Peer review and preprint policies are unclear at most major journals

Thomas Klebel[1*], Stefan Reichmann[1], Jessica Polka[2], Gary McDowell[3], Naomi Penfold[2], Samantha Hindle[4], Tony Ross-Hellauer[1]

[1]Graz University of Technology, Institute of Interactive Systems and Data Science, Graz, Austria

[2]ASAPbio, San Francisco, CA, United States of America

[3]Lightoller LLC, Chicago, IL, United States of America

[4]bioRxiv, Cold Spring Harbor Laboratory, NY, United States of America

*Corresponding Author

E-Mail: thomas.klebel@tugraz.at

Abstract

Clear and findable publishing policies are important for authors to choose appropriate journals for publication. We investigated the clarity of policies of 171 major academic journals across disciplines regarding peer review and preprinting. 31.6% of journals surveyed do not provide information on the type of peer review they use. Information on whether preprints can be posted or not is unclear in 39.2% of journals. 58.5% of journals offer no clear information on whether reviewer identities are revealed to authors. Around 75% of journals have no clear policy on co-reviewing, citation of preprints, and publication of reviewer identities. Information regarding practices of Open Peer Review is even more scarce, with <20% of journals providing clear information. Having found a lack of clear information, we conclude by examining the implications this has for researchers (especially early career) and the spread of open research practices.

Introduction

Scholarly publishing, as the steward of the scientific record, has a great deal of power to steer researcher practices. Despite emergent trends towards greater openness and transparency in all areas of research (McKiernan et al., 2016; Walker & Da Silva, 2015), publication practices of academic journals can remain something of a black box for authors and readers (Jette, 2018). Processes of editorial handling and peer review are usually hidden behind curtains of confidentiality or anonymity. But worse, journal policies which should orient authors and readers as to the editorial standards employed by individual journals, including what the general type of peer review system is or whether preprinting manuscripts is allowed, have been suggested to be often unclear (Castelo-Baz et al., 2015; Chawla, 2018; Nambiar et al., 2014). Unclear policies, for example regarding copyright or licensing, could expose researchers to unnecessary risk (Chawla, 2018). Lack of clarity of policies would also make it difficult for authors to find publishers with desirable practices, and even slow the appreciation among authors that different approaches are possible. Finally, opacity impedes our ability to track the prevalence of emerging policies, inhibiting understanding of how common and well-accepted those policies are.

This study aims to investigate the clarity of policies of major academic journals across academic disciplines regarding peer review and preprinting.

Consider the case of a graduate student wanting to preprint their manuscript. The graduate student is concerned about publishing in a recognised journal, one they deem "high impact" so that they can make progress in their career. They may have to submit to several journals before their work is accepted for publication. Will preprinting preclude publication in any of these journals? The majority of researchers are disincentivised from preprinting if a journal does not accept preprinted submissions (59% of 392 respondents to ASAPbio survey, 2016, https://asapbio.org/survey). In reality, the majority of preprints posted to arXiv and bioRxiv end up being published in a range of journals (Klein et al., 2019; Sever et al., 2019), and the graduate student can easily look up whether a specific journal would accept preprinted submissions using SHERPA/RoMEO[1]. The acceptance and adoption of preprints varies between disciplines: while established in several fields of physics (Elmore, 2018; Ginsparg, 2016), computer science, and mathematics, adoption in the life sciences (e.g. Balaji & Dhanamjaya, 2019; Chiarelli et al., 2019; Teixeira da Silva, 2017), chemistry, medicine (Johansson et al., 2018; Peiperl & Editors, 2018), and the social sciences and humanities is lower, and this may affect how many journals explicitly encourage or allow preprinted submissions. Further, some journals may specify the type of preprint they allow: the specific server(s) it may be posted to, the licence used for the preprint, whether (and which) different versions may be posted, and what types of blog or media coverage of the preprint would constitute an unacceptable breach of any journal press embargo. Varied and vague policies restrict author choices, and any constraints become more complicated with each additional journal considered. Furthermore, policies vary not only in their substance, but also in where they are communicated: sometimes they can be found under the instructions to authors, other times in many more obscure locations, and not unfrequently spread over

[1] http://www.sherpa.ac.uk/romeo/search.php

several web pages. The path of least risk and resistance to the graduate student may simply be to not preprint at all.

The situation is more difficult if the researcher wants to select journals based on practices for which there are no databases, such as peer review practices (at least for journals that do not partner with Publons). If our graduate student prefers to submit to a journal that will anonymously publish the content of peer reviews (believing that these will be more constructive, well-prepared, and professional), they must assemble a list of candidate journals identified by word-of-mouth or by searching across multiple journal websites for policies that are often difficult to find. Various forms of innovation grouped under the umbrella term "Open Peer Review" (Ross-Hellauer, 2017) result in a bewildering range of novel models for peer review. Especially for early career researchers, orienting themselves in this environment and understanding what is required of them can be a confusing process.

Finally, consider a graduate student deliberating whether or not to help their advisor with a peer review. They might want to know if a journal allows such *co-reviewing* and whether the review form enables them to be acknowledged when that review is submitted: in a recent survey, 82% of early-career researchers think it is unethical for PIs to submit peer review reports without naming all contributors to the report, and yet 70% of co-reviewers have contributed to peer review without any attribution (McDowell et al., 2019). The only way to find out if the journal process helps the graduate student's peer-review contributions to be recognised at present is to either contact the journal directly or to find someone with experience reviewing there.

TRANsparency in Scholarly Publishing and Open Science Evolution (TRANSPOSE) is a new initiative that addresses these issues. The TRANSPOSE initiative has created a database of journal policies for (1) open peer review, (2) co-reviewer involvement, and (3) preprinting.[2] Here we undertake a closer investigation of a subset of journals to systematically taxonomize and analyse their peer review and preprinting policies as stated in journals' author guidelines. We surveyed 171 major academic journals, drawn from the top-100 overall and top-20 per discipline of Google Scholar Metrics. The specific aims of the present study are to (1) systematically analyse the publicly available policies for preprinting and peer review of a corpus of highly cited journals, (2) assess the clarity and explicitness of policies, and (3) provide evidence for best-practice recommendations. While desirable, policies are not located conveniently in a limited number of uniform documents in many cases. All journals in our sample make some form of author guidelines publicly available. However, availability does not make for understandability.

Results

Policy Clarity

Within our sample, unclear policies are the norm, rather than the exception. Figure 1 displays all major aspects that were investigated, sorted by the proportion of clear policies within the sample. Overall, 54 out of 171 journals surveyed (31.6%) do not provide information on which

[2] https://transpose-publishing.github.io

type of peer review (double blind, single blind, not blinded, or other) is used. Information on whether preprints can be posted or not is similarly common, with 67 journals (39.2%) having no clear policy in this regard. There is no clear information on whether reviewer identities are revealed privately to the authors for 100 out of 171 journals (58.5%). Three quarters of journals in our sample have no clear policy with respect to whether co-reviewing is allowed, whether preprints can be cited or if reviewer identities are published. All other aspects (listed in Figure 1) are even more unclear, with 80% to 90% of journals giving no clear information on their website.

