
Marjorie Bardiau, University of Liège, 0000-0002-1042-5788
Christophe Dony, University of Liège, 0000-0001-7421-5993

Abstract

Little attention has been devoted to whether the Impact Factor (IF) can be considered a responsible metric in light of bibliodiversity. This paper critically engages with this question in measuring the following variables of IF journals included in the 2021 Journal Citation Reports and examining their distribution: publishing models (hybrid, Open Access with or without fees, subscription), world regions, language(s) of publication, subject categories, publishers, and the prices of article processing charges (APC) if any. Our results show that the quest for prestige or perceived quality through the IF brand poses serious threats to bibliodiversity. The IF brand can indeed hardly be considered a responsible metric insofar as it perpetuates publishing concentration, maintains a domination of the Global North and its attendant artificial image of mega producer of scholarly content, does not promote linguistic diversity, and de-incentivizes fair and equitable open access by entrenching fee-based OA delivery options with rather high APCs.

1. Introduction

The Impact Factor (IF) is a bibliometric indicator that has been the subject of much criticism, especially in the current state of advocacy for more responsible metrics. Various key reform-oriented texts concerned with the advancement of research assessment such as the San Francisco Declaration on Research Assessment (DORA 2013), the Leiden Manifesto (Hicks et al. 2015), the Metric Tide (Wilsdon et al. 2015), or the Agreement on Reforming Research Assessment (Science Europe 2022), for example, have called to stop using the IF in research evaluation, notably because of its calculation-related deficiencies such as skewed citation distribution (Baum 2011; Larivière et al. 2016; Vanclay 2011). These and other texts have also alerted about the normative uses of the IF and its adverse effects on the research ecosystem. These concerns and criticisms include unfair comparisons between fields with different citations practices (Editors 2006; Garfield 2006; Archambault and Larivière 2009), perverse incentives (Wilsdon et al. 2015; Hicks et al. 2015) including citation gaming (Ritchie 2020; Siler and Larivière 2022), the amplification of journal status endowment through article processing charges (APCs) inflation (Demeter and Istratii 2020; Siler and Frenken 2020), and the use of the IF as a proxy for perceived quality or excellence (Leclercq 2016; Morales et al. 2021; Pontika et al. 2022; Wouters et al. 2015), despite evidence to the contrary (Brembs, Button, and Munafò 2013; Brembs 2018; Chavarro, Ràfols, and Tang 2018; Dougherty and Horne 2022).

Despite the prevalence of the IF in the research ecosystem, including its influence in research assessment and rewards systems (Moher et al. 2018; Morales et al. 2021; Rice et al. 2020; Pontika et al. 2022), research has so far failed to propose a comprehensive analysis of bibliodiversity as it can relate to the IF brand. Bibliodiversity is key in maintaining a diverse, inclusive, and equitable research landscape that can “accommodate the different workflows, languages, publication outputs, and research topics that support the needs and epistemic pluralism of different research communities” (Shearer et al. 2020). A lack of bibliodiversity can result in homogenized and hegemonic trends in the ways research is carried out, made visible, and produced or disseminated (Shearer and Becerril-García 2021; 2021). This impacts who can
access, read, and benefit from research (Pölönen, Syrjämäki et al. 2021; Pölönen, Guns et al. 2021). Taking its cue from this characterization of bibliodiversity as a multifaceted and structuring concept of scholarly communications, this paper examines the bibliodiversity of the IF brand in measuring the following variables of IF journals included in the 2021 Journal Citation Reports (n=12391): publishing models (hybrid, Open Access with or without fees, subscription), world regions, language(s) of publication, subject categories, publishers, and the prices of article processing charges (APC) if any.

Measuring these variables and examining their distribution can provide valuable insights into the IF brand, in particular how it might be said to reject bibliodiversity in reinforcing "monopoly, monoculture, and high prices" in scholarly communications (Shearer et al. 2020). In turn, these insights may help various stakeholders in scholarly communications, research assessment, information literacy and open science to further critically engage with the IF and how its brand may actually jeopardize the development of a healthy, diverse, and sustainable scholarly communications ecosystem.

2. Literature Review

Bibliodiversity is a term that can refer to a variety of realities and practices. The International Alliance of Independent Publishers defines bibliodiversity as "cultural diversity applied to the world of books" (International Alliance of Independent Publishers 2014). In practice, this is envisioned as "a complex, self-sustaining system of storytelling, writing, publishing, and other kinds of production of oral and written literature" which contributes to more sociocultural diversity and equity (International Alliance of Independent Publishers 2014). For libraries, providing access to a diversity of outputs and formats is a core principle embedded in both advocacy actions and collection development strategies (e.g. American Library Association (ALA) 2006; International Federation of Library Associations (IFLA) and UNESCO 2012). In the world of scholarly communications, the concept is even more wide-ranging.

