ORIGINAL RESEARCH

Gatekeeping should be conserved in the open science era

Hugh Desmond[1,2]

Abstract

The elimination of gatekeepers for scientific publication has been represented as a means to promote the core moral values of open science, including democratic decision-making and inclusiveness. I argue that this framing ignores the reality that gatekeeping is a way of structuring prestige hierarchies, and that without gatekeeping, some other structuring would be needed: the flattening of prestige hierarchies is not possible given scientists' need to navigate information overload. I consider two potential restructurings of prestige hierarchies, one based on citation count and the other on search algorithm rank. These are shown to simply reintroduce status biases and hierarchies in ways that either do not further the open science ideals of democracy and inclusiveness, or else involve some de facto gatekeeping. Gatekeeper elimination should not be thought of as an intrinsic part of the open science movement. In fact, insofar as gatekeeping is guided by professional ideals of impartiality and diligence, it can be thought of as an ally of open science values.

1 Introduction

Nobody can deny that scientific gatekeeping, as it is commonly practiced, often falls disappointingly short of its ideals of impartiality, diligence, and competence. However, this is not a new realization. The biases and inefficiencies of peer and editorial review were documented over a half-century ago (Zuckerman & Merton, 1971), and the list of shortcomings only seems to have grown since then (Lee et al., 2013). What *is* novel today is that it is has become possible to publish without gatekeeping, through online publication. Without a need to adhere to the physical limitations of print publication,

✉ Hugh Desmond
hugh.desmond@gmail.com

1   Department of Philosophy, Leibniz Universität Hannover, Hannover, Germany

2   Department of Philosophy, University of Antwerp, Antwerp, Belgium

　Springer

one of the old rationales for prepublication peer review—to select only the best articles to fill the limited number of pages of a print journal—has evaporated. Gatekeeper elimination has become technologically viable. However, would it also be desirable for science?

The open science movement, or at least some of its more ambitious advocates, state unambiguously that it would be. A gatekeeper-less science is a more *open* science:

> Picture a situation in which scientists would be able to publish all their thoughts, results, conclusions, data, and such as they occur, openly and widely available to everybody. The Internet already provides tools that could make this possible (microblogs, blogs, wikis, etc.). (…) Knowledge could flow quickly, regardless of institutions and personal networks. Research results could be published as they occur. There would be no need to wait until results are complete enough to support a full paper. (Bartling & Friesike, 2014, p. 8)

This passage exemplifies what I call the "open science rationale for gatekeeper elimination": the practice of gatekeeping is antithetical to the core values of transparency, democracy, and inclusiveness, and should be eliminated or minimized as much as possible. This rationale will be laid out in more detail later, but its core is implied by the very term of "gatekeeper". A "gatekeeper" calls to mind an individual who decides on who may or may not enter through the gates of a city. Gatekeeping thus, per definition, means selective exclusion. Decisions about exclusion are also not necessarily democratic: it is primarily the gatekeepers who decide, not the city's population. And transparency and accountability are optional for gatekeepers: the gatekeepers need not justify their decisions.

It is important to acknowledge from the outset that the push for gatekeeper elimination remains very much a minority position, one that major policy documents do not currently call for.[1] The rhetoric of open science thus is far from being actualized in science policy. However, the *rationale*—the line of reasoning—supporting this position is not self-evidently invalid. Gatekeeping *does* intuitively seem opaque, undemocratic, and exclusionary, and the normative case for its elimination does seem accordingly persuasive (for examples, see e.g. Bartling & Friesike, 2014; Persic et al., 2021; Heesen & Bright, 2021). Are gatekeepers then to be considered as anachronisms in twenty-first century science?

The strategy of this paper is to make a different diagnosis of what is perceived as problematic about gatekeeping: gatekeeping can give rise to *prestige hierarchies* and biases that are antithetical to ideals of democracy and inclusiveness. This allows for a view on the matter that is more grounded in anthropology and psychology. Prestige biases, in general, are social learning biases that arise spontaneously in humans as a way to deal with information overload (Henrich & Gil-White, 2001), and as such play an important function that cannot simply be eliminated. Mainly based on this consideration, I propose that it is not realistic to expect that prestige differentials would be flattened if gatekeepers were eliminated. The question therefore is: in the

---

[1] A National Academy of Sciences report focuses on efforts to reform existing review structures (NAS, 2018, p. 151) and the closest a European Commission report comes is in stating that "early sharing of different kinds of research outputs should be encouraged".

absence of gatekeeping, how should prestige differentials be restructured, and how would open science ideals be promoted?