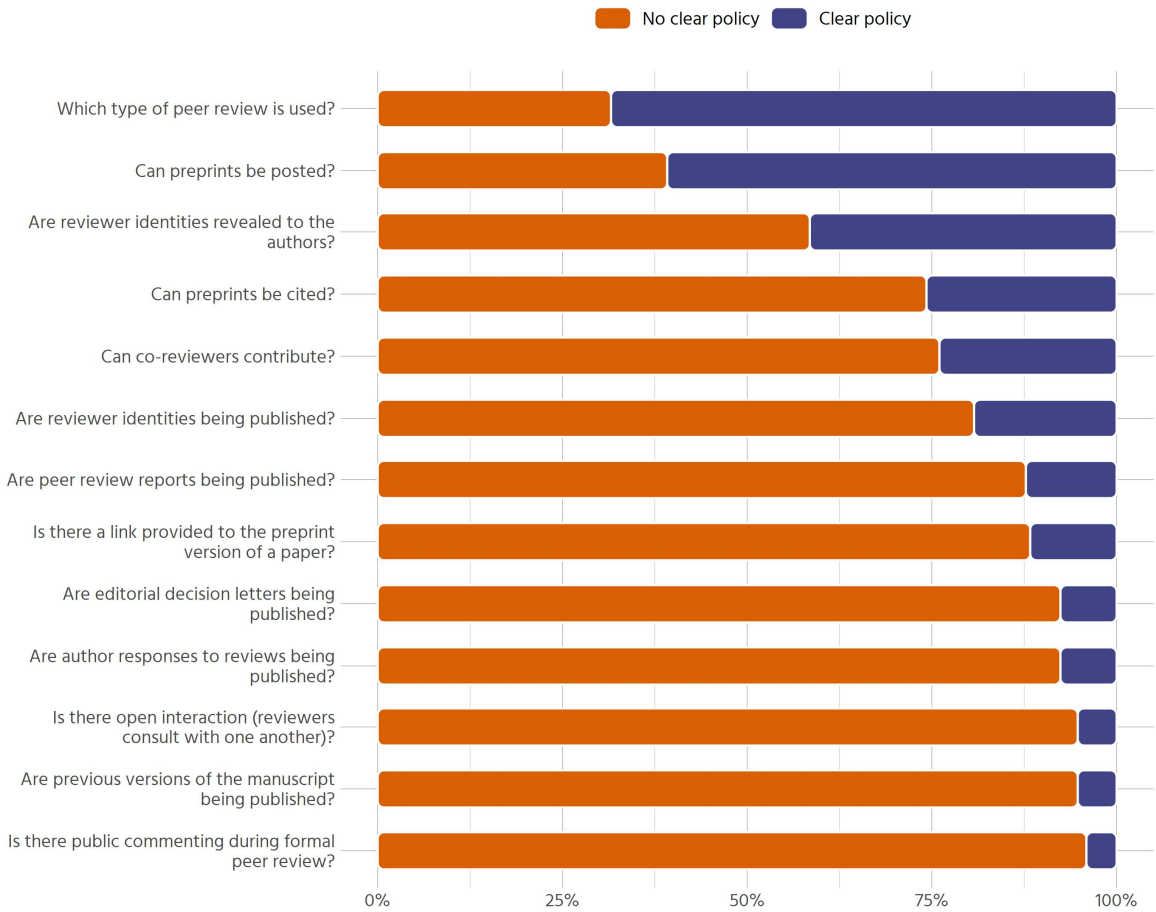

Figure 1: Overall Clarity of Policies (n = 171)

Regarding policy clarity, there is substantial variation between disciplines and publishers. This gives rise to many relevant questions: In what ways are policies related to each other? Do journals that allow co-reviewing also allow preprints? Is there a gradient between journals that encourage open research, and others that don't? Or are there certain groups of journals, open in one area, reluctant in another and maybe unclear in a third? To answer these questions, we employ Multiple Correspondence Analysis (MCA).

Results indicate that the different aspects of open research policies go hand in hand (Figure 2A, Table A1). Journals with clear policies on posting preprints tend to also give clear information on whether coreviewing is accepted, which type of peer review is used, and whether reviewer identities are revealed to the authors. On the other hand, journals with unclear policies in one

area more often than not have unclear policies in the other areas. Dimension 1 (horizontal) in figure 2A represents this gradient from journals with above average clear policies to journals whose policies are less clear than the average. This first dimension accounts for 72.2% of total variance, while the second dimension only accounts for 4.1% of total variance. The second dimension is thus of relatively small importance and should only be interpreted with caution (see also de Leeuw (2006, p. 121)). It mainly represents journals that have clear policies on co-reviewing and unclear policies on posting preprints on the bottom, with the complementary journals on top.

Turning to differences between disciplines and publishers, Figure 2B displays them projected onto the first dimension. Journals from the life sciences and earth sciences are well above average regarding clarity of policies, with journals from physics & mathematics, chemical & materials sciences and health & medical sciences being slightly above average. Journals from engineering & computer science are slightly below average, followed by journals from the social sciences, and humanities, literature & arts. Journals from business, economics & management have the least clear policies of our sample. The publishers of the journals sampled reflect the disciplinary differences. Journals from Springer Nature and the Royal Society of Chemistry are well above average in regard to policy clarity. While the American Chemical Society represents the average of journals sampled, Elsevier, IEEE, and the group of other publishers are below average in regard to clarity of policies.