In this context, bibliodiversity indeed pertains to editorial methods, discoverability practices, publishing trends and infrastructures, and issues of affordability and sustainability (see Berger 2021; Shearer et al. 2020; Shearer and Becerril-García 2021), in particular how the latter can relate to the politics of open access (cf. Alborno, Okune, and Chan 2020; Chan and Costa 2005; Gray 2020). The 2017 Jussieu Call for Open Science and Bibliodiversity, for instance, argues that developing a scholarly publishing ecosystem by privileging a fee-based open access model may "slow if not check the advent of bibliodiversity" (Baun et al. 2017). Similarly, Monica Berger describes fee-based open access as a colonizing force and enterprise in the landscape of scholarly publishing (Berger 2021). A similar view is shared by Shearer and Becerril-García who contend that bibliodiversity can contribute to the decolonization of scholarly communications (Shearer and Becerril-García 2021). In their respective articles, both Berger as well as Shearer and Becerril-García point to various Latin American initiatives of community-owned and non-commercial infrastructures or publishing platforms such as Latindex, Redalyc and SciELO, all of which have enhanced the creation, discoverability, prevalence and significance of locally-produced research (Aguado-López et al. 2012; Becerril-García and Aguado-López 2018; Packer 2009; 2020). In so doing, these initiatives have shown a commitment to the "sustainable, anticolonial ethos" of bibliodiversity (Berger 2021), that is, they constitute a viable alternative to the domination of international English-language journals published by a handful of large

---

1 The Journal Citation Report (JCR) features journals included in the Science Citation Index Expanded (SCIE) and/or the Social Sciences Citation Index (SSCI).
commercial publishers (Larivière, Haustein, and Mongeon 2015), which have increasingly shifted to pay-to-publish open access options delivery methods in recent years (Butler et al. 2022).

Bibliodiversity thus runs throughout all the production and dissemination mechanisms and practices of the scholarly communications ecosystem. But this wide scope makes it particularly difficult to propose a comprehensive analysis of how bibliodiversity is articulated in a given corpus. This is why studies usually only examine the state of one main bibliodiversity-related variable in particular bibliographic databases, journals’ lists, or subsets of thereof in focusing on, for example, coverage or scope limitations as they can relate to subjects (Aksnes and Sivertsen 2019; Tennant 2020; Khanna et al. 2022; Visser, van Eck, and Waltman 2021; Guns and Holowiecki 2022), geographic representation (Tennant 2020; Khanna et al. 2022), linguistic diversity (Balula and Leão 2019; Kulczycki et al. 2020; Larivière and Riddles 2021; Vera-Baceta, Thelwall, and Kousha 2019), publisher concentration (Larivière, Haustein, and Mongeon 2015; Butler et al. 2022), or questions of affordability and models and how they tie in with the politics of open access (Bosman et al. 2021; Khoo 2019; Klebel and Ross-Hellauer 2022; Morrison et al. 2022; Butler et al. 2022).

This paper complements the above mentioned research in adopting a more holistic approach to bibliodiversity as it can be applied to the IF brand by measuring the previously mentioned bibliodiversity data points of journals included in the 2021 JCR. In so doing, it provides a blueprint for a reproducible method of “measuring back” which echoes the decolonial ethos of bibliodiversity and can be reused for other journals’ lists or publishing brands.

3. Methods and materials

The initial dataset used in this study was created by collating data from the 2021 SCIE, SSCI and JCR lists that are freely available on Clarivate’s website.2 The following information for individual journal entries was obtained from these lists: journal title, issn(s), publisher, publisher’s address, publisher country, language(s) of publication, the Web of Science Categories (WoS), and the index(es) that the journal is part of (SCIE, SSCI, or both).

This dataset was then enriched with information regarding publishing models and APC prices, if any, using freely available datasets or lists from publishers and other sources, namely: Walt Crawford’s dataset “Gold Open Access 6: 2015-2020” (Crawford 2021), a journal metadata file produced by DOAJ3, and APC price lists from various publishers. Information regarding journal models and possible PAC was also manually checked and retrieved from publishers and journals’ websites.

This dataset is openly available on Zenodo 10.5281/zenodo.7683743 (Bardiau, Marjorie and Dony, Christophe 2023).

