The X Factor: Open Access, New Journals, and Incumbent Competitors*

W. Benedikt Schmal†

DICE, HHU Düsseldorf & Walter Eucken Institute, Freiburg

February 13, 2024

The academic publishing market is highly lucrative and controlled by a few dominant players. This paper evaluates to which extent a journal's reputation poses a barrier to entry for new and competing journals and publishers. I leverage a quasi-experimental variation created by Elsevier to measure citation differentials – a core indicator for market success. Relying on the editorial process of their 'parent journals,' the publisher Elsevier launched 'X journals,' fully open-access derivatives of their established 'parents.' In parallel, Elsevier continued to offer an open-access option for publications in the latter outlets. Exploiting this threefold variation, a significant and adverse effect on citations exists for the novel journals. This 'X factor' represents a non-negligible entry barrier for potential competitors. One possible way to mitigate this could be strict open-access requirements for publications supported by research grants. My findings highlight the challenges to more open access and stronger competition in the academic publishing market.

JEL Classification: *D43, L13, L17, L86*
Keywords: *Open Access, Academic Publishing, Market Entry Barriers, Elsevier, Citations, Mirror Journals, Research Grants, Transformative Agreements*

---

*I am grateful for the hospitality at the Department of Management, Strategy, and Innovation (MSI) at KU Leuven, where this project has been developed. I am further thankful for valuable feedback from Oliver Budzinski, Justus Haucap, Ulrich Herb, Leon Knoke, Maikel Pellens, Michael Rose, Chris Snyder, Thomas Steger, Leonardo Urrutia, Elena Veretennik, Ludo Waltman, and Rainer Widman as well as from participants at REGIS Summer School 2023, the RISE6 workshop at Max Planck Institute for Innovation and Competition, the CWTS Research Seminar at Leiden University, the Leipzig University Lunchtime Seminar, the Eucken Institute's internal seminar, and the DICE PhD research workshop. All errors are my own. I gratefully acknowledge financial support from the German Research Foundation DFG (235577387/GRK 1974).

†Walter Eucken Institute, Goethestraße 10, Freiburg 79100, Germany. schmal@dice.hhu.de

1 Introduction

Open access to scientific research is a long-cherished dream that may become a reality in the digital age. There are no longer physical barriers to disseminating new findings. Public bodies such as the US White House propose open science to "provide access to the results of the nation's taxpayer-supported research, accelerate discovery and innovation, promote public trust, and drive more equitable outcomes."[1]

Nevertheless, researchers remain skeptical regarding the quality of such publications. If one enters the query "are open access" into the search engine Google (see Figure F1 in the appendix), the autocomplete function suggests completing it as 'are open access journals peer-reviewed/bad/free/credible/cited more.' Not only does the reliability of open access publications remain unclear, but the proposals of the search engine also address the often stated 'citation advantage' (see, e.g., Wang et al., 2015) of open access publications due to the absence of paywalls. In any case, open access has become a large business model for commercial publishers.

I utilize the unique quasi-experimental situation of Elsevier temporarily operating fully open-access derivatives of established journals. It allows me to gain novel insights into the market entry barriers for new publishers and journals. By studying citation differentials between access types, I can show that incumbent journals possess high market power, which makes it unlikely that the oligopolistic market structure of this multi-billion Dollar market will change soon. Distinctive of my data is that I can net out any variation in quality while investigating 70 outlets and more than 120,000 publications across several disciplines. By that, it is hitherto the first paper to entirely eliminate the quality confounder in research examining access types of academic research. Furthermore, the paper illustrates the anti-competitive effects of the stellar importance of a journal's reputation for authors. It is the main hurdle for breaking up the oligopolistic market composition in scientific publishing.

[1] See the announcement of the Biden-Harris administration, https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research, published on January 11, 2023.

The scientific publisher Elsevier launched in 2019 a new open-access option by creating open-access forks of leading journals – the so-called 'X journals.' The distinctive factor of these outlets was that *"[t]he editorial and peer review process is identical for the parent journal and the OA mirror journal. . . . During the submission process and just before acceptance, authors can choose whether to publish in the parent journal or the OA mirror journal . . . [which] have the same title as their parent journal, distinguished by the letter X"* (Harrison, 2019). The journals selected to receive a 'twin' usually operated with subscriptions but offered authors to publish an accepted article with open access for an extra fee ('hybrid' journals). Furthermore, researchers could always store their papers in freely accessible repositories.

The complex setting provides me with a quasi-causal setting as I can compare *two* types of open access with restricted access requiring a subscription to access the paper. Using Poisson regressions to compute the publication option's impact on a paper's citation count, I find no difference between publications in incumbent journals with any kind of open access to them and restricted access but a significant negative relation between a paper being published in an X mirror journal and its citations. The findings reemphasize the relevance of a journal's reputation for both the recipients and, ultimately, the authors. Given the enormous number of journals nowadays, a simple 'X' might be already enough to create the impression of a different outlet. It may decrease the willingness of authors to submit to such journals. While Elsevier already discontinued many X journals, this may apply even more strongly to newly set up open access journals that shall pave the way to open science because they do have not only different names but also other editorial boards than the leading (hybrid) journals.

This paper is related to earlier work on the effect of open access on citations. Many publications refer to the early findings of Lawrence (2001) and Eysenbach (2006) who identify the so-called 'open access citation advantage.' However, these papers rather study online compared to print availability. Already a randomized controlled trial (Davis et al., 2008), as well as an early literature review (Craig et al., 2007), do not find any citation advantage anymore. McCabe and Snyder (2014, 2021) introduce journal fixed-effects to control for unobserved heterogeneity in the outlets and found these effects to control for most of the citation advantage found in earlier studies. They also identify a strong positive effect for the best journals but a significantly negative effect for weaker outlets. Furthermore, the effects of open access to the final versions of a paper in a journal are blurred by the easy and broad dissemination of working papers, pre-, and postprint versions via online repositories such as SSRN or arXiv (McCabe & Snyder, 2015).

This paper adds to the literature as it investigates differences between fully and hybrid open-access papers as well as papers requiring a subscription while completely netting out quality differences in the journals. Related to my approach is the study of Wang et al. (2015) about the switch of *Nature Communications* from restricted to open access. However, their work only compares point estimate averages without statistical testing. My paper investigates the availability effect of open access but also the reputation effect of a journal on citations. It further examines the relevance of the geography of affiliations as well as funding for choosing open access and the number of citations a paper receives. I identify an individual citation abatement for less well-established X journals, which strengthens the position of the large incumbent commercial publishers with their large stock of settled journals. However, strict open-access requirements of prestigious grants might not only enable a larger audience to read novel and meaningful findings but could also strengthen new market entrants among the journals, as my analysis highlights.

The remainder of this paper is structured as follows. In Section 2, I sketch the theoretical background of the dichotomy of X and parent journals and propose an economic mechanism of researcher behavior. Section 3 provides descriptive statistics and describes my empirical approach. Section 4 shows my results. In Section 5, I discuss the economic implications of my findings. Section 6 concludes.

2 Theoretical Background

The positioning of the 'X' journals: In general – and without the special case of X journals, there exist three types of journals by access type:

I Subscription-based journals that one can only access with a license that has to be acquired individually or, in most cases, by an institution. Examples of this are the journals of the American Economic Association, e.g., the *American Economic Review*.

II Hybrid journals that require a subscription but offer the authors to purchase open access to their paper for a fixed price. Many established journals nowadays have such a business model. An example of that is *Research Policy*. Here, the default is the requirement of a subscription to access a publication, but authors may purchase open-access to their paper by paying a fee of 3,710 USD + VAT.[2]

III Fully or gold open-access journals, in which every paper is published with open access by default and not upon a special order as in a hybrid journal.[3] An example is interdisciplinary natural sciences journal *PLoS One*.

Aside from open access to the final publication via hybrid or gold open-access, there exists also the option of green open-access, i.e., free access to a version of the paper stored in a freely accessible repository or on a website.[4] The benefit for readers is that they do not need to have a subscription or purchase the paper they want to read. The downside is that it is less convenient to search for a freely accessible version on a platform different from the journal or publisher website. Furthermore,

[2]See `https://www.elsevier.com/journals/research-policy/0048-7333/open-access-options`, last checked July 10, 2023.

[3]'Gold' or 'full' open access refers to final publications without any access barriers. In contrast, 'green open access' means that for a final publication behind a paywall, a copy or earlier version is uploaded on a publicly available repository. While papers published with open access in a hybrid journal can also be considered gold open access, gold open access *journals* exclusively publish papers with open access in contrast to hybrid journals.

[4]See, for example, Schmal (2023) for more details on the open-access 'colors.' Examples for such repositories are arXiv, ResearchSquare, or SSRN.

they have to evaluate whether the 'green' version is the same or similar to the final version published in the journal.

The only source of revenue for fully open-access journals is publishing papers, as they cannot sell subscriptions by construction. It raises the gloomy incentive for editors to accept additional papers or for a publisher to pressure the editorial boards to accept more papers, which inevitably lowers the average quality of the published papers (McCabe & Snyder, 2005; Armstrong, 2015). Elsevier's introduction of X journals ties in with hybrid and fully open-access journals. While the existing parent journals usually were hybrid journals requiring a subscription but offering optional open access, the X derivatives became fully open-access journals. While there is no official reason why Elsevier introduced this second open-access channel, strong suggestive evidence exists in the institutional environment and the announcement by Harrison (2019). In September 2018, the so-called 'cOAlition S' proposed 'Plan S' to push forward open access in academic publications. It was supported, among others, by the European Commission and the European Research Council, both major funders of research.[5] Principle #8 of 'Plan S' states that "The Funders do not support the 'hybrid' model of publishing."[6] Shortly later, Elsevier launched the first pilots for its X journals. One reason for their introduction has been that "[t]he OA mirror journals have not been launched with any one funding body in mind, but over the last two years, we have seen an increase in funders focusing on fully gold OA journals. We therefore hope that the OA mirror journals will provide another option that authors and funding bodies can consider."(Harrison, 2019)

Even though neither this quote nor the whole article refers explicitly to 'Plan S,' it stands to reason that Elsevier attempted to circumvent the funding ban for open access in hybrid journals by establishing fully open-access clones that fall under the conditions of the initial 'Plan S' and similar approaches. Already the first revision

[5]See the press release on September 4, 2018, `https://www.coalition-s.org/coalition-s-launch/`, last checked July 10, 2023.

[6]See `https://www.coalition-s.org/plan_s_principles/`, last checked July 10, 2023.

of ‘Plan S’ in May 2019 specifically pointed out that mirror journals are considered hybrid journals and are, therefore, excluded from funding.[7] It is likely one reason why the X derivatives never succeeded in the number of publications. Nevertheless, their existence provides me with a unique setting of three different access types that arguably have the same quality due to the same journal scope and, more importantly, the same editorial board.

After introducing the ‘X’ option, scholars interested in publishing their work with open access could choose between the full open access ‘X’ derivative or open access within the subscription-based parent journal. Figure 1 sketches this. I label the combination of the parent journal and its mirror ‘journal compound’ as it aggregates the three types of access: *restricted access* and *subscription-based open access* within the parent journal, and *X open access* within the mirror X journal. As cited above, editorial boards and the peer review were the same for the parent and the mirror journal. By that, the variation between restricted access and subscription-based open access publication is accessibility. The variation between restricted access and X open access is accessibility and a different name (the added X), including a different ISSN identifier, which also led to different citation metrics. The only difference between hybrid and X open access is the variation in the journal’s name and the quantitative reputation measures (as between RA and XOA), because the qualitative assignment of reputation was meant to be adopted from the established parent journal.

As reported by Asai (2023), the prices for X and subscription-based open access were nearly always the same in the initial year 2019 but varied considerably afterwards. Elsevier stopped the X experiment after a short period for many journals and returned step by step to only using the hybrid publication model (Shortliffe & Peleg, 2019) and discontinuing many of the X derivatives. While a couple of X journals still exist in the ‘tradition’ of a mirror journal, others have separated from their

[7] See for details `https://www.coalition-s.org/rationale-for-the-revisions/`, published May 31, 2019, last checked July 10, 2023.

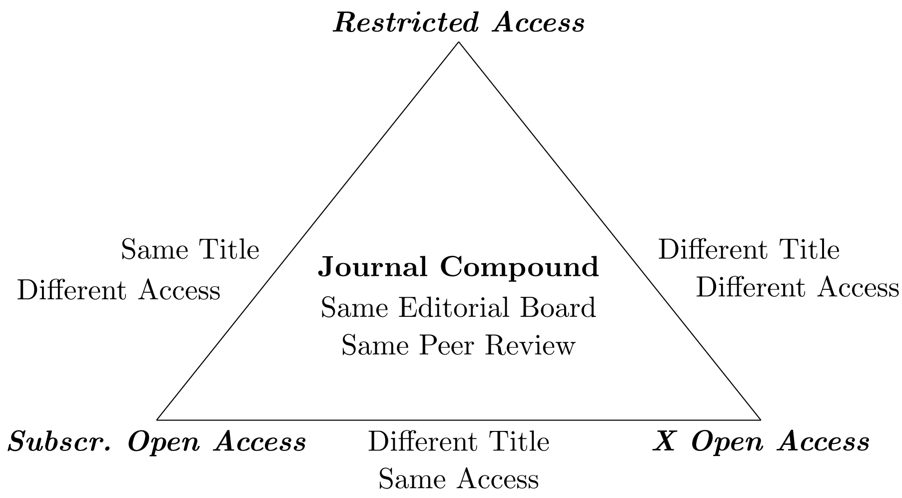


Figure 1: Structure of the three access options within a journal compound

parent journals by electing new editors – even though a notable overlap between the editorial boards often remains.

Economic mechanism: Researchers have a dual role in academic publishing. On the one hand, they are authors of papers. On the other hand, they are readers. As authors, they want to publish their papers – optimally in a highly respected journal in their field or else in the highest journal possible with respect to the scope and methods of their paper.[8] As recipients, it is essential to read material of a high quality that benefits their work and that they can rely on. By construction, only a tiny share of papers can appear in top-ranked publications that everyone in a field knows, so many publications appear in journals that may be less established and known within a field or across fields. When referring to such papers, researchers must ensure that it is work of decent quality.

As there exist thousands of journals, researchers need heuristics to accelerate the evaluation of a paper's outlet when conducting, for example, a literature review. An evident approach is an evaluation of the author or group of authors. For the group of

[8]There exist alternative incentives such as generating a lot of public attention or publishing many papers in a short period of time. Both may lead to journal choices that do not only take into account the ranking of a journal. However, it is reasonable to assume that a journal's reputation is for many researchers likely to be the strongest incentive to publish there.

persons a researcher personally knows, evaluating their ability and the corresponding (expected) quality of the paper is relatively easy. However, when personal relations do not exist, researchers may use other approximations to evaluate the inherent ability of a researcher and the derived quality of their paper. One indicator is an author's affiliation, as many consider a publication of researchers from a prestigious university as a high-quality publication simply based on the institutional reputation. While this is a granular approach, it comes with high arbitrariness in the cut-off decisions.[9] To avoid this I use a broader but clear-cut evaluation based on geography.