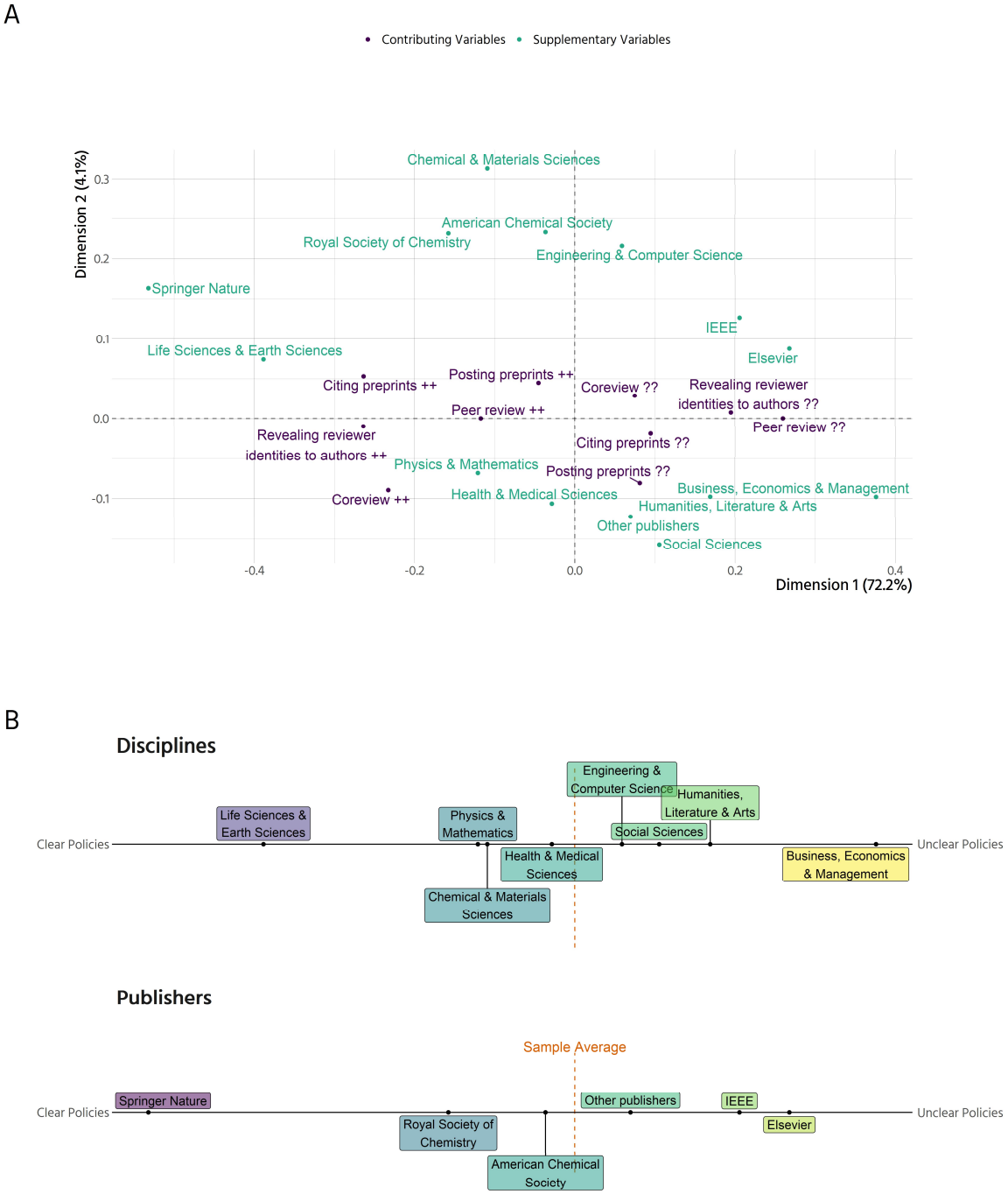

Figure 2: The Landscape of Open Research Policies

(A) Graphical display of a Multiple Correspondence Analysis. The contributing variables are the basis for the model and determine the layout of the space. "++" means that there is a clear policy, "??" that there is no clear policy. Disciplines and publishers were added as supplementary (passive) variables and have no impact on the space. Dimension 1 (horizontal) explains 72.2% of the variance, Dimension 2 explains 4.1% of the variance in the contributing variables.

(B) The supplementary variables from (A) projected onto the horizontal axis. Journals from disciplines and publishers with policies that are clearer than the average journal in our sample are located on the left, journals with less clear policies than the average on the right.

Peer Review

Availability of information on the type of peer review used by a journal is mixed (Figure 3A). Overall, 54 out of 171 journals (31.6%) do not provide clear information about their peer review process. For those journals with clear information, the most common peer review policy is single blind peer review (29.8%), followed closely by double blind peer review (26.9%). Some journals offer the option for authors to choose whether to use single or double blind peer review – for example, the Nature journals have a single-blind process as default but allow authors to choose to be double-blind if preferred. These cases have been coded as "Other", accounting for the majority of titles belonging to this category. 1% of journals ("The BMJ" and "The Cochrane Database of Systematic Reviews") do not anonymize authors or reviewers during the review process.

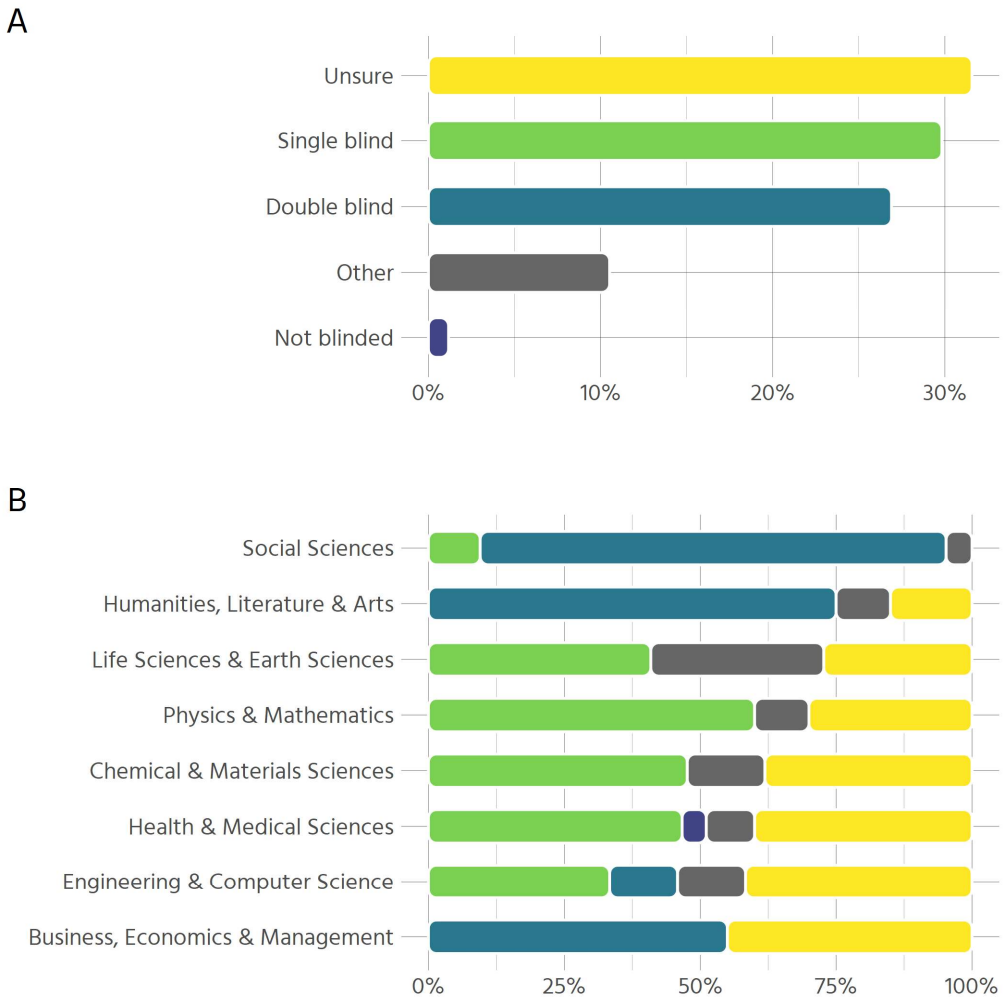

Figure 3: Type of peer review employed by journals

(A) Type of peer review used overall (n = 171) (B) Type of peer review used by disciplines (n = 193)

However, there are major differences between disciplines (Figure 3B). In the social sciences, humanities, and business, double blind peer review is the norm, while the natural sciences rely more heavily on single blind peer review. Among all disciplines, business, economics & management display the highest proportion of unclear policies, with social science and humanities being very clear and the remaining disciplines somewhere in between.

Open Peer Review

Information on open peer review (OPR) is similarly scarce (Figure 4A) across the sample. The survey included questions on select dimensions of open peer review, e.g. whether a journal publishes peer review reports, editorial decision letters or previous versions of the manuscript, whether it offers public commenting during the peer review process, and similar questions. More than 50% of journals surveyed do not provide any information on these aspects of OPR. No journal in our sample allows public commenting during formal peer review. Other forms of openness are similarly rare. With the exception that some journals state that they may reveal reviewer identities to authors, information on the other aspects is either not specified or OPR is not practiced by more than 95% of journals.