Hereafter, detail is provided for the data standardization and structuration methods used for the following data variables: publishing models, APC prices, subject categories, publishers’ names, world regions, and language(s) of publication.

---

2 Lists were downloaded on July 29, 2021. They can be downloaded for free after logging in at https://mjl.clarivate.com/collection-list-downloads.

3 This DOAJ metadata file was downloaded on August 13, 2021 on https://doaj.org/docs/public-data-dump.
3.1 Publishing models

Different criteria were used and articulated with a boolean logic to identify the publishing models of journals, especially for OA journals. Four main models were identified: fully fee-based OA journals (OA-APC), full OA journals not requiring fees (OA-no fees), hybrid, and subscription. The criteria used to categorize journals into these models are defined below.

**OA with APC:**
- Journals are included in the DOAJ or GOA6 and mention the presence of fees; OR
- Journals are included in APC prices lists of publishers and are explicitly labeled as OA, Gold OA, or full-OA journals; OR
- All of the journal’s content can be accessed for free and publishing in said journal can only be achieved in exchange of an APC, regardless of the type of license that may be used or whether said journal refers to or uses the phrase “Open Access”; OR
- Journals have transitioned to full APC-based OA and all of their content since their flipping date is published OA.

**OA without APC:**
- Journals are included in the DOAJ or GOA6 and explicitly mention the absence of fees, i.e. fees = 0 $; OR
- All of the journal’s content can be accessed for free and it is explicitly mentioned that publishing in said journal does not require the payment of APC, regardless of the license that may be used or whether said journal refers to or uses the phrase “Open Access”; OR
- Journals have transitioned to no-fee OA or have adopted the “Subscribe to Open” model but the content predating this flipping moment may still only be available behind a paywall.

**Hybrid:**
- Subscription journals which offer OA publishing only as an option through which articles “are immediately free to read under an open license” that is granted “in exchange for an article processing charge (APC)” (Piwowar et al. 2018), including journals labeling themselves as “transformative”; OR
- Journals are included in APC price lists of publishers and are explicitly labeled as Hybrid or Optional OA journals, regardless of the license that may be used;

**Subscription:**
- The source’s content is only available through subscription or purchase options; it may contain some Bronze OA, i.e. “free to read content on the publisher’s page, but without an [sic] clearly identifiable license” (Piwowar et al. 2018); OR
- Journals use a delayed- or embargoed OA model, i.e. there may be an OA-moving barrier but the latest content can only be accessed through subscription or purchase options.

---

4 Definitions of OA journals may differ according to the subsets or features of OA used for identification (cf. Piwowar et al. 2018; Suber 2012). Although “OA journals are most often defined by their inclusion in the Directory of Open Access Journals” (DOAJ), the directory is not exhaustive. According to a recent global study on no-fee (i.e. Diamond) open access journals, “[t]he number of OA diamond journals is high (estimated to 29,000) but only a third of them are registered in DOAJ” (Bosman et al. 2021). As a result, several other criteria were used to determine the OA status of journals.

5 The label “Subscribe to Open” for these journals is included in the “Additional information” column of our dataset (Bardiau, Marjorie and Dony, Christophe 2023).
It was sometimes not possible to ascertain the publishing models of some journals, or their possible use of APC or lack thereof. As a result, two other categories were used: “unidentified” (n=19), and “OA-possible presence of fees unknown” (n=304).

3.2 APC Prices

In order to ensure valid comparison, APC prices were all collected in $US for US authors without any taxes or any membership or society discounts for what comes closest to “traditional” research articles in terms of output type, without taking into account potential submission fees or extra fees for coloured pages and/or images, tables, and figures. When different types of licenses were offered, prices for a CC-BY license were collected. When APC varied according to length, the same limit was used as that of Crawford in GOA6 (Crawford 2021, 2), i.e. 10 pages. When APC prices were not mentioned in $US, prices were converted using the exchange rate of the date the said price was collected. When journals only charged authors within the country of publication, those charges were used (converted in $US if necessary).

The label “unidentified” was used for APC prices of full-OA or hybrid journals whose APCs could not be identified (n=172).

3.3 Unique subject categories

The WoS uses over 250 subject categories to classify its indexed journals. To facilitate data analysis for the present study, Milojević’s reclassification scheme of WoS content into broad subject areas was used (Milojević 2020).