For this geographical distinction, I use the emerging separation between the so-called 'global North' and the 'global South,' which implies a distinction between developed countries with high GDP per capita (North) and developing countries (South). The distinction may also reflect reputation of and tacit trust in the higher education systems and, by that, in the quality of the research output. All things equal, it may be easier for publications from the global North to gain interest and, subsequently, citations relative to publications coauthored by researchers based in the global South. In addition, I distinguish along the lines of the World Bank's income classification. Lastly, I use the age of the university system as separating factor and approximation for trust in the research output.

Instead of looking at the characteristics of the authors, a conceivable alternative is evaluating the journal in which the paper in question is published. To do so, journal rankings such as the Journal Impact Factor (JIF) may serve as a quick and convenient way to get an impression of an outlet's reputation, as such measures may carry more information than the mere count of citations of a paper does (Laband, 2013; Osterloh & Frey, 2020) – even though critics argue that metrics such as the JIF are essentially constructed based on aggregating citation counts, which carries its own problems (Wooding, 2020). Nevertheless, the number of citations might not fully reflect the actual rigor of a paper but also how topical its content is or

[9] For example, it is unclear how many universities are considered as leading or how to capture the composition of author groups in terms of top-, mid- or low-tier institutions.

whether it contributes to a highly debated issue. Moreover, in their role as authors, researchers use rankings as well. Here, they use it to decide where to publish their work (Śpiewanowski & Talavera, 2021). Hence, a higher journal ranking is likely to increase the credibility of a research paper, which, in turn, should foster citations.

A third driver of citations is the availability of research. One can only read work that is accessible. Hence, for every paper published with access restrictions, the researcher must have a compatible license to read the paper.[10] This might not always be the case, especially for less essential journals or those with a strong regional focus. Open access may enable more researchers to read and cite one's work in such situations. It should, therefore, be an upward driver of citations, regardless of the journal's quality.

$$Citations = f(paper\ quality,\ journal\ reputation,\ accessibility)$$

One can aggregate the sketched drivers in a simple form that describes citations as a function of a paper's underlying quality, which is often difficult to observe directly.[11] Furthermore, it is a function of the reputation of the journal it is published in. Last, it is considered to be a function of accessibility. Clearly, all three factors increase the number of citations if they go up. If a paper is published under an open-access license, it increases its availability.

The journal's reputation channel is more ambiguous. In my setting, the distinctive factor is the overall reputation correlated with the open access *type*. Here, a 'birth defect' of Elsevier's X journals plays a significant role. Due to their (assumed) intention to circumvent the ban of hybrid journals from open-access funding, the publisher needed to establish independent journals with their own identifying ISSN number that differs from the respective parent journal. This, however, also im-

[10] A possible but illegal workaround of paywalls is using predatory repositories such as 'sci-hub.'

[11] In addition, it is, of course, debatable what 'quality' actually means. Ellison (2002) suggests a distinction between 'q' and 'r' quality, where the former means the impact of the main ideas carried in the paper. The latter encompasses the other aspects that may be typically requested by referees, such as the technical rigor of a paper, e.g., robustness checks of empirical estimations or generalizations of theoretical models.

plies an independent journal impact factor and many other metrics. Put differently, although the journals borrowed their names (except for the additional 'X,' the editorial board, and the rigor in peer review from their established incumbent parents, they were formally new journals. Hence, they had no long history of publications, no journal impact factor, and a name that also varies at least slightly.

Adding an 'X' to the name might appear negligible at first sight. Looking at the discipline of economics, one quickly sees that small differences may have large implications. While the *Economic Journal* – hosted by the British Royal Economic Society – is a leading outlet, this does not necessarily apply to the '*Economia Journal*' from the Latin American and Caribbean Economic Association. The *Journal of Economics and Statistics* is hardly similar to the *Review of Economics and Statistics*. The *Eurasian Economic Review* has different standing than the *European Economic Review*. These and many other examples sound highly similar and often share the same abbreviations but differ conspicuously in scope, method, and rankings. In a world with thousands of similar-sounding journals, adding an 'X' is not just a further letter but might imply a wholly different journal.

3 Descriptive Statistics and Empirical Strategy

Data and Descriptive Statistics: The core of the data forms a list of 35 journal compounds, i.e., the pairs of the established 'main' journals and its 'X' fork. The set is based on the list used by Asai (2023) as well as on a manual search of Elsevier's journal library. I have retrieved the metadata of all publications in these compounds from 2018 until 2022 directly from the Scopus database using the *pybliometrics* wrapper of Rose and Kitchin (2019). It carries the benefit that Elsevier hosts all journals and the database. Thus, it is likely that the publication records are highly accurate. Starting with 128,364 publications, I only proceed with the paper types 'article,' 'review,' and 'note.' The three categories account for 97.1% of all records.

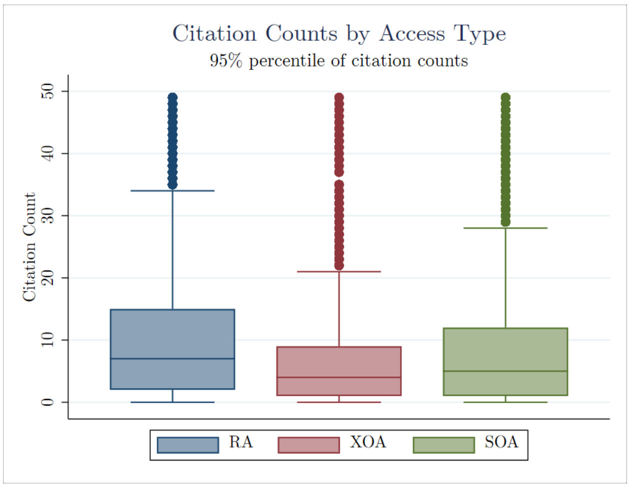
Furthermore, letters to the editors, errata, and editorials have other functions than disseminating novel research, so I abstain from including them. In total, I use 123,939 publications in 70 journals or else 35 journal compounds.[12]

Year	Restricted Access	X Journal OA	Subscription OA	Total
2018	21,114	11	1,941	23,066
2019	21,560	714	2,063	24,337
2020	21,568	565	2,358	24,491
2021	21,807	516	3,301	25,624
2022	21,779	698	3,944	26,421
Total	107,828	2,504	13,607	123,939

Table 1: Number of publications by access type and year

Table 1 shows the distribution of the papers across the five years included in the analysis. There are few X journal observations in 2018, when Elsevier experimented with the new format, followed by a steep increase in 2019 when the X journals officially entered the market. Open access (green or hybrid-gold) to publications in the incumbent parent journals (SOA) occurs much more often. The lion's share of publications in the present sample is still published without open access, though.Table 2 shows all 35 journal compounds and the number of publications differentiated by the publishing type. It also lists whenever an X journal has been discontinued in the time window of the analysis or whether it separated from its parent journal in a way that it got its own editorial board and scope. I obtained the publication data from March 20-30, 2023.

[12]I manually verified that the journal compounds continue to share the same editorial board, aims and scope, and peer review process. I remove those records of X journals that separated from their parent journals. This affects *Atmospheric Environment X* from 2021 on, *Biosensors and Bioelectronics X* from 2020 on, *Food Chemistry X* from 2022 on, and the *Journal of Asian Earth Sciences X* from 2020 on. In total, I lose 702 observations by that. Furthermore, the *International Journal of Pharmaceutics* and *Vaccine* both have X derivatives that do not share the same editorial board but boards with extremely high overlap. I keep them in the sample but run a robustness check that shows that excluding both journals does not qualitatively affect my results.



The 95th percentile of the overall distribution of citation counts corresponds to 49 citations. Meaning of the abbreviations: RA = restricted access, SOA = subscription-based open access, XOA = X journal open access.

Figure 2: Boxplots for the number of citations by access type

The core variable of interest is the number of citations a publication has received. The mean citation rate is 13.89, the median is 7, the minimum value is, obviously, 0, and the maximum is 5,946 for a paper by Harris et al. (2019) about a global data platform for medical research. Figure 2 highlights that the differences in the number of citations do not only vary in the average count but also the overall distribution, as one can draw in particular from the inter-quartile range and the upper adjacent values. Radicchi et al. (2008) find an overall 'universality of citation distributions' across fields, meaning that the accumulation of citations varies across fields, but the overall pattern is similar. Waltman et al. (2012) show in a reconsideration that this universality holds for many but not (almost) all fields. Nevertheless, they find that one comes closer to the initial hypothesis once papers without citations are excluded. I keep the zero-cited papers in my data but conduct a robustness check excluding them, confirming the results.

Compound	RA	XOA	SOA	Total	Notes
Analytica Chimica Acta	4,068	27	318	4,413	
Atmospheric Environment	2,605	96	483	3,184	X sep. in 2021
Biosensors and Bioelectronics	3,821	10	325	4,156	X sep. in 2020
Chaos, Solitons and Fractals	3,710	51	158	3,919	
Chemical Engineering Science	3,629	98	211	3,938	X disc. in 2023
Chemical Physics Letters	3,968	25	176	4,169	X disc. in 2021
Computers and Graphics	563	8	86	657	X disc. in 2019
Contraception	677	83	164	924	
Cytokine	1,373	36	172	1,581	
Ecological Engineering	1,314	15	164	1,493	X disc. in 2021
Energy Conversion and Mgmt.	5,535	307	413	6,255	
European Journal of Obstetrics...	2,003	134	150	2,287	
Food Chemistry	10,581	120	883	11,584	X sep. in 2021
Gene	3,451	30	218	3,699	X disc. in 2021
Intl. Journal of Pharmaceutics*	4,203	125	490	4,818	
Journal of Asian Earth Sciences	1,631	8	58	1,697	X sep. in 2020
Journal of Biomedical Informatics	91	20	896	1,007	X disc. in 2020
Journal of Computational Physics	3,152	68	262	3,482	X disc. in 2023
Journal of Dentistry	817	13	201	1,031	X disc. in 2020
Journal of Hydrology	5,241	93	658	5,992	
Journal of Non-Crystalline Solids	2,572	103	181	2,856	
Journal of Pediatrics	2,033	23	1,202	3,258	
Journal of Structural Biology	473	72	159	704	
Materials Letters	9,083	127	232	9,442	
Nutrition	1,222	9	174	1,405	X disc. in 2021
Optical Materials	4,732	164	126	5,022	
Research Policy	609	3	220	832	X disc. in 2020
Resources, Cons. and Recycling	2,060	29	452	2,541	X disc. in 2020
Respiratory Medicine	333	24	877	1,234	fully disc. in 2021
Sleep Medicine	1,471	46	188	1,705	
Toxicon	1,006	112	137	1,255	
Vaccine*	3,230	204	1,824	5,258	
Veterinary Parasitology	822	29	121	972	X disc. in 2020
Water Research	4,504	127	844	5,475	
World Neurosurgery	11,246	65	383	11,694	X sep. in 2022

RA = restricted access, SOA = subscription-based open access, XOA = X journal open access.
* marks journal compounds with highly similar but not equal editorial boards for X and parent journal. 'disc.' is the abbreviation for 'discontinued,' 'sep.' abbreviates 'separated.'

Table 2: Number of publications by journal compound and access type

Empirical Strategy: I apply Poisson regressions to regress the number of citations, which are count data, of a particular paper on a categorical variable for the access option and further covariates.[13] First, I distinguish between restricted and open access in general by pooling subscription-based and X journal open access. In the main specification, I use a categorical specification with three outcomes – restricted subscription-based access (RA), open-access to a paper in a parent journal (SOA), and X open-access (XOA) – to estimate the effect of the two open-access options

[13]In the case of overdispersion, a negative-binomial specification might appear preferable. However, a Poisson estimation bears the significant advantage that its coefficients remain consistent even if the distributional assumptions are violated. The negative-binomial specification requires much stronger distributional assumptions and becomes inconsistent when they are not met.

relative to restricted access.

Regarding the control variables, I follow the approach of McCabe and Snyder (2014), who add age and age^2 to capture the nonlinear time dimension of citations where I define $age = 2023 - publication\ year$. One shortcoming of my data is that I can only use cumulative citations per paper and no citation windows that analyze the number of citations within a fixed time window. Wang (2013) finds a window of three years to be most informative in most cases. I address this by two sets of regressions that study solely the year 2019 for a three-year citation window or else the year 2020 for the two-year equivalent.

I also control for the number of authors as more authors may lead to more citations since the paper might be more visible. I also add a journal fixed effect but use the journal compound, i.e., the combination of parent and X journal, as the journals are rooted in different disciplines and within these disciplines in different relative reputational quantiles. I cluster the standard errors also on the journal compound level. Robustness checks with bootstrapped standard errors confirm this choice of clustering. I also compute selected regressions using ordinary least squares (OLS) instead of Poisson. Again, my findings are qualitatively the same. To assess the relation between third-party funding and open access, I conduct binomial and multinomial logit regressions using access types as dependent variables.

Overall, the identifying assumption of my setting relies on the proposition that the establishment of the X option of publishing occurred rather randomly. As mentioned beforehand, the introduction of the 'Plan S' together with the suggested criteria for funding open access options, was unrelated to the journals and based on general developments in the publishing market. Thus, the *timing* of the introduction of the X journals can, indeed, be considered idiosyncratic.

In contrast, the journals that received an X companion appear to have been chose consciously by Elsevier. Table A1 in the appendix shows the list of journals, their journal-specific H Index, and the quartile within the SCImago Journal ranking for each field.[14] Inspecting this list, Elsevier focused on journals leading in their fields. The set of journals seems widely homogeneous in their relative quality.

Lastly, X journals might have been chosen by different types of researchers compared for the incumbent parent journals, even though aims, scope, and peer review were the same. I cannot rules this out but it does not invalidate the findings as they address the competition implications, which apply also or especially if author characteristics varied.

4 Results

4.1 Baseline results

This section presents the results for the relationship between the number of citations and the type of access to a paper. Every table presents a battery of regression results in which I step-wise add regressors. In general, the abbreviation OA refers to both types of open access in contrast to restricted access (RA). XOA refers to open-access publications in X journals, while SOA refers to open access in subscription-based parent journals. In any case, I restrict my sample to the X journals and their parent companions, i.e., there are no journals in my sample that did not have an X twin at some point.

