliably), and because hard copies of journals are a product that can be sold to libraries, journals strive to maintain the traditional, hard-bound version of each issue.

There are many advantages to digital-only publication. First, it reduces carbon emissions. Second, it eliminates costs associated with producing and distributing hard copies. Third, it reduces the need for arbitrary page or word limits (Gerring, Cojocaru 2020). Fourth, papers can be published as soon as copy-editing is completed, minimizing delay between acceptance and publication and eliminating the confusing "first view" (pre-publication publication), which renders two publication dates for each article (once in first-view and once in the final publication). Fifth, the delicate synchronization of publication deadlines and content for "volumes" and "issues" can be dispensed with.

Sixth, color graphics are supported without extra cost. This allows for clearer, more impactful figures including photographs, maps, drawings, paintings, and other visualizations that depend upon a full palette. This would be a boon for work based on visual material, aka the "digital humanities."

Finally, digital publication opens the way for dynamic formats that capitalize on software already common on web sites – maps with features that can be turned on and off, videos, music, time-lapse pictures, and so forth. Executable books, which offer a truly interactive experience, run on Python code, often employing a software called Jupyter Notebooks (Matuschak, Nielsen 2019; Rule et al. 2019). There is no reason, in principle, not to incorporate these elements into an article if they are supported by open source software or commercial software licensed to the Consortium. One of the benefits of a large IT team (which we assume the Consortium will have) is that they could serve as consultants for innovative publication formats.

IV. DIMENSIONS OF ASSESSMENT

Social science work is diverse and contributes to knowledge in different ways. It follows that an adequate method of assessment must be multi-dimensional.

This is implicitly recognized by many journals, who ask reviewers to fill out a short survey summarizing a few of the elements that might distinguish a publishable paper from one that does not meet that journal's standards. Our proposal extends this effort to include a larger set of questions – and therefore, we hope, a more complete and nuanced assessment.

Our approach is similar to the "grade" envisioned by Nosek and Bar-Anan (2012: 233), except that it is multidimensional: many features are assessed and there is no overall grade. By specifying multiple dimensions we alleviate the problem of aggregation faced in the current peer review process, where multiple dimensions must be reduced to one or several. Research suggests that disaggregation of uncorrelated elements allows for quicker and more assured judgments (Arkes et al 2010; Arkes et al. 2006). Over time, reviewers will presumably gain familiarity with the dimensions and will apply them quickly and easily.

We are not sure exactly how many dimensions of assessment deserve to be recognized, or how each dimension should be articulated. This element of the proposal will require extensive deliberation and will need to be pilot-tested. Preliminarily, we offer the following questions.

1. *Style.* How tight is the organization? Is it too long or too short? Are there redundancies to be eliminated or ambiguities to be resolved? How felicitous is the prose?
2. *Literature review.* How accurate is the author's portrayal of the literature on her topic? Is the literature fairly characterized, or caricatured (as "straw men")? Is important work on the topic cited? Do the cited studies represent the full set of extant studies?
3. *Theoretical coherence.* Is the theory well-constructed? This is not about whether the theory is true; it is about the formal properties of the theory. Are the scope-conditions clearly laid out and logical (as opposed to arbitrary)? Is the theory internally consistent, concise, and linked up to other theories (if relevant to the question at hand)? If the theory is causal, are the causal mechanisms clearly laid out and plausible?
4. *New data.* Is new data provided in this study? How useful is it likely to be for other researchers (at work on the same or different projects)? How much of a contribution is the data, in and of itself?
5. *Measurement.* Is the chosen measurement instrument(s) a good representation of the theory/hypothesis being tested (construct validity)? Are there significant problems of measurement validity or reliability?
6. *Design.* How strong is the research design as a test of the main theory/hypothesis? How likely is it that the result of the test will be valid?
7. *Internal validity.* How likely is it that the main conclusion of the paper, as interpreted by the author, is true for the studied case or sample? Equally important, have potential threats to internal validity been discussed and honestly appraised?
8. *External validity.* Is the main argument likely to be true for the larger population of theoretical interest? Or, alternatively, does the sample incorporate the entire population of theoretical interest? Have potential threats to external validity been discussed and honestly appraised? Implicit in this question is whether the scope-conditions of the theory are clear. If they are not, the paper deserves a low score for external validity.
9. *Relative validity.* How strong are the claims to validity (internal and external) relative to other studies of the same subject and relative to what might have been accomplished (with a reasonable input of time and resources)?
10. *Novelty.* How new and different is the main argument relative to extant work on the subject?
11. *Methodological contribution.* Is there a significant methodological contribution? That is, does the study develop a new methodology (or an innovation on an established methodology) and is that methodology potentially fruitful for other work? What is its potential scope of application?
12. *Data transparency.* Has the author done as much as can be expected – given practical and ethical constraints – to make background materials (data, code, et al.) accessible to end-users who might wish to better understand the empirics, reproduce, or replicate the study?
13. *Breadth of appeal.* How broad is the appeal of this paper? Assuming that the claims are true (bracket your skepticism), what share of political scientists are likely to find this argument interesting or informative? How much would it move their priors, and how much would they care about the topic? (Evidently, one could as the reviewer to speak only of him/herself, and we suppose that most will do so. However, it seems better to ask them to contemplate their brethren, as the latter seems more likely to reflect the larger population in this small sample, which is prone to small-sample errors.)

Clarifications

Each of the foregoing questions is scored on a scale from 0 to 10.

For Registered Reports, a few of these questions would have to await the final version of the paper (after the research has been conducted).

Since articles have different objectives, some of these evaluative dimensions may not apply. (For example, a theory article may have no empirical material, obviating questions about empirics and methodology.) The editorial team will need to decide this, prior to being sent out for review.

Editorial teams could choose to add questions to this survey if they feel additional questions are scoreable (quantitatively) and are important for evaluating work in their field. These questions would not appear (unless requested) in the review process for other editorial teams.

We hold out for further consideration the possibility that *authors* would fill out the same score sheets as reviewers, perhaps at the beginning of the process, as a gauge for how much agreement there is about the strengths and weaknesses of a manuscript.

Confidence

For each score, reviewers also report a level of *confidence* – from 0 (not at all confident) to 10 (completely confident). This is important, given that reviewers are likely to know more about some aspects of a manuscript than about others. Note that some reviewers may be chosen for their methodological skills, their knowledge about the theory, or their substantive knowledge of the area under investigation. Under the circumstances, one can imagine they might feel differentially confident about different questions.

At one extreme, a reviewer may feel she has no basis for judgment at all, registering "0" confidence. At another extreme, a reviewer may feel that she knows an area well enough to register a very high level of confidence. This rightly influences the weight assigned to their score.

It also encourages reviewers to offer a score along all dimensions, even if they are not very confident about what that score should be. There is no excuse for leaving a question blank, as one can indicate a low degree of confidence where one is highly uncertain.

V. MEASUREMENT MODEL (MM)

A core objective of any system of peer review is to offer clear and unbiased signals to producers and consumers of social science about the quality of work that is published. Unfortunately, there is only so much that we can do to encourage reviewers to offer unbiased reviews. After all, reviewers have different standards. Moreover, the review process has certain structural features that – at least in some circumstances – inhibit dispassionate deliberation and honest scoring. We suspect that many biases are unconscious.

However, a large pool of reviewers and a structured system of scoring affords the possibility of enlisting methods from the field of measurement to adjust scores so that some biases (namely, those that can be measured) are mitigated.[7] It also allows us to use patterns of ratings to norm scores across reviewers, to measure and adjust for reviewer reliability, and to provide assessments of uncertainty around reviewer evaluations.

[7] This builds on earlier work by Riggs and Wilensky (2001).

In this section, we introduce a measurement modeling framework (MM) designed to provide useful summary metrics of reviewer evaluations, facilitate editorial teams' ability to identify appropriate reviewers for new submissions, and to improve researchers' ability to find relevant work and to understand relationships between papers across the discipline.

While we envision this framework as a powerful feature of the Consortium publishing model, we want to emphasize that it is not fundamental to its operation. The MM provides a tool that can help editors, authors, and readers; but this tool should feature only as one part of a human-driven process of paper evaluation, understanding, and development. While the MM will produce potentially useful metrics of paper quality, situate a paper within literature and authorship networks, and help editors to identify qualified reviewers, we envision it as an adjunct to other measures.

For some Consortium users, the products of the MM may be viewed as crucial "value-added"; for others, it will be an extra bell or whistle; and for still others, it may be seen as a misguided attempt to quantify an inherently qualitative process of understanding and evaluation. This is entirely consistent with our intent and reflects the flexibility of the Consortium publishing platform. We also envision an open-data model that will allow others to roll their own versions of the modeling framework to satisfy a wide range of needs for metrics.

A Hierarchical, Network-Integrated, Topic Situated, Graded Response Model

Scientific publication is a process whereby a network of scientists—linked by educational experience, co-authorship, and expertise within and across sub-disciplines—produce and evaluate a network of writings—linked by authorship, citation, feedback networks, and topics. Both of these networks are hierarchical. For instance, scholars of Congress are nested within legislative studies and American politics, and some may be further nested within broader networks of scholars of institutions, either in the American context, or more broadly. Some such scholars may share membership in other sub-disciplinary communities, both methodological communities like experimental methods, and substantive communities like political psychology. Many communities have both substantive and methodological character. Similarly, papers are hierarchically situated within a complicated structure of topics. Both of these networks—*scholarly communities*, and *paper topics*—are complex and individual scholars (papers) may be partially situated within a number of communities (topics). That is, scholars (papers), may have graded or partial membership within different communities (topics). These networks are structurally related to one another. Certain topics are associated, more or less, with certain scholarly communities. As a result, knowledge about the scholarly communities to which a paper's authors belong will help predict the topics of the paper they write, and vice versa.

The review process is necessarily situated within this hierarchical structure. A reviewer's evaluation of a paper—itself composed of a certain set of topics—reflects the communities to which she belongs. In turn, the extent to which a particular evaluation—produced by a member belonging to a particular set of scholarly communities—would reflect an assessment by a scholar hailing from another scholarly community is inherently unclear. One implication of this structure is that there is almost certainly no such thing as a rating of paper quality, on the dimensions described in §IV, that holds across an entire scientific discipline. At worst, this concept is undefined; at best it is a highly uncertain construct. At the same time, reviewer evaluations are probably not only meaningful at the lowest level of the sub-disciplinary hierarchy. Scholarly communities are nested, and nested groups of communities have links to other hierarchical community clusters, so reviews produced by a member of one community can tell us something, if only roughly, about how a member of a related community would evaluate a paper.

We propose a modeling framework that takes the complex structure of the scientific paper writing and evaluation process into account. It integrates research on combining graded evaluations

from multiple experts (Johnson and Albert 1999, Pemstein et al. 2019) with work on topic modeling (Blei, Ng, and Jordan 2003) and network analysis (Newman, Watts and Strogatz 2002), and work that bridges topic models with network analysis (Sun, Han and Yu 2009) and item response theory (Gu et al. 2014). The resulting mouthful—a "hierarchical network-integrated, topic-situated, graded response model"—will capture the basic structure of the publication process. Using information about scholars' professional relationships, natural language processing of their papers that they produce, and both scholars' and papers' citation patterns, it will use random graph and latent and topic modeling techniques to map the structure of both the discipline's professional communities and the papers that people within these communities produce. Reviewer evaluations—the graded response part of the model—will be situated within this network structure. In particular, each reviewer's position within the network of scholars, relative to the submitted paper's position within the topic network, will inform the estimation of measures of rater strictness and reliability. The model could learn, for instance, the extent to which reviewers from one scholarly community provide reliable reviews for papers on topics associated with another community, and which sorts of scholars tend to be especially hard, or easy, on which sorts of papers. The mathematical details of such a model are beyond the scope of this paper. We therefore restrict ourselves to outlining the key inputs and outputs of this approach.

MM Input

As envisioned, the MM will synthesize the following information:

1. *Scholar (author or reviewer) characteristics*
    a. *PhD institution.* One's PhD granting institutions shapes one as a scholar and is highly informative about the scholarly communities to which one belongs.
    b. *PhD adviser(s).* While students do not always study the same things as their advisors, there tends to be strong overlap in student-adviser scholarly community.
    c. *Year of PhD.* Scholarly communities and scientific topics are dynamic and change over time. While individuals evolve, different generations of scholars nonetheless often exhibit systematic variation in how they fit into disciplinary sub-communities.
    d. *Co-authors.* The model can learn a lot about network structure from co-authorship relationships.
    e. *Citation patterns.* Scholars' career-long histories of citation, both from and to their papers, provide a wealth of information about their place in scholarly networks.
    f. *Publications.* Scholars will be linked with the papers that they have published, which will be classified by the topic modeling portion of the framework. Knowing which sets of topics a scholar has published on can help the model to place the scholar within disciplinary communities.
    g. *Demographics.* Various demographics, such as age, gender, nationality, and country of residence are predictive of the topics that people focus on, and the disciplinary communities within which they operate. They may also predict biases in the review process. For instance, even when authorship is technically blinded, unconscious biases could influence the evaluations of reviewers who know an author's identity. Note that we propose using all of the author characteristics that we describe here, to model *only* patterns of scholarly community membership, and the context of reviewer evaluations, not as predictors of paper quality. We include these variables place people into

communities, and to help predict and alleviate reviewer bias, **not** to algorithmically bake patterns of bias into scores.