Using this reclassification scheme brought about some challenges. First, journals can be assigned different WoS categories, which are not “explicitly hierarchical, even though some subject categories can be considered as part of other, broader ones” (Milojević 2020). Most journals assigned with more than one WoS category point to the same broad area after reclassification, but this was not true in some cases. Nevertheless, the arbitrary decision was made to assign a broad area based on the reclassification of the first WoS category appearing in Clarivate’s data. Second, the WoS category scheme contains the “Multidisciplinary Sciences” label that has no equivalent in Milojević’s reclassification scheme because it focuses on article-level reclassification, for which the label "Multidisciplinary Sciences" makes little sense. Since this study focuses on journals, this WoS category was kept as a broad area, thereby updating Milojević’s scheme up to 15 unique broad areas. Finally, we found three Wos categories that did not have any match in Milojević’s reclassification scheme, namely “Development

---

6 Out of these 19 journals, 10 have ceased or stopped their journal publishing activities. This information is included in the “Additional information” column of our dataset. 7 For reasons of transparency and reproducibility, information regarding the type of output for which APC prices were collected can be found in the “Output type” column of our dataset.

8 These criteria are similar to those Crawford used for GOA6 (Crawford 2021, 2).

9 When APC prices varied according to a word limit, the limit of 5000 words was used.

10 Wayback machine links with a 2021 date stamp were used or created to that effect as well as reproducibility purposes whenever technically possible. This was not possible for Taylor & Francis journals requiring APCs as prices for these can now only be collected through a search module within a webpage of the publisher’s website. For similar technical reasons, 0,82% journals (n=102) point to 2022 wayback machine links for models and possible APC data.

11 For a general description of this classification system and its heuristics, see (Pudovkin and Garfield 2002).

12 This might induce a slight bias in the reclassification as WoS subject categories appear in alphabetical order.
“Social Sciences” broad area was used to reclassify the journals tagged with “Development Studies” or “Regional & Urban Planning” as a first WoS category because of their semantic similarity with the WoS category “Planning and Development”, which falls into Milojević’s broad area “Social Sciences”. Journals assigned with “Quantum Science & Technology” as a first WoS category all contained at least one other WoS category which was used to reclassify those journals with a broad area label from Milojević’s scheme.

### 3.4 Standardized Publishers' names and Publisher's ensembles

Many scholarly and scientific societies, institutions, or unions are client organizations of commercial publishers, whose platforms they use to distribute content and sometimes manage publishing processes. In the WoS data used to build our initial dataset, journals are associated with these client organizations or with publishers’ imprints and brands, some of which include many name variants or are abbreviated.

Several strategies were used to clean up and standardize the names of most publishers. First, the list of publisher imprints in Langham-Putrow and Enriquez's worksheet was used to reclassify data regarding imprints and brands into larger standardized publisher names (Langham-Putrow and Enriquez 2022)\(^\text{13}\). Second, common abbreviations or name variants were searched and assigned a standardized name. Finally, the label “Other” was assigned to all remaining journals without a standardized name (n=2711).

To facilitate the analysis of results, these standardized publishers’ names were then mapped onto a scheme of three categories: the Oligopoly (Elsevier, Sage, Springer Nature, Taylor & Francis, and Wiley), the next 20 most common publishers\(^\text{14}\) in the dataset (after the Oligopoly and omitting publishers tagged with the label “Other”), and the category of “Other publishers” (all the rest).

### 3.5 World Regions

On the basis of the WoS country data included in the initial dataset, a “World Region” value was assigned to journals using UNICEF’s regional classifications scheme\(^\text{15}\).

### 3.6 Languages

Six categories of languages of publication were used to facilitate analysis: English, Multilingual including English (i.e. journals accepting papers in English in addition to at least one other language), German, Spanish, French, and Other language(s) (i.e journals accepting papers in at least one other language that is not included in the previous categories).

Two challenges had to be faced when dealing with the standardization of data regarding language(s) of publication. First, this data was missing for 88 journals in the WoS data. As a

---

\(^{13}\) See Appendix in Langham-Putrow’s “Lesson 2.4: Cleaning up publisher name variants” (2022).


\(^{15}\) See [https://data.unicef.org/regionalclassifications/](https://data.unicef.org/regionalclassifications/).
result, data was added manually for these journals after checking their websites.\textsuperscript{16} Second, the Wos information regarding language(s) of publication is not always consistently structured. While for some journals, multiple languages are enumerated, for some others the label “Multi-Language” is used (n=715). As a result of this discrepancy, every journal tagged with this “Multi-Language” label was considered to be a journal using English in addition to another language of publication. Though this might induce a slight bias in the data, the rationale behind this decision is based on the fact that when they are enumerated, languages of publication of journals almost always explicitly include English.