My argument is that, among the potentially plausible alternatives, the implied prestige hierarchies are not consistent with ideals of democracy and inclusiveness. I consider two open science strategies to deal with information overload—identifying prestige with citation count, or with search engine rank. I discuss how the latter strategy has already been shown to involve novel forms of gatekeeping (and so is not really an alternative to gatekeeping), and how the former could potentially be realized with gatekeeping but incentivizes research behaviors (such as "rentiership") that are clearly antithetical to open science values. The upshot is the erection of de facto barriers to inclusiveness and democracy, resulting in the open science rationale for gatekeeper elimination to in fact appear as unsound upon closer consideration.

In this way, this paper adds to a growing number of cautious voices regarding open science reforms. For instance, recent editorials have begun cautioning against preprint publication, the oldest of open science practices (Schalkwyk et al., 2020; Sheldon, 2018). These authors argue that preprints can promote confusion by circulating multiple versions of the same paper, and by encouraging the public to have inappropriate levels of credence in flawed preprints. Some such shortcomings were on display during the Covid-19 pandemic, when studies were rushed to publication despite being plagued by problems (Jung et al., 2021; Ziemann et al., 2022).

The paper is structured as follows. After laying out the open science rationale for eliminating gatekeepers in more detail in Sect. 2, in the subsequent section I discuss the structural features that support and promote prestige heuristics, and argue that gatekeeper elimination cannot conceivably eliminate prestige hierarchies but only lead to their restructuring. Sections 4 and 5 examine one alternative where prestige is defined in terms of citation metrics, while Sect. 6 examines the use of search engine rank as a way for scientists to decide on the value or prestige of publications. Both alternatives come up short, and in the final section I suggest how gatekeeping can in fact be aligned with open science ideals and that therefore our default stance should be to conserve gatekeeping.

2 The open science rationale for eliminating gatekeepers

"Open science" is a term that is currently being used in a number of different ways. In one sense, "open science" simply refers to scientific practices that promote transparency, especially to the ones made possible by online communication. These practices include the sharing of preprints (Taubes, 1993), pre-registering protocols (Nosek & Lakens, 2014), posting post-publication peer review (Knoepfler, 2015), data sharing (Gewin, 2016), and transparency about author contribution or even breaking up authorship into component roles (Habgood-Coote, 2021; Vasilevsky et al., 2021). All these practices promote transparency and allow for increased Mertonian communism. They also reduce, to varying degrees, the importance of getting gatekeeper approval for sharing scientific research.

Such open science *practices* are not antithetical to traditional gatekeeping. In fact, the practice of "open peer review", where reviewer identities are revealed and reviewer

reports are published, represents exactly a form of peer review that integrates open science practices. Open peer review thus could be thought of as a next step in the history of gatekeeping, which has always coevolved with communication technology.[2]

However, in a second sense, open science is also a vision about what science *should* be -- not merely what it *can* be. In this sense, open science is not merely about promoting particular practices, but is a normative view on what values should govern the various activities that comprise scientific research. This paper will especially focus on three values often associated with open science: transparency, inclusiveness, and democracy (see Persic et al., 2021). The question for this paper is whether it makes sense to expect that the elimination of gatekeeping would promote these values.

On a superficial level, it would seem obvious that doing away with gatekeepers would improve transparency, inclusiveness and democracy. After all, the function of gatekeeping is to exclude some voices in a non-democratic way. It is a primarily technocratic form of decision-making by an unelected "expert" where transparency is strictly optional. Sometimes gatekeepers give a justification for their acceptance or rejection, but sometimes they do not.

The proposal of gatekeeper elimination could be evaluated according to epistemic norms: would it promote knowledge, or scientific progress? Often open science reforms and values are advocated for in this way. For instance, some argue that by increasing diversity and inclusiveness, ideas will be shared more rapidly and research groups will consider a wider diversity of viewpoints, thereby benefiting their research (see e.g. Nature, 2018). Another line of reasoning points to reproducibility problems that hamper the reliability of research: open science practices such as protocol pre-registrations or data sharing will help prevent such problems and thus benefit the trustworthiness of science (see e.g. Munafò et al., 2017).

However, there is also a considerable moral component to the open science movement. This is what will be of interest in this article. The concept of open science (e.g., the dichotomy between "open" and "closed") is part of the same family as notions such as "open government" where transparency and inclusiveness are also foregrounded (Brezzi et al., 2021; OECD, 2021).[3] Both trace back to the term "open society", originally coined by Henri Bergson ([1932] 1963) and popularized by Karl Popper ([1945] 2011). In Popper's version, open societies are based on democracy, citizen participation and free public exchange, and are to be contrasted with totalitarian societies. Not surprisingly then, "open science" inherits to a significant extent the normative

---

[2] This is a fascinating and largely forgotten history (see Spier 2002). What is today seen as "traditional gatekeeping" was at one time a progressive innovation made possible by technological innovation: the Xerox photocopier in 1959. This allowed submitted manuscripts to be copied and sent out to potential peer-reviewers. Prior to that, the review process was carried out by the editor and a committee of assistant editors who received carbon copies. (This type of editorial review, in turn, was only possible thanks to the invention of the typewriter in the 1890s.).