As revealing reviewer identities privately to authors is the only aspect of open peer review that is explicitly allowed by a substantive number of journals (23.4%), we examine it separately for each discipline (Figure 4B). Whereas the social sciences, humanities and business journals' policies do not mention revealing reviewer identities to authors, this is not unusual in the natural sciences, at least on an optional basis (many journals offer referees the opportunity to sign their reviews).

Figure 4: Aspects of open peer review

(A) Aspects of open peer review across all journals in the sample (n = 171). Categories: Yes, Conditional (i.e., is true if other conditions apply), Optional (i.e., either author or reviewer can choose but not mandatory), No, Not Specified (i.e., information not found online)

(B) Results on whether reviewer identities are revealed to the authors, even if they are not published. (n = 193)

Co-Review

Information on co-review policies is not uniformly available: 87 out of 171 journals (50.9%) have an explicit co-review policy. There are notable disciplinary differences (Figure 5). In the

life and earth sciences, health & medical sciences as well as physics & mathematics more than a quarter of journals explicitly permit contributions from co-reviewers, whereas in the humanities, chemical & materials sciences, and in business, economics & management only around 10% do.

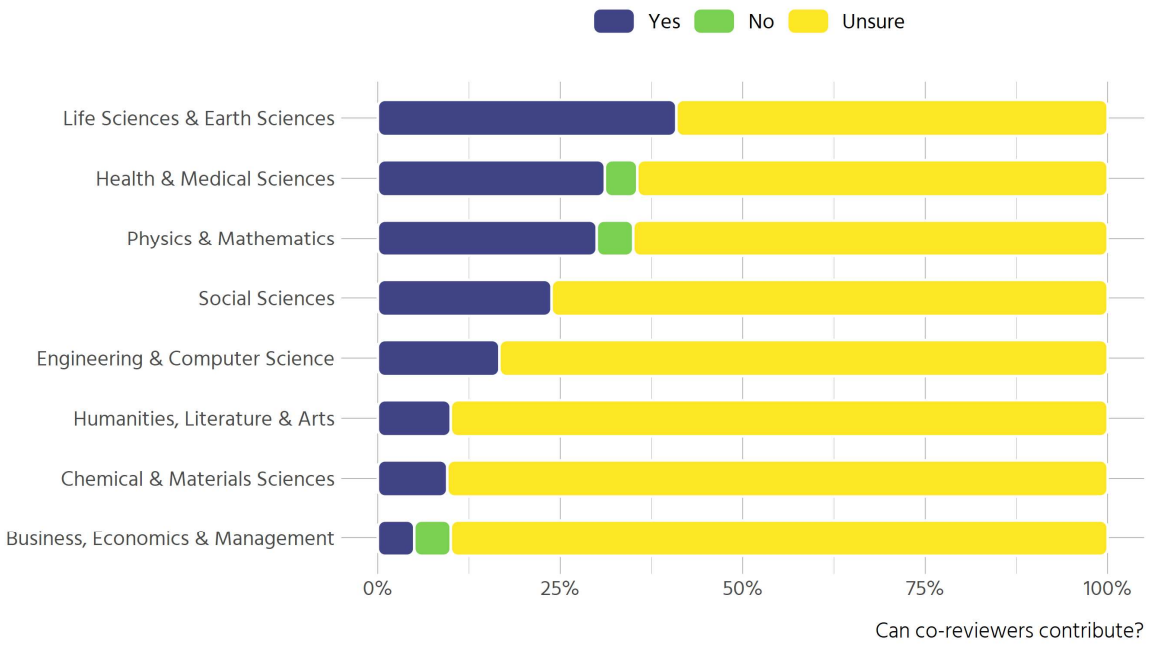

Figure 5: Clarity of co-review policies (n = 171)

To obtain a more nuanced view of the policies, we analysed their content via text mining. Table 1 displays the most frequent terms of the distinct policies (n = 35), sorted by the proportion of policies that contain a given term. The most prominent themes that emerge are:

- Individuals with varying stakes regarding peer review: editor, colleague, collaborator, student, author, peer.
- Confidentiality as a central principle.
- Important elements of scholarly publishing: manuscript, journal, review.
- Verbal forms pertaining to relationships between the individuals: inform, involve, consult, discuss, disclose, share.

Table 1: Propensity of terms in co-review policies

Term	Term frequency	Proportion of policies that contain term
review	100	93%
manuscript	43	75%
editor	33	73%
confidenti	26	63%
not	24	60%
inform	19	51%
colleagu	18	49%
student	14	34%
discuss	12	32%
involv	12	32%
consult	12	32%
permiss	11	31%
disclos	12	29%
author	11	29%
peer	10	29%
journal	10	28%
share	9	25%
collabor	10	24%
advic	8	23%
ident	8	23%

Journals stress the importance of "maintaining confidentiality" through "not shar[ing]" or disclosing information, neither to "junior researchers" and "laboratory colleagues" nor to "graduate students". Even if the policies do not explicitly forbid or allow the involvement of other researchers, in many cases they mandate the reviewer to first obtain permission from the editor in case they want to involve someone else in their review. The editor's prominent role can also be observed by the terms' frequent appearance in the policies: almost three quarters of all policies mention the term "editor". In the majority of cases, policies state that one must "obtain permission from the journal editor" to show the manuscript to others or that co-reviewing is not permitted "unless previously agreed with the editor".

Preprints

Policies for posting or citing preprints are more common within our sample compared to open peer review or co-review policies. 120 out of 171 journals (70.2%) state that they allow some form of preprints. Most of them (39.2% of the total sample) however only allow preprints before peer review while 22.8% do not have a preprint policy.

Similar to our results on peer review, preprint policies vary considerably between disciplines (Figure 6A). While in the life sciences & earth sciences 91% of all journals allow preprints in some way, in the humanities only 45% do. The natural sciences in general tend towards allowing preprints only on first submission (before peer review). Journals from the social sciences, the humanities and from business, economics and management generally either have no preprint policy at all or are more diverse in regard to preprint version, also allowing preprints after peer review.

A complementary aspect of the acceptance of preprints is whether they can be cited. The majority of journals (57.3%) do not specify whether this is possible. Unclear policies on how to cite preprints (e.g. in the references or only as footnotes in the text) are also quite common (15.2%). Where citations of preprints are allowed, this is possible in the references for 78% of journals, with some journals restricting citations of preprints to the text (14%).

Preprint policies with respect to citations again vary greatly between disciplines (Figure 6B). Policies permitting citation of preprints are more common in the natural sciences, with 55% of all journals in the life and earth sciences allowing citations to preprints either in the text or in the reference list. In contrast, the social sciences and humanities largely have unclear policies or no policies at all regarding whether preprints can be cited or not.

Besides posting and citing preprints, we surveyed other aspects of preprint policies as well: whether there is information on which licenses are permitted for the preprint, or if there is scoop protection, e.g. if a preprint will still be considered for publication even if a competing work is published in another journal after the date of preprinting. Further aspects were whether a published paper includes a link to the preprint version, what type of media coverage of the preprint is permitted and if there is a policy on community review for preprints. Overall, guidance on these issues is rarely provided: 72.5% of journals provide no information on permitted media coverage and 88.3% of journals provide no information on whether the publication will include a link to the preprint. 94.7% of journals provide no guidance on which license is permitted for the preprint, 98.2% give no information on scoop protection, and 98.2% of journals give no indication whether public comments on preprints will have any effect on manuscript acceptance.