4. Results

Fig.1 to Fig.4 show the diversity of the IF brand per publishers’ ensembles and one of the following variables: world regions, subjects, language(s) of publications, or publishing model.

Fig.1 shows little geographic diversity and indicates high levels of publisher concentration, especially in Western Europe and North America. These two regions produce 84.61\% (n= 10485) of all IF journals included in the 2021 JCR. All other world regions are underrepresented, especially Easter and Southern Africa and West and Central Africa, which respectively account for 0.45\% (n=56) and 0.03\% (n=4) of all journals. Fig.1 also shows the global reach of the Oligopoly (i.e. Elsevier, Sage, Springer Nature, Taylor & Francis, and Wiley) as it is present in all world regions, albeit in varying degrees. The top 25 publishers, which consist of the Oligopoly and the next 20 publishers with the biggest portfolio, account for 74.97\% (n= 9290) of all journals. Western Europe and North America publish 89.54\% (n= 8318) of journals produced by these 25 publishers. In contrast, other world regions show lower levels of publisher concentration as they present bigger, albeit variable shares of journals tagged in the “Other publishers” category, ranging from 34.17\% (n=41) in the Middle East and North Africa to 89.45\% (n=212) in Latin America and the Caribbean.

\textbf{Fig.1: Distribution of journals per world region and publishers’ ensemble}

\textsuperscript{16} For the sake of transparency and reproducibility, this information is recorded in the dataset column “Added Language” (Bardiau, Marjorie and Dony, Christophe 2023).
Fig. 2 indicates high levels of publisher concentration for the top 25 publishers in all subject categories. Across all subjects, the Oligopoly represents 58.8% (n=7286) while the next 20 publishers with the biggest portfolio account for 16.2% (n=2004) of all journals. In six subject categories, the Oligopoly’s shares of journals reach levels higher than the overall average: Chemistry (61.2%; n=366), Geosciences (60.7%; n=417), Medical Sciences (59.2%; n=1847), Professional fields (63.3%; n=656), Psychology (70.8%; n=330), and Social Sciences (63.5%; n=606). When adding the next 20 publishers with the biggest portfolio to the Oligopoly category, levels of publishers concentration reach levels higher than their combined overall average in six subject categories: Astronomy (81.8%; n=54), Chemistry (83.6%; n=500), Computer sciences (78.1%; n=383), Medical sciences (77%; n=2435), Professional fields (78.1%; n=809), Psychology (86.3%; n=402), and Social sciences (79.6%; n=760).

Fig. 3 shows that English prevails as the only language of publication for a majority of journals (89.4% of all journals; n= 11077) and that the top 25 publishers publish a majority of English-only journals (78.6%, n=8699). In contrast, the category of “Other publishers” somewhat levels the playing field for more linguistic diversity as publishers that do not belong to the top 25 publishers produce more than 50% of journals in all language categories except English- and German-only titles: Other language(s) (87.2%; n=109), Spanish (82.02%; n=73), French (58.7%; n=37), and Multilingual including English (50.1%; n=469).

The proportions of journals per language category fluctuate very little according to subjects or world regions. It is nevertheless worth noting that the least linguistically diverse category is that of Computer sciences, of which 97.34% (n=477) of titles are English-only journals. Worth noting as well is the fact that the most linguistically diverse world region is Latin America and the Caribbean, with only 32.07% (n=76) of English-only journals.
Fig. 4 indicates that the hybrid publishing model is dominant (67.97% of all journals; n=8422) and that the top 25 publishers privilege hybrid and fully OA-APC journals. Together, these 25 publishers produce 90.14% (n=7592) of all hybrid journals and 60.25% (n=1115) of all full OA fee-based journals. In contrast, the category of “Other publishers” is more diverse in terms of publishing models; it accounts for 79.89% (n=608) of all OA journals that do not require fees and represents 59.8% (n=610) of all journals using a subscription model. This category of other publishers is also the one with the most journals for which the possible presence of fees could not be identified (97.4%; n=295).
The distribution of publishing models does not vary much according to subjects but fluctuates more per world region (Fig. 5). Fig. 5 shows that Western Europe and North America privilege hybrid and OA-APC publishing models. Taken together, these models account for 89.6% of all journals in Western Europe and 84.3% of all journals in North America, with the hybrid model representing respectively 73.1% and 76.3% of journals in these regions. In contrast, other regions show more diverse publishing models albeit in varying degrees. The subscription model is prevalent in Eastern Europe and Central Asia (48.3%; n=173), while in Latin America and the Caribbean, the most common model is that of OA without fees (42.2%, n=100). Proportionally speaking, Western Europe and North America also have the smallest shares of OA journals that do not require fees, respectively 5.1% (n=310) and 2.2% (n=96).