2. *Paper characteristics*
    a. *Paper text.* The paper text itself will inform the topic classification model.
    b. *Paper citations.* For papers that have been published, we will collect citation counts and lists. The counts will serve as one metric of paper quality while the pattern of citations will inform the network structure of the latent topic hierarchy.
    c. *Paper bibliography.* The model will use reference patterns to inform the paper's location within the latent topic hierarchy.
    d. *Paper journal/ editorial team.* The model will incorporate information on papers published both as part of the Consortium and in other outlets. Thus, it will include existing work in the topic modeling process. The identities of existing journal outlets and our envisioned editorial team approach provide information about where a paper fits into the disciplinary topic structure.
    e. *Paper year.* Collecting temporal information allows us to model dynamic change in both the hierarchy of scholarly communities and the paper topics.
3. *Review process characteristics*
    a. *Reviewer scores.* Reviewers will provide graded evaluations, on an ordinal scale, of each dimension described in §IV.
    b. *Reviewer confidence.* Because reviewers enter a confidence score for each (substantive) score, the MM may give greater weight to scores that are associated with greater (self-assessed) confidence, because the model can learn the relationship between confidence and reliability (which could be negative). Confidence scores can also be situated relative to reviewer's personal characteristics. That is, the model could learn that reviewers in certain scholarly communities, or particular educational backgrounds or demographics, tend to be over-, or under-confident, on average
    c. *Reviewer behavior.* The existence (for reviewers who only need provide scores) and length of reviewer comments provide information about the extent to which a given reviewer is dedicating time and effort. One could also, in principle, also use natural language processing techniques to evaluate the tone of the review, and include measures based on such an analysis.
    d. *Dyadic characteristics.* Typically, editors try to avoid reviewers who are close to one another and thus not in a position to give a dispassionate review – perhaps even engaged in a quid pro quo agreement to favorably review each other's work. However, this is hard to judge, and always a matter of degrees. Likewise, there is usually a tradeoff – the most dispassionate reviewers probably know least about the subject. A better approach may be to incorporate these dyadic elements into the MM's estimation of how strict a reviewer is when paired with a particular paper, e.g., whether the author and reviewer have ever shared institutional affiliations (in grad school or subsequently), whether and how often they have coauthored, whether they live in the same city or state, whether they are co-nationals, whether they are close in age, and so forth. Insofar as favoritism derives from closeness, many of these features are measurable. Much of this information will be summarized by placement within the disciplinary network, but we may include direct measures of closeness to augment the network analysis.

e. *Dyadic history*. It can be problematic if authors review one another's work repeatedly and if those reviews are more favorable than their reviews of other similar papers. This may be a sign of a formal or informal cartel. This information can be incorporated into the MM.

f. *Treatment of reviewer's work*. Each reviewer answer several questions about how the reviewed manuscript treats their own work. (1) Is your work cited? (Y/N). If yes, (2) how extensively? (a) minimally (e.g., one or two citations), (b) moderately (several citations or short discussion), (c) long discussion. (3) How is your work is treated by the author? (a) favorably, (b) neutrally, (c) critically. The reviewer should register her view of these matters, which are – especially in the last instance – somewhat subjective. This response is confidential, and thus not published along with scores and comments.

g. *Knowledge of author*. Reviewers often have an idea of who wrote the paper they are reviewing, and it is plausible that this speculative guess – or, in some cases, sure knowledge – influences their review. Note that bias may exist even if the reviewer's guess about the author's identity is incorrect. Thus, we pose the following question to each reviewer: "If you were forced to guess about the identity of the author of the paper you are reviewing, in what category would you place this person?" Options include: (a) distant (never met), (b) professional colleague (e.g., probably met at conferences), (c) close colleague (e.g., have worked together on projects or sat on committees), (d) friend. We will also ask the authors to provide a measure of how confident they are that they know the identity of the author(s).

h. *Review evaluations*. We will allow editors and authors to score the helpfulness of written reviews. These scores can help inform the model's estimation of reviewer reliability and strictness.

MM Output

The MM will provide the following outputs:

1. *Scholarly community network map*. The model will produce a hierarchical random graph representation of how members of the Consortium, and scholars who have published work in political science, or adjacent, journals,[8] fit into the discipline. One can leverage this information to identify communities within the discipline, to determine how central individual scholars are to particular sub-communities, or the discipline as a whole, to examine the similarity, in terms of community membership, between subsets of scholars, and so on. It represents a way to visualize the professional communities within the discipline. It also provides a powerful tool for identifying potential collaborators, panel organizers, potential hires, or other scholars who might provide one with useful feedback on work.
2. *Paper topic map*. The model will also cluster papers into a hierarchical set of topics and provide evaluations of the extent to which each paper in the input data—both Consortium papers and publications in existing journals—fits into each latent topic. This information could provide an invaluable tool for improving literature searches and for understanding how particular work speaks to the rest of the discipline.

[8] The model would incorporate a large corpus of existing journal publications into its input. While the model would have access to more information about Consortium members than non-members, it could nonetheless include non-members in the community model.

3. *Scholar topic map.* The model will allow for feedback between the scholar network and paper topic sub-models. This means that one can ask how a particular scholar's work fits into the topic hierarchy. One could examine the extent to which different sorts of scholars specialize and identify the scholars most associated with a given scholarly topic. Topic information could also help inform the interpretation of the scholar community network, making it easier to label the various sub-communities evident in the learned random graph.
4. *Paper community network map.* The model will also allow us to situate papers within scholarly communities. Thus, one can use the model to identify papers that are especially central to particular scholarly communities. This information, while informed by citation practices, will go beyond simple citation counts, and can allow analysts to examine not only how influential a paper is, but which communities it influences, and which papers are widely or narrowly situated. This information could be valuable for users interested in understanding how papers fit into the discipline, and in identifying opportunities to bridge disparate communities.
5. *Reviewer reliability.* The graded response portion of the model will estimate reliability scores for reviewers, based on their patterns of agreement with other reviewers. The model will use a Bayesian empirical prior approach to incorporate the various scholar characteristics that we describe in the preceding section, to inform these estimates. These reliability scores will be topic specific. Because reviewers lack the expertise to review any paper in any sub-literature, such scores will be attached to estimated topics and certainty around reliability estimates will be contingent on information. For example, the model might learn, with great certainty, that a certain formal theorist is a highly reliable reviewer of formal papers but place a large confidence interval on its estimate of her reliability as a reviewer of ethnography. These reliability measures could help inform the selection of reviewers by editorial teams—one would prefer to select reviewers who have a history of reliable evaluation of the topic(s) in the paper under consideration. Reviewers will also be able to view their own reliability estimates—estimates will be linked to individuals by an anonymous id, and therefore not to public names—and to learn the extent to which their reviews exhibit consistency with their peers.
6. *Reviewer difficulty.* The model will estimate difficulty parameters for each reviewer, again relying on an empirical prior approach to incorporate scholar-specific and author-reviewer dyad characteristics. These difficulty parameters are measures of strictness, or how hard a particular reviewer is. Like the reliability parameters, they will be nested within topics. Unlike reliability parameters, they will also be community-specific, to allow the model to capture potential biases driven by dyadic factors. Again, empirical priors will allow us to incorporate directly measured individual and dyadic characteristics into their estimation. Like reliability parameters, these parameters will help editors identify appropriate reviewers and could be used by scholars to evaluate their reviewing practices, and potentially adjust if they found themselves to be too strict or too lenient. The network-contingent aspect of these parameters will also aid in the identification of publishing cartels.
7. *Paper ratings.* The model will produce scores for each dimension listed in §IV, by aggregating the scores provided by individual reviewers. Standard graded response modeling techniques will adjust scores for reviewer reliability and difficulty. In other words, the model will place greater weight on more reliable reviewers' ratings and will take reviewer strictness into account. These scores will be produced at the community level. The certainty around any hierarchical version of a given score will depend on the set of communities, and topics, represented by the reviewers and paper in question. Some communities will have a lot of overlap with other communities, others will not. The model will be able to provide reasonably certain estimates of scores situated within the communities strongly represented by reviewers. Scores for

communities that are less well represented by the reviewers will be less certain, but the degree of uncertainty will reflect how strongly linked the represented communities are to the community of interest. The model will therefore not provide scores in the conventional sense. Rather, it will estimate how a representative reviewer, contextualized within a given community—or mixture of communities—will view the paper. This provides observers with multiple angles from which to evaluate a paper's scores. An editorial team might, for instance, evaluate a paper situated in communities that it considers its main audience and one could build tools, based on estimated model parameters, to facilitate appropriate reviewer selection to achieve a precise estimate for a given set of target communities. A potential reader might evaluate the estimated score from the vantage point of her estimated mixture of community memberships. A hiring committee might see how a paper scores in the community of scholars in which it is looking to hire. But, importantly, scores will be inherently contingent on the point of view, and will be more certain when the mixture of reviewers represents the point-of-view of interest, reflecting what we consider a fundamental reality of the review process.

Discussion

The MM will update in real time. Using a Bayesian approach, the model could update sequentially instead of running from scratch with each iteration. The platform will make raw and MM scores available to authors, editors, and reviewers at each round of review. In this fashion, the author knows what the raw score will mean in the final adjusted score. In particular, a low raw score may translate into a moderate or even a high adjusted score for a tough reviewer. Scores for each published work will change in small ways with each iteration of the MM. However, as the body of work contained in the Consortium database increases, these perturbations will reduce in magnitude.

The proposed elements of the MM model allow one to control for several perceived biases in the peer review process. Others could be added, so long as they are measurable. Of course there is a danger of over-correcting. The perceived gains from certain aspects of the MM must be weighed against the costs, and each element of the model must be tested so that the community of users understands how much impact, and what sort of impact, each element has. We anticipate running several versions of the MM, each with different assumptions. The community of political scientists should have an open debate about which version to regard as the most authoritative.

VI. IMPLICATIONS FOR ASSESSMENT

In this section, we discuss the implications of the Consortium for the assessment of work submitted for review.

Plagiarism

The current journal-centered system of peer review has a problem with plagiarism. The independent operation of hundreds of journals encourages acts of plagiarism, makes it difficult to identify when such acts occur, and provides an uncoordinated response on those rare occasions when they are discovered.

Consider, first, that plagiarism is most likely, and most damaging, when an idea or dataset is appropriated *prior* to publication. Post-publication plagiarism is less likely – because the idea is already

in the public domain – and less worrisome since the author with the earlier publication date can prove that she got their first.

We have demonstrated that it takes a long while for ideas to reach publication under the current system of peer review, and many papers never reach publication. During this time, ideas (and perhaps even datasets) circulate as working papers, conference papers, talks, and as submissions to journals. Since first submissions are often unsuccessful, many editors and many more reviewers will see a paper before it is published. Each viewer has an opportunity to steal (Taichman 2017). It is important to recognize that theft is not always intentional. Once one has read a paper one's view of that topic is inevitably affected, sometimes in ways one is scarcely aware of. One cannot delete one's memory banks. And while a paper is under review it is difficult or impossible to cite it; during this period, which may last for several years, the manuscript exists in a liminal space. It is a system designed to encourage plagiarism.

By contrast, the Consortium brings papers quickly through the pipeline. Moreover, they are already accepted for publication at the time they go out for review, so there is little payoff to would-be plagiarists, and no reason not to cite the paper if they wish to incorporate ideas from it in their own work.

The existing system of peer review is also ill-adept at catching plagiarism when it occurs. Since there are many avenues to publication, each operating independently, the plagiarist has many venues for publication. All it takes is one hit.

By contrast, in the Consortium all submissions (even if desk-rejected) are maintained in a database, so all new submissions can be checked for redundancy using plagiarism detection software. These checks can easily be extended to work published elsewhere (outside the Consortium).

Finally, the current system of peer review is incapable of formulating and implementing a clear and consistent policy towards plagiarism. Essentially, each journal, and each editor, exercises her judgment about how to respond (Rosser 2014).