Figure 6: Posting and citing of preprints

(A) Results on whether a preprint can be posted, and which version is allowed (n = 193). (B) Results on whether preprints can be cited (n = 193)

Discussion

Clarity of journal policies

Our results suggest that policies regarding various aspects of scholarly publishing are very often unclear. Even the most basic kind of information – which type of peer review a journal uses – could not be found on the website in more than 30% of journals. Information on all other aspects we investigated is even harder to find. Whether preprints can be posted is unclear in 39.2% of journals, whether reviewer identities are revealed to the authors is unclear in 53.5% of journals. All other aspects are unclear in at least 75% of journals sampled. This is problematic, since it hinders the uptake of open research practices on several fronts. Authors might be reluctant to post or cite preprints if they cannot be sure how this will impact their submission. Reviewers might be disinclined to sign their reviews or involve junior colleagues in writing the review if they do not know how editors will handle these cases.

We found that there is a gradient between journals that have clear policies on the different aspects of open research practices and other journals with unclear policies. This gradient is roughly structured along the distinction between SSH and STEM disciplines. Since open research practices are as yet less common in the SSH, it should come as no surprise that journals have no or unclear policies. The other side of the gradient is marked by disciplines from the natural sciences where, generally speaking, open research practices are more common (Balaji & Dhanamjaya, 2019; Chiarelli et al., 2019; Elmore, 2018; Ginsparg, 2016; Johansson et al., 2018; Peiperl & Editors, 2018; Teixeira da Silva, 2017).

An alternative explanation to the lack of clear policies might also be that a given practice (e.g. double blind peer review) is so common in certain disciplines that specific policies are not put in place or not communicated transparently. One of our findings helps to illustrate this point. Recall Figure 4B, where we investigated whether reviewer identities are revealed to authors, even if they are not made public. The high proportion of journals within SSH that are categorised as "Not specified" might be surprising, given that most of them conduct double blind peer review. One could thus infer that reviewer identities are not revealed to the authors. This inference however is the root problem: there is no clear policy. Reviewers might sign their review or not; what the authors receive is at the editor's discretion.

Peer review and co-review

We found that 31.6% of journals in our sample don't offer clear information on which type of peer review they employ. This is in line with Utrobičić et al. (2014) who studied editorial structures and peer review policies in Croatian journals indexed in Web of Science, finding a lack of transparency of publicly available information for authors on peer review processes. There are ongoing debates e.g. in medical journals how this situation might be amended (e.g. Castelo-Baz et al., 2015; Sprowson et al., 2013). Increasing availability of information regarding the editorial procedure might be beneficial for journals themselves, since disclosure of information about the editorial and peer-review process correlates with authors' perceptions of a high-quality peer review process and the journal rejecting hoax papers (Wicherts, 2016).

The highly influential role of editors in what practices are acceptable or prohibited and how certain policies might be implemented has been investigated with regard to peer review (Resnik & Elmore, 2016). This can be extended to the issue of co-reviewing. 50.9% of journals in our sample have an explicit co-review policy. Analysing the respective policies revealed that many of them reference confidentiality as a core principle. If a manuscript is to be shown to or discussed with another researcher, reviewers have to ask the editor for permission in the majority of cases. This is problematic, since co-reviewing and ghostwriting is very common among early career researchers, and in practice permission is not asked for from the editor, but the manuscript is shared anyway (McDowell et al., 2019). Early career researchers will likely hesitate to contact the journal's editor if their superior asks them to help with or write a review, and in turn the invited reviewer will, upon the submission of the review, consider omitting the participation of the co-reviewers as the lesser sin compared to not having asked permission to do so – or simply may not consider naming of co-reviewers as necessary, in the absence of clear journal policies surrounding co-reviewing. In addition, the contribution of early career researcher co-reviewers might be prohibited by informal editorial policy or it might go unnoticed, since acknowledging the efforts made by multiple reviewers is very rare in general.

Preprints

Researchers generally feel they must publish in community-recognised journals for career progression and evidence of productivity. As a consequence, whether a journal regards preprints as prior publication or not is an important policy factor, as posting a preprint of a manuscript might effectively forestall publication in a journal. Additional considerations where authors may expect clarity include preprint licensing, which version can be uploaded to which server(s), and whether preprints can be cited (and if so, how). All of these matter for the individual author as well as the use of preprints in a discipline in general. We found that 29.8% of journals sampled do not offer clear information on whether preprints can be posted online, before and/or after submission to the journal. The majority of journals (57.3%) do not specify whether citing preprints is possible. Information on which licences are appropriate for the preprint, if media coverage for the preprint would preclude the manuscript from being published in the journal or whether the publication will include a link to the preprint cannot be found in at least 70% of journals sampled.

The policies' content varies by discipline. For example, in the humanities only 45% of journals explicitly allow authors to post preprint versions of their manuscript, while in the life and earth sciences 91% do. Our results in this regard support previous work on disciplinary cultures and differential propensity to accept preprints (e.g. Fry et al., 2016). In the social sciences, publication patterns and citation cycles differ markedly from those in the natural sciences, e.g. citation cycles are generally much longer (Fleck, 2013), reducing the efficacy of preprinting. Furthermore, the social sciences and humanities operate on vastly different conceptions of originality (Guetzkow et al., 2004), placing different strains on publication processes.

In summary, we find that policies regarding various aspects of scholarly publishing are very often unclear or missing. This is not to say that policies should be an iron cage, with no flexibility for editorial decisions. Professional judgement is an important part of performing the tasks of an editor. However, uncertainty for authors and reviewers alike is unconstructive. If

there is no guidance on whether certain practices are encouraged or prohibited, submitting and reviewing for journals becomes a minefield that is not easily navigated. This might further hinder scholarly participation from early career researchers who are less accustomed and aware of certain norms in their field.

Data and Methods

We used the Google Scholar Metrics service[3] (GSM) to compile a list of the top 100 publications (journals) ordered according to their five-year h-index metric as of 13th October 2018. The five-year h-index "is the largest number h such that h articles published in [the last 5 complete years] have at least h citations each" (Google Scholar Metrics, 2019). In addition, we took the top 20 results from each of the 7 broad subcategories offered by GSM: Business, Economics & Management; Chemical & Material Sciences; Engineering & Computer Science; Health & Medical Sciences; Humanities, Literature & Arts; Life Sciences & Earth Sciences; Physics & Mathematics; Social Sciences. Results were returned on 13th October 2018 (although the GSM about page notes these results are based on "our index as it was in July 2018"). These lists were copied over to a spreadsheet where the journal titles were compiled and de-duplicated, with information retained about their relative position in one or more of the top 100 and 7 sub-categories. The full list is available at https://zenodo.org/record/3627116.