Fig. 4: Distribution of journals per publishing model and publishers’ ensemble
Fig. 5: Distribution of journals per world region and publishing model

Fig. 6 to Fig. 8 below focus on APC-related information for hybrid and APC-based OA journals per publisher ensemble (Fig. 6 and Fig. 7) and world regions (Fig. 8). Fig. 6 shows that the top 25 publishers produce more journals in higher APC price ranges than the category of “Other Publishers”. This is especially true for hybrid journals. These top 25 publishers produce 72.25% (n=5485) of their hybrid journals with APC prices higher than the $US 2800 threshold, which corresponds to the highest peak for the top 25 publishers in Fig. 6. In contrast, other publishers only produce 31.57% (n=262) of their hybrid journals within the same APC price ranges, i.e. from $US 2801 onwards. This category of “Other Publishers” also produces more OA-APC journals in lower APC price ranges than the top 25 publishers: 71.8% (=527) of their journals operating with this model are situated below the $US 1200 threshold, while the top 25 publishers only produce 9.06% (n=101) of their OA-APC journals within the same price ranges (from $US 1 to $US 1200).
Fig. 6 shows a statistical summary of APC prices for hybrid and OA-APC journals per publishers’ ensembles, with boxplots showing the first quartile, median, and the third quartile of the values. This figure confirms higher median APC prices for the top 25 publishers hinted at in Fig. 6. For hybrid journals, the median APC price is $US 3000 for the Oligopoly, $US 3255 for the next 20 publishers, and $US 2380 for the category of Other Publishers. The difference in median APC prices for OA-APC journals between the Other publishers and the two other categories of publishers’ ensembles is even more marked: $US 2290 for the Oligopoly, $US 1916 for the next 20 publishers, and $US 684 for the Other Publishers.

Fig. 8 shows a statistical summary of APC prices with box plots for hybrid and OA-APC journals per world region. Median APC prices of hybrid journals approximate $US 3000 in all world regions, except in Eastern Europe and Central Asia. Median APC prices of OA-APC journals fluctuate much more per world region as they are below $US1000 in all world regions but three: East Asia and Pacific ($US 1100), North America ($US 2000), and Western Europe ($US 2000).
Fig. 7: Distribution of APC prices with box plots for hybrid and OA-APC journals per publishers’ ensembles.

Fig. 8: Distribution of APC prices with box plots for hybrid and OA-APC journals per world region.
5. Discussion

The IF brand indicates a continued domination of the oligopoly (Elsevier, Sage, Springer Nature, Taylor & Francis, and Wiley) in the scholarly publications system, with even higher shares than the ones reported in Larivière, Haustein, and Mongeon’s seminal 2015 study of publisher concentration in the Web of Science content, which averaged slightly above 50% in 2013. These variations can in part result from differences in methodology and scope. While Larivière et al.’s study examined documents throughout the entire WoS content, our study looks at publisher concentration at the journal level and only within a subset of the WoS content (i.e. the SCIE and/or SSCI indexes). Next to the oligopoly, our results also show that the IF brand favors well-established and historical publishers and scholarly societies (ACS, APA, BMJ, CUP, De Gruyter, Emerald, OUP, Wolters Kluwer), or more recent publishers with scaling-up capacities and infrastructures that specifically target international audiences (Bentham, Frontiers, MDPI, IOP Publishing, World Scientific Publishing). This is most likely due to several factors, including historical reasons, discoverability, indexing, and visibility issues, as well as the curation process and inclusion criteria used by the WoS, which notably require that all titles and abstracts must be translated into English to be included in one of its indexes. In any case, encouraging publishing in IF journals contributes to the market consolidation of particular commercial publishers that are mainly located in the Global North.