I begin with the central part of the analysis: The distinction between open access in subscription-based journals and open access in their 'X twins.' Table 3 presents the findings of my main specification. Once I control for age and the number of coauthors as well as journal compound fixed effects, I detect a citation *disadvantage* of -0.371 for papers published in X journals, which translates into a decrease by 31% or nearly one-third relative to papers published in the subscription-based main journals. Given an average citation count of 13.96, this 'X factor' implies a reduction of 4.33 citations for an X journal paper relative to a restricted access

[14] See Guerrero-Bote and Moya-Anegón (2012) for a technical description of this metric.

publication in a parent journal. This value corresponds to the coefficient of -3.818 estimated in the OLS specification in the sixth column of Table 3, which is slightly smaller but indistinguishable from the computation based on the Poisson regression as $F(1,34) = 0.21, p = 0.65$ for a Wald test of $\beta_{XOA} \overset{!}{=} -4.33$. I can rule out equality of the coefficients for the two types of open access as a Wald test of $\beta_{XOA} \overset{!}{=} \beta_{SOA}$ in the Poisson specification of column 5 leads to a test statistic of $\chi^2(1) = 52,81,\ p = 0.000$ and for the OLS specification to $F(1,34) = 10.59,\ p = 0.0025$.

	Poisson	Poisson	Poisson	Poisson	Poisson	OLS
XOA	-0.565***	-0.574***	-0.305***	-0.374***	-0.371***	-3.818**
	(0.107)	(0.135)	(0.052)	(0.039)	(0.038)	(1.117)
SOA	-0.144	-0.234***	-0.036	-0.021	-0.019	-0.517
	(0.119)	(0.058)	(0.053)	(0.055)	(0.050)	(0.620)
age			0.421***	1.288***	1.292***	8.518***
			(0.011)	(0.068)	(0.068)	(1.112)
age^2				-0.132***	-0.132***	-0.474***
				(0.009)	(0.009)	(0.102)
#authors					0.016***	0.202***
					(0.002)	(0.040)
constant	2.660***	2.755***	1.352***	0.145	0.044	-5.407*
	(0.132)	(0.004)	(0.040)	(0.121)	(0.127)	(2.511)
journal FE	NO	YES	YES	YES	YES	YES
N	123,939	123,939	123,939	123,939	123,922	123,922

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$. Dependent Variable: Number of citations in levels. Standard Errors in parentheses clustered on the journal compound level (i.e., X journals and main journals together). Alternative specification with bootstrapped standard errors in Table A2 in the appendix. Coefficients for X journal and open access publications relative to the reference category of subscription based publications in the main journals.

Table 3: Regression results for the impact of the different access options

Regarding open access in an incumbent parent journal (SOA), one can see that – with one exception in column 2 – open access publications do *not* differ in the number of citations in my sample. In contrast, not only are the standard errors too large for statistical significance, but the coefficients are also close to zero. These findings confirm earlier evidence but contradict the shared wisdom of an open-access citation advantage.[15] Even though the regression without any covariates but including the

[15]Investigating solely hybrid gold open access to papers in the incumbent journals and not every type of open access, I am able to find a small but significant positive effect on citations relative to restricted access. It accounts for approximately 1/3 to 2/5 of the disadvantage for 'X' publications in absolute terms.

journal compound fixed effects leads to a significant negative effect also for this type of open access (column 2), any significance vanishes once adding a linear covariate for age. It is perfectly reasonable as my dataset started in 2018 when the first X journal 'pilots' were run. In the following years, especially in 2019-2021, X journals were a substitute for subscription-based open access such that a larger share of SOA publications is relatively young, given that in 2021 and 2022, Elsevier discontinued many X journals. In response, subscription-based open access became the only publishing option without access barriers. The median publication year reflects the higher share of younger publications of those with hybrid open access: While it is 2020 for X journals, it is 2021 for subscription-based open access. As the parent journals are all highly ranked and well-established in their disciplines, many institutions might have subscribed these outlets in any case. So, the access advantage for (SOA) papers in these journals diminishes.

The triangular setting of both restricted and open access within the subscription-based parent journals and open access in the X derivatives likewise allows digesting the 'X factor' of appearing in a novel journal from the open access effect. Holding quality constant not only econometrically but actually through the same editorial boards and processes, the citation disadvantage for X journal publications should be entirely related to the new name and independent citation and impact measures of the outlets. From a citation perspective, assigning the X forks new ISSN numbers was a mistake, leading to a different journal impact factor and journal ranking. Together with the new name, X journals may have appeared as novel market entrants that challenge the incumbent hybrid journals rather than complement them.

Citation Window Computations: Plenty of bibliometric research has evaluated and discussed the relevance of citation windows when working with citation data, e.g., Glänzel and Garfield (2004), Abramo et al. (2011), Campanario (2011), and Wang et al. (2015). Essentially, it is the question of when to evaluate a paper's impact properly, such that one does not truncate a notable amount of yet-to-be-made citations in the future. Reviewing the literature carefully, two or three years after publication, the accumulated number of citations should sufficiently represent a paper's relevance as measured in citations. Put differently, the average paper should have exceeded its citation peak within one of the two time windows.

	3 year citation window Publications in 2019			**2 year citation window** Publications in 2020		
	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
XOA	-0.418***	-0.309***	-0.309***	-0.353*	-0.303***	-0.301***
	(0.099)	(0.064)	(0.063)	(0.141)	(0.084)	(0.083)
SOA	0.052	-0.034	-0.032	-0.074	-0.150*	-0.149*
	(0.196)	(0.059)	(0.052)	(0.150)	(0.068)	(0.063)
#authors			0.020***			0.016***
			(0.004)			(0.003)
constant	3.001***	3.133***	3.022***	2.734***	2.814***	2.721***
	(0.149)	(0.002)	(0.024)	(0.142)	(0.003)	(0.019)
journal FE	NO	YES	YES	NO	YES	YES
N	24,337	24,337	24,332	24,491	24,491	24,488

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$. Dependent Variable: Number of citations in levels. Standard Errors in parentheses clustered on the journal compound level. Coefficients for XOA and SOA publications relative to the reference category of restricted access publications. coefficients of age and age^2 omitted by construction.

Table 4: Impact of the different access options using quasi citation windows

As mentioned beforehand, to evaluate the number of citations within a citation window, one does need the timing of the citations. The Scopus data I use in this paper do not provide this information but solely gather the number of citations accumulated until the bibliometric data have been retrieved from the database. I did this in late March 2023, so I consider the end of 2022 for calculating citation windows as sufficiently precise, as even the first quarter of the year has yet to be over. Nevertheless, I still face the problem of having only accumulated data. I address it by constructing two 'quasi windows.' Using only the publications in 2019, I essentially have a three-year window for this subset of papers. Equivalently for 2020, I have a two-year 'quasi window.' As I use only one year of observations in both regressions, the covariates for age become obsolete. In addition, in both years, nearly all X journals were still active.

I provide the results for the regressions of these two quasi-citation windows in Table 4. For the three-year window on the left, one can see that the results are qualitatively the same as in the baseline regression. Also, the values of the coefficients are very similar to those shown in the main specification in Table 3. In contrast, those of the two-year window on the right, only relying on the observations from 2020, find not only a significantly negative effect for XOA publications but also a significant and negative effect for open-access publications in the established parent journals. The coefficients of the regressions, including journal fixed effects, are both significant on the 5% level but have half the size of those for the X journals. One has to treat this carefully, but it serves as evidence that open access might increase citations in the medium and long run or, reversely, a citation disadvantage may disappear over time, probably due to better availability than papers with restricted access. It does not apply to X journals, which may reemphasize that reputation disadvantages hardly disappear over time and that this is a fundamental disadvantage for papers published in such journals.

4.2 Extensions

Geographic differences of the authors: To better understand the underlying factors of the disparities in citations, I further extend my analysis by looking at differences between researchers being based at institutions at the global North (developed countries) and the global South (developing countries). Here, I broadly follow the country classification of the United Nations Conference on Trade and Development (UNCTAD).[16] The 'global North' indicator comes along with a broad

[16]See `https://unctadstat.unctad.org/EN/Classifications.html`. It states that *"The developing economies broadly comprise Africa, Latin America and the Caribbean, Asia without Israel, Japan, and the Republic of Korea, and Oceania without Australia and New Zealand. The developed economies broadly comprise Northern America and Europe, Israel, Japan, the Republic of Korea, Australia, and New Zealand."* In addition, I compute various robustness checks using the described alternative distinctions based on income and how established the higher education systems are in appendix B.

understanding of leading higher education systems instead of taking a narrow view of looking only at the top 100 universities (or similar). Nevertheless, one should remember that the analysis looks at the *within* journal compound level, i.e., the same editorial boards handled the papers, and they appear within the same journal compound.

	Broad definition		*Narrow definition*		
	Global North	Other	Global North	Other	Total
RA	12,759	85,510	3,907	94,362	98,269
XOA	582	1,676	168	2,090	2,258
SOA	3,403	8,129	1,034	10,498	11,532
Total	16,744	95,315	5,109	106,950	112,059

The 'global North category includes all publications up to nine authors that have a majority of coauthors affiliated with an institution based in a country of the global North (broad definition) or else all publications up to nine authors where all coauthors are affiliated with an institution based in a country of the global North (narrow definition). 'Other' must not be set equal to 'global South' as it also comprises papers with equal shares of North/South authors (broad definition). Likewise for the narrow definition, 'other' also encompasses papers with a majority of authors from the global North but without unanimity.

Table 5: Publications from countries in the 'global North' by access type

To classify a publication as stemming from the 'global North,' I rely on two methods. First, I set $\mathbb{1}_{global\ North} = 1$ if a *majority* of authors have an affiliation with a Western institution (broad definition). Alternatively, I code a paper as 'global North' only if *all* authors come from an institution in a developed global North country. For obvious reasons, this considerably decreases the sample, as Table 5 shows. For higher tractability, I restrict my sample to publications with up to nine authors. This covers 90.42% of all publications.

The robustness check with bootstrapped standard errors on the LHS of Table A9 in the appendix shows that the base SOA coefficient also becomes fully significant under this specification. However, the point estimate still is much smaller in its absolute size than the XOA coefficient. Also, the interaction effect between coming from the global North and publishing in with SOA outnumbers the negative SOA base effect in absolute terms. Hence, while open-access publications suffer from a

	Poisson	OLS	log OLS	Poisson	OLS	log OLS
	Global North: broad definition			Global North: narrow definition		
XOA	-0.352***	-4.035**	-0.325***	-0.354***	-4.021**	-0.323***
	(0.048)	(1.264)	(0.045)	(0.048)	(1.277)	(0.045)
SOA	-0.077	-1.345*	-0.061	-0.078	-1.318*	-0.056
	(0.048)	(0.611)	(0.040)	(0.046)	(0.599)	(0.040)
North	-0.079*	-0.976	-0.095*	-0.154**	-1.811*	-0.142**
	(0.034)	(0.631)	(0.040)	(0.053)	(0.762)	(0.052)
XOA × North	0.045	1.392	0.110	0.316	4.512*	0.340**
	(0.101)	(1.306)	(0.063)	(0.172)	(2.010)	(0.111)
SOA × North	0.209***	2.692**	0.142**	0.225	3.057	0.203*
	(0.059)	(0.969)	(0.040)	(0.138)	(1.559)	(0.088)
age	1.274***	8.277***	1.063***	1.262***	8.395***	1.068***
	(0.070)	(1.112)	(0.063)	(0.069)	(1.152)	(0.063)
age^2	-0.129***	-0.433***	-0.103***	-0.128***	-0.449***	-0.104***
	(0.009)	(0.110)	(0.009)	(0.009)	(0.112)	(0.009)
#authors	0.024**	0.253*	0.032***	0.022**	0.227*	0.030***
	(0.009)	(0.104)	(0.007)	(0.008)	(0.096)	(0.007)
constant	0.032	-5.230*	0.029	0.062	-5.331*	0.036
	(0.146)	(2.520)	(0.099)	(0.140)	(2.611)	(0.101)
journal FE	YES	YES	YES	YES	YES	YES
N	112,059	112,059	100,302	100,424	100,424	89,944

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Dependent Variable: Number of citations in levels (columns 1, 2, 4, 5), or logs (columns 3, 6). Standard Errors in parentheses clustered on the journal compound level. Alternative specification with bootstrapped standard errors in Table A9 in the appendix.

Table 6: Impact of the affiliation countries of authors: Global North

citation disadvantage in general, this does not hold anymore if a sufficient amount of authors from, on average, better endowed institutions in the global North are involved.

Switching from the broad to the narrow definition of a 'global North' paper, one can see notable differences between the specifications. Now, all authors are affiliated with a institution from a country being part of the global North as defined by UNCTAD. First, one can see on the RHS of Table 6 that the plain covariate for a paper coauthored by a team entirely from the global North is significantly negative in the Poisson specification. Thus, papers from such universities are cited less often than their 'global South' or 'mixed-North-South' counterparts.

Given the general strength of higher education systems in the more developed global North of the world, how is this reasonable? While I cannot provide abundant evidence, one reasonable hypothesis is that the geographical and academic proximity of authors and editors since editors appear to have some home bias (Bethmann et al., 2023; Rubin et al., 2023). In addition, Colussi (2018) shows that any professional relationship between an author and an editor starkly increases the likelihood of being published in a particular journal. Moreover, Rose and Shekhar (2023) find that strong networks of academic advisers accelerate the careers of their graduate students. Aside from academia, jury voting in entertainment contests seems biased by proximity to the candidates (Budzinski et al., 2023).

All things equal, it implies that the papers of such known contributors may be of lower quality. While personal relations might be the most robust shifter for the probability of acceptance, other dimensions of proximity, such as the editor being familiar with the institution of the submitter, might be another driver of acceptance. In total, these issues imply an acceptance advantage for submissions from institutions in the global North, which, in turn, could explain the citation disadvantage, which might reveal their lower quality (on average).

Second, I detect a significant interaction effect for publications with open access in the novel X journals stemming from authors from the global North. The effect is significant on the 5% and 1% level in the OLS and the log OLS specification, and has a p-value of $p = 0.066$ in the Poisson specification.[17] The interaction effect outnumbers the negative baseline effect for publications from these countries regardless of the access type. Further, it comes close to the citation penalty for X journal open-access papers.

The disadvantage of publishing in such an outlet vanishes for authors from the global North. It is odd under the hypothesis that publications from the global North might be slightly inferior. In light of the likely lack of reputation of the novel X journals, the absence of access barriers combined with the overall reputation of established research institutions in rich, developed countries may successfully tackle

[17] The bootstrap on the RHS of Table A9 in the appendix shows that the Poisson interaction coefficient becomes significant on the 5% level for this standard error computation.

this market entry barrier as the work of academics from globally well-established universities may help establish new journals as additional market participants. Consider the extreme example of a Nobel prize winner publishing in such an X journal. It is more than likely that the name on top of the paper would 'overrule' the disadvantages stemming from the outlet the paper is published in.