We acknowledge several limitations of this approach. Firstly, GSM does not enable browsing by subject area for non-English-language titles. This naturally means that our lists do not properly represent non-English language titles. Moreover, by focusing on "high-impact" titles, we can assume our sample is biased towards titles that are better resourced financially, which can be assumed to have more developed policies in place than their less well-resourced counterparts. Hence, this landscape scan cannot represent the totality of the journal landscape. In addition, this is based on non-open data: the criteria for inclusion and exclusion in the Google Scholar index are opaque and non-reproducible (Giustini & Boulos, 2013). This study, however, does not aim at a complete picture of all journals across all domains, regions and languages – rather to scope the policies of a limited number based on their perceived prominence to global scholarly communities, and with a corpus that is manageable for qualitative investigation and classification. The h-index has been subject to critique regarding its use as a measure of scientific impact (Kreiner, 2016). Here, however, we are clear that it is used only as a proxy for visibility within scientific communities. A further difficulty with this approach is that taking only the top 20 journals in each category further impacts the representativeness of this sample. Levels of citations vary widely not only between broad categories of research, but also within specific disciplines and subdisciplines (Hutchins et al., 2016), and the number of journals sampled does not scale with the total number of journals or researchers in those areas. Again, we here acknowledge this limitation as an artefact of the pragmatic need to compile a corpus small enough to allow qualitative interrogation but large enough to include at least some data on differences across broad categories of research. We of course encourage further replications of this analysis at subdiscipline level.

[3] https://scholar.google.co.uk/citations?view_op=top_venues&hl=en

Data collection

De-duplication returned a list of 171 journals. Each title was then assigned to two assessors who applied a standardised data-collection instrument and protocol to determine what information is publicly available online regarding peer review and preprint policies at each journal. The first round of data-collection took place between 2018-11-21 and 2019-02-15 and the second round between 2019-04-11 and 2019-04-24. In a third round between 2019-04-24 and 2019-04-28 data from the two assessors was cross-checked, resolving any discrepancies. The data-collection instrument is available at https://zenodo.org/record/3627116. The aim was to mirror the experience of a researcher who might wish to find this information online. Search began from the journal website, and internal links followed from there. No secondary sources were used (e.g., assessor's prior knowledge; external databases; contact with journal editorial staff). An alternative strategy was to use web keyword search (via Google) using, for example "[journal name] AND 'peer review' OR 'pre-print' OR 'preprint' OR 'working paper'", or, in the case of co-reviewing policies, "[journal name] AND 'confidentiality'". The second assessor checked the first assessor's answers and revised or challenged based on their own interpretation of the information found online. Disputes were then adjudicated by two authors (JP & TRH) who reviewed the second-round edits in a third and final round. Note here that we do not claim that our dataset collects all possible information which could have been found online for these journal policies. Information can be spread widely over a confusing number of journal- and/or publisher-level pages. Hence, there is the possibility that some information was not captured despite two rounds of review.

After the third round of review, the collected data were imported to R and cleaned for further analysis. This involved unifying categories for plotting and merging with data from GSM on disciplinary area. The approach taken to create the sample of journals led to a few journals having no subdiscipline: some journals like "Gut" were within the top 100 journals, but not within any of the subdisciplines. This is because the h5-index varies widely between subdisciplines. Figure 7A shows the top-20 journals of each discipline.

The missing categorisations were added in a second step, to facilitate analysis of all journals that distinguishes by discipline. To this end, we scraped all disciplines and subdisciplines from GSM on 18th of June 2019 and matched those to our data.

As stated, the criteria for inclusion into the Google Scholar rankings are opaque and non-reproducible. For example, it is possible for a journal to be included in different disciplines, e.g. "Physics & Mathematics" along with "Engineering & Computer Science". It is however also possible for a journal to be included in a subdiscipline, and not in the parent discipline, despite having a higher h-index than all journals listed in the parent discipline.[4]

[4] As of 2019-12-20, the "Journal of Cleaner Production" is listed in the social sciences under "sustainable development" (https://scholar.google.at/citations?view_op=top_venues&hl=en&vq=soc_sustainabledevelopment). But it is not listed under the parent category (https://scholar.google.at/citations?view_op=top_venues&hl=en&vq=soc).

The nature of our selection means that 22 out of 171 journals are assigned to two disciplines. All results that distinguish between disciplines are therefore based on 193 cases. The inclusion criteria further mean that disciplines are not represented equally within the sample. Since about one quarter of the top 100 journals belong to the health and medical sciences, the sample is slightly skewed in that direction (Figure 7B).

Figure 7: Sample characteristics

(A): The distribution of journal's h5-indices across disciplines (n=171). (B) Number and proportion of journals sampled by discipline (n=193)

Regarding practices of open access, only 8 of 171 journals are listed in the Directory of Open Access Journals (DOAJ) and can thus be considered fully open access.

Methods

Data analysis was done in R (R Core Team, 2019), with the aid of many packages from the tidyverse (Wickham et al., 2019). The analysis of the policies generally follows two directions: first, whether clear policies can be found, and second, what their content is.

To investigate clarity on policy, we selectively recoded variables with regard to whether certain policies were clear or not, thus omitting the subtle differences within the policies (i.e. "which version of a preprint can be cited" was simplified for whether the policy was clear (references allowed in text, reference list or not allowed) versus unclear (unsure about policy, no policy and other)). "Clarity", or more precisely the "explicitness", of author guidelines therefore have been operationalized pragmatically as whether a reasonably well-versed researcher would be able to *locate* and *understand* a given journal's regulations on peer review, preprints, and co-reviewing in a reasonable amount of time. It should be noted that this represents an analytic categorization which is not necessarily reflected in the conceptualizations employed/relevance ascribed by journals. However, we expect any assessor with reasonable practical knowledge of academic publishing to be able to reproduce the data collection procedure based on the assessment framework described in the section "Data Collection". It should be noted, though, that conducting the data collection procedure again will lead to partly different results, since the policies under scrutiny are subject to change.

After recoding for clarity, we analysed the variables via Multiple Correspondence Analysis (Greenacre and Nenadic, 2018), which lets us explore the different policies jointly (Greenacre and Blasius, 2006) and thus paint a landscape of open research policies among journals. It should be noted that this procedure is strictly exploratory. We are exploring possible associations between the policies, not testing any hypotheses.

We included five active categories in our model. All were recoded in terms of whether there was a clear policy on:

- Type of peer review (double blind, single blind, not blinded, or other)
- Co-reviewing
- Revealing reviewer identities to authors
- Posting preprints
- Citing preprints

The geometric layout of the space displayed in figure 2A is determined by these five active categories. Interpretation of the points displayed is achieved by projecting them onto the axes. Furthermore, only statements regarding the sample average are possible. If a given journal is far away from zero, it indicates that this journal is more clear/unclear than the rest of the sample, but not that the journal's policies are completely clear/unclear in absolute terms. To further illuminate some of the results, the disciplinary areas and the five most common publishers were added as passive categories. They have no influence on the geometric layout but allow us to draw conclusions on which policies are more prevalent in one area or another.

To investigate the policies' contents, the main analytical approach was to create displays of cross tabulations with ggplot2 (Wickham, 2009). When reporting percentages from these cross

tabulations, we report percentages with one decimal for the full sample (171 or 193 journals (e.g. 23.3%)). When reporting disciplinary differences (n = 20-45), we report percentages without decimals (e.g. 23%).

Co-review policies were further analysed via text mining. Due to the prevalence of publisher-level policies for many journals in the sample, there are 35 distinct policies on co-review in our dataset (compared to 87 policies in total[5]). Since the policies are generally rather short in length, the analysis is somewhat limited with respect to insights we can gain from automated procedures. To extract meaningful information we first removed common words from the English language (via the list of stop-words from the tidytext package (Silge and Robinson, 2016), except for the word "not", which is relevant since some policies state that it is *not* appropriate to share information with students or colleagues). The resulting list contains 886 words in total. For a simple overview, the words were stemmed to reduce similar but not identical versions of certain words (like editor/editors).