By contrast, in the Consortium we envision the creation of an Ethics committee that can deal in a comprehensive and consistent fashion with the issue – which is, after all, a complex one (especially with respect to self-plagiarism). Given the size of the Consortium it could probably afford to enlist legal counsel, if that becomes necessary. In any case, editorial teams should not be left alone to wrestle with ethical and disciplinary issues.

Journal fraud and incompetence

Currently, scholarly journals are unregulated. Anyone can start a journal, operate it in whatever fashion they see fit, and publish whatever they wish. This system works adequately at the top of the hierarchy, where journals are supervised by editorial boards (and sometimes sponsoring associations or universities), and where they have an established reputation and an incentive to preserve it.

At lower levels, however, there is little oversight, little in the way of public knowledge about the journal, and little incentive to establish and preserve the journal's reputation. Some journals are cash vehicles, where a position on the journal is sold to authors, a form of vanity publishing – but not one that is easy to discern to end-users. Even where editors are conscientious, they may not have the means to establish effective oversight over the review process. Note that where it is difficult to find reviewers, any reviewers – including those suggested by the author – may suffice. Some research environments seem to be especially conducive to academic fraud (Fanelli, Costas, Larivière 2015).

The Consortium cannot solve all of these problems. But it can serve an oversight function, and this is quite a bit more than currently occurs in the dark corners of the publishing world. An advantage of the Consortium database is that one can check the credentials of editors prior to joining the Consortium and on an intermittent basis thereafter. Those without established academic track

records may be regarded as suspect. As a large operation, Consortium could afford to devote some staff resources to sniffing out potential cases of fraud. And it could develop algorithms designed to detect fraud – e.g., when reviews for a paper appear on the same day, when more than one reviewer is unknown (no previous history as a reviewer or author), or when newly registered reviewers do not have Orcid accounts and claim positions at universities that do not match other entries in the database.

Oversight is enhanced because there are ready mechanisms of enforcement. Offending editorial teams can be suspended or replaced. Offending authors and reviewers can be put on notice, or shut out from the Consortium entirely. In this manner, the Consortium could assure a degree of credibility that is currently lacking from the peer review process.

Reviewer disagreement and the problem of stochastic error

Studies of the peer review process show that there is a very modest correlation across reviews of the same paper. Frequently, reviewers disagree about whether a paper should be published (Bornmann, Mutz, Daniel 2010; Cicchetti 1991; Ernst, Saradeth, Resch 1993; Marsh, Jayasinghe, Bond 2008; Rothwell, Martyn 2000).

On its face, this sort of disagreement is discrediting, and the journal's decision to accept or reject seems highly arbitrary. We regard this as a strong prima facie argument in favor of setting a low threshold to publication, as we propose to do. In addition, we want to consider some features of the traditional review process that may impede consensus, or prevent us from appreciating the consensus that exists. In each case, the process envisioned for the Consortium should vitiate the problem.

Disagreement among reviewers may reflect different standards of adequacy. There are tough graders and easy graders. Fabrizio Zilibotti (2014: 106), former editor of the *Review of Economic Studies* and the *Journal of the European Economic Association*, observes: "Some people are hyper-critical, and would never recommend any paper for publication. Their recommendations range between flat and borderline negative recommendations…There are as well 'soft' referees who view all papers as interesting and worthy of publication."[9] In principle, the editors of a traditional journal would be able to judge these matters, and some editors with lengthy experience come to learn the personal styles and predilections that each reviewer brings to the table. However, this is a rare skill, and virtually impossible to apply when a reviewer pool is large and diverse. In a MM, so long as each reviewer has a track record, it is easy to incorporate her "fixed effect" in an adjusted score.

Disagreement among reviewers may reflect different priorities. Some may prioritize internal validity while others prioritize external validity, for example. This issue is also easy to solve if we can effectively disaggregate the various dimensions of evaluation, as we also propose to do (§IV).

Disagreement among reviewers may reflect reviewer choices about what to focus on in their review of a manuscript. Fiske and Fogg (1990: 591) find that "In the typical case, two reviews of the same paper had no critical point in common. … [T]hey wrote about different topics, each making points that were appropriate and accurate. As a consequence, their recommendations about editorial decisions showed hardly any agreement." In this respect, the review process is reminiscent of the parable of the blind men and the elephant. This sort of pseudo-disagreement can be solved by forcing reviewers to focus on the same dimensions of an article in their scores, as we propose to do. Different priorities would thereby be distinguished from actual disagreements.

Disagreement among reviewers may reflect varying familiarity with the substance or methods of the paper under review. This is registered in the MM by the reviewer's self-reported confidence in

[9] Zilibotti (2014: 107) adds, "Guessing referee personalities from general traits is less easy than one might expect. Some people who are hypercritical in seminars and have a mixed reputation in terms of social skills are very constructive referees, and some friendly, soft-spoken researchers turn cruel when acting as anonymous referees."

each of her scores and incorporated into the MM by situating ratings with respect to scholarly communities and paper topics.

Disagreement among reviewers may reflect a lower overall familiarity – across all reviewers – with the substance or methods of the paper under review. Note that in the traditional review process, the expertise of reviewers is likely to decline as a paper goes through multiple submissions and rejections. For a given paper, there are a limited number of highly qualified reviewers, and these individuals are likely to be identified by the first journal to which an article is submitted. Typically, these reviewers decline to review the same manuscript a second time. It follows that the more times a paper is reviewed the lower the quality of the ultimate reviewers is likely to be. In the Consortium, the review process is limited to one iteration so there is no risk of reviewer attrition. There are also more incentives for reviewers to accept invitations to review and to perform their task in a conscientious fashion. Insofar as reviewers are more qualified and more motivated, his should reduce inter-rater disagreement.

Disagreement among reviewers in the current system, as measured by standard statistical tests, is exaggerated by the small number of reviewers – generally, two or three. To combat the problem of small-sample variability the Consortium would enlist six reviewers, and an additional three reviewers (chosen by algorithm) who attend to the matter of breadth. While this is still a small number, doubling the number of reviewers should enhance levels of agreement and reduce stochastic error in the overall result.

Despite all we have said in the foregoing paragraphs, it should be acknowledged that there is no way to entirely vanquish error. As with all social phenomena, the process of peer review is subject to mistakes. It is, after all, a judgment. In social-scientific studies we accept the inevitability of error, representing it with a confidence interval or some other measure of uncertainty. (If the research is qualitative this uncertainty will be expressed in prose.) The acknowledgment of uncertainty, and the attempt to measure it, is one of the defining features of science.

Unfortunately, no measure of uncertainty is publicly registered in the traditional peer review process. The author is privy to the reviews of her work, which offer a rough indication of the degree of consensus reached by the reviewers. But the readers of a published paper have no indication of the extent of inter-reviewer agreement, or what they disagreed about.

An essential component of the Consortium is the publication of *confidence intervals* along with point estimates (for the various dimensions of reviewer evaluation). These confidence intervals reflect the degree of agreement/disagreement among the reviewers, most importantly, but also other aspects that might affect uncertainty such as the number of reviewers, their reliability, and so forth. In addition, readers may peruse the reviewer reports so they have a better sense of what issues were vetted and where they may have failed to reach agreement. As we mention above, the MM will provide scores that are contextual, and measured with respect to some (mixture of) scholarly communities. Different cross-sections of reviewers provide different points of view, and the MM is designed to reflect this reality. In turn, the model will provide estimates of confidence that are a reflection of points of view, and confidence will reflect the extent to which evaluations produces by members of certain scholarly communities (fail to) reflect the evaluations of members of other communities, and the model's information about cross-community evaluations.[10]

The uncertainty of a score is almost as important as the score itself. End-users should take seriously point estimates with tight confidence intervals and should be skeptical of point estimates

[10] This is simply a matter of data availability. Some sets of communities will include members that review one another's work, while other communities will be more isolated. The model will fall back on vague prior information when it doesn't have evidence about how members of one community evaluate papers drawing on certain topics and report extreme uncertainty about scores from such disparate points of view.

with very broad confidence intervals. They should also take the vantage point of the reviewers into account and reflect on how well such reviewers might capture their own point of view. On both fronts, a report from the editorial team supervising the review process may be especially helpful. One can also imagine a post-publication review process that revisits especially contentious reviews in an attempt to gain greater clarity.

Biases

For any system of peer review, we must consider the sort of biases – conscious and unconscious – it establishes. Most of these problems exist in the current journal-centered system of peer review, so the standard we ought to be considering is not whether they could be entirely overcome but whether they could be ameliorated.

One concern is the *cartel*. Here, an editorial team conspires to promote their own work and the work of their students and friends, or to those who do a certain brand of work, giving high marks to this in-group and low marks to everyone else. This is easy to achieve simply by finding the "right" reviewers for each paper. However, it is sustainable only for a short while under the Consortium. Note that – unlike the traditional review system – all scores are public, so patterns of nepotism are easy to spot. Those who feel that the scoring system is rigged can submit their work elsewhere, or form their own team. In this scenario, the team is reduced to insiders. As a group of insiders, the cartel is sustainable. However, the MM adjusts each score according to a reviewer's history. Someone with a history of high scores will see her scores adjusted downward. So, if members of a cartel are reviewing each other's work and assigning high scores, over time these scores will be adjusted downward. The only way to maintain the value of their scores is to review papers outside the cartel and give those papers low scores. However, in this setting one imagines that cartel reviewers would be out of synch with other reviewers of those papers, which lowers their reliability and lessens their weight in the MM. The MM also situates reviews within the network structure of the discipline, allowing it to capture a process whereby reviewers adjust their strictness based on paper topic, or the communities of a paper's authors. Moreover, it is unlikely they will continue to be asked to review papers outside their niche. For these reasons, cartels will have a tendency to implode.

A second concern are *quid pro quo* agreements. Here, individual authors conspire to give each other positive reviews. Alternatively, less formal arrangements may arise in which members of a network of friends and collaborators give high scores to each other without any formal conspiracy. This behavior is more subtle, but is still fairly easy to detect with the MM, and thus properly discount. Because the reports of all reviewers are contained in a single database one can develop algorithms that detect favoritism, e.g., by comparing the scores that A gives to B, and B gives to A, with the scores they give to all other submissions. The current journal-based system offers no mechanism for uncovering this sort of favoritism.

A third concern is *forum-shopping*. If authors wish to maximize their adjusted scores they will choose the editorial team that is likely to give them the highest (adjusted) scores. It is not clear, however, how they would achieve this. Suppose an author has conducted a study with an experimental design. The author may request an editorial team that publishes a lot of experimental work, thinking that they will appreciate her work. However, reviewers for that team may have high standards. In order to obtain a high (adjusted) score the study must be viewed as superior to other experimental studies. Alternatively, the author may request an editorial team that does not have a lot of experience with experimental studies. They are not likely to accept it, but if they do it is not clear what they will make of it. They may regard it positively, or they may not like it at all (since this is not the sort of work that they do). In any case, we suspect that work will sort itself into editorial teams in a logical fashion – precisely because this is public information and readers will detect obvious cases of forum-shopping.

An experimental study that is published by an editorial team known for interpretivism will be regarded askance, even if it achieves a high score.

If it is the case that ex post evaluation is beginning to displace ex ante evaluation (as discussed below), then authors ought to be more concerned primarily with the quality of the articles that they publish than their scores or the editorial team that sponsors each publication. Accordingly, they ought to send each paper to the editorial team that is likely to give them the most useful feedback, even if that team has a "tough" reputation.

A final consideration are the biases that, according to many studies, arise from the reviewer's knowledge of the identity of the author (§I). Some biases are fairly easy to correct. If we assume that the quality of manuscripts is not higher or lower for men and women, on average, we can include the gender of the author in the MM to correct for potential biases in the review process, as we propose to do. We cannot do the same for factors such as university ranking or country of residence, as these may be correlated with quality. However, we can easily test to see if these factors are correlated with scores in a systematic fashion. If it is discovered that papers emanating from Africa are reviewed less favorably than papers emanating from Europe we could suggest to authors from Africa that they refrain from posting their papers prior to submission so their identity is not obtainable from a web search.

Breadth

A common complaint of the current peer review system is that it privileges work that is narrow, catering to a specific subfield, methodology, or perhaps even a small set of people within a subfield who work on a particular topic.[11] Shared preferences are often reinforced by networks among friends, mentors and their former students, and PhD programs. There is of course nothing wrong with like-minded people gathering together to promote their work. A key advantage of the Consortium is that it eases the organizational and financial challenges of doing so. However, it is also important to know whether the work promoted by that group is widely regarded as interesting.

One safeguard is the MM, which adjusts the scores for each reviewer by the scores that reviewer has given across all other reviews. So, if members of the niche group review only work in their niche, and assign high scores to all such work, their scores will be adjusted downward, as discussed.