For robustness, I apply two alternative ways for the geographical distinction. The first one is based on income as defined by the World Bank. Here, I separate those with 'high income' from all other countries. The rationale behind it is that countries with high income are able to spend more on their higher education systems, which should lead to higher quality of the institutions and, on average, to better and more trustworthy research output. The second approach is based on the fact that the higher education system as we know it emerged from medieval Western Europe and was also established in the colonies that were created by the European empires. Due to the fact that research institutions in Western Europe, Australia, Canada, and the US are so old and established, this is another potential driver of trust in these institutions. The computations for both approaches are presented in appendix B and qualitatively tie in with the previous findings.

The impact of funding: As mentioned in the theoretical background of the formation of the X journals, funding organizations are also interested in research funded by their grants being published with open access. Furthermore, third-party funding might be an additional source of money to pay for open-access fees. At that point, it is important to clarify what funding means here. Using metadata on publications from Scopus, I am able to access 'funding' variables that contain what researchers report on funding that they obtained in addition to being employed at their institution. The quality of these variables depends, for apparent reasons, on whether researchers correctly and fully report their funding sources. As funders often make it mandatory that work (partially) financed by their grants is earmarked with a referral to the funder and mentioning a grant number, I consider this source widely reliable and potential errors idiosyncratic.

Funding Type	$\mathbb{1}_{funding} = 1$	$\mathbb{1}_{funding} = 0$	Total
General Funding	74,490	48,999	123,939
EU Funding	1,951	121,988	123,939
US Funding	5,133	118,806	123,939

The category 'General Funding' includes all papers that report some kind of funding by reporting a grant or project number. The category 'EU funding' includes all publications that report the European Commission, the European Research Council, or the Horizon 2020 scheme of the European Commission as funding entity. All three are part of the European Union. The category 'US Funding' includes all publications that name the National Science Foundation or the National Institutes for Health as funding entities.

Table 7: Number of funded papers by funding type

Table 7 presents the number of papers that report having received any kind of funding in the first row. Here, a peculiarity of my data comes into play. One can already see that a majority of papers report external funding. I measure that by counting those papers that report a funding number or ID. In contrast, even more papers (86,240) report a funding entity. Investigating the data in more detail, the disparity between the latter value and the one reported in Table 7 stems mostly from researchers reporting universities as funders without mentioning a specific funding number. It is unclear whether researchers just report their university as it employs them. As said beforehand, this variable relies entirely on the self-reporting of the submitting authors. To rule out mere acknowledgments of employers, I code $\mathbb{1}_{funding} = 1$ only if a funding ID is clearly mentioned.

Besides this rather broad definition of funding, which leads to the high count of funded publications as shown in the top row of Table 7, more restrictive definitions are useful. In particular, I redefine the indicator for funding in a way that it turns one only for funding from the European Commission (EC), the European Research Council (ERC), or the Horizon 2020 (H2020) scheme, a major research funding program with a size of some 80 billion EUR that has been issued by the European

Commission.[18] Constructing the funding indicator alternatively for the US, it only turns one if research is funded by the National Science Foundation (NSF) or the National Institutes for Health (NIH). These institutions are leading funding bodies, and receiving a grant from such funders is usually considered a precious signal for a researcher about their underlying and unobserved ability. Hence, it is reasonable to argue that such grants correlate with researchers who can produce high-quality research. As shown in Table 7, the two narrower variables do not code all observations without funding as zero but also those with funding from other sources than those mentioned. To ensure that this mechanical effect does not drive my results, I conduct robustness checks for all regressions that use these funding variables such that I exclude all observations that have received funding from any other source. Even though the sample size diminishes substantially, the results are qualitatively quite similar.[19]

	Logit	Logit
$\mathbb{1}_{funding}$	-0.2514	-0.1526
	(0.2223)	(0.1054)
constant	-1.7547***	-2.3418***
	(0.2722)	(0.0858)
journal FE	NO	YES
N	123,939	123,939

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Binary indicator that turns one if a paper has been published with open access. Standard Errors in parentheses clustered on the journal compound level.

Table 8: The effect of funding on the probability of open-access

In the first step, I present a parsimonious logit regression that tests whether a paper that received any type of funding is subsequently more often published with

[18]See `https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-2020_en`, last checked July 10, 2023.

[19]The construction of the two smaller funding indicator variables, namely for EU funding and US funding, slightly conflicts with the modeling of the general funding variable. For the former, I identify funding entities, which, by construction, has to be done using the funder's name instead of the funding ID variable. This leads to 166 cases (8.5%) where the EU funding indicator turns one, but a funding ID is missing (which I use to set the general funding indicator $\mathbb{1}_{funding} = 1$) and 224 cases (4.4%) for the corresponding US case. As one does not simply mention such prestigious institutions without having a relation to them, I choose to keep these values in my regressions.

open access, i.e., I regress an indicator for open access on an indicator of whether a paper has received funding as reported by the authors. Table 8 presents the respective results. The left column disregards journal fixed-effects and the right one includes them. One can see that the fixed effect does not make a meaningful difference. In both cases, funding is *not* related to a higher tendency of the researchers to publish their supported work with open access.

	Multinomial Logit *Reference: RA*
	XOA
$\mathbb{1}_{funding}$	-0.3892***
	(0.1139)
constant	-4.7021***
	(0.0881)
journal FE	YES
	SOA
$\mathbb{1}_{funding}$	-0.1005
	(0.1246)
constant	-2.4658***
	(0.1022)
journal FE	YES
N	123,939

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Binary indicator that turns one if a paper has been published with open access. Standard Errors in parentheses clustered on the journal compound level.

Table 9: The effect of funding on the probability of open-access by type

Separating the open-access variable by type, i.e., into SOA and XOA, and using a multinomial logit model, Table 9 shows that the likelihood of a paper being published with XOA relative to restricted access when having received funding is significantly lower. At the same time, the effect on SOA is indistinguishable from zero. Hence, for the broad definition of funding, there even exists a disadvantage for the novel X journals. In the second step, I apply the two more selective indicators, that only include the very prestigious funding bodies from the European Union and the United States. Table 10 presents the findings for the logit regressions with the two more restrictive funding indicators. The left side shows the results for the funding variable that captures only the EU schemes, and the right side presents those of the

US schemes. Other than the null effect for funded research in general, the effects for the selected leading research funding entities from the EU indicate that these grants lead to a much higher uptake of open access.

	EU: ERC, EC & H2020		US: NSF & NIH	
	Logit	Logit	Logit	Logit
$\mathbb{1}_{funding}$	1.2410***	1.3119***	0.9644***	0.4753
	(0.1780)	(0.1104)	(0.2446)	(0.3142)
constant	-1.9303***	-2.5115***	-1.9559***	-2.4889***
	(0.1890)	(0.0055)	(0.1883)	(0.0170)
journal FE	NO	YES	NO	YES
N	123,939	123,939	123,939	123,939

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Binary indicator that turns one if a paper has been published with open access. Standard Errors in parentheses clustered on the journal compound level. Table A10 in the appendix presents results excluding observations with funding from other sources.

Table 10: The effect of funding on the probability of open-access: Effects for selected funding bodies from the EU and the US

It is not a coincidence but part of the grant requirements. For example, the 'Annotated Model Grant Agreement' of the Horizon 2020 program states in Article 29.2 (p. 245) that *"Each beneficiary must ensure open access (free of charge, online access for any user) to all peer-reviewed scientific publications relating to its results. In particular, it must: (a) as soon as possible and at the latest on publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications;"*[20] The criteria of the US National Science Foundation sound highly similar: *"NSF requires that either the version of record or the final accepted manuscript in peer-reviewed scholarly journals ... be deposited in a public access compliant repository designated by NSF."*[21] It encompasses both green and hybrid-gold open access. However, I cannot detect a positive uptake of open access among research papers funded by the two US institutions. The funding coefficient becomes insignificant once I control for

[20]See `https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf`, Version 5.2 of the agreement, date: June 26, 2019. Last checked July 10, 2023.

[21]See NSF document NSF18-041, Frequently Asked Questions (FAQs) for Public Access, `https://www.nsf.gov/pubs/2018/nsf18041/nsf18041.pdf`, last checked July 10, 2023.

journal fixed effects.

Both frameworks do not rule out an exact copy of a paper to be published as a working paper without the layout of the journal publication. Nevertheless, it might be more convenient for many researchers to publish it under an open-access license in the first place. For risk-averse researchers, publishing the final publication with open access may also be preferable to using parent journals without open access and uploading a copy somewhere, as there may remain some uncertainty about whether this is sufficient to comply with the rules.

	Multinomial Logit	
	EU	US
	Reference: RA	
	XOA	
$\mathbb{1}_{funding}$	1.2255***	0.3883
	(0.2017)	(0.2178)
constant	-5.0542***	-5.0319***
	(0.0105)	(0.0109)
journal FE	YES	
	SOA	
$\mathbb{1}_{funding}$	1.3312***	0.4895
	(0.1130)	(0.3424)
constant	-2.5940***	-2.5712***
	(0.0057)	(0.0187)
journal FE	YES	YES
N	123,939	123,939

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Categorical variable that turns one if a paper has been published with X open access, and two with subscription based open access. Reference category: Restricted access. Standard Errors in parentheses clustered on the journal compound level. Table A11 provides a robustness check that mutually excludes all other funding scheme from the regressions.

Table 11: The effect of funding on the probability of open-access by type – EU & US

The differences between the two systems also become visible when looking at the multinomial logit estimators for the impact of funding on the choice of the two open access types relative to publishing behind a subscription paywall. Table 11 shows no effect of funding on any open access type for US funding. Quite the opposite, papers funded by the European Union schemes appear more often in both X journals and as open access in hybrid journals.

Whatever the actual reason for the differences between the EU and the US funding is, the strong focus of the EU on open access seems to pay off. It leads me to the question of whether open access also pays off in terms of citations of the funded papers. Hottenrott and Lawson (2017) and Yan et al. (2018) find that funded research generally receives more citations than publications without specific third-party funding.[22] I cannot confirm their findings with my sample, as Table 12 demonstrates. Just as funding (in its general specification) is unrelated to open access, it does not correlate with a higher count of citations.

	Poisson	OLS	log OLS
XOA	-0.359***	-3.905**	-0.300***
	(0.048)	(1.197)	(0.054)
SOA	-0.051	-0.641	-0.013
	(0.111)	(1.145)	(0.078)
Funding	-0.004	-0.341	0.032
	(0.036)	(0.488)	(0.030)
XOA×Funding	-0.023	0.112	-0.018
	(0.068)	(1.051)	(0.061)
SOA×Funding	0.050	0.200	-0.016
	(0.111)	(1.237)	(0.076)
⋮	⋮	⋮	⋮
Constant	0.047	-5.151	0.066
	(0.115)	(2.578)	(0.105)
journal FE	YES	YES	YES
N	123,922	123,922	110,984

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Number of citations in levels (columns 1, 2), or logs (column 3). Standard Errors in parentheses clustered on the journal compound level.

Table 12: Regression results for citations addressing funding

However, this only holds for the whole set of funding sources. The situation changes again when I replace the general funding indicator with the two indicators that capture either the leading EU or US funders. Table 13 displays the results for that. As always in this study, the X journals have a significant citation disadvantage. And as before, publications with open access in hybrid parent journals do not have any citation advantage compared to restricted access. However, when including the

[22]Bryan and Ozcan (2021) identify an increase in citations from patents for open access papers funded by the NIH.

funding indicator variable that only captures grants from the EC, ERC, and H2020, a notable interaction exists between the three European schemes and citations in subscription-based open access. In particular, papers supported by EU funding and published with open access in an incumbent journal receive substantially more citations, as the coefficient of +0.276 in column 1 (and the respective coefficients for the linear and log OLS specifications) suggest. Another noteworthy observation is the fact that the baseline coefficient for EU funding is significantly negative, i.e., research funded by the three mentioned EU schemes, on average, receives fewer citations than research either unfunded or funded by other bodies. It corresponds to the negative coefficient of authors entirely affiliated with institutions based in the global North. Again, it is probably easier for funded research to become accepted at a journal due to the reputation of the grants and the likely proximity between authors and editors.

	European Union			United States		
	Poisson	OLS	log OLS	Poisson	OLS	log OLS
XOA	-0.376***	-3.844**	-0.314***	-0.372***	-3.868**	-0.314***
	(0.039)	(1.117)	(0.050)	(0.037)	(1.125)	(0.047)
SOA	-0.032	-0.640	-0.032	-0.007	-0.389	-0.024
	(0.052)	(0.652)	(0.041)	(0.053)	(0.654)	(0.041)
Funding (EU/US)	-0.078**	-0.884	-0.036	-0.022	-0.449	0.004
	(0.029)	(0.589)	(0.031)	(0.041)	(0.584)	(0.040)
XOA×Funding	0.148	1.184	0.064	0.017	0.914	0.031
	(0.131)	(1.494)	(0.103)	(0.113)	(1.869)	(0.104)
SOA×Funding	0.276***	3.394*	0.200**	-0.104	-1.290	0.006
	(0.082)	(1.449)	(0.063)	(0.068)	(1.003)	(0.048)
⋮	⋮	⋮	⋮	⋮	⋮	⋮
journal FE	YES	YES	YES	YES	YES	YES
N	123,922	123,922	110,984	123,922	123,922	110,984

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Number of citations in levels. Standard Errors in parentheses clustered on the journal compound level. Alternative specifications are provided in the appendix: Table A12 presents bootstrapped standard errors instead of the clustered standard errors displayed here. Table A13 presents the results for the regressions that mutually exclude all other funding schemes.

Table 13: Regression results for citations addressing funding from the EU and the US

Crossing the Atlantic, I cannot identify such an interaction effect for funding from the US entities NSF and NIH as the right part of Table 13 demonstrates. I also cannot replicate the negative baseline coefficient for these publications. It corresponds to the fact that I also do not find a higher tendency of research funded by the NSF or NIH to be published under open-access conditions as final publications. I draw two things from that finding. Other than earlier research, it does not seem the case that funding fosters citations, but funding that fosters open access may turn the open access obligation into an open access citation advantage. Why is that, given that could hardly detect such an advantage beforehand? The open access requirements of the European funding schemes seem to shift publications to open access that would have likely appeared under restricted access otherwise. These publications may attract interest either because of their quality, their research question, or a combination of both. Open access now allows a broader audience to read and cite this work.