We used the package visdat (Tierney, 2019) to explore the data at the beginning of analysis, and used ggrepel (Slowikowski, 2019) to design comprehensible figures with labels. All data and code, including a reproducible version of the results section, is available at https://zenodo.org/record/3627116.

References

Balaji, B. P., & Dhanamjaya, M. (2019). Preprints in scholarly communication: Re-imagining metrics and infrastructures. *Publications*, *7*(1). https://doi.org/10.3390/publications7010006

Castelo-Baz, P., Leira-Feijoo, Y., Seoane-Romero, J. M., Varela-Centelles, P., & Seoane, J. (2015). Accessibility to editorial information in Oral and Maxillofacial Surgery journals: The authors' point of view. *Journal of Cranio-Maxillofacial Surgery*, *43*(7), 1078–1081. https://doi.org/10.1016/j.jcms.2015.06.005

Chawla, D. S. (2018). Scientists struggle with confusing journal guidelines. *Nature*. https://doi.org/10.1038/d41586-018-07326-9

Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Preprints and Scholarly Communication: Adoption, Practices, Drivers and Barriers. *F1000Research*, *8*, 971. https://doi.org/10.12688/f1000research.19619.1

de Leeuw, J. (2006). Nonlinear Principal Component Analysis and Related Techniques. In M. J. Greenacre & J. Blasius (Eds.), *Multiple correspondence analysis and related methods* (pp. 107–134). Chapman & Hall/CRC.

[5] During data collection, investigators manually copied relevant parts of the policies to our dataset. This inhibited detecting duplicates, since for the same policy different parts might have been copied or abbreviated. To identify duplicates, TK compared the policies with the distance metric "Jaccard" and then manually went through the most similar ones. Selection for deleting duplicates was done by keeping the version with more text to retain as much information as possible.

Elmore, S. A. (2018). Preprints: What Role Do These Have in Communicating Scientific Results? *Toxicologic Pathology*, *46*(4), 364–365. https://doi.org/10.1177/0192623318767322

Fleck, C. (2013). The Impact Factor Fetishism. *European Journal of Sociology*, *54*(2), 327–356. https://doi.org/10.1017/S0003975613000167

Fry, J., Spezi, V., Probets, S., & Creaser, C. (2016). Towards an understanding of the relationship between disciplinary research cultures and open access repository behaviors. *Journal of the Association for Information Science and Technology*, *67*(11), 2710–2724. https://doi.org/10.1002/asi.23621

Ginsparg, P. (2016). Preprint Déjà Vu. *The EMBO Journal*, *35*(24), 2620–2625. https://doi.org/10.15252/embj.201695531

Giustini, D., & Boulos, M. N. K. (2013). Google Scholar is not enough to be used alone for systematic reviews. *Online Journal of Public Health Informatics*, *5*(2), 214. https://doi.org/10.5210/ojphi.v5i2.4623

Google Scholar Metrics. (2019, November 15). *Google Scholar Metrics*. https://scholar.google.at/citations?view_op=top_venues&hl=en&vq=en

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is originality in the humanities and the social sciences? *American Sociological Review*, *69*(2), 190–212.

Hutchins, B. I., Yuan, X., Anderson, J. M., & Santangelo, G. M. (2016). Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level. *PLOS Biology*, *14*(9), e1002541. https://doi.org/10.1371/journal.pbio.1002541

Jette, A. M. (2018). Opening Up the Black Box of Peer Review. *Physical Therapy*, *98*(4), 207–208. https://doi.org/10.1093/ptj/pzx126

Johansson, M. A., Reich, N. G., Meyers, L. A., & Lipsitch, M. (2018). Preprints: An underutilized mechanism to accelerate outbreak science. *PLOS Medicine*, *15*(4), e1002549. https://doi.org/10.1371/journal.pmed.1002549

Klein, M., Broadwell, P., Farb, S. E., & Grappone, T. (2019). Comparing published scientific journal articles to their pre-print versions. *International Journal on Digital Libraries*, *20*(4), 335–350. https://doi.org/10.1007/s00799-018-0234-1

Kreiner, G. (2016). The Slavery of the h-index—Measuring the Unmeasurable. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00556

McDowell, G. S., Knutsen, J. D., Graham, J. M., Oelker, S. K., & Lijek, R. S. (2019). Co-reviewing and ghostwriting by early-career researchers in the peer review of manuscripts. *ELife*, *8*, e48425. https://doi.org/10.7554/eLife.48425

McKiernan, E. C., Bourne, P. E., Brown, C. T., & Buck, S. (2016). How open science helps researchers succeed. *ELife*, *5*. https://doi.org/10.7554/eLife.16800

Nambiar, R., Tilak, P., & Cerejo, C. (2014). Quality of author guidelines of journals in the biomedical and physical sciences. *Learned Publishing*, *27*(3), 201–206. https://doi.org/10.1087/20140306

Peiperl, L., & Editors, on behalf of the P. M. (2018). Preprints in medical research: Progress and principles. *PLOS Medicine*, *15*(4), e1002563. https://doi.org/10.1371/journal.pmed.1002563

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Resnik, D. B., & Elmore, S. A. (2016). Ensuring the Quality, Fairness, and Integrity of Journal Peer Review: A Possible Role of Editors. *Science and Engineering Ethics*, *22*(1), 169–188. https://doi.org/10.1007/s11948-015-9625-5

Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research*, *6*, 588. https://doi.org/10.12688/f1000research.11369.2

Sever, R., Roeder, T., Hindle, S., Sussman, L., Black, K.-J., Argentine, J., Manos, W., & Inglis, J. R. (2019). bioRxiv: The preprint server for biology. *BioRxiv*, 833400. https://doi.org/10.1101/833400

Slowikowski, K. (2019). *ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2."* https://CRAN.R-project.org/package=ggrepel

Sprowson, A. P., Rankin, K. S., McNamara, I., Costa, M. L., & Rangan, A. (2013). Improving the peer review process in orthopaedic journals. *Bone & Joint Research*, *2*(11), 245–247. https://doi.org/10.1302/2046-3758.211.2000224

Teixeira da Silva, J. A. (2017). Preprints: Ethical hazard or academic liberation? *KOME*, *5*(2), 73–80. https://doi.org/10.17646/KOME.2017.26

Tierney, N. (2019). *visdat: Preliminary Visualisation of Data*. https://CRAN.R-project.org/package=visdat

Utrobičić, A., Marušić, Matko, & Marušić, Ana. (2014). Composition of editorial boards and peer review policies of Croatian journals indexed in Web of Science and Scopus. *European Science Editing*, *40*(2), 31–33.

Walker, R., & Da Silva, P. (2015). Emerging trends in peer review-a survey. *Frontiers in Neuroscience*, *9*, 169. https://doi.org/10.3389/fnins.2015.00169

Wicherts, J. M. (2016). Peer Review Quality and Transparency of the Peer-Review Process in Open Access and Subscription Journals. *PLOS ONE*, *11*(1), e0147913. https://doi.org/10.1371/journal.pone.0147913

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Author ORCIDs

Thomas Klebel: 0000-0002-7331-4751

Stefan Reichmann: 0000-0003-1544-5064

Jessica Polka: 0000-0001-6610-9293

Gary McDowell: 0000-0002-9470-3799

Naomi Penfold: 0000-0003-0568-1194

Samantha Hindle: 0000-0002-3708-3546

Tony Ross-Hellauer: 0000-0003-4470-7027

Acknowledgements

We thank the 2018 Scholarly Communication Institute (Chapel Hill, NC) for supporting the early phases of this project.