To recognize work that has broader ambitions we include a question on the reviewer survey about a paper's anticipated breadth of appeal across the discipline (§IV). This question is answered not only by the six expert reviewers but also by three reviewers chosen by algorithm from subfields outside the article's subfield.

In addition, we situate each article within a network-topic analysis of the entire discipline (§V). In this fashion, we can show where the author and the reviewers fit within political science, i.e., how central or peripheral they are, and in which parts of the sub-discipline they most clearly fit.

Of course, one might take the position that questions of breadth are better answered ex post – e.g., by a paper's citation count – than ex ante. However, some papers have potential for broader impact but are never discovered by a larger audience. They are lost in the shuffle. Others gain citations for reasons other than their intrinsic value, e.g., because the author is well-known or is an especially effective promoter of her work. For these reasons, it seems important to try to measure the breadth

[11] Hal Varian (1997) relates a story: "It is said that a ghost named 'Pedro' haunts the bell tower at Berkeley. The undergrads make offerings to Pedro at the Campanile on the evening before the exam. Pedro, it is said, was a graduate student in linguistics who wanted to right his thesis on Sanskrit. In fact, it was a thesis about one word in Sanskrit. And, it was not just one word, but in fact was on one of this word's forms in one of the particularly obscure declensions of Sanskrit. Alas, his thesis committee rejected Pedro's topic as 'too broad.'"

of a paper's appeal ex ante as a clue to a paper's potential impact ex post. The MM also uses both ex ante and ex post measures, such as citations, allowing it to adjust initial estimates as more information about a paper's post-publication reception becomes available. Readers looking for work on a topic with broad appeal will have a mechanism for identifying that work.

In these respects, the Consortium offers a counterbalance to the forces of specialization in science, making sure that integrative work is recognized.

Informativeness

Under the current journal-centered system of evaluation only one piece of information about the quality of a paper is available at publication: the status of the publishing journal. However, the status of a journal, whether measured by "journal impact factors" or overall reputation, is an extremely crude indicator of the quality of a particular article. Note that the decision to publish involves a great many parameters. A paper may have been accepted because it addresses a new question, develops a new theory, applies a new method, employs a new set of empirical data, has a strong (quasi-experimental) test, is deemed appropriate for the journal (along whatever parameters the editorial board judges these things), or some combination of the above. Thus, the binary decision to publish is not very informative, and journal rankings are hard to interpret and open to abuse (Heneberg 2016; Larivière et al. 2016).

In the Consortium, editorial teams are likely to function in much the same fashion as journals now do. (Indeed, they may be run by the same editorial teams that formerly operated as independent journals.) Teams will craft distinctive identities, with specific methodological and substantive preferences. They will attract varying numbers of submissions, and from these we can calculate the preference-orderings provided by submitting authors (1-5). They will have varying rates of desk-rejection and post-review rejection (though the latter must remain below 20%). And they will have varying citation counts. All of this will be calculated in a uniform fashion (so that statistics are comparable) and will be readily available so that end-users have full information upon which to make judgments about journal quality. Indeed, this information will be more complete and more reliable than that which currently is available about journals, who have been accused of ramping up their journal impact factors by hook and by crook (e.g., by pressuring authors to cite papers published in their journal or by excluding certain categories of article from their impact factor).

In addition, the Consortium offers a wealth of information about each published article. Various dimensions of evaluation are distinguished (§IV) and a 0-10 score is assigned to each. Along with raw scores, the Consortium offers scores adjusted by the MM, which corrects for some common biases, as discussed. A confidence interval expresses how certain that estimate is. Editorial teams may also decide to add their scores and comments, adding another layer of information.

End-users can choose to focus on a particular dimension(s) of particular interest. For a meta-analysis, one might restrict the sample to studies with strong internal validity. For a literature review focused on theoretical developments in a field one might limit the sample to studies that are judged to be theoretically innovative, and which the MM situates within a key topic or scholarly community of interest. And so forth.

To skeptics, these efforts may appear as yet another doomed venture in academic metrics (Espeland, Sauder 2016; Muller 2018; Osterloh, Frey 2015; Wilsdon et al. 2015). However, those who don't like, or don't trust, the metrics offered by the Consortium have another option available. They may ignore them entirely – focusing instead on the reviews themselves, and the author's responses, which (unlike the traditional review process) are made public. They may also follow the thread of post-publication commentary contained in the digital community space. Words can replace numbers, for those who believe that this is a more informative way of judging quality.

In summary, for purposes of assessment – whether it is at the level of an individual article, an individual scholar, an editorial team, or a genre or research area – the Consortium offers considerably more information, and more systematic information, than the current protocol.

Post-publication assessment

Although pre-publication peer review plays an indispensable role in scholarly research, it cannot be expected to provide a complete assessment of the value of academic work. Some aspects of this assessment must be ex post (after publication). This includes (a) replications, (b) deliberation by the scholarly community, and (c) scholarly impact.

Replications are encouraged by the Consortium's low bar to publication, which should allow any replication of a published work to appear as a stand-alone publication – so long as it is implemented in a rigorous and non-prejudicial fashion. The Consortium could also provide a uniform way to link replications to the publication that they replicate, and one another.

Ongoing deliberation is encouraged by the Consortium's digital community space. Note that the current journal-centered system offers no organized and accessible venue for commenting on published work after it is published. Commentary appears in some academic blogs, sometimes with attribution and sometimes without. More commonly, it occurs in an evanescent fashion in emails, hallways, and classrooms. There is no permanent or public record, so all of this deliberation is lost – and much of it, we suspect, is not very deliberative. Moreover, the author has no way of responding; indeed, she may be entirely unaware.

The Consortium's digital community space would allow anyone to add their voices, with real-name attribution based on Consortium membership, to the discussion of a published paper, providing a focal point for ongoing debate and discussion, allowing the scholarly community to digest, critique, and extend the points made by the original author and reviewers of the manuscript. This process recognizes that social science is sometimes flawed, always uncertain, and always open to improvement. By providing a space for further reflection we allow the process of evaluation and deliberation to continue indefinitely in a fully public forum.

Measurement of scholarly impact is also facilitated by the Consortium. Note that one of the advantages of a unified, open-access repository of articles is that one can monitor the number of times each paper is accessed and downloaded.

Of course, the Consortium is not intended to provide all aspects of scholarly assessment. We recognize the ongoing importance of conventional measures of impact such as citation counts (Gerring, Karcher, Apfeld 2020). We also recognize the important role of prizes (e.g., those issued by APSA sections), blog posts (e.g., those that highlight certain streams of research), and other forms of recognition. We can envision a world in which ratings issued by various scholarly bodies identify what they view as important work (Nosek, Bar-Anan 2012: 232). In much the same way that interest group ratings of members of Congress provide a signal to voters, scholarly bodies with varying substantive agendas and methodological tastes might offer signals to end-users about the value of studies published by the Consortium.

In summary, there is a lot more that we can do as a scholarly community to provide signals about published work that will help the broader community – especially those who are not specialists – to find what they are looking for. The Consortium offers a step in that direction, but it is surely not the entire solution.

VII. BROADER RAMIFICATIONS

In this section, we consider the broader ramifications of the Consortium. This includes (a) the time required for editors, reviewers, and authors as part of the process of review and publication, (b) the role of reviewers and editors in the review process, (c) the volume and diversity of papers that reach publication, (d) the achievement of meritocratic standards that can be applied internationally, (e) dissemination, (f) employment decisions (e.g., promotion and tenure), (g) ethics, and (h) learning about the production of knowledge in political science.

Time

The process of peer review and publication is time-consuming for all concerned – authors, reviewers, editors. Under the current system, the average article is submitted to multiple journals before it finds a home (if it finds a home). Each of these submissions imposes a laborious process of review, sometimes with multiple rounds. Often, it takes several years for a manuscript gain entry to a journal and some additional time for it to be published. Papers that never find their way to publication also consume a good deal of time if they have gone through the review process at several journals.

By setting a low bar to publication, the Consortium limits the number of submissions to one (though there is a possibility of resubmitting a desk-rejected manuscript if the Oversight committee offers that option). Our assumption is that papers that cannot pass this low bar are probably not worth pursuing and authors will not bother submitting them to journals.

After acceptance at the Consortium, a paper receives nine reviews – many more than the two or three that is common for journals. However, three of those reviewers are asked only to answer a single question about the disciplinary appeal of a paper; this can be done after skimming the abstract, introduction, and conclusion, a five-minute task. Furthermore, since each paper receives only one draw from the reviewer pool, the average number of reviews per paper is fewer than under the current system.

Editors must enlist reviewers. However, we anticipate that this time-consuming process will be streamlined as the Consortium database offers suggestions for who might be qualified and available, and incentives for them to accept.

Editors can choose how engaged they wish to be in the review process, e.g., whether they wish to write one of the reviews that accompanies the manuscript. If an editorial team wishes to maintain a highly selective acceptance rate they will need to spend a good deal of time agonizing about which papers to accept and which to reject. That is their choice. In either case, they do not need to write a decision letter if they do not wish to do so. For editorial teams with a less selective approach, not much work is required in order to process a manuscript.

Now, let us consider the time-line from initial (first) submission to publication. Here is what we envision:

Day	Action
1	Submission.
7	Accept or reject. If the former, send to editorial team.
10	Issue invitations to reviewers.
70	Reviews complete.
77	Author resubmits. Revised article and responses sent to reviewers.
137	Reviews complete.
144	Author submits responses to reviews. Article sent to copy-editor.
151	Copy-editor completes work. Article sent to author for responses.
155	Author completes responses. Article sent to type-setter.
160	Type-setter completes work. Proofs sent to author.
165	Author completes review of proofs. Proofs sent to type-setter.
170	Article published.

Admittedly, this is an optimistic scenario. Sometimes, reviewers do not complete their assignments in a timely fashion. Sometimes, they drop out entirely – in which case the managing editor will need to decide whether it makes sense to try to recruit a replacement or go with the reviews in hand. Sometimes, authors take a long time revising prior to resubmission. These holdups are similar in the traditional journal-centered system of review.

Hamermesh (1994: 158-9) finds that the primary cause of delays in the peer review process stem from reviewers who accept the invitation to review but do not fulfill their obligation in a timely fashion. Since the Consortium process starts with six reviewers, editors have more room for maneuver. They may decide to complete the review process with five reviewers if one is AWOL. Little is lost so long as the five reviews are conducted in a conscientious fashion and are not widely disparate in their opinions of the work under review. Similarly, the system can reward timely reviews, or punish late ones. If we treat reviewers who miss deadlines identically to those who refuse to review in the first place, or even more punitively—say by treating a late review as two declines—we could strongly incentivize scholars who wish to publish with the Consortium respect their review commitments. More lenient incentivization approaches could still provide substantially better, and clearer, incentives for timely review than the current system.

In any case, we expect that the timeline envisioned for the Consortium – from initial submission to publication – to be substantially shorter than that experienced at most traditional journals. By limiting the number of rounds, by using algorithms from the database to identify reviewers, by employing incentives possible only with a unified system of review to motivate reviewers, and by publishing as soon as a paper is finalized, the timeline is shortened.

Most important, multiple rounds of submission and rejection are avoided, saving time for all concerned. Note that the timespan of greatest significance is that which elapses from *first* submission to final publication.

Peer review as collaboration

The current review system is oriented toward reaching a decision about whether or not to publish. Editors and reviewers for the journal must guard the journal's reputation by rejecting sub-standard scholarship, preserving scarce space in the journal for only the highest quality work. Top journals in the social sciences reject more than 90% of submissions. Naturally, an adversarial relationship develops in which authors clamor at the gates while editors and reviewers poor hot wax from the ramparts. Frequently, reviews consist of "things that the author got wrong" or "ways that the study is deficient."

Under the Consortium, the goal is to publish everything that adds something to the sum-total of human knowledge. Accordingly, the bar for desk-reject is set fairly low and everything sent out for review is published (unless the author and a majority of reviewers decide otherwise). Moreover, there is no obligation for an author to kowtow to reviewer suggestions that she does not agree with. It is the "author's cut" that appears in print.

The role of the review process is therefore quite different. Editors and reviewers offer advice to the author so that the paper can be improved. Editors and reviewers, who presumably have strong ties to the field or subfield in which the paper falls, have an incentive to make each paper as strong as it can be so that the field can move forward. This also prompts reviewers to focus not only on absolute standards (e.g., Is there a high level of internal validity?) but relative standards (Is internal validity as high as could reasonably be achieved given the research question of interest?). It is up to the author whether or not to accept advice received from the review process.