The US case is an attractive counterfactual in this domain as there exists no shift towards open access, i.e., there is no inflow of papers with a high probability of having high quality or relevance and, consequentially, no reaction in the citations. Hence, it seems to be this shift towards open access of strong publications in the first place that overcomes the challenge that researchers publishing critical work either do not care for open access or even consciously reject it, potentially due to the publication fees or doubts regarding the trustworthiness of open access even in hybrid journals.

It is important to acknowledge that only the interaction effect with incumbent subscription-based open access is positive for the EU funding, even though researchers are more likely also to publish their work in X journals. Hence, the open access obligation removes access barriers, leading to higher citation counts. However, it cannot overcome the reputation disadvantage of recently established new journals.

Robustness Checks: A large set of robustness checks in the appendix backs my baseline findings and extensions. Table A2 uses the same setting as in my main specification shown in Table 3 but computes the standard errors based on bootstrapping with 250 replications instead of clustering them. Tables A3 and A4 compute the same regressions restricting the dataset to its 99$^{\text{th}}$ or else 95$^{\text{th}}$ percentile (in terms of citations) to ensure that outliers do not drive the results. Table A5 estimates negative-binomial regressions as an alternative to the used Poisson specification here. In all cases, the results vary only slightly and remain qualitatively the same.

In Table A7 in the appendix, I show that the results also remain the same when excluding the two journal compounds without having editorial boards that are the same for parent and mirror journal, namely the *International Journal of Pharmaceutics*, *Vaccine*, and their respective X equivalents. Table A8 presents the baseline results without considering uncited papers as suggested by Waltman et al. (2012). The results are qualitatively the same, and the point estimates are highly similar.

Last, the majority of my observations fall into the global COVID-19 pandemic. It affected researchers in many dimensions. First, it triggered a massive response in scholarly output on the disease and its implications (Haghani & Bliemer, 2020). Further, it may have adversely affected researchers in their productivity by the non-pharmaceutical interventions such as kindergarten, school, and university closures (Kwon et al., 2023) as well as by direct effects of a cured COVID-19 infection on one's own bodily constitution (Fischer et al., 2022). While I cannot directly address the indirect effects on productivity, I can avoid that my results are driven by the highly upward-pushed interest in medical publications. To do so, I rerun the main regression excluding nine medical and pharmaceutical journals. Table A6 lists the excluded outlets and shows that my findings qualitatively remain unaffected and are quantitatively highly similar to the coefficients of my principal regression.

5 Economic Implications

In the present sample of journals and publications, I cannot detect a citation advantage for open access *within* incumbent journals but a significant disadvantage for novel and relatively unknown journals, even if the quality is arguably the same. This core finding of my analysis is not only of academic interest but bears critical economic implications for the scientific publishing market. It is considerably large, with an annual size of some 19bn GBP or 21.5bn EUR (Buranyi, 2017).[23] Further, its function as a distributor and platform for communicating novel insights gives it extraordinary importance.

Academic journals can be considered two-sided markets (McCabe & Snyder, 2007; Jeon & Rochet, 2010). They embody strong market power as prestigious outlets attract more submissions, which should lead to a higher quality of published submissions. This again attracts more attention and recognition from readers. Due to the tremendous amount of publications nowadays, many researchers only have a brief look at the journal a paper is published in to evaluate its (assumed) quality and the (assumed) ability of the authors. This mechanism makes it hard to set up new journals as they need to gain recognition from authors and readers (who are, in many cases, the same people). Furthermore, according to Schmal et al. (2023) male researchers tend to seek reputation through their publications, so they might be less likely to shift to such newly established outlets. Even though the extension on funding has shown that prestigious grants increase the likelihood of a paper being published in an X journal, it does not seem to pay off in citations. While non-mainstream journals suffer from this problem as well (Chavarro et al., 2017), they do not compete with the leading outlets by definition as they specifically cover their own niches.

Reversely, new market entrants that might challenge the leading incumbent pub-

[23] The GBP-EUR conversion was done using the exchange rate of June 27, 2017, when the article was published, see `https://www.exchangerates.org.uk/EUR-GBP-spot-exchange-rates-history-2017.html`, last checked July 10, 2023.

lishing houses might suffer from the ‘X factor’ that their newly established outlets must have a different title, necessarily a different ISSN identifier, and, therefore, are not established by definition and cannot have a presentable journal impact factor or other merits. This first-mover advantage for the established parent journals or else second-mover disadvantage for the X newcomers is often present among platforms.

While my findings empirically confirm the two-sided market hypothesis, they also reemphasize the competition issue by that. Together with the contract-based finding that the big transformative agreements may strengthen the large publishers regardless of the attractiveness of their journals (Schmal, 2023), my results further challenge the hope of many researchers for a shift towards more open science and less market power of the leading commercial publishers. There exist examples such as the *Journal of the European Economic Association* that had been launched after a dispute between the society and the publisher Elsevier (Bolton et al., 2003). Furthermore, there exist two additional branch journals of the *Journal of Political Economy* (JPE), namely the *JPE: Microeconomics* and the *JPE: Macroeconomics*. The American Economic Association launched several *American Economic Journals*, top-notch derivatives of the leading *American Economic Review* that cover different economic subfields.

However, the number of publications, e.g., for the *American Economic Journal: Economic Policy* has been some 50 papers in the past three years (Luttmer, 2022). This is just a drop in the ocean of annual publications in economics, management, and adjacent fields (Schmal et al., 2023). At the same time, the number of academic publications grows in the long run with an annual rate of 4.1% (Bornmann et al., 2021). Schmal et al. (2023), detect in their study covering the years since 2015 an even higher annual growth rate of 5-7%. Thus, initiatives such as the *American Economic Journals* do not even cover the annual growth in publications, let alone a substantial change. Therefore, a significant shift from journals hosted by commercial publishers to those of non-profit societies and university presses has not happened, and there is no sign of such a change shortly. The disadvantages for researchers to publish in newly established, unknown, unranked, and rather obscure outlets are non-negligible. Semi-successful attempts, such as the *Berkeley Electronic (B.E.)* journals, fuel further doubts as to whether new journals will establish themselves.[24]

On the other hand, the significantly positive interaction term between X journal open access and all coauthors being affiliated with institutions from the global North suggests that the citation penalty for new journals in the market can be overcome if enough well-established scholars give these new participants an initial stimulus as citations may boost the perceived quality of a journal, which will further strengthen its position in the market. The same holds for prestigious grants that push publications on supported research projects in specific journals.

Regarding the estimated extensions, my results suggest that open access is rewarded more if a paper stems from authors with university affiliations in developed countries. This is a disadvantage for researchers from developing countries, who already face the challenge of lower funding and less-developed academic networks.[25] While open access, by definition, helps financially disadvantaged academics when accessing research, it may become a hurdle when publishing their own research as open access.

A last economic implication leads back to the early raised concern by McCabe and Snyder (2005) that open access journals are incentivized to accept more papers due to the business model based on publications instead of subscriptions. If researchers realize that open access does not pay off in terms of citations for their publications in strong journals, they might focus on publishing weaker work with open access. It would make open access a substitute for quality instead of a complement, as in the case of top publications. It would not only reinforce the subjacent

[24]For example, the *B.E. Journal of Macroeconomics* reached its peak SJR of 1.447 in 2008. Since then, it collapsed to 0.217, 15% of the all-time high. See `https://www.scimagojr.com/journalsearch.php?q=8300153213&tip=sid`, last checked July 10, 2023.

[25]The geographic diversity of authors particularly in Elsevier's X journals has been already examined by Smith et al. (2021).

quality concerns regarding open access but also lock out the general public and developing countries from research not idiosyncratically but rather from leading work. This argument is in favor of transformative agreements as they make any publication in any journal of an eligible publisher open access by default (see, e.g., Haucap et al., 2021; Schmal, 2024) such that the issue above may not play a role anymore.

What speaks against these contracts is that they make it more attractive for eligible researchers to publish in the included journals. These are usually established outlets so that they may keep researchers away from new market participants. It is particularly problematic as 89.8% of all transformative agreements are closed between publishers and countries from the global West. Even more serious is that contracts with countries not from the global West cover only 2.9% of all publications estimated to be published under such agreements.[26] As I could show beforehand, authors from the global North as well as in long.standing higher education systems in the global West do not suffer from a citation disadvantage in newly set up journals. Funding bodies can also significantly push papers towards fully open-access outlets. Hence, they could be a core driver in establishing a higher level of competition in the publishing market by strategically requiring submissions to these journals and not hybrid-gold or green open-access to publications in established journals. However, the transformative agreements will encourage them to stick with those often well-established publishers.

6 Conclusion

My analysis of the unique setting of two open access options alongside the restricted-access publishing option does not only reject the existence of such an advantage when studying within journal variation across 70 journals and 35 compounds but

[26] The numbers are computed based on the transformative agreement registry of the ESAC initiative, see `https://esac-initiative.org/about/transformative-agreements/agreement-registry/`. Last database update: May 12, 2023. Every contract is counted separately, for example, the German 'DEAL' agreements are listed twice, once for Wiley and once for Springer Nature.

also detects what I subsume the ‘X factor,’ a significant decrease in citations for open access publications in newly launched journals even though they rely on the editorial boards of their parent journals and should be, thus, qualitatively indistinguishable from their counterparts.

While this disadvantage for non-established open access is prominent, it diminishes among publications from authors affiliated with institutions from the global North relative to the papers from other countries. Hence, researchers reading these publications might perceive open access to papers from often well-established universities as a complement to research quality but rather a substitute for quality when it comes to open access to papers from non-Western institutions, even though that happens *within* the same journals with the same editorial boards and peer review processes.

A way to foster competition is likely to be via clearly specified open access requirements of grants. As the extension on funding has shown, publications supported by EU funding are more often published under an open access license. Funding bodies should, therefore, consider whether they tighten the requirement of publishing results not only with open access but within a fully open access journal, i.e., ruling out hybrid outlets. This corresponds to the suggestion of Schmal (2023) to introduce shades in the color scheme of structuring types of open access in a way that full open access to a paper has one shade if it is published in a hybrid journal and another one if it is published in a fully open access journal.

Future research should take into account a longer time span and, if possible, a broader set of journals that covers not only different publishers but also different quality ranges. As the studied setting is highly unique, it remains an open question whether this trifold scenario can be investigated with different journals but revisiting it in a few years might contribute a further understanding of the long-run effects of open access in incumbent and new journals relative to subscription-based outlets.

My results have non-negligible implications for the ongoing changes in the academic publishing market, especially concerning the 'transformative agreements' between often large publishers and many university consortia. While the primary demand by the universities – all papers being published as open access immediately – is satisfied by construction, it makes publishing in the journals of these publishers c.p. more attractive even though newly established competing journals already suffer from a citation disadvantage as my results for Elsevier's X journals have shown.

The comparative advantage of hybrid open access compared to newly established gold open access journals may strengthen the position of the already large publishing houses. In the medium and long run, this mechanism could harm market entrants and prevent smaller players. It should lead to further concentration and less competition in the market for academic publishing.

References

Abramo, Giovanni, Tindaro Cicero, and Ciriaco Andrea D'Angelo (2011). "Assessing the varying level of impact measurement accuracy as a function of the citation window length". *Journal of Informetrics* 5.4, 659–667.

Armstrong, Mark (2015). "Opening Access to Research". *The Economic Journal* 125.586, F1–F30.

Asai, Sumiko (2023). "Authors' choice between parent and mirror journals of Elsevier". *Learned Publishing* 36.2, 299–306.

Bethmann, Dirk, Felix Bransch, Michael Kvasnicka, and Abdolkarim Sadrieh (2023). "Home bias in top economics journals". *IZA Discussion Paper*. No. 15,965.

Bolton, Patrick, Alan Krueger, and Roberto Perotti (2003). "Statement by the Editors". *Journal of the European Economic Association* 1.1, iv.

Bornmann, Lutz, Robin Haunschild, and Rüdiger Mutz (2021). "Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases". *Humanities and Social Sciences Communications* 8.1, 1–15.

Braun, Tibor, Wolfgang Glänzel, and András Schubert (2006). "A Hirsch-type index for journals". *Scientometrics* 69.1, 169–173.

Bryan, Kevin A and Yasin Ozcan (2021). "The impact of open access mandates on invention". *Review of Economics and Statistics* 103.5, 954–967.

Budzinski, Oliver, Sophia Gaenssle, and Daniel Weimar (2023). "Disentangling Individual Biases in Jury Voting: An Empirical Analysis of Voting Behavior in the Eurovision Song Contest". *Available at SSRN 4343866*.

Buranyi, Stephen (2017). "Is the staggeringly profitable business of scientific publishing bad for science?" *The Guardian*. date of publication: 27/06/2017.

Campanario, Juan Miguel (2011). "Empirical study of journal impact factors obtained using the classical two-year citation window versus a five-year citation window". *Scientometrics* 87.1, 189–204.

Chavarro, Diego, Puay Tang, and Ismael Ràfols (2017). "Why researchers publish in non-mainstream journals: Training, knowledge bridging, and gap filling". *Research Policy* 46.9, 1666–1680.

Colussi, Tommaso (2018). "Social Ties in Academia: A Friend Is a Treasure". *Review of Economics and Statistics* 100.1, 45–50.

Craig, I. et al. (2007). "Do open access articles have greater citation impact?A critical review of the literature". *Journal of Informetrics* 1.3, 239–248.

Davis, Philip M. et al. (2008). "Open access publishing, article downloads, and citations: randomised controlled trial". *BMJ (Clinical research ed.)* 337, a568.

Ellison, Glenn (2002). "Evolving standards for academic publishing: A q-r theory". *Journal of Political Economy* 110.5, 994–1034.

Eysenbach, Gunther (2006). "Citation advantage of open access articles". *PLoS biology* 4.5, e157.

Fischer, Kai, J. James Reade, and W. Benedikt Schmal (2022). "What cannot be cured must be endured: The long-lasting effect of a COVID-19 infection on workplace productivity". *Labour Economics* 79, 102281.