Similarly, promoting IF journals as publishing venues is bound to hamper the development of multilingual or non-English content in research given the very high shares of IF-journals accepting papers in English (>90%). This result is in line with a previous study showing the proportions of documents in English at 95.37% in the WoS database (Vera-Baceta, Thelwall, and Kousha 2019, 1806). By comparison, a study analyzing the 25,671 active journals employing the open-source publishing platform Open Journal Systems (OJS) reports a proportion of journals using English as a main language at 49.7% (Khanna et al. 2022). This comparison shows the imbalance of English-only journals in our dataset while pointing to a rescaling possibility of the use of languages other than English in scholarly communications. Of course, the use of a lingua franca such as English can be beneficial for reasons of global dissemination. But it should not prevail to the detriment of research in other languages, which can, as the Helsinki Initiative on Multilingualism in Scholarly Communication reminds us, benefit society “beyond academia”, particularly when it is related to issues of “heritage, culture, and society”. In fact, a “balanced multilingualism” in research (Sivertsen 2018) can keep “locally relevant research alive” as well as potentially “create localized impacts” (Kulczycki et al. 2020, 1371–72).

Several reasons can be advanced as to why the IF brand does not currently achieve this balanced multilingualism by privileging English. First, the IF was historically created and “specifically to cater to the needs of US librarians” (Archambault and Larivière 2009, 637). It later developed through “reshap[ing] international science in favour of both the US and the English language” (Archambault and Larivière 2009, 644). Second, much of the globalization and internationalization of research since the second half of the 20th century onwards has made English increasingly hypervisible (Alatas 2022; Galloway and McKinley 2021; Wilkinson and Gabriëls 2021). Third, the increasing use of English-centered databases such as Scopus and WoS in research assessment may influence the publishing agenda of researchers, who “may choose to move away from locally relevant research toward decontextualized approaches of interest to English-language audiences (Lopez Pineiro and Hicks 2015)” (Kulczycki et al. 2020,
Finally, the low proportion of OA journals not requiring fees (6.1%; n=761 – cf. Fig. 4) may also in part explain the current English-centeredness of the IF brand. In a recent study on the state of Diamond OA journals (i.e. OA journals not requiring fees), it has been observed that Diamond OA journals “are much more multilingual” than “APC-based” journals (Bosman et al. 2021, 7). The previously referred to study of OJS journals seems to indicate a comparable result with higher shares of non-English journals and an overall proportion of 84.2% journals using an OA Diamond model (see Khanna et al. 2022).

The very low share of no-fees OA journals observed here (6.1%) is particularly questionable and worrying when compared to the proportion of APC-based OA journals (14.9%; n=1849). This ratio of OA journals with or without fees indeed sharply contrasts with that of the DOAJ, which features more than 12,959 no-fees OA journals and only 5,886 OA journals requiring APCs. Even if all OA-journals for which the presence of fees could not be confirmed (n=303) were considered to be Diamond OA journals, the share of OA-APC journals would still by far surpass the share of Diamond OA journals in the present study.

The IF brand can thus be said to privilege a minority model (OA-APC) over a fairer, more equitable, and more frequent OA publishing model. This can be in part explained by the prevalence of big commercial publishers of the Global North (cf. Fig. 5). Even so, however, the share of OA-APC journals reported here entrenches a publishing model supporting exclusionary politics of publishing, which discriminate against researchers from low-income countries, independent scientists, or other scholars unable to finance APCs (Asai 2020; Babini et al. 2022; Demeter 2020; Klebel and Ross-Hellauer 2022). This is particularly true when said APC prices, especially as they are practiced by the top 25 publishers and journals of the Global North (cf. Fig.7 and Fig.8), surpass by far the actual costs to publish scholarly articles (Grossmann and Brembs 2021). This APC barrier, in turn, adds another bias in the international research landscape which yet again privileges particular voices such as senior and male academics in contributing to the hypervisible scholarly record (Albornoz, Okune, and Chan 2020; Matheka et al. 2014; Olejniczak and Wilson 2020; Segado-Boj, Martín-Quevedo, and Prieto-Gutiérrez 2018; Siler and Frenken 2020; Smith et al. 2021). Finally, the prevalence of the OA-APC model manifest in the IF brand also raises questions of sustainability since it has been demonstrated that there exists a positive relationship between APC prices and journals with a high IF (Demeter and Istratii 2020; Siler and Frenken 2020), which encourages APC inflation.

The IF-brand also de-incentivizes fair, equitable, and sustainable open access publishing by privileging the hybrid publishing model, which most journals rely on according to our data (cf. Fig. 4). Again, this may be due to the significant numbers of big commercial publishers within the IF list, as the top 25 publishers indeed publish more than 90% of all hybrid journals. Whatever the reasons behind the number of hybrid journals, their sheer number contributes to the same exclusionary politics as journals using an OA-APC model, even more so given that the APC levels of hybrid journals reported here (cf. Fig.7 and Fig.8) confirm previous analyses showing that they are more expensive than those practiced by OA-APC journals (Laakso and Björk 2016; Pinfield, Salter, and Bath 2016; Solomon and Björk 2016; Butler et al. 2022). These high prices may also play a role in perpetuating the hybrid model as well as the possibility for publishers to double dip, i.e. to charge institutions for subscriptions and authors for fees for optional OA publishing (Eve 2014; Pinfield, Salter, and Bath 2016; Butler et al. 2022).