Not only is this process more pleasant, it may also have positive repercussions for the quality and quantity of work produced in the discipline. Note that people are generally more responsive to positive cues than to negative cues. This is especially likely to be the case for those who have secured tenure or who work as independent scholars or in other positions where they have no strong incentive to publish. To incentivize these scholars to maintain an active research profile the discipline needs to offer positive reinforcement (where warranted). Authors should feel they are making a contribution to scholarship and given due recognition for their efforts.

Our willingness to publish everything that is not desk-rejected is not simply a matter of protecting authors' self-esteem. It is also, we believe, a better incentive structure for producing useful knowledge. In the current system, reviewers play a determinative role in the decision to publish. In this capacity, they can make demands that an author is effectively unable to refuse if she wishes to publish in that journal (which we shall assume is an important step in her career). The reviewer thus occupies the position of a veto-holder, especially at top journals where a strong signal to reject a manuscript, even if only from one reviewer, may doom that manuscript.

There are four potential problems here. First, the reviewer probably does not know as much about the subject and the data as the author. Second, since the reviewer has no skin in the game she may not have thought very deeply about the subject at issue. Third, the reviewer's interests (in protecting her turf) or pre-commitments (in a particular theory or method) may impair her ability to judge a matter dispassionately. Fourth, the reviewer may offer a suggestion for revision that fundamentally distorts the author's work. Preston McAfee (2014: 61), long-time editor of the *American Economics Review*, attests…

> Referees [sometimes] offer specific advice that push papers away from the author's intent. It is one thing for a referee to say, 'I do not find this paper compelling because of X,' and another thing entirely to say that the referee would rather see a different paper on the same general topic.

Authors sometimes find themselves in a compromising situation. Frey (2003: 206) writes,

> Survival in academia depends on publications in refereed journals. Authors only get their papers accepted if they intellectually prostitute themselves by slavishly following the demands made by anonymous referees who have no property rights to the journals they advise.

Arguably, the end-result is that the quality of papers are worsened, or at least watered-down. McAfee (2014: 61) argues, "This system encourages authors to submit papers that are crafted not for publication but for survival in the revision process."

We believe that leaving the final prerogative in the hands of the author is the best way to advance the cause of truth. The important caveat is that reviewers may exercise their right – indeed, are under an obligation to exercise their right – to assign the scores they feel are appropriate to the manuscript and to explain themselves in prose. If reviewers believe the author has made a mistake, or that a work is of low quality, they should say so. This way, disputes reach the public eye, where they can be further adjudicated.

Volume and diversity

Lowering barriers to publication would presumably stimulate the publication of more studies, about more topics, and with a wider diversity of approaches, than presently appear in the annals of political science.

First, more studies with unexciting findings would be published, combating the current tendency to discriminate against work with null results or results that confirm standard wisdom. This would go some way toward combatting the file-drawer problem, giving us a less biased knowledge base upon which to gauge the probable truth of various hypotheses.

Second, more studies with speculative arguments and findings would be published, combating another supposed tendency of the current journal-based system of review, which is alleged to discriminate against unconventional ideas and ideas that contravene reviewers' theories. This should make it easier to publish work that is at an early stage of development, and this in turn should speed the progress of science. Vale (2015: 13441) points out,

> there is…merit in getting new ideas and key experiments published with reasonable speed, even if they are incomplete. Once in the public domain, the collective power of the scientific enterprise can take effect, and the ideas can be tested and advanced further, not only by the original researchers but also by other investigators as well. Once results are published, other scientists can see connections with their own work, perform new experiments that the original investigators might never do, and also emerge with new ideas. Overall, putting new results and ideas in the public domain is good for science and serves the mission of the funding agencies that seek to advance research overall.

Vale's immediate concern is the field of biology, but these observations seem to extend equally to the social sciences.

Third, more studies of non-traditional subjects might be entertained. As it stands, political science journals impose boundaries on what they consider to be appropriate topics for their journal. If there is no journal in place that considers your topic topical, you are out of luck. Since "politics" is a hard thing to define, there is a certain arbitrariness to these boundaries and we have a sense that they may be excluding some interesting and important work. By contrast, economics journals impose no topical boundaries on what may be published (though they do seem to impose theoretical and methodological boundaries on how those subjects may be approached). In any case, the Consortium would presumably offer opportunities to expand the definition of politics, and perhaps even to undertake subjects that are not considered political at all but nonetheless important. Among such non-traditional subjects are papers that attempt to synthesize knowledge about a subject (Gerring 2020), papers that are purely theoretical (Clarke, Primo 2012), papers that are methodological in a non-standard sense (e.g., they are about qualitative methods, epistemology, or the production of knowledge), papers that introduce a new dataset or a new set of qualitative empirical materials, papers that are descriptive rather than causal, reports from the field (e.g., a recent election in some country around the world), papers that address a lay audience (e.g., policymakers), and papers whose goal is primarily didactic (e.g., directed at students of political science).

Fifth, space limitations on what might be considered eligible for publication would presumably be relaxed (Gerring, Cojocaru 2020). In the Consortium, all articles are published on-line, so arbitrary limits do not arise from the obligation to print and mail hard copies. Copy-editing and type-setting are expensive, but these logistical elements of the traditional publication process may be dispensed with, as discussed in Section IX. One thing that cannot be dispensed with is the willingness of reviewers to spend time on a manuscript. If it is too long, they may demur. In any case, the issue of length limits would be left to editorial teams, which may choose to impose limits, or not. We can imagine that teams might take different approaches to this question, with some being amenable to longer submissions. Some teams might even specialize in book-length publications.

Before quitting this subject it is important to acknowledge the downside of increasing the volume and diversity of published work: the Consortium would publish some studies that add nothing, or perhaps even detract from, our knowledge of the world.

How serious would this problem be? It would be a serious problem if authors aim only for quantity, i.e., maximizing the total number of articles published (e.g., by chopping up big ideas into little ones and/or repeating ideas and analyses across several, highly redundant articles). We don't think this is very likely, however.

First, the Consortium allows editors to view all the author's prior work and subjects each new work to an algorithm that should alert editors to situations of serial publishing. By contrast, the current system of journal review is highly fragmented, and editors and reviewers often unaware if an author submits papers to multiple outlets that are similar in content.

Second, the Consortium offers strong signals about quality. These take the form of published reviews as well as reviewer scores. Insofar as these signals matter, authors are incentivized to publish high-quality work and to enhance and preserve their mean score across all articles published in the Consortium. It is our impression that quality matters more than quantity in the current system of scholarly assessment (e.g., for hiring, promotion, and salary adjustment); we see no reason to suppose that this trend would change under the Consortium.

Even if there *is* more debris in the Consortium than in the current journal-centered system of peer review and publication we do not see this as a major problem. In the production of knowledge suppressing true claims seems like a more serious flaw than admitting false or redundant claims. We would prefer errors of commission to errors of omission.

Of course, this depends upon our ability to sort through the pile of published work, separating the wheat from the chaff. The scores accompanying each published article should help with that, as will post-publication assessments (addressed above).

Meritocracy and internationalization

The current system of journal-centered peer review and publication probably favors established authors and authors with ties to top journals. It very explicitly favors articles published in top journals, meaning that studies published in lower-ranked journals may not receive the attention they deserve.

In making these criticisms we are not advocating for an egalitarian system in which all articles are assumed to have equal merit. Rather, we are advocating for a system that gives every article, and every author, an equal shot – a publishing meritocracy.

Let us begin with the current hierarchy of journals. Although this hierarchy may be partially reconstituted as a hierarchy of editorial teams within the Consortium, they would be forced to compete on terms that are more nearly equal.

First, there are no distinctions between print and on-line journals. Likewise, the appearance and availability of articles published by editorial teams within the Consortium are identical. The only

distinguishing characteristics are the titles of the editorial teams, along with the names of the editors, reproduced on the front page of each article, and rejection rates.

Second, let us consider the reviewing pool, an issue researched by Hamermesh (1994) for the field of economics. Currently, economics journals rely heavily on authors who have published in their pages. Top journals are able to draw on the talents of reviewers occupying the top rungs of the economics discipline, which presumably enhances the quality of their published work, while lower-ranked journals must depend upon the talents of lower-ranked reviewers. Journal editors give preference to steady reviewers in their choice of which manuscripts to publish. As an ecological system, these factors work to enforce the Matthew Effect: to those who have, more shall be given. We have no reason to believe that things are substantially different in political science, though it is true that top journals are not monopolized by top departments as they are in economics.

Under the Consortium, an algorithm offers suggestions for who might serve as reviewers for a paper. Editors are free to disregard these suggestions, but at least the algorithm is designed not to replicate standard editorial biases. (One can also check, over time, to see if some editorial teams are more likely to deviate from the algorithm than others, a tip that they might be tilting toward a favored in-group.) Reviewer invitations are sent from headquarters, with no indication of the editorial team, so reviewers cannot pitch their reviews to a particular editorial team. Moreover, incentives established for reviewers do not give them much leeway to reject assignments they do not like. Accordingly, we can anticipate that the services of top reviewers will be available on a more equal basis to all editorial teams and it will be harder for top editorial teams and top reviewers to sort.

Finally, let us consider the advantages of established authors – those with name recognition in the discipline, usually those with senior status in a highly ranked department. Under the current system, these authors seem to be treated more favorably by editors. This may be because they know each other, or because editors feel pressured to boost their journal impact factor (JIF) and their overall reputation. High-profile authors are more likely to be cited in subsequent work. Under the Consortium, editorial teams do not need to worry unduly about the citation counts of articles published under their purview. The business model is quite different. One can hope that the citation counts of articles displaces the citation counts of editorial teams as a marker of success. (Surely, it is more important to have written a highly cited article than to have published with a highly cited editorial team.)

Human capital

A core objective of any system of peer review and publication is to maximize the use of human capital across a discipline. Human capital is a scarce resource, and perhaps especially so in the field of political science. If one were to count the number of individuals with PhDs in political science and related fields (e.g., international relations, international studies, public policy) it would fall considerably short of the number of PhDs in economics, psychology, biomedical sciences, or engineering (NSF 2015). Likewise, if one were to count the number of individuals employed in jobs where political science research is valued and rewarded it would be a much smaller number than in most other fields. For the most part, research in political science is conducted by political scientists with teaching positions within universities. By contrast, in many other fields there are professionals working outside academe who conduct ongoing research on some aspect of their PhD subject, e.g., economists working for banks, corporations, or governments, or chemists working for pharmaceutical firms.

To enhance progress in political science we must make the best use possible of the limited human capital available. It should be emphasized that we do not view political science as a discipline where progress can be achieved by a few geniuses or a small cadre of great minds working at elite institutions. Creativity is hard to identify, so the more minds are enlisted in thinking about a task the

more opportunities there are for fundamental breakthroughs. Likewise, progress is not simply a matter of coming up with good ideas; those ideas must be tested in a rigorous fashion, which entails iterated studies in different settings using standardized protocols (Dunning et al. 2019). Cumulation of knowledge cannot occur without numerous replications.

It follows that the machinery of peer review and publication must be oriented toward mobilizing a veritable army of scholars. So far as we can tell, the current system is not up to the job. The traditional system of journal-centered publishing operates with a *guild* format. Editors, reviewers, and authors generally know each other personally. Often, they are educated in the same universities and attend the same annual conferences. Because most professional activity revolves around the top journals and the top departments, consensus is likely to develop about the worthiness of a manuscript prior to publication through a series of workshops, conferences, and talks. Once it goes under review, the editor can call upon her network of friends and colleagues for reviews, and lean on recalcitrant reviewers who don't carry their weight or don't perform their duties in a timely fashion. Small networks have their value.

However, the guild format also contributes to an insider/outsider cleavage that undermines the credentials of a supposedly meritocratic discipline and minimizes incentives for those on the periphery, who may not have equal access. Likewise, as the field of political science grows, and as more and more academics embrace the goal of research, the system of peer review and publication must be able to accommodate an increasing flow of manuscripts. Most of this growth is likely to occur outside the traditional bastions of North America and Europe.

In this growing and diversifying universe the guild model breaks down – not simply because of the numbers problem but also because people living on different sides of the world are unlikely to know each other or to have repeated interactions with each other via email and skype. Coordination through informal channels will not work with such a large and diverse population.

By contrast, the Consortium, which handles most tasks algorithmically, should be able to handle the growth and internationalization of political science. Likewise, new editorial teams are easy to form, while their administration and their output can be monitored. There is less risk emanating from "dark corners" of the publishing world. Additionally, because tasks are handled through a web-based system in the Consortium, face-to-face meetings and skype meetings are unnecessary, rendering the problem of transport and communication across time-zones irrelevant – or nearly so (some editors may prefer to chat in a live setting rather than correspond).