Glänzel, Wolfgang and Eugene Garfield (2004). "The myth of delayed recognition". *The Scientist* 18.11, 8–9.
Guerrero-Bote, Vicente P and Félix Moya-Anegón (2012). "A further step forward in measuring journals' scientific prestige: The SJR2 indicator". *Journal of Informetrics* 6.4, 674–688.
Haghani, Milad and Michiel C. J. Bliemer (2020). "Covid-19 pandemic and the unprecedented mobilisation of scholarly efforts prompted by a health crisis: Scientometric comparisons across SARS, MERS and 2019-nCoV literature". *Scientometrics* 125.3, 2695–2726.
Harris, Paul A. et al. (2019). "The REDCap consortium: building an international community of software platform partners". *Journal of Biomedical Informatics* 95, 1–10.
Harrison, Peter (2019). "What are mirror journals, and can they offer a new world of open access?" https://www.elsevier.com/connect/what-are-mirror-journals-and-can-they-offer-a-new-world-of-open-access, Published April 4, 2019, last checked July 10, 2023.
Haucap, Justus, Nima Moshgbar, and W. Benedikt Schmal (2021). "The impact of the German 'DEAL' on competition in the academic publishing market". *Managerial and Decision Economics* 42.8, 2027–2049.
Hottenrott, Hanna and Cornelia Lawson (2017). "Fishing for complementarities: Research grants and research productivity". *International Journal of Industrial Organization* 51, 1–38.
Jeon, Doh-Shin and Jean-Charles Rochet (2010). "The Pricing of Academic Journals: A Two-Sided Market Perspective". *American Economic Journal: Microeconomics* 2.2, 222–255.
Jon, Jae-Eun, Jenny J. Lee, and Kiyong Byun (2014). "The emergence of a regional hub: comparing international student choices and experiences in South Korea". *Higher Education* 67.5, 691–710.
Kwon, Eunrang, Jinhyuk Yun, and Jeong-han Kang (2023). "The effect of the COVID-19 pandemic on gendered research productivity and its correlates". *Journal of Informetrics* 17.1, 101380.
Laband, David N. (2013). "On the Use and Abuse of Economics Journal Rankings". *The Economic Journal* 123.570, F223–F254.
Lawrence, S. (2001). "Free online availability substantially increases a paper's impact". *Nature* 411.6837, 521.
Luttmer, Erzo F. P. (2022). "Report of the Editor: American Economic Journal: Economic Policy". *AEA Papers & Proceedings* 112, 709–720.
McCabe, Mark J. and Christopher M. Snyder (2005). "Open access and academic journal quality". *AEA Papers & Proceedings* 95.2, 453–458.
McCabe, Mark J. and Christopher M. Snyder (2007). "Academic Journal Prices in a Digital Age: A Two-Sided Market Model". *The B.E. Journal of Economic Analysis & Policy* 7.1.
McCabe, Mark J. and Christopher M. Snyder (2014). "Identifying The Effect Of Open Access On Citations Using A Panel Of Science Journals". *Economic Inquiry* 52.4, 1284–1300.

McCabe, Mark J. and Christopher M. Snyder (2015). "Does Online Availability Increase Citations? Theory and Evidence from a Panel of Economics and Business Journals". *Review of Economics and Statistics* 97.1, 144–165.

McCabe, Mark J. and Christopher M. Snyder (2021). "Cite unseen: Theory and evidence on the effect of open access on cites to academic articles across the quality spectrum". *Managerial and Decision Economics* 42.8, 1960–1979.

McCullough, Rachel (2022). "Scopus filters for Open Access type and Green OA full-text access option". `https://www.elsevier.com/connect/what-are-mirror-journals-and-can-they-offer-a-new-world-of-open-access`, Published January 13, 2022, last checked December 1, 2023.

Osterloh, Margit and Bruno S. Frey (2020). "How to avoid borrowed plumes in academia". *Research Policy* 49.1, 103831.

Radicchi, Filippo, Santo Fortunato, and Claudio Castellano (2008). "Universality of citation distributions: toward an objective measure of scientific impact". *Proceedings of the National Academy of Sciences (PNAS)* 105.45, 17268–17272.

Rose, Michael E. and John R. Kitchin (2019). "pybliometrics: Scriptable bibliometrics using a Python interface to Scopus". *SoftwareX* 10, 100263.

Rose, Michael E. and Suraj Shekhar (2023). "Adviser Connectedness and Placement Outcomes in the Economics Job Market". *Labour Economics*, 102397.

Rubin, Amir, Eran Rubin, and Dan Segal (2023). "Editor home bias?" *Research Policy* 52.6, 104766.

Rüegg, Walter (1992). "Foreword: The University as a European Institution". In: *A history of the university in Europe 1: Universities in the middle ages.* Ed. by Hilde De Ridder-Symoens. Cambridge, UK: Cambridge University Press, xix–xxvii.

Schmal, W. Benedikt (2023). "The 'Must Stock' Challenge in Academic Publishing: Pricing Implications of Transformative Agreements". *mimeo.*

Schmal, W. Benedikt (2024). "How Transformative are Transformative Agreements? Evidence from Germany Across Disciplines". *Scientometrics.* forthcoming.

Schmal, W. Benedikt, Justus Haucap, and Leon Knoke (2023). "The role of gender and coauthors in academic publication behavior". *Research Policy* 52.10, 104874.

Shortliffe, Edward H. and Mor Peleg (2019). "Retirement of JBI's mirror journal". *Journal of Biomedical Informatics* 100S, 103846.

Smith, Audrey C. et al. (2021). "Assessing the effect of article processing charges on the geographic diversity of authors using Elsevier's "Mirror Journal" system". *Quantitative Science Studies* 2.4, 1123–1143.

Śpiewanowski, Piotr and Oleksandr Talavera (2021). "Journal rankings and publication strategy". *Scientometrics* 126.4, 3227–3242.

Waltman, Ludo, Nees Jan van Eck, and Anthony F. J. van Raan (2012). "Universality of citation distributions revisited". *Journal of the American Society for Information Science and Technology* 63.1, 72–77.

Wang, Jian (2013). "Citation time window choice for research impact evaluation". *Scientometrics* 94.3, 851–872.

Wang, Xianwen, Chen Liu, Wenli Mao, and Zhichao Fang (2015). "The open access advantage considering citation, article usage and social media attention". *Scientometrics* 103.2, 555–564.

Wooding, Steven (2020). “Heuristics, not plumage: A response to Osterloh and Frey’s discussion paper on ‘Borrowed plumes’”. *Research Policy* 49.1, 103871.

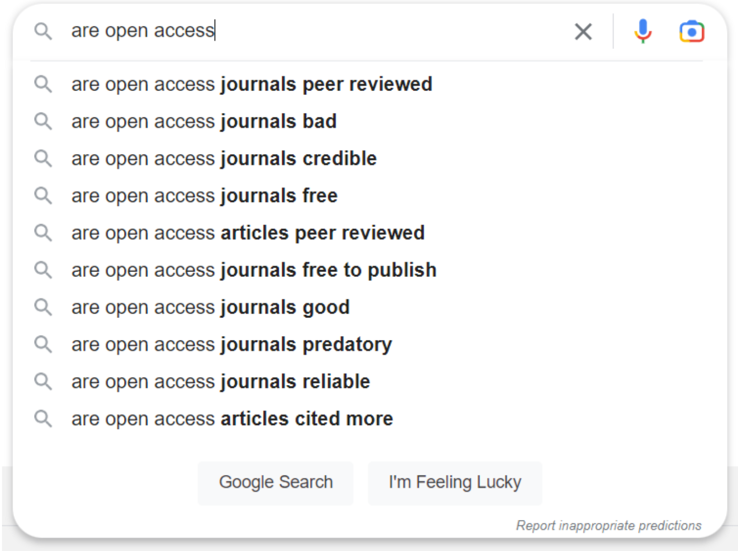
Yan, Erjia, Chaojiang Wu, and Min Song (2018). “The funding factor: a cross-disciplinary examination of the association between research funding and citation impact”. *Scientometrics* 115.1, 369–384.

Appendix A

Journal Compound	H Index	Journal Rank
Analytica Chimica Acta	224	q1
Atmospheric Environment	270	q1-q2
Biosensors and Bioelectronics	222	q1
Chaos, Solitons and Fractals	160	q1
Chemical Engineering Science	280	q1
Chemical Physics Letters	248	q2
Computers and Graphics	79	q2
Contraception	110	q1
Cytokine	130	q1-q2
Ecological Engineering	150	q1
Energy Conversion and Mgmt.	232	q1
European Journal of Obstetrics...	111	q2
Food Chemistry	302	q1
Gene	188	q1-q2
Intl. Journal of Pharmaceutics	244	q1
Journal of Asian Earth Sciences	146	q1
Journal of Biomedical Informatics	121	q1
Journal of Computational Physics	275	q1
Journal of Dentistry	130	q1
Journal of Hydrology	260	q1
Journal of Non-Crystalline Solids	188	q1-q2
Journal of Pediatrics	227	q1
Journal of Structural Biology	156	q1
Materials Letters	164	q1-q2
Nutrition	156	q1-q2
Optical Materials	113	q2
Research Policy	271	q1
Resources, Cons. and Recycling	170	q1
Respiratory Medicine	134	q1
Sleep Medicine	141	q1
Toxicon	140	q3
Vaccine	205	q1
Veterinary Parasitology	138	q2
Water Research	354	q1
World Neurosurgery	106	q2

H Index computed on the journal level. Journal rank: quartile within the SCImago Journal Ranking from 2018 to 2022 in the main research category as reported by SCImago. q1 means that a journal is in the top quartile of a certain discipline during 2018-2022. The H index is computed on the journal level (see Braun et al., 2006, for the conceptual idea) and from 2023. Due to the concave functional form of this measure, it should be highly similar to its past values.

Table A1: Journal reputation



Autocomplete Suggestion generated by entering the query "are open access" into the search box of `https://www.google.com/` without being logged into a Google account using the browser Google Chrome Version 113.0.5672.93 (64-Bit). Day of the search: May 11, 2023. Results might vary slightly with different specifications.

Figure F1: Google autocomplete for the search query "are open access"

	Poisson	Poisson	Poisson	Poisson	Poisson	OLS
XOA	-0.565***	-0.574***	-0.305***	-0.374***	-0.371***	-3.818***
	(0.033)	(0.035)	(0.030)	(0.033)	(0.033)	(0.290)
SOA	-0.144***	-0.234***	-0.036*	-0.021	-0.019	-0.517*
	(0.037)	(0.018)	(0.018)	(0.019)	(0.018)	(0.250)
age			0.421***	1.288***	1.292***	8.518***
			(0.003)	(0.020)	(0.022)	(0.233)
age^2				-0.132***	-0.132***	-0.474***
				(0.003)	(0.003)	(0.041)
#authors					0.016***	0.202***
					(0.001)	(0.024)
constant	2.660***	2.755***	1.352***	0.145***	0.044	-5.407***
	(0.005)	(0.023)	(0.021)	(0.037)	(0.042)	(0.495)
journal FE	NO	YES	YES	YES	YES	YES
N	123,939	123,939	123,939	123,939	123,922	123,922

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Number of citations in levels. Standard errors in parentheses bootstrapped with 250 replications. Coefficients for X journal and open access publications relative to the reference category of subscription based publications in the main journals.

Table A2: Regression results X journals and open access combined: Bootstrapped SEs

	Poisson	Poisson	Poisson	Poisson	Poisson	OLS
XOA	-0.498***	-0.500***	-0.261***	-0.329***	-0.327***	-3.082**
	(0.096)	(0.116)	(0.046)	(0.036)	(0.035)	(0.862)
SOA	-0.206*	-0.246***	-0.060	-0.044	-0.042	-0.664
	(0.100)	(0.057)	(0.052)	(0.054)	(0.048)	(0.538)
age			0.390***	1.234***	1.239***	8.119***
			(0.009)	(0.052)	(0.052)	(1.096)
age^2				-0.130***	-0.130***	-0.573***
				(0.007)	(0.007)	(0.111)
#authors					0.015***	0.180***
					(0.002)	(0.028)
constant	2.541***	2.678***	1.394***	0.236*	0.137	-4.191
	(0.120)	(0.004)	(0.034)	(0.095)	(0.098)	(2.107)
journal FE	NO	YES	YES	YES	YES	YES
N	122,721	122,721	122,721	122,721	122,704	122,704

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Number of citations in levels. Standard errors in parentheses clustered on the journal compound level. Alternative specification excluding the citation counts exceeding the 99th percentile. Coefficients for X journal and open access publications relative to the reference category of subscription based publications in the main journals.

Table A3: Regression results: 99% computation

	Poisson	Poisson	Poisson	Poisson	Poisson	OLS
XOA	-0.418***	-0.431***	-0.240***	-0.307***	-0.304***	-2.372***
	(0.082)	(0.090)	(0.036)	(0.030)	(0.028)	(0.522)
SOA	-0.185*	-0.222***	-0.060	-0.045	-0.043	-0.477
	(0.089)	(0.060)	(0.057)	(0.058)	(0.051)	(0.456)
age			0.354***	1.149***	1.154***	7.099***
			(0.013)	(0.050)	(0.050)	(0.895)
age^2				-0.124***	-0.124***	-0.603***
				(0.006)	(0.006)	(0.112)
#authors					0.015***	0.161***
					(0.002)	(0.022)
constant	2.328***	2.538***	1.402***	0.341***	0.240*	-2.535
	(0.099)	(0.004)	(0.046)	(0.100)	(0.102)	(1.444)
journal FE	NO	YES	YES	YES	YES	YES
N	117,745	117,745	117,745	117,745	117,729	117,729

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Number of citations in levels. Standard errors in parentheses clustered on the journal compound level. Alternative specification excluding the citation counts exceeding the 95th percentile. Coefficients for X journal and open access publications relative to the reference category of subscription based publications in the main journals.

Table A4: Regression results 95% computation

	Neg Bin	Neg Bin	Neg Bin	Neg Bin	Neg Bin
XOA	-0.565***	-0.500***	-0.302***	-0.374***	-0.366***
	(0.107)	(0.096)	(0.054)	(0.045)	(0.044)
SOA	-0.144	-0.236**	-0.073	-0.032	-0.036
	(0.119)	(0.078)	(0.069)	(0.071)	(0.066)
age			0.504***	1.427***	1.430***
			(0.014)	(0.062)	(0.061)
				-0.151***	-0.151***
				(0.009)	(0.009)
#authors					0.024***
					(0.006)
constant	2.660***	2.752***	1.055***	-0.082	-0.232*
	(0.132)	(0.005)	(0.041)	(0.095)	(0.109)
journal FE	NO	YES	YES	YES	YES
ln(α)	0.424***	0.231***	-0.127	-0.192**	-0.201**
	(0.066)	(0.053)	(0.070)	(0.070)	(0.068)
N	123,939	123,939	123,939	123,939	123,922

$^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$. Dependent Variable: Number of citations in levels. Standard errors in parentheses clustered on the journal compound level. Alternative specification using a negative-binomial setting instead of Poisson. Coefficients for X journal and open access publications relative to the reference category of subscription based publications in the main journals.