[3] For purposes of this paper, I will understand inclusiveness as not unjustly excluding voices in the scientific community. Democracy (in line with Christiano & Bajaj, 2022) is understood as the principle that the opinions of all scientists should be given equal weight when collectively deliberating on the quality of scientific ideas. On this definition, democracy entails inclusiveness, but not vice versa since an opinions of all scientists can be included (i.e., given some weight in collective deliberation) without all opinions given equal weight.

rationale for "the open society": the implication is that rational-critical, transparent, and inclusive deliberation is as crucial for good science as it is for democracy.

Today's open science advocates have quite straightforwardly adopted Popper's joining of moral values to epistemic values. Consider for instance:

> In a break from our traditionally closed science systems, the open science movement has vowed to make the scientific process more *transparent*, more *inclusive* and more *democratic*. The culture of sharing, at the core of open science, nurtures synergies and avoids duplication of scientific effort, leading to research that is conducted more quickly and efficiently, easier to scrutinize and, therefore, of *higher quality*. (Persic et al., 2021, p. 16)

In this way, open science advocates not only hold that transparency, inclusiveness, and democracy go hand in hand with scientific progress or higher quality research, but in fact are a way (perhaps even *the* way) of achieving scientific progress and quality.

This paper departs from the Popperian line in two ways. First, it will not assume that moral values necessarily promote scientific epistemic values. For instance, in certain scenarios, non-transparency, exclusion, and a lack of democracy could turn out to be maximally beneficial for scientific progress. Would then open science advocates embrace such policies? In general, scientific progress can clash with moral values (most poignantly illustrated by the ethics of clinical trials: Nuremberg Code, 1947; WMA, 2013), and in particular, it would seem that open science moral values could at least sometimes clash with the epistemic values of truth, knowledge, or scientific progress. Open science—or at least, for the purposes of this paper—is to be understood as primarily opposed to *closed* science, not to stagnant or non-progressive science.

This separation of moral and epistemic values will allow this paper to examine the impact of gatekeeper elimination on the former without examining the impact on the latter. The reason for doing so is that it is difficult to state with any confidence what would be the general impact of the values of transparency or inclusiveness on scientific progress. Perhaps eliminating gatekeepers would lead to the "quick flowing of knowledge" as Bartling and Frieseke claim (Bartling & Friesike, 2014, p. 8). It is also quite possible it would lead to the quick flowing of bad science, through over-publication and the chasing of scientific fashions (this is part of the rationale for the "slow science" movement (Stengers, 2018)).

The second departure from the Popperian line is the different stance taken on gatekeeping. For Popper's intellectual descendants, *gatekeeping* seems to play a structurally similar role to the power structures (e.g., religiously or ideologically biased actors in society) that had an interest in preventing the open society from being realized. This paper will take a different approach: gatekeeping is simply one way among many to organize prestige differentials.

3 Prestige heuristics

What is prestige? Although the word harks back to the Latin word for illusion, *prestigium,* it has come to denote something with very real effects. For our purposes here,

prestige can be defined as an indicator of the extent to which a manuscript's communicated content is deemed worth learning. Prestige is conferred freely, and it guides social learning (Henrich & Gil-White, 2001). Prestige, at least for purposes here, can be thought of in a subjectivist sense as *perceived value.*

Different entities can possess prestige. Publications can possess prestige, such as "canonical works" that everyone in a subfield is familiar with. Journals can be more or less prestigious, according to whether they tend to select prestigious manuscripts. Individuals can be prestigious, for instance if they tend to write valuable manuscripts. And institutions can be prestigious to varying degrees, for instance if they tend to house prestigious individuals. In general then there can be many factors that enhance prestige (e.g., the prestige of an institution can contribute to the prestige of an individual scientist).

In particular, the label of "peer-reviewed" tends to add to the prestige of the manuscript. This label is perceived as indicating the trust-worthiness of the publication, and perhaps its scientific quality more generally. As we will discuss later, a high citation count (and to a lesser degree, a good ranking in a database search) can also contribute to the prestige or perceived value of a publication.