Now let us consider the alternatives. If the traditional journal system is expanded to encompass a truly global constituency one can envision two scenarios. In one scenario, journals are founded in every country or region of the world, replicating each other's missions. Such a system would be unwieldy and its output difficult to judge. What is one to make of an article published by a *Nigerian Political Science Review*? In another scenario, existing journals – headquartered almost exclusively in the US and Europe, and bearing names highlighting their parochialism (e.g., *The American Political Science Review*) – would try to integrate scholars from all over the world. This does not seem very plausible. For one thing, these journals are already overstretched, and show no signs of interest or capacity in expanding their pages to accommodate an increased flow of articles. Nor are there signs that these vehicles can accommodate editorial teams from around the world, although the APSR has made a gesture in this direction.

Dissemination

Under the current system, access to journals is restricted to those with a university affiliation or a very good public library. Under the proposed system, access is unrestricted. Being more accessible, the Consortium should enhance the influence of political science, making it a truly public endeavor. It

should also help to internationalize the discipline of political science, making publications available to those in poor countries and in areas distant from universities and public libraries. All that is necessary to access materials in the Consortium is an electronic device that connects to the internet.

Dissemination of published work could also be made more efficient, even for those with access to gated journals. Under the current fragmented system, hundreds of journals may publish work related to your area of interest – not to mention an even larger collection of working papers and web sites. Consequently, it is difficult to stay abreast of developments.

In a unified system like the Consortium, which publishes everything that passes a low threshold, it should be much easier to follow the flow of knowledge – even if that flow is more like a hydrant than a spigot. To manage the flow, end-users could specify specific areas of interest, specific editorial teams, or articles that cite their own work. Algorithms could be developed that suggest work that might be relevant to them.

End-users could request immediate notification (by email) of publications that pertain to their chosen topics or scholarly communities. Alternatively, relevant publications could be stored in one's profile (if one is a member of the Consortium), to be accessed whenever convenient.

Personnel decisions

The Consortium would presumably have some impact on personnel decisions – hiring, promotion, tenure, and salary. The major difference relative to the current system is that the Consortium offers a great deal more information for gatekeepers to consider.

The traditional journal system offers only one bit of information – the reputation of the journal where a paper is published. Under the Consortium, committees could judge the reputation of the editorial team that publishes each paper. But they could also examine the scores – raw and adjusted – received by a paper along various dimensions (§IV). They could easily examine where papers, and their authors, sit within the network of disciplinary communities and topics. And they could read the reviewer reports and the author's response. They could also read the report issued by the editorial team, if any. This should give committees a good sense of a candidate's strengths and weaknesses, and her contribution to the field. Impact may be measured by traditional citation counts and also by page views and downloads from the Consortium system.

Committees who wish to evaluate a candidate's willingness to provide public goods could examine the candidate's reviewing record – the number of reviews she has conducted, the number of times she has declined an invitation to review, and her reviewer score (§III, "Motivating reviewers").

Clearly, the Consortium offers a lot more information to gatekeepers attempting to decide whom to hire and promote, and how much to compensate them. Importantly, this information is *comparable*. The MM-adjusted scores for one article can be compared to the scores received by another article, at least at the level of a given scholarly community or topic. And because written comments from reviewers are obtained in a process that is identical across the Consortium (reviewers do not know which editorial team they are reviewing for), their reviews are also comparable. By contrast, it will never be entirely clear how to weight the value of publications in different journals. Is an APSR worth two PRQ's? Or three? Or one? These are not answerable questions.

Finally, because papers wend their way through the review process much more quickly in the Consortium than in the traditional journal-centered system there is a longer record to peruse. Note that a quick publication schedule means that citation counts also accrue quickly. Once a paper is published other published papers may appear with citations to it in the succeeding year or two. By contrast, in the traditional journal-centered system it takes multiple years for a citation record to become meaningful as a measure of impact. Accordingly, academics at the beginning of their careers do not have a track record that can be evaluated.

Learning about the production of knowledge

There is a lot we don't know about the production of knowledge in political science (Elman et al. 2020). Is there demonstrable scientific progress? Where is progress most marked, and across which dimensions? To what extent do different subfields share methodological values (as revealed by their judgments about specific papers)? How consequential are the epistemological divides? More generally, how should we understand disagreement among scholars in the evaluation of a study? How frequent is this disagreement, and how severe? What sort of studies elicit the most agreement, or disagreement? Is consensus (or dissensus) growing over time?

The Consortium provides a database with which these – and many other – questions might be evaluated. Naturally, there are important research design issues to resolve, and some problems of descriptive and causal inference may not be fully resolvable.

Even so, we suspect that the Consortium would offer fertile ground for addressing some of the philosophy of science and history of science questions that bedevil the discipline. This research may suggest some practical ways in which research productivity – and scientific progress more generally – could be improved.

It may also feed back into improvements in the peer review process. Jefferson et al. (2007: 12) laments, "Given the widespread use of peer review and its importance, it is surprising that so little is known of its effects." Tennant et al. (2017: 7) notes that we don't know very much about the effectiveness of peer review, "as the data required to [understand it] are often not available." Accordingly, it is "very hard to measure and assess the quality, standard, and consistency of peer review not only between articles and journals, but also on a system-wide scale in the scholarly literature." To fully exploit the potential of the Consortium, one could ask permission of members to serve as subjects in periodic experiments where various features of the review process are altered in order to gauge the effects.

VIII. IMPLEMENTATION

If readers are convinced that the Consortium offers an improvement over the current system of peer review and publication we must ponder the practical question of how such a system might be implemented. Suboptimal equilibria are often highly resistant to change. However, there seems to be considerable frustration with the status quo, as outlined in §I. This may augur for far-reaching change.

In this section, we consider several practical issues. First, there is a question of how the proposal might be greeted by political science journals and by the political science community. Then, there are challenges of initiating such a wholesale change in existing, long-established practices. Finally, there are financial problems to be dealt with; the Consortium must have a viable business plan.

Reception by journals

Political science journals may view this proposal as an onslaught; after all, it questions their continued existence. An alternate – and in our view, more accurate – reading is that it is an invitation for journals to continue doing their job within a new umbrella. Note that any group may form a new editorial team at any time. So, if a journal wishes to join the Consortium it may do so, bringing along its name (which may have to be altered somewhat for copyright reasons), its editorial team, and its constituency of authors, reviewers, and readers.

A journal integrated in this fashion into the Consortium is relieved of many of the costs involved in maintaining a journal while continuing its work in much the same fashion as it has in the past. This means that the intellectual continuity of the journal is preserved and can be extended into the future (so long as the flow of submissions does not dry up). Although commercial publishers are sure to oppose a move into the Consortium, the staff who conduct the intellectual work of these journals (i.e., their editors, editorial board, reviewers, and contributing authors) – not to mention the readers (who no longer have to pay for access) – may view the proposition positively.

Since the role of editors in the Consortium is very similar to their role in the traditional journal, we expect that whatever factors motivate political scientists to provide this public good would continue to operate. Let us underline a few key points.

First, although editors in the Consortium work for an editorial team rather than an independent journal the team would have a name, an identity, and a mission – just as journals do. So, if titles matter – and surely they do – being an editor should bring honor and prominence to the bearer of that title. Second, editors have less busywork (organizational and logistical work) than editors of a traditional journal, as these tasks are handled by the Consortium. Third, editors spend less time identifying reviewers and chasing recalcitrant reviewers, as the Consortium database offers suggestions for who might serve as good reviewers and incentives for them to accept and complete their tasks in a timely manner. Fourth, editors retain the ability to accept or reject articles; they can control their brand.

Finally, although the Consortium would not have money to compensate editors (at least not initially), many sources of funding are fungible. Journals, or individual editors, who receive support from their university, from a scholarly association (e.g., APSA), or from a private foundation, could carry that support into the Consortium. Relieved of many of the costs of publication, funders may have extra funds that can be redirected to other uses (e.g., increased stipends for editors or staff).

Granted, a few editors receive funding from commercial presses, and this money would no longer be available if the journal reconstitutes as an editorial team. It is our hope that universities would redirect some portion of the money they save from purchasing journals to the upkeep of the Consortium, as discussed below; if so, some portion could be allocated to editors. An equitable compensation system would presumably rest primarily on the volume of submissions each team handles (as part of a full review). The team could be in charge of divvying up the money among editors and staff.

Reception by authors

For authors, the Consortium offers a quick path to publication (for all papers that pass the threshold of acceptance). It also offers a larger set of reviews than is common in the traditional peer review process – six rather than two or three. We have every reason to believe that the quality of those reviews will be as good as, or better than, those that might be received through the current journal-centered process. After all, it is the same people doing the reviewing, and they have stronger incentives to accept invitations to review and to conduct conscientious reviews than they do under the current system.

There will inevitably be situations where author and reviewer disagree – perhaps because the reviewer is a priori ill-disposed to the findings, the theory, or the methodology of a paper, or because she does not understand the paper, which may in turn be because she has not taken the time to read it carefully. Under these circumstances, the aggrieved author may wish to withdraw her paper and start anew with a different set of reviewers. This is akin to the traditional review process, where the author retains a right of refusal and where negative reviews usually result in the journal rejecting a paper. The problem with this procedure is that papers must then be reviewed multiple times (representing a loss

of efficiency) with the quality of reviewers declining (on average) with each iteration, and some papers that should see the light of day never do (a file-drawer problem).

From the author's perspective, the redeeming feature of the Consortium is that she can control the product – her article. She also has an opportunity to respond to reviewers, who must explain themselves and must respond to the author's counter-critiques. This process should force reviewers to reconsider in situations where they cannot defend their position. And in situations where disagreements persist, they are presumably *real* disagreements – not signals of ignorance or laziness on the part of the reviewer. As such, it is appropriate to air them publicly so that the community of scholars can decide for themselves. If the author is sure of her position, she should appreciate the opportunity to put her arguments before the public.

Reception by various groups within the discipline

Although it is premature to evaluate the response of the political science community to this initiative, it may be helpful to anticipate some of the interests and values at stake, which are likely to color the way in which particular subsets of the discipline respond.

Let us begin with those who thrive under the current journal-centered system of review – political scientists with successful publication records in top journals. This subset of authors presumably has the most to lose by any change to the status quo. However, it is not clear that they will lose anything at all. It seems likely that the current system of hierarchically ranked journals will be reproduced within the Consortium in the form of ranked editorial teams, who exercise their discriminating taste by rejecting manuscripts at a high rate. There is no reason that APSR, AJPS, and JOP cannot reconstitute themselves as editorial teams, for example. Alternatively, because there are no space constraints on how many articles a single team can publish, a single large editorial team (e.g., an "APSR" with an expanded quota of published articles) could perform the function of all existing general-purpose journals.

Let us now consider the reception of the Consortium for those who have been less successful under the current publishing regime. For this group, who have struggled to get their work published – or at least, published in the top journals – the current initiative may be welcome. Note that the low threshold for publication in the Consortium is intended to reach out to all work that offers some value. Note also that editorial teams may organize along any sort of methodological or substantive dimension, offering hope for authors who feel that their brand of work is not valued by existing journal outlets. Note, finally, that all papers are published in the same on-line system – with the same type-setting, formatting, and level of accessibility. Invidious comparisons among journals would be somewhat diminished. This should serve to focus attention on the content of published papers, where it belongs, rather than the venue.

As a third consideration, let us address epistemological differences. Some aspects of Consortium probably have greater appeal to those with a "positivist" orientation (defined here as the view that social science ought to aim for the same goals as natural science). However, the Consortium may also have much to offer those who adopt a different vision of the study of politics. It is fair to say that political theorists and those with an interpretivist or post-structuralist bent are marginalized in the current journal-centered system of publication. To be sure, they have their own journals, and they have a small bit of real estate in some general-interest journals (depending upon its editorial team). However, there are many fewer venues for this sort of work than for work with a positivist orientation. Moreover, the balkanization of the field into different journals reinforces existing epistemological divides. Insofar as marginalized research communities aspire to a broader hearing they might welcome a system that places their work on equal footing.

Finally, let us consider the impact of the Consortium for those at the beginning of their academic careers, i.e., PhD students and assistant professors. The current system of journal-based peer review and publication is a trying one for those facing tight deadlines. Nowadays, doctoral students are expected to have at least one paper accepted for publication prior to applying for top jobs. This is difficult, as most doctoral students do not have a publishable paper until the final years of their graduate program, on the eve of their launch into the job market. As we have observed, it takes a long while for a paper to wend its way through the publication process, especially if it is not accepted at the first venue (which is normal). Assistant professors must assemble their tenure file in five or six years, which is not a lot of time to work through a second major project (after their dissertation). Typically, papers do not count for much unless they have received the imprimatur of a press. Thus, both graduate students and pre-tenure professors face a tough choice between starting at the top and working their way down the journal hierarchy – which maximizes their chances of landing a spot at a prestigious journal – and sending their paper to a journal that is more likely to publish it (without multiple rounds of reviews).