Table A5: Regression results: Negative-Binomial specification

	Poisson	Poisson	Poisson	Poisson	Poisson	OLS
XOA	-0.522***	-0.558***	-0.298***	-0.369***	-0.366***	-4.270**
	(0.125)	(0.164)	(0.061)	(0.044)	(0.042)	(1.311)
SOA	-0.010	-0.244***	-0.022	-0.008	-0.012	-0.621
	(0.118)	(0.048)	(0.036)	(0.038)	(0.038)	(0.531)
age			0.423***	1.294***	1.299***	9.094***
			(0.012)	(0.078)	(0.078)	(1.274)
age^2				-0.133***	-0.133***	-0.506***
				(0.010)	(0.010)	(0.116)
#authors					0.023***	0.311***
					(0.004)	(0.074)
constant	2.700***	2.755***	1.344***	0.135	-0.015	-7.374*
	(0.151)	(0.003)	(0.046)	(0.138)	(0.151)	(3.050)
journal FE	NO	YES	YES	YES	YES	
N	100,431	100,431	100,431	100,431	100,426	100,426

$^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$. Dependent Variable: Number of citations in levels. Standard errors in parentheses clustered on the journal compound level. Coefficients for X journal and open access publications relative to the reference category of subscription based publications in the main journals. Excluding medicince journals that may have been affected by COVID-19, namely Contraception, the European Journal of Obstetrics & Gynecology and Reproductive Biology, Cytokine, Gene, International Journal of Pharmaceutics, the Journal of Dentistry, the Journal of Pediatrics, Respiratory Medicine, Sleep Medicine, and Vaccine, and all of their X derivatives.

Table A6: Regression results excluding medicine journals

	Poisson	Poisson	Poisson	Poisson	Poisson	OLS
XOA	-0.538***	-0.560***	-0.308***	-0.383***	-0.379***	-4.268**
	(0.113)	(0.152)	(0.057)	(0.042)	(0.041)	(1.225)
SOA	-0.114	-0.248***	-0.053	-0.039	-0.034	-0.829
	(0.132)	(0.062)	(0.057)	(0.059)	(0.053)	(0.616)
age			0.421***	1.296***	1.301***	8.755***
			(0.011)	(0.073)	(0.073)	(1.192)
age^2				-0.133***	-0.133***	-0.501***
				(0.010)	(0.010)	(0.108)
#authors					0.016***	0.218***
					(0.003)	(0.046)
constant	2.665***	2.756***	1.351***	0.134	0.026	-5.880*
	(0.140)	(0.004)	(0.042)	(0.129)	(0.135)	(2.697)
journal FE	NO	YES	YES	YES	YES	
N	113,863	113,863	113,863	113,863	113,851	113,851

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Number of citations in levels. Standard errors in parentheses clustered on the journal compound level. Coefficients for X journal and open access publications relative to the reference category of subscription based publications in the main journals. Excluding International Journal of Pharmaceutics, Vaccine, and their X derivatives.

Table A7: Regression results excluding compounds not sharing the same editorial board

	Poisson	Poisson	Poisson	Poisson	Poisson	OLS
XOA	-0.482***	-0.493***	-0.274***	-0.336***	-0.334***	-4.156***
	(0.091)	(0.124)	(0.050)	(0.039)	(0.038)	(1.079)
SOA	-0.080	-0.189***	-0.013	0.002	0.000	-0.271
	(0.097)	(0.043)	(0.034)	(0.035)	(0.034)	(0.537)
age			0.380***	1.095***	1.099***	9.319***
			(0.011)	(0.054)	(0.055)	(1.174)
age^2				-0.108***	-0.108***	-0.578***
				(0.007)	(0.007)	(0.114)
#authors					0.012***	0.178***
					(0.003)	(0.047)
constant	2.762***	2.816***	1.532***	0.522***	0.443***	-6.199*
	(0.116)	(0.003)	(0.039)	(0.096)	(0.105)	(2.700)
journal FE	NO	YES	YES	YES	YES	YES
N	110,990	110,990	110,990	110,990	110,984	110,984

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Number of citations in levels excluding 12,949 observations without citations. Standard errors in parentheses clustered on the journal compound level. Coefficients for X journal and open access publications relative to the reference category of subscription based publications in the main journals.

Table A8: Regression results excluding uncited papers

	Poisson	OLS	log OLS	Poisson	OLS	log OLS
	broad definition			narrow definition		
XOA	0.352***	-4.035***	-0.325***	-0.354***	-4.021***	-0.323***
	(0.038)	(0.370)	(0.024)	(0.041)	(0.369)	(0.026)
SOA	-0.077***	-1.345***	-0.061***	-0.078***	-1.318***	-0.056***
	(0.017)	(0.217)	(0.012)	(0.018)	(0.237)	(0.012)
Global North	-0.079***	-0.976**	-0.095***	-0.154***	-1.811***	-0.142***
	(0.022)	(0.322)	(0.011)	(0.032)	(0.381)	(0.018)
XOA × Global North	0.045	1.392	0.110*	0.316*	4.512***	0.340***
	(0.074)	(0.767)	(0.045)	(0.140)	(1.261)	(0.093)
SOA × Global North	0.209***	2.692***	0.142***	0.225**	3.057***	0.203***
	(0.046)	(0.624)	(0.021)	(0.079)	(0.873)	(0.040)
age	1.274***	8.277***	1.063***	1.262***	8.395***	1.068***
	(0.018)	(0.230)	(0.010)	(0.016)	(0.215)	(0.010)
age^2	-0.129***	-0.433***	-0.103***	-0.128***	-0.449***	-0.104***
	(0.003)	(0.041)	(0.002)	(0.003)	(0.040)	(0.002)
#authors	0.024***	0.253***	0.032***	0.022***	0.227***	0.030***
	(0.003)	(0.034)	(0.002)	(0.002)	(0.031)	(0.002)
constant	0.032	-5.230***	0.029	0.062	-5.331***	0.036
	(0.036)	(0.486)	(0.021)	(0.035)	(0.463)	(0.022)
journal FE	YES	YES	YES	YES	YES	YES
N	112,059	112,059	100,302	100,424	100,424	89,944

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Dependent Variable: Number of citations in levels (columns 1, 2, 4, 5), or logs (columns 3, 6). Standard errors in parentheses bootstrapped with 250 replications.

Table A9: Impact of the affiliation countries of authors: Global North – bootstrapped SEs

	EU: ERC, EC & H2020 Logit	Logit	US: NSF & NIH Logit	Logit
$\mathbb{1}_{funding}$	1.0721***	1.1905***	0.7698*	0.3541
	(0.2683)	(0.1073)	(0.3075)	(0.3270)
constant	-1.7614***	-1.9535***	-1.7614	-1.8125
	(0.2724)	(0.0207)	(0.2724)	(0.0727)
journal FE	NO	YES	NO	YES
N	50,560	50,560	53,742	53,742

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Dependent Variable: Binary indicator that turns one if a paper has been published with open access. Standard errors in parentheses clustered on the journal compound level. The regression for EU funding excludes all observations that have received any kind of funding as defined in Table 7. In addition, it excludes those observations that have received US funding but do not mention a funding number. The regression for US funding excludes all observations that have received any kind of funding as defined in Table 7. In addition, it excludes those observations that have received EU funding but do not mention a funding number.

Table A10: The effect of funding on the probability of open-access: Effects for selected funding bodies from the EU and the US – excluding observations with funding from any other source

	Multinomial Logit	
	EU	US
	Reference: RA	
	XOA	
$\mathbb{1}_{funding}$	0.9051***	0.1714
	(0.2375)	(0.1871)
constant	-4,0324***	-4.0317
	(0.0428)	(0.0367)
journal FE	YES	
	SOA	
$\mathbb{1}_{funding}$	1.2639***	0.3850
	(0.1124)	(0.3676)
constant	-2.0948***	-1.9300***
	(0.0228)	(0.0835)
journal FE	YES	YES
N	50,560	53,742

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Categorical variable that turns one if a paper has been published with X open access, and two with subscription based open access. Reference category: Restricted access. Standard errors in parentheses clustered on the journal compound level. The regression for EU funding excludes all observations that have received US funding as defined in Table 7. The regression for US funding excludes all observations that have received EU funding as defined in Table 7.

Table A11: The effect of funding on the probability of open-access by type – EU & US – excluding observations with funding from any other source

	European Union			United States		
	Poisson	OLS	log OLS	Poisson	OLS	log OLS
XOA	-0.376***	-3.844***	-0.314***	-0.372***	-3.868***	-0.314***
	(0.032)	(0.271)	(0.021)	(0.033)	(0.284)	(0.021)
SOA	-0.032	-0.640**	-0.032**	-0.007	-0.389	-0.024*
	(0.018)	(0.240)	(0.010)	(0.021)	(0.282)	(0.011)
Funding (EU/US)	-0.078*	-0.884	-0.036	-0.022	-0.449	0.004
	(0.034)	(0.608)	(0.023)	(0.029)	(0.381)	(0.017)
XOA×Funding	0.148	1.184	0.064	0.017	0.914	0.031
	(0.134)	(1.585)	(0.099)	(0.151)	(1.319)	(0.105)
SOA×Funding	0.276***	3.394**	0.200***	-0.104	-1.290	0.006
	(0.075)	(1.300)	(0.045)	(0.085)	(1.162)	(0.034)
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Constant	0.043	-5.416***	0.089***	0.043	-5.410***	0.090***
	(0.040)	(0.481)	(0.019)	(0.042)	(0.516)	(0.019)
journal FE	YES	YES	YES	YES	YES	YES
N	123,922	123,922	110,984	123,922	123,922	110,984

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Dependent Variable: Number of citations in levels (columns 1, 2, 4, 5), or logs (columns 3, 6). Standard errors in parentheses bootstrapped with 250 replications.

Table A12: Regression results addressing funding from the EU and the US – bootstrapped SEs

	European Union			United States		
	Poisson	OLS	log OLS	Poisson	OLS	log OLS
XOA	-0.344***	-4.223**	-0.312***	-0.346***	-4.254**	-0.313***
	(0.053)	(1.311)	(0.059)	(0.053)	(1.336)	(0.060)
SOA	-0.006	-0.544	-0.031	-0.011	-0.620	-0.024
	(0.096)	(0.920)	(0.071)	(0.102)	(0.997)	(0.073)
Funding (EU/US)	-0.075+	-0.920	-0.007	-0.017	-0.535	0.029
	(0.045)	(0.894)	(0.046)	(0.050)	(0.776)	(0.045)
XOA×Funding	0.118	1.197	0.047	0.015	1.199	0.023
	(0.135)	(1.663)	(0.099)	(0.116)	(1.850)	(0.108)
SOA×Funding	0.260*	2.572	0.174+	-0.054	-0.667	-0.004
	(0.118)	(1.604)	(0.088)	(0.100)	(1.245)	(0.070)
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Constant	-0.070	-3.114	0.188	-0.058	-2.871	0.180
	(0.129)	(2.839)	(0.119)	(0.116)	(2.686)	(0.112)
journal FE	YES	YES	YES	YES	YES	YES
N	50,544	50,544	43,554	53,726	53,726	46,357

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Number of citations in levels (columns 1, 2, 4, 5), or logs (columns 3, 6). Standard errors in parentheses clustered on the journal compound level. Alternative specification with bootstrapped standard errors in Table A14 below. The regression for EU funding excludes all observations that have received US funding as defined in Table 7. The regression for US funding excludes all observations that have received EU funding as defined in Table 7. Due to the smaller sample size, I also report significance on the 10% level using a +.

Table A13: Regression results addressing funding from the EU and the US – excluding observations with funding from any other source

	European Union			United States		
	Poisson	OLS	log OLS	Poisson	OLS	log OLS
XOA	-0.344***	-4.223***	-0.312***	-0.346***	-4.254***	-0.313***
	(0.050)	(0.439)	(0.032)	(0.052)	(0.462)	(0.030)
SOA	-0.006	-0.544	-0.031*	-0.011	-0.620+	-0.024
	(0.039)	(0.405)	(0.016)	(0.036)	(0.367)	(0.017)
Funding (EU/US)	-0.075*	-0.920	-0.007	-0.017	-0.535	0.029
	(0.033)	(0.667)	(0.027)	(0.028)	(0.420)	(0.018)
XOA×Funding	0.118	1.197	0.047	0.015	1.199	0.023
	(0.142)	(1.629)	(0.103)	(0.149)	(1.359)	(0.098)
SOA×Funding	0.260**	2.572+	0.174***	-0.054	-0.667	-0.004
	(0.082)	(1.426)	(0.049)	(0.061)	(0.816)	(0.035)
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Constant	-0.070	-3.114***	0.188***	-0.058	-2.871***	0.180***
	(0.057)	(0.692)	(0.036)	(0.058)	(0.632)	(0.039)
journal FE	YES	YES	YES	YES	YES	YES
N	50,544	50,544	43,554	53,726	53,726	46,357

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent Variable: Number of citations in levels (columns 1, 2, 4, 5), or logs (columns 3, 6). Standard errors in parentheses bootstrapped with 250 replications. The regression for EU funding excludes all observations that have received US funding as defined in Table 7. The regression for US funding excludes all observations that have received EU funding as defined in Table 7. Due to the smaller sample size, I also report significance on the 10% level using a +.

Table A14: Regression results addressing funding from the EU and the US – excluding observations with funding from any other source

Appendix B

Alternative geographical distinctions

For geographical distinction, I apply the binary separation of the world in a developed 'global North' and a developing 'global South.' In this appendix, I present additional evidence by using two different ways to separate affiliation countries of the author groups of the publications, namely based on income and age of the higher education system. Both variables shall capture the reputation of the research output of researchers based at institutions in countries with either high income or a longstanding tradition of university-based research and education.

High Income Countries:
The first one is based on the World Bank's country classification by income[27] using data from the bank's fiscal year 2020, which includes information up to the end of 2019. This is to avoid distortions from the COVID-19 pandemic and its economic consequences. The identification of this distinction builds upon the consideration that high-income countries can spend more on their higher education systems.

	Broad definition		*Narrow definition*		
	High Income	Other	High Income	Other	Total
RA	13,634	84,635	4,203	94,066	98,269
XOA	600	1,658	174	2,084	2,258
SOA	3,561	7,971	1,098	10,434	11,532
Total	17,795	94,264	5,475	106,584	112,059

The 'high income' category includes all publications up to nine authors that have a majority of coauthors affiliated with an institution based in a high income country as defined by the World Bank (broad definition) or else all publications up to nine authors where all coauthors are affiliated with an institution based in a high income country (narrow definition).

Table B1: Publications from high income countries by access type

[27] See `https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html`.