Conflicts of Interest

Gary McDowell works at a for-profit that provides consulting services to organizations addressing issues concerning early career researchers. Samantha Hindle is Content Lead at bioRxiv, a preprint server for the biological sciences. Tony Ross-Hellauer is Editor-in-Chief of the journal "Publications" (ISSN 2304-6775).

Author Contributions

Conceptualization: Jessica Polka, Gary McDowell, Tony Ross-Hellauer

Data curation: Thomas Klebel, Stefan Reichmann, Jessica Polka, Gary McDowell

Formal analysis: Thomas Klebel

Funding acquisition: Tony Ross-Hellauer

Investigation: Thomas Klebel, Stefan Reichmann, Jessica Polka, Gary McDowell, Naomi Penfold, Samantha Hindle, Tony Ross-Hellauer

Methodology: Thomas Klebel, Jessica Polka, Gary McDowell, Tony Ross-Hellauer

Project administration: Tony Ross-Hellauer

Resources: Jessica Polka

Supervision: Tony Ross-Hellauer

Visualization: Thomas Klebel

Writing – original draft: Thomas Klebel, Stefan Reichmann, Jessica Polka, Gary McDowell, Tony Ross-Hellauer

Writing – review & editing: Thomas Klebel, Stefan Reichmann, Jessica Polka, Gary McDowell, Naomi Penfold, Samantha Hindle, Tony Ross-Hellauer

Appendix

Table A1: Numerical output from Multiple Correspondence analysis

Variable	mass	inertia	k=1	correlation with dim 1	contribution to dim 1	k=2	correlation with dim 2	contribution to dim 2
Coreview ??	151	48	75	586	33	29	87	86
Coreview ++	49	148	-233	586	103	-90	87	267
Posting preprints ??	72	118	81	305	18	-81	302	320
Posting preprints ++	128	66	-45	305	10	45	302	178
Citing preprints ??	147	52	95	861	52	-19	35	37
Citing preprints ++	53	144	-264	861	143	53	35	102
Revealing reviewer identities to authors ??	115	95	195	728	171	7	1	4
Revealing reviewer identities to authors ++	85	129	-264	728	231	-10	1	6
Peer review ??	62	138	260	853	164	0	0	0
Peer review ++	138	62	-117	853	74	0	0	0
Business, Economics & Management			376	936		-98	64	
Chemical & Materials Sciences			-109	108		313	892	
Engineering & Computer Science			59	70		216	930	
Health & Medical Sciences			-29	67		-107	933	

Humanities, Literature & Arts	169	749	-98	251
Life Sciences & Earth Sciences	-388	965	74	35
Physics & Mathematics	-121	757	-68	243
Social Sciences	105	310	-157	690
American Chemical Society	-36	24	233	976
Elsevier	268	903	88	97
IEEE	206	727	126	273
Royal Society of Chemistry	-158	317	231	683
Springer Nature	-532	915	163	85
Other publishers	70	243	-123	757

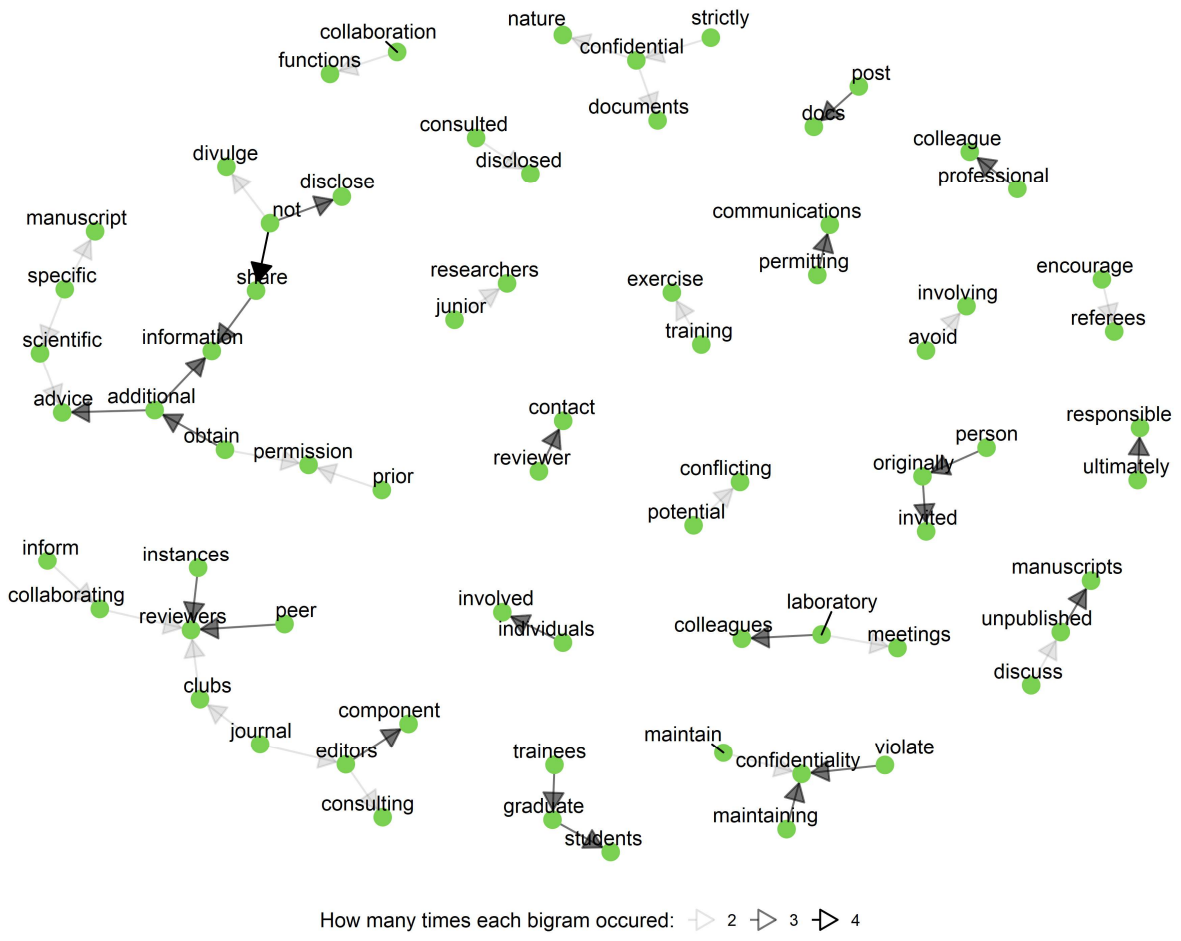

Figure A1: Directed bigram graph of co-reviewing policies

Displayed are bigrams for all terms in the co-reviewing policies, after removal of stop-words (the word “not” was not removed, see the “Methods” section). When creating bigrams, the text is split into pairs of words (for example the sentence "All humans are equal" becomes "All humans", "humans are", "are equal"). The most prominent bigrams were "peer -> review" and "review -> process". To look at the strength of other associations, the term "review" was removed from the figure. The most frequent associations in the figure are depicted by bold arrows.