---

17 As of January 13th 2023.
The significant shares of OA-APC and hybrid journals observed in the IF brand (cf. Fig. 4) should also raise further sustainability and equity concerns as the number of articles delivered through these methods has increased over the last few years. In his yearly studies of Gold Open Access output based on DOAJ data, Crawford has shown that the minority of OA-APC journals included in the DOAJ produce a majority of the total output of OA articles, i.e. OA articles published in either Diamond or fee-based OA journals included in the DOAJ (see Crawford 2021; 2022). In 2020, the minority of OA-APC journals produced 65% of the overall output of OA articles (Crawford 2021, 3), a proportion which increased to 69% in 2021 (Crawford 2022, 2). As for hybrid journals, several studies have shown an uptake of hybrid articles published by the Oligopoly. Laakso and Björk (2016) reported a number of hybrid OA articles doubling every year between 2007 and 2013. More recently, Butler et al. (2022) demonstrated a “slight increase” of hybrid OA articles “from 29% in 2015 to 32.4% in 2018” (Butler et al. 2022, 14). This increase of OA articles through the hybrid model should nevertheless be put into perspective as it does not seem to be significant enough to allow journals to transition to full OA in a reasonable timeframe (Piwowar, Priem, and Orr 2019), thus prolonging the unsustainability and inequity of the model, notably as it is implemented through so-called transformative agreements (see Farley et al. 2021; Nous 2021)

The fact that journals included in other WoS indexes will soon receive an IF (Clarivate 2022) might possibly slightly reduce the significance of some of the hegemonic and oligopolistic trends discussed above. Journals included in the Arts and Humanities Citation Index (AHCI) might, for example, improve the representation of the Humanities within the IF brand. This index might also improve the linguistic diversity of the IF brand since multilingual publishing “is an ongoing practice in many SSH research fields regardless of geographical location, political situation, and/or historical heritage” (Kulczycki et al. 2020, 1371). The addition of journals from “more than 3,000 publishers, many of which are smaller publishers from the developing world” (Clarivate 2022), might also reduce the share of English-only journals with an IF, especially if these publishers show more linguistic diversity, as does the category of “Other publishers” in this study (cf. Fig. 3). These publishers may also contribute to reducing the domination of the Global North (cf. Fig.1) and trends of publisher concentration (cf. Fig.1 and Fig.2). Finally, the assignment of an IF to almost 9,000 journals may increase the share of OA journals by 8% in the JCR as per Clarivate’s press release (Clarivate 2022). But it is unknown what proportion of the OA journals will be APC-based.

6. Conclusion

The quest for prestige or perceived quality through the IF brand poses serious threats to biodiversity in scholarly communications. On the whole, our results indeed show that the IF brand perpetuates publishing concentration, maintains a domination of the Global North and its attendant artificial image of mega producer of scholarly content, does not promote linguistic diversity, and de-incentivizes fair and equitable open access by entrenching fee-based OA delivery options with rather high APCs.

Given the prevalence of the IF in research assessment and its ties to issues of professional advancement or career development (Berenbaum 2019; Casadevall and Fang 2014; Fyfe et al. 2017; Gingras 2016; Larivière and Sugimoto 2019; Leclercq 2016; Moher et al. 2018; Morales et al. 2021; Rice et al. 2020; Verma 2015), this problematic state of biodiversity should be properly reckoned with by the scholarly community to minimize the adverse effects of journals’ lists on the diversity of the publishing landscape and its future developments. In terms of
research, this implies a further and regular examination of how the IF and other journals’ lists used for evaluation tackle bibliodiversity over time. In terms of advocacy and policy-making, this implies that bibliodiversity issues be addressed in research training and research evaluation programs and reforms. A first step in this decolonial endeavor may be to recognize the increasing “intertwining of research assessment and open science” (Curry 2022, 100), which notably transpires from the UNESCO recommendation on open science (UNESCO 2021).

Acknowledgments

The authors would like to thank Hollie White for her assistance in the conceptualization of this study and Elisa Grégoire, Sabrina Lekhder, and David Tredinnick for their contribution in data collection.

Works Cited