	Poisson	OLS	log OLS	Poisson	OLS	log OLS
	broad definition			narrow definition		
XOA	-0.358***	-3.986**	-0.325***	-0.359***	-3.978**	-0.322***
	(0.047)	(1.254)	(0.045)	(0.047)	(1.266)	(0.045)
SOA	-0.073	-1.306*	-0.059	-0.076	-1.310*	-0.055
	(0.048)	(0.610)	(0.040)	(0.045)	(0.596)	(0.039)
High Income	-0.061	-0.745	-0.078*	-0.143***	-1.695*	-0.119**
	(0.032)	(0.589)	(0.035)	(0.037)	(0.623)	(0.037)
XOA×High Income	0.055	1.068	0.097	0.289	4.152*	0.320**
	(0.087)	(1.142)	(0.061)	(0.170)	(1.940)	(0.113)
SOA×High Income	0.180**	2.368*	0.122**	0.213	2.831	0.170
	(0.060)	(0.949)	(0.038)	(0.133)	(1.468)	(0.084)
age	1.273***	8.275***	1.062***	1.265***	8.390***	1.068***
	(0.070)	(1.112)	(0.063)	(0.070)	(1.157)	(0.063)
age^2	-0.129***	-0.432***	-0.103***	-0.128***	-0.451***	-0.104***
	(0.010)	(0.110)	(0.009)	(0.009)	(0.112)	(0.009)
#authors	0.024**	0.257*	0.032***	0.022**	0.232*	0.031***
	(0.009)	(0.106)	(0.007)	(0.008)	(0.096)	(0.007)
constant	0.030	-5.253*	0.028	0.060	-5.253	0.034
	(0.148)	(2.527)	(0.100)	(0.142)	(2.614)	(0.101)
journal FE	YES	YES	YES	YES	YES	YES
N	112,059	112,059	100,302	99,739	99,739	89,331

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Dependent Variable: Number of citations in levels (columns 1, 2, 4, 5), or logs (columns 3, 6). Standard errors in parentheses clustered on the journal compound level. Alternative specification with bootstrapped standard errors in Table B3 below.

Table B2: Impact of the affiliation countries of authors: High Income Countries

	Poisson	OLS	log OLS	Poisson	OLS	log OLS
	broad definition			narrow definition		
XOA	-0.358***	-3.986***	-0.325***	-0.359***	-3.978***	-0.322***
	(0.041)	(0.335)	(0.026)	(0.037)	(0.345)	(0.027)
SOA	-0.073***	-1.306***	-0.059***	-0.076***	-1.310***	-0.055***
	(0.017)	(0.230)	(0.013)	(0.019)	(0.233)	(0.012)
High Income	-0.061**	-0.745*	-0.078***	-0.143***	-1.695***	-0.119***
	(0.022)	(0.304)	(0.010)	(0.035)	(0.387)	(0.018)
XOA×High Income	0.055	1.068	0.097	0.289*	4.152***	0.320***
	(0.075)	(0.728)	(0.051)	(0.136)	(1.218)	(0.091)
SOA×High Income	0.180***	2.368***	0.122***	0.213**	2.831***	0.170***
	(0.051)	(0.684)	(0.024)	(0.074)	(0.760)	(0.041)
age	1.273***	8.275***	1.062***	1.265***	8.390***	1.068***
	(0.018)	(0.228)	(0.011)	(0.018)	(0.217)	(0.011)
age^2	-0.129***	-0.432***	-0.103***	-0.128***	-0.451***	-0.104***
	(0.003)	(0.042)	(0.002)	(0.003)	(0.039)	(0.002)
#authors	0.024***	0.257***	0.032***	0.022***	0.232***	0.031***
	(0.002)	(0.034)	(0.002)	(0.003)	(0.034)	(0.002)
constant	0.030	-5.253***	0.028	0.060	-5.253***	0.034
	(0.040)	(0.462)	(0.021)	(0.037)	(0.436)	(0.022)
journal FE	YES	YES	YES	YES	YES	YES
N	112,059	112,059	100,302	99,739	99,739	89,331

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Dependent Variable: Number of citations in levels (columns 1, 2, 4, 5), or logs (columns 3, 6). Standard errors in parentheses bootstrapped with 250 replications

Table B3: Impact of the affiliation countries of authors: High Income Countries – bootstrapped SEs

Established Higher Education Systems:
An alternative distinction is using a more narrow but slightly outdated distinction between 'Western' and 'non-Western' countries. The modern understanding of a university as a higher education institution emerged from medieval Western Europe and is a genuine European creation (Rüegg, 1992), which got exported during the age of colonization to what became the United States, Canada, and Australia. These countries are also among the leading economic powers globally, spend heavily on their higher education systems, and consist of open societies that ensure freedom of research. The endowment, the institutional environment, and the high standards for their higher education institutions lead to the strong reputation of Western universities. For example, Australia, France, Germany, the United Kingdom, and the United States host more than half of all exchange students, while more than half stem from Asian countries (Jon et al., 2014). Thus, although the 'global West' definition might appear slightly outdated, it captures well the location of long-established research infrastructures. In this robustness check, I use the member states of the European Union between the years 1995 and 2003, when it had its largest size in *Western* Europe before expanding to post-Soviet countries in 2004 and later on. Additionally, I include Australia, Canada, and the United States as the largest non-European Western countries, which also have well-established university systems due to the colonization by the British and French.

	Broad definition		Narrow definition	
	West = 0	West = 1	West = 0	West = 1
RA	98,710	9,119	105,104	2,725
XOA	2,038	466	2,369	135
SOA	10,836	2,770	12,786	820
Total	111,584	12,355	120,259	3,680

Table B4: Publications from Western countries with established higher education systems by access type

	Poisson	OLS	log OLS	Poisson	OLS	log OLS
	broad definition			narrow definition		
XOA	-0.361***	-3.983**	-0.325***	-0.362***	-3.983**	-0.324***
	(0.045)	(1.234)	(0.044)	(0.044)	(1.240)	(0.045)
SOA	-0.061	-1.131	-0.053	-0.062	-1.130*	-0.050
	(0.046)	(0.564)	(0.039)	(0.043)	(0.549)	(0.038)
Estd. HE	-0.025	-0.203	-0.049	-0.083*	-0.954	-0.073*
	(0.039)	(0.621)	(0.039)	(0.037)	(0.592)	(0.031)
XOA×Estd. HE	0.064	1.151	0.116	0.204	3.480	0.310**
	(0.104)	(1.250)	(0.070)	(0.167)	(1.913)	(0.112)
SOA×Estd. HE	0.173*	2.286*	0.127**	0.130	2.053	0.158
	(0.074)	(1.074)	(0.045)	(0.134)	(1.374)	(0.095)
age	1.273***	8.269***	1.062***	1.263***	8.327***	1.065***
	(0.070)	(1.109)	(0.063)	(0.069)	(1.159)	(0.063)
age^2	-0.129***	-0.432***	-0.103***	-0.128***	-0.441***	-0.104***
	(0.009)	(0.110)	(0.009)	(0.009)	(0.113)	(0.009)
#authors	0.025**	0.273*	0.033***	0.024**	0.260*	0.032***
	(0.009)	(0.103)	(0.007)	(0.008)	(0.098)	(0.007)
constant	0.021	-5.378*	0.018	0.053	-5.299	0.025
	(0.146)	(2.540)	(0.100)	(0.143)	(2.665)	(0.102)
journal FE	YES	YES	YES	YES	YES	YES
N	112,059	112,059	100,302	104,138	104,138	93,259

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Dependent Variable: Number of citations in levels (columns 1, 2, 4, 5), or logs (columns 3, 6). Standard errors in parentheses clustered on the journal compound level. Alternative specification with bootstrapped standard errors in Table B6 below.

Table B5: Impact of the affiliation countries of authors: Established Higher Education Systems

	Poisson	OLS	log OLS	Poisson	OLS	log OLS
	broad definition			narrow definition		
XOA	-0.361***	-3.983***	-0.325***	-0.362***	-3.983***	-0.324***
	(0.035)	(0.325)	(0.025)	(0.038)	(0.336)	(0.024)
SOA	-0.061**	-1.131***	-0.053***	-0.062**	-1.130***	-0.050***
	(0.020)	(0.250)	(0.013)	(0.021)	(0.256)	(0.012)
Estd. HE	-0.025	-0.203	-0.049***	-0.083*	-0.954	-0.073***
	(0.030)	(0.431)	(0.013)	(0.040)	(0.498)	(0.022)
XOA×Estd. HE	0.064	1.151	0.116*	0.204	3.480**	0.310**
	(0.085)	(0.780)	(0.055)	(0.150)	(1.302)	(0.097)
SOA×Estd. HE	0.173**	2.286**	0.127***	0.130	2.053*	0.158***
	(0.058)	(0.776)	(0.026)	(0.086)	(0.938)	(0.047)
age	1.273***	8.269***	1.062***	1.263***	8.327***	1.065***
	(0.018)	(0.227)	(0.010)	(0.017)	(0.217)	(0.011)
age^2	-0.129***	-0.432***	-0.103***	-0.128***	-0.441***	-0.104***
	(0.003)	(0.042)	(0.002)	(0.003)	(0.040)	(0.002)
#authors	0.025***	0.273***	0.033***	0.024***	0.260***	0.032***
	(0.002)	(0.032)	(0.001)	(0.002)	(0.031)	(0.001)
constant	0.021	-5.378***	0.018	0.053	-5.299***	0.025
	(0.038)	(0.480)	(0.021)	(0.035)	(0.437)	(0.022)
journal FE	YES	YES	YES	YES	YES	YES
N	112,059	112,059	100,302	104,138	104,138	93,259

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Dependent Variable: Number of citations in levels (columns 1, 2, 4, 5), or logs (columns 3, 6). Standard errors in parentheses bootstrapped with 250 replications

Table B6: Impact of the affiliation countries of authors: Established Higher Education Systems – bootstrapped SEs

Differentiation by open-access types

In the present paper, I investigate different access types and their relation to citations. Publications in restricted access papers are straightforward. For X journals, gold open access, i.e., free access to the published version on the journal's official website, is the crucial feature. Further versions may exist in repositories, i.e., additional green open access, but as gold open access is more convenient, this should have a negligible impact on citations, if at all. For subscription-based open access, the case is slightly more entangled. Here, open access can be granted via the journal, i.e., hybrid open access to the final publication in a journal. Otherwise, open access can be made possible by uploading the paper to a freely accessible repository, the so-called green open access. In contrast, the final journal publication can still require a subscription. By doing so, the authors can circumvent potential APCs as sharing a paper on public repositories is usually free of charge (see, e.g., arXiv, ResearchSquare, or SSRN). The downside is that readers need to look for such a version in case they are on the publisher's website – even though search engines such as Google Scholar often link to both the journal's website and repositories if uploads are available there. This switch of platforms may be costly for potential readers.

Access Type	Frequency	Share
Restricted Access	107,828	87.00%
X Open Access	2,504	2.02%
Green/Bronze Open Access (SOA)	6,676	5.39%
Hybrid Gold Open Access (SOA)	6,931	5.59%
Total	123,939	

Table B7: Publications distinguished by open-access type among SOA publications

The last option is 'bronze' open-access, something between gold and green open-access and less well-defined than the other two 'colors.' Examples are final journal publications that do not face subscription paywalls solely due to the publisher's goodwill and not due to contractual arrangements, as is the case for hybrid or gold open-access. However, Elsevier, which runs the Scopus database from where I receive my data, also labels publications as 'bronze' if they are republished in special archives or repositories, which has sufficient overlap with the repository-based green open access (McCullough, 2022). For this reason, I excluded all 'non-hybrid-gold' open-access publications from the SOA group of publications in this additional analysis. Table B7 presents the split within the SOA category, which now encompasses the two alternatives 'green/bronze open-access' and 'hybrid gold open-access' separately. One can see that hybrid gold accounts for approximately half of all SOA publications. And green or bronze open access for the other half.

	Poisson	Poisson	Poisson	Poisson	Poisson	OLS
XOA	-0.565***	-0.562***	-0.293***	-0.364***	-0.361***	-3.708**
	(0.107)	(0.138)	(0.053)	(0.038)	(0.038)	(1.133)
SOA	-0.089	-0.190**	0.075	0.143***	0.139***	1.235*
(hybrid-gold)	(0.125)	(0.068)	(0.047)	(0.042)	(0.042)	(0.503)
age			0.420***	1.304***	1.307***	8.805***
			(0.011)	(0.070)	(0.070)	(1.160)
age^2				-0.134***	-0.134***	-0.511***
				(0.009)	(0.009)	(0.106)
#authors					0.013***	0.167***
					(0.003)	(0.044)
constant	2.660***	2.755***	1.352***	0.121	0.037	-5.674*
	(0.132)	(0.003)	(0.041)	(0.125)	(0.129)	(2.613)
journal FE	NO	YES	YES	YES	YES	YES
N	117,263	117,263	117,263	117,263	117,251	117,251

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Dependent Variable: Number of citations in levels. Standard Errors in parentheses clustered on the journal compound level (i.e., X journals and main journals together). Alternative specification with bootstrapped standard errors in Table B9 below. Coefficients for X journal and *hybrid* open access publications in parent journals relative to the reference category of subscription based publications in the main journals. SOA publications with 'green' or 'bronze' open-access excluded.

Table B8: Regression results for the impact of the different access options

Table B8 presents the regression results for this adjusted dataset. One can see that the central insight for the newly established X journals remains unaffected. They continue to face a severe disadvantage in terms of citations relative to the incumbent journal publications with restricted access. In contrast, using only gold open-access for the SOA publications, one can see a significant benefit regarding citations gained. Hence, hassle-free open access to publications in established journals may be advantageous for authors, which increases the hurdle for new and competing journals that try to establish themselves as an alternative to the incumbents.

	Poisson	Poisson	Poisson	Poisson	Poisson	OLS
XOA	-0.565***	-0.562***	-0.293***	-0.364***	-0.361***	-3.708***
	(0.038)	(0.038)	(0.032)	(0.033)	(0.031)	(0.294)
SOA	-0.089**	-0.190***	0.075**	0.143***	0.139***	1.235***
(hybrid-gold)	(0.029)	(0.028)	(0.025)	(0.026)	(0.024)	(0.352)
age			0.420***	1.304***	1.307***	8.805***
			(0.003)	(0.018)	(0.017)	(0.228)
age^2				-0.134***	-0.134***	-0.511***
				(0.003)	(0.003)	(0.041)
#authors					0.013***	0.167***
					(0.001)	(0.016)
constant	2.660***	2.755***	1.352***	0.121***	0.037	-5.675***
	(0.005)	(0.022)	(0.022)	(0.032)	(0.032)	(0.450)
journal FE	NO	YES	YES	YES	YES	YES
N	117,263	117,263	117,263	117,263	117,251	117,251

$^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$. Dependent Variable: Number of citations in levels. Standard errors in parentheses bootstrapped with 250 replications. Coefficients for X journal and *hybrid* open access publications in parent journals relative to the reference category of subscription based publications in the main journals. SOA publications with 'green' or 'bronze' open-access excluded.

Table B9: Regression results for the impact of the different access options