In this light, the advantage of the Consortium is pretty clear. Any paper that would be published in an old-style journal would most likely clear the desk-reject hurdle and would therefore become published in a highly compressed time-frame (§VII). A few weeks from submission the paper could be listed on a CV as "forthcoming." Reviews and scores would appear several months later. And there is no need to face an agonizing choice between maximizing and satisficing.

Granted, there are some costs for early adopters. We can anticipate a period of adjustment in which gatekeepers (hiring committees, tenure committees, and the like) do not know how to regard publications emanating from the Consortium. But if the Consortium hits the road running, with a large volume of work published in its first year, it should become apparent pretty quickly how to evaluate its products. This leads into our next topic.

Startup challenges

The startup challenge for our proposal is that it cannot be instituted properly in a piecemeal fashion. Until high-quality work appears regularly in the Consortium, and until established senior scholars adopt it, the Consortium is likely to be regarded as a low-quality venue, a last resort for papers that have been rejected by traditional venues. Worse, there are costs for those who submit high-quality work to a (perceived) low-quality outlet: their work will not receive the attention it deserves and their career may experience untoward consequences. The initial incentive structure is highly unfavorable.

To solve the startup challenge, and to deal with adverse selection, we offer two possibilities.

The first involves an on-line petition. Signing the petition commits the signer to send future papers (for three years following the official start-date) to the Consortium and to conduct reviews of papers submitted to the Consortium. The inauguration of the Consortium is postponed until at least fifty percent of APSA members sign the petition. If this point is never reached, no action is taken; effectively, the discipline will have voted against it. In this fashion, no time, money, or reputations are wasted. But if a critical mass is reached, the Consortium can move ahead. And with demonstrated support from half of APSA's membership it should be possible to motivate potential sponsors (including APSA itself).

The second approach focuses on existing journals. To operate effectively, the Consortium would need to integrate a large number of smaller journals or a few large journals. One could establish a threshold – a lower bound above which we believe that the Consortium could realize economies of scale and offer added value – and then wait to see whether enough journals sign up to warrant the venture.

These two approaches could be pursued simultaneously, with the impetus to move forward triggered by passage of one or the other benchmark.

Once the Consortium is initiated it would be in direct competition with established journals. To ensure that researchers do not employ the Consortium as a dumping-ground for low-quality work one might ask authors who submit to the Consortium to forswear submitting to journals (except in the unlikely event that their submission is first rejected by the Consortium).

Over time, if it can be demonstrated that reviews garnered through the Consortium are as useful and on-point as reviews garnered through the traditional journals, and if it can be shown that publication through the Consortium brings as much acclaim and attention as publishing in traditional journals, there would be no reason to bother with sanctions. Presumably, the greater availability of open-access publication will enhance citations to work published in Consortium relative to gated journals.

Finances

In order to handle all the features envisioned in this proposal, and all the traffic that a discipline-wide system would attract, a sophisticated web-based platform would need to be developed, maintained, and modified over time. One might begin with open-source software available through PubSweet (https://coko.foundation/product-suite/). Even so, the Consortium will require a lot of programmer hours. In addition, copy-editing and type-setting would be costly.

Online publishing tools offer the possibility of dispensing with type-setting. Latex can be converted into any number of formats using Overleaf. MS Word is clunkier, but with a little work on the author's part could be converted into a uniform publishing style.

Copy-editing cannot be automated, but it could be offloaded on authors. Authors could enlist private editors to help them improve their prose – especially important for non-native speakers. Under the current system, journals offer only minimal copy-editing so this is not as big departure from the status quo as it may seem.

Dispensing with in-house copy-editing and type-setting, on the model of open-access repositories like SSRN would save a *lot* of money and also reduces the time lag between acceptance and publication.

Since the Consortium is envisioned as an umbrella organization, encompassing the entire discipline of political science, it should benefit from economies of scale. And, by cutting out the publisher, and the necessity of hard-copy journals, two major costs of the current system would be avoided. Over time, libraries would be relieved of purchasing costly subscriptions. So, we anticipate significant cost savings, overall, to the Consortium.

The problem is that these cost savings are realized by libraries and universities, not by scholars or by the Consortium. This means that we must think hard about how to construct a viable business plan.

We can envision four financing models, which might be employed singly or in combination.

First, universities might be asked to re-direct some portion of the savings they realize from journal subscriptions to the Consortium.

Second, authors might be charged for each submission. This would also serve as a screening device, discouraging frivolous submissions.

Third, governments or philanthropic bodies might subsidize the venture – at the very least, the startup costs for the venture. Although NSF funding is limited, other governments, including the EU, may be interested, especially as they have expressed dismay with the gated nature of scientific knowledge and are looking for ways to economize on library and university expenses. This would have the added benefit of internationalizing the initiative.

Fourth, APSA and other associations might assume financial responsibility, perhaps in conjunction with the closure of one or more existing journals, so that this commitment is cost-neutral.

IX. REFERENCES

Akst, J. 2010. "I hate your paper." *The Scientist Magazine* 24, 36–41.

Altman, Micah, Marguerite Avery. 2015. "Information wants someone else to pay for it: laws of information economics and scholarly publishing." *Information Services & Use* 35.1-2: 57-70.

Arkes, H. R., C. Gonzalez-Vallejo, A.J. Bonham, Y-H. Kung, N. Bailey. 2010. "Assessing the merits and faults of holistic and disaggregated judgments." *Journal of Behavioral Decision Making*, 23, 250-270.

Arkes, H. R., V.A. Shaffer, R.M. Dawes. 2006. "Comparing holistic and disaggregated ratings in the evaluation of scientific presentations." *Journal of Behavioral Decision Making*, 19, 429-439.

Armstrong, JS, R. Hubbard. 1991. "Does the need for agreement among reviewers inhibit the publication of controversial findings?" *Behav Brain Sci* 14(1):136–137.

Azar, Ofer H. 2006. "The academic review process: How can we make it more efficient?." *The American Economist* 50.1: 37-50.

Bagues, M., M. Sylos-Labini, N. Zinovyeva. 2017. "Does the gender composition of scientific committees matter?" *Am. Econ. Rev.* 107 (4), 1207–1238.

Beall, J. 2018. "Predatory journals exploit structural weaknesses in scholarly publishing." *4open*, 1, 1.

Bergstrom, Theodore C. 2001. "Free labour for costly journals?." *Journal of Economic Perspectives* 15.4: 183-198.

Binfield, P. 2013. "Open access megajournals – have they changed everything?" Creative Commons Blog, 23 October, available at: http://creativecommons.org.nz/2013/10/open-accessmegajournals-have-they-changed-everything/ (accessed 29 December 2015).

Björk, B.-C., P. Catani. 2016. "Peer review in megajournals compared with traditional scholarly journals: does it make a difference?" *Learned Publishing*, Vol. 29 No. 1, 9-12.

Blei, David M, Ng, Andrew Y., and Jordan, Michael I. 2003. "Latent Dirichlet Allocation." Journal of Machine Learning Research 3: 993-1022

Bohannon, J. 2013. "Who's afraid of peer review?" *Science* 342, 60–65.

Borer , Douglas A . 1997. "The Ugly Process of Journal Submissions: A Call for Reform." *PS: Political Science and Politics* 30 (3): 558 –60.

Bornmann, Lutz, Rüdiger Mutz, Hans-Dieter Daniel. 2010. "A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants." *PloS one* 5.12 (2010): e14331.

Butler, D. 2013. "The dark side of publishing." *Nature* 495(7442), 433–435.

Chetty, R., S. Emmanuel, S. Laszlo. 2014. "What policies increase prosocial behavior? an Experiment with referees at the journal of public economics." *Journal of Economic Perspectives*, 28(3), 169e188.

Cicchetti, D. V. 1991. "The reliability of peer review for manuscript and grant submissions: a cross-disciplinary investigation." *Behavioral and Brain Sciences* 14:119-135.

Clarke, Kevin A., David M. Primo. 2012. *A model discipline: Political science and the logic of representations*. Oxford: Oxford University Press.

Copiello, Sergio. 2018. "On the money value of peer review." *Scientometrics* 115.1: 613-620.

Daniel, Hans-Dieter. 1993. *Guardians of science: Fairness and reliability of peer review*. VCH.

Djupe, Paul A. 2015. "Peer reviewing in political science: New survey results." *PS: Political Science & Politics* 48.2: 346-352.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan Hyde, Craig McIntosh, Gareth Nellis. 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. Cambridge: Cambridge University Press.

Durning, Steven J., David P. Sklar, Eric W. Driessen, Lauren A. Maggio. 2019. "'This Manuscript Was a Complete Waste of Time': Reviewer Etiquette Matters." *Academic Medicine* 94, no. 6: 744-745.

Ellison, G. 2002. "The slowdown of the economics publishing process." *Journal of Political Economy* 110(5), 947–993.

Elman, Colin, John Gerring, James Mahoney (eds). 2020. *The Production of Knowledge: Enhancing Progress in Social Science*. Cambridge: Cambridge University Press.

Ernst, E., T. Saradeth, K.L. Resch. 1993. "Drawbacks of peer review." *Nature*, 363(6427), 296.

Espeland, Wendy Nelson, Michael Sauder. 2016. *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. New York: Russell Sage Foundation.

Fanelli, Daniele, Rodrigo Costas, Vincent Larivière. 2015. "Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity." *PloS one* 10.6: e0127556.

Ferguson, C., A. Marcus, I. Oransky. 2014. "The peer review scam." *Nature* 515, 480–482.

Ferreira, Catarina, Guillaume Bastille-Rousseau, Amanda M. Bennett, E. Hance Ellington, Christine Terwissen, Cayla Austin, Adrian Borlestean et al. 2016. "The evolution of peer review as a basis for scientific publication: directional selection towards a robust discipline?." *Biological Reviews* 91, no. 3: 597-610.

Fiske, D. W., L. Fogg. 1990. "But the reviewers are making different criticisms of my paper: Diversity and uniqueness in reviewer comments." *American Psychologist*, 45(5), 591–598.

Florian, Răzvan Valentin. 2012. "Aggregating post-publication peer reviews and ratings." *Frontiers in computational neuroscience* 6: 31.

Franco, Annie, Neil Malhotra, Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345 (6203): 1502–5. doi:10.1126/science.1255484.

Frey, Bruno S. 2003. "Publishing as prostitution? Choosing between one's own ideas and academic success." *Public Choice* 116.1-2: 205-223.

Gans, Joshua S. 2017. *Scholarly publishing and its discontents*. Toronto: Core Economic Research.

Gans, Joshua S., G.B. Shepherd. 1994. "How are the mighty fallen: Rejected classic articles by leading economists." *J Econ Perspect* 8(1):165–179.

Gerring, John. 2011. "How Good is Good Enough? A Multidimensional, Best-Possible Standard for Research Design." *Political Research Quarterly* 64:3 (September) 625-36.

Gerring, John. 2019. "Proposal: *The American Political Science Review*." Downloadable at https://utexas.app.box.com/s/gnifxxnbiq4dmvv15uhj8aq0xasg3d4h

Gerring, John. 2020. "Comprehensive Appraisal." In Colin Elman, John Gerring, & James Mahoney (eds), *The Production of Knowledge: Enhancing Progress in Social Science* (Cambridge University Press, forthcoming).

Gerring, John, Lee Cojocaru. 2020. "Length Limits." In Colin Elman, John Gerring, & James Mahoney (eds), *The Production of Knowledge: Enhancing Progress in Social Science* (Cambridge University Press, forthcoming).

Gerring, John, Sebastian Karcher, Brendan Apfeld. 2020. "Impact Metrics." In Colin Elman, John Gerring, & James Mahoney (eds), *The Production of Knowledge: Enhancing Progress in Social Science* (Cambridge University Press, forthcoming).

Godlee, Fiona, Tom Jefferson (eds). 2003. *Peer review in health sciences*. London: BMJ Books.

Gould, Thomas HP. 2012. *Do we still need peer review?: An argument for change*. Scarecrow Press.

Grayson, L. 2002. "Evidence based policy and the quality of evidence: Rethinking peer review." London: Centre for Evidence Based Policy and Practice (ESRC).

Gu, Y., Sun, Y., Jiang, N., Wang, B., and Chen, T. 2014. "Topic-factorized ideal point estimation model for legislative voting network." In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining: 183-192.