Prestige hierarchies and open science. At first sight, the existence of prestige hierarchies between scientists sits uneasily with the moral values of open science. How so? By indicating what is worth learning, prestige differentials help direct the flow of *attention* and information through the scientific community. Moreover, because the success of scientists' careers is bound up with how much value is attached to their ideas, prestige is a precious resource and hence a powerful incentive for scientists. Thus, prestige differentials mean that the voices and opinions of some scientists are given more weight than the voices and opinions of others: not all scientists are treated equally in the process of collective scientific deliberation. This means that some scientists have more *power* than others to influence the course of the scientific debate (e.g., what questions get attention, what possible types of solution are primary entertained). Or to put it in the language of inclusiveness and democracy: prestige differentials mean that some voices receive such little attention as to be de facto excluded and to preclude genuine democratic deliberation.

Gatekeeping is but one factor that helps shape prestige differentials. However, it is an important one. Gatekeeping involves a small group of people (editors, peer-reviewers) deciding on whether to accord some degree of prestige to a scientific communication (article, letter, commentary). The impact factor of the journal (a metric for journal prestige more generally) can have a large impact on the perceived prestige of a publication, and can lead to a larger number of citations. If the journal is sufficiently prestigious, then a single publication can even improve perceptions of the author's prestige, to the extent that career opportunities (funding applications, job applications, prizes, collaborations) can be materially improved.

From the perspective of open science ideals, gatekeeping seems to be especially problematic because it refers to a non-transparent, undemocratic, and exclusionary process—and yet it is a significant contributing factor to the prestige of a publication or individual. The abolishing of gatekeeping would clearly flatten this aspect of prestige differentials, since all manuscripts would be published in online repositories. Without gatekeeping, the scientific community as a whole—rather than a small number of

gatekeepers—could consider and evaluate all manuscripts, not just those preselected by gatekeepers.

Can prestige differentials be eliminated? It is important to emphasize that the elimination of prestige differentials is neither what is at stake, nor a remotely plausible scenario. Prestige heuristics are a social learning strategy, grounded in the rationale that social learning is more effective when the learner pays disproportionate attention to skillful or competent persons. Cultural evolutionists argue that social learning has been crucial for human evolution (Boyd & Richerson, 1985; Herrmann et al., 2007), and there is evidence that prestige biases are relatively deeply engrained in human cognition (Anderson et al., 2015; Atkisson et al., 2012).[4] Even three-year-olds recognize high-prestige individuals and pay disproportionate attention to them (Chudek et al., 2012).

A social learning strategy is adaptive when agents cannot *independently* distinguish between features in the environment that are useful for their goals, and features that are not so useful or are even harmful. The classic example is that of European explorers stranded in the wilderness: they typically did not know how to extract nutrients from the environment and would have died if not aided by indigenous tribes (Henrich, 2016). To what challenges in the scientific environment are prestige heuristics potentially adapted?

One challenge for scientists consists in information overload. Too much is published to be read, let alone studied, and scientists need indicators on what to pay attention to. In 2014, 2.5 million articles were published (Ware and Mabe 2015), and at the estimated growth rate of 3%, this translates to almost 9000 articles published per day in 2023. By this crude metric alone, it is clear that only an exceedingly small fraction of published articles can be read by any single researcher, and that a drastic selection must be made. A prestige heuristic is a quick and easy way to select which articles to heed.

Another challenge involves interactions between scientists, whether in collaboration or through peer-review. If a scientist interacts with a scientist from another sub-field, the first may often not be able to directly gauge the competence or trustworthiness of the second. Collaboration would not be possible without collaborators having a certain degree of trust in each other's competence (Desmond, 2021). Even in the case of peer-review—a process where the reviewer and the author often possess the same expertise—reviewers must be able to assume a basic trustworthiness in the author, because they typically cannot check for outright fraud or falsification (Crocker & Lynne Cooper, 2011). An editor or collaborator might rely on some basic markers of prestige, some formal like having a PhD, others informal such as being able to write according to the expected standards of the domain in question. Hence, given the ineliminable need to trust relatively unknown colleague scientists, the need for prestige indicators will be similarly ineliminable. This is a different type of "information overload", that does not reflect the impossibility of being able to read everything, but rather the impossibility of being able to know everyone.

---

[4] In this paper I will use "bias" and "heuristic" interchangeably. Admittedly, there are some differences between the two—for instance, a bias is often assumed to be unconscious whereas an agent may consciously adopt a heuristic—but these differences are not relevant here.

This is a very general, coarse-grained argument for why it is unlikely that the presence of exclusionary and undemocratic prestige heuristics can be engineered away by reform of science policy, no matter how ingenious. Even if one source of prestige bias—gatekeeping—were to be eliminated, the need to make expedient judgments of others' trustworthiness would