Hadavand, Aboozar, Daniel S. Hamermesh, Wesley W. Wilson. 2020. "Is Scholarly Refereeing Productive (At the Margin)?" Working Paper 26614

Hamermesh, Daniel S. 1994. "Facts and myths about refereeing." *Journal of Economic Perspectives* 8.1: 153-163.

Hasan, Zubair. 2018. "Academic sociology: The alarming rise in predatory publishing and its consequences for Islamic economics and finance." *ISRA International Journal of Islamic Finance* 10.1: 6-18.

Heneberg, Petr. 2016. "From Excessive Journal Self-Cites to Citation Stacking: Analysis of Journal Self-Citation Kinetics in Search for Journals, Which Boost Their Scientometric Indicators." PloS One 11 (4): e0153730. https://doi.org/10.1371/journal.pone.0153730.

Hengel, Erin. 2017. "Publishing while Female. Are women held to higher standards? Evidence from peer review." Cambridge Working Paper Economics: 1753.

Hojat, Mohammadreza, Joseph S. Gonnella, Addeane S. Caelleigh. 2003. "Impartial judgment by the "gatekeepers" of science: fallibility and accountability in the peer review process." *Advances in Health Sciences Education* 8.1: 75-96.

Horn, Sierk A. 2016. "The social and psychological costs of peer review: Stress and coping with manuscript rejection." *Journal of Management Inquiry* 25.1: 11-26.

Horrobin, DF. 1990. "The philosophical basis of peer review and the suppression of innovation." *JAMA* 263(10):1438–1441.

Jang, D., S. Doh, G.-M. Kang, D.-S. Han. 2016. "Impact of alumni connections on peer review ratings and selection Success rate in national research." *Sci. Technol. Human Values* 42 (1), 116–143.

Jefferson T, P. Alderson, E. Wager, F. Davidoff. 2002. "Effects of editorial peer review: a systematic review." *JAMA* 287: 2784–6.

Jefferson, Tom, Melanie Rudin, Suzanne Brodney Folse, Frank Davidoff. 2007. "Editorial peer review for improving the quality of reports of biomedical studies." *Cochrane Database of Systematic Reviews* 1.

Jiang Y, R. Lerrigo, A. Ullah, M. Alagappan, S.M. Asch, S.N. Goodman, et al. 2019. "The high resource impact of reformatting requirements for scientific papers." *PLoS ONE* 14(10): e0223976

Johnson, Valen E. & James H. Albert. 1999. Ordinal data modeling. New York: Springer.

Larivière, Vincent, Veronique Kiermer, Catriona J. MacCallum, Marcia McNutt, Mark Patterson, Bernd Pulverer, Sowmya Swaminathan, Stuart Taylor, and Stephen Curry. 2016. "A Simple Proposal for the Publication of Journal Citation Distributions." BioRxiv, July, 062109. https://doi.org/10.1101/062109.

Lawrence, Peter A. 2003. "The politics of publication." *Nature* 422.6929: 259.

Liao, Tim. 2020. "Peer Review." In Colin Elman, John Gerring, James Mahoney (eds), *The Production of Knowledge: Enhancing Progress in Social Science* (Cambridge: Cambridge University Press).

Lindsey, D. 1988. "Assessing precision in the manuscript review process: A little better than a dice roll." *Scientometrics* 14(1-2):75–82.

Link, A. M. 1998. "US and non-US submissions: an analysis of reviewer bias." *Journal of the American Medical Association*, 280(3), 246–247.

Mackie, Christopher D. 1998. *Canonizing economic theory*. Armonk, NY: M.E. Sharpe.

Mahoney, MJ. 1977. "Publication prejudices: An experimental study of confirmatory bias in the peer review system." *Cognit Ther Res* 1(2):161–175.

Marsh, H.W., U. W. Jayasinghe, N. W. Bond. 2008. "Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability." *American Psychologist*, 63, 160–168.

Matuschak, Andy, Michael Nielsen. 2019. "How can we develop transformative tools for thought?", https://numinous.productions/ttft, San Francisco.

McAfee, R. Preston. 2014. "Edifying editing." In Michael Szenberg and Lall Ramrattan (eds), *Secrets of Economics Editors* (Cambridge: MIT Press): 53-66.

Muller, Jerry Z. 2018. *The Tyranny of Metrics*. Princeton: Princeton University Press.

National Science Foundation, National Center for Science and Engineering Statistics. 2015. "Doctorate Recipients from U.S. Universities: 2014." Special Report NSF 16-300. Arlington, VA. Available at www.nsf.gov/statistics/2016/nsf16300/.

Newman, M. E. J., Watts, D. J., and Strogatz, S. H. 2002. "Random graph models of social networks." Proceedings of the National Academy of Sciences 99 (suppl 1): 2566-2572.

Nosek, Brian A., Daniel Lakens. 2014. "Registered Reports: A Method to Increase the Credibility of Published Results." *Social Psychology* 45 (3): 137–41.

Nosek, Brian A., George Alter, George C. Banks, Denny Borsboom, Sara D. Bowman, Steven J. Breckler, Stuart Buck et al. 2015. "Promoting an open research culture." *Science* 348, no. 6242: 1422-1425.

Nosek, Brian A., Jeffrey R. Spies, Matt Motyl. 2012. "Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability." *Perspectives on Psychological Science* 7.6: 615-631.

Nosek, Brian A., Yoav Bar-Anan. 2012. "Scientific Utopia: I. Opening Scientific Communication." *Psychological Inquiry* 23: 217-243.

Osterloh, M., B.S. Frey. 2015. "Ranking games." *Evaluation Review*, 39(1), 102–129.

Paltridge, Brian. 2017. *The discourse of peer review*. London: palgrave macmillan.

Pemstein, Daniel, Marquardt, Kyle L., Tzelgov, Eitan, Wang, Yi-ting, Medzihorsky, Juraj, Krusell, Joshua, Miri, Farhad, and von Römer, Johannes. 2019. "The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data." University of Gothenburg, Varieties of Democracy Institute: Working Paper No. 21, 4th edition.

Pencavel, John. 2014. "Journals, editors, referees, and authors: Experiences at the Journal of Economic Literature." In Michael Szenberg and Lall Ramrattan (eds), *Secrets of Economics Editors* (Cambridge: MIT Press) 83-96.

Peters, D.P., S.J. Ceci. 1982. "Peer-review practices of psychological journals: The fate of published articles, submitted again." *The Behavioural and Brain Sciences* **5**:187–255.

Pressman, Steven. 2014. "Econ Agonistes: Navigating and Surviving the Publishing Process." In Michael Szenberg and Lall Ramrattan (eds), *Secrets of Economics Editors* (Cambridge: MIT Press): 297-310.

Rennie, Drummond. 2003. "Innovation and peer review." In Godlee, Jefferson (eds), *Peer review in health sciences* (London: BMJ Books) 76-90.

Research Information Network (RIN). 2008. "Activities, costs and funding flows in the scholarly communications system in the UK." http://www.rin.ac.uk/system/files/attachments/Activites-costs-flows-report.pdf

Riggs, Tracy, Robert Wilensky. 2001. "An algorithm for automated rating of reviewers." *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*. ACM.

Rodriguez, Marko A., Johan Bollen, Herbert Van de Sompel. 2006. "The convergence of digital libraries and the peer-review process." *Journal of Information Science* 32.2: 149-159.

Rosser, J. Barkley, Jr. 2014. "Tales from the Editor's Crypt: Dealing with True, Uncertain, and False Accusations of Plagiarism." In Michael Szenberg and Lall Ramrattan (eds), *Secrets of Economics Editors* (Cambridge: MIT Press): 311-28.

Rothwell, P. M., C.N. Martyn. 2000. "Reproducibility of peer review in clinical neuroscience: is agreement between reviewers any greater than would be expected by chance alone?" *Brain*, 123, 1964-1969.

Rule A, A. Birmingham, C. Zuniga, I. Altintas, S-C. Huang, R. Knight, et al. 2019. "Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks." *PLoS Comput Biol* 15(7): e1007007. https://doi.org/10.1371/journal.pcbi.1007007

Sandstrom, U., M. Hallsten. 2008. "Persistent nepotism in peer-review." *Scientometrics*, 74(2), 175e189.

Shatz, David. 2004. *Peer review: A critical inquiry*. Rowman & Littlefield.

Shepherd, George B. (ed). 1995. *Rejected: Leading economists ponder the publication process*. T. Horton and Daughters.

Sigelman, Lee, Marcia Lynn Whicker. 1987. "Some implications of bias in peer review: a simulation-based analysis." *Social Science Quarterly* 68.3: 494.

Siler, Kyle, David Strang. 2017. "Peer review and scholarly originality: Let 1,000 flowers bloom, but don't step on any." *Science, Technology, & Human Values* 42.1: 29-61.

Siler, Kyle, Kirby Lee, Lisa Bero. 2015. "Measuring the effectiveness of scientific gatekeeping." *Proceedings of the National Academy of Sciences* 112.2: 360-365.

Smith, Richard. 2006. "Peer review: a flawed process at the heart of science and journals." *Journal of the royal society of medicine* 99.4: 178-182.

Spezi, Valerie, Simon Wakeling, Stephen Pinfield, Claire Creaser, Jenny Fry, Peter Willett. 2017. "Open-access mega-journals: the future of scholarly communication or academic dumping ground? A review." *Journal of documentation* 73, no. 2: 263-283.

Staddon, Jessica, Richard Chow. 2008. "Detecting reviewer bias through web-based association mining." *Proceedings of the 2nd ACM workshop on Information credibility on the web*. ACM.

Stafford, T. 2018. "Reviews, Reviewers, and Reviewing: The "Tragedy of the Commons" in the Scientific Publication Process." *Commun. Assoc. Inf. Syst.*, 42, 25.

Starbuck, William H. 2005. "How much better are the most-prestigious journals? The statistics of academic publication." *Organization Science* 16.2: 180-200.

Starbuck, William H. 2016. "60th anniversary essay: How journals could improve research practices in social science." *Administrative Science Quarterly* 61.2: 165-183.

Starck, J. Matthias. 2017. *Scientific Peer Review*. Wiesbaden, Germany: Springer Spektrum.

Stitzel, Brandli, Gary A. Hoover, William Clark. 2018. "More on plagiarism in the social sciences." *Social Science Quarterly* 99.3: 1075-1088.

Suls, Jerry, Rene Martin. 2009. "The Air We Breathe: A Critical Look at Practices and Alternatives in the Peer-Review Process." *Perspectives in Psychological Science* 4:1, 40-50.

Sun, T., J. Han, Gao J. and Yu, Y. 2009. "iTopicModel: Information Network-Integrated Topic Modeling," *Ninth IEEE International Conference on Data Mining*: 493-502.

Taichman, D. 2017. "Misuse of Received Manuscripts by Peer Reviewers: A Cross-Sectional Survey."

Tennant, J.P., J.M. Dugan, D. Graziotin, et al. 2017. "A multi-disciplinary perspective on emergent and future innovations in peer review." *F1000Research* 6:1151.

Teplitskiy, Misha, Daniel Acuna, Aïda Elamrani-Raoult, Konrad Körding, James Evans. 2018. "The sociology of scientific validity: How professional networks shape judgement in peer review." *Research Policy* 47, no. 9: 1825-1841.

Tomkins, A., M. Zhang, W.D. Heavlin. 2017. "Reviewer bias in single- versus double-blind peer review." *Proceedings of the National Academy of Sciences*, 114(48), 12708–12713.

Vale, Ronald D. 2015. "Accelerating scientific publication in biology." *Proceedings of the National Academy of Sciences* 112.44: 13439-13446.

Varian, Hal R. 1997. "The Future of Electronic Journals." Paper presented at the Conference on Scholarly Communication and Technology (Atlanta, GA, April 24-25).

Walker, Richard, Pascal Rocha da Silva. 2015. "Emerging trends in peer review—a survey." *Frontiers in neuroscience* 9: 169.

Wenneras, C., A. Wold. 1997. "Sexism and nepotism in peer-review." *Nature* 387:341–3

Wilsdon, J., L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill, B. Johnson. 2015. *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. London: SAGE.

Wu, Guangyu, Derek Greene, Barry Smyth, Pádraig Cunningham. 2010. "Distortion as a validation criterion in the identification of suspicious reviews." In *Proceedings of the First Workshop on Social Media Analytics*, pp. 10-13. ACM.

Zilibotti, Fabrizio. 2014. "Memoirs of an editor." In Michael Szenberg and Lall Ramrattan (eds), *Secrets of Economics Editors* (Cambridge: MIT Press) 97-118.

Zuckerman, H., Robert K. Merton. 1971a. "Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system." *Minerva*, 9(1), 66–100.

Zuckerman, H., Robert K. Merton. 1971b. "Sociology of refereeing." *Physics Today*, 24(7), 28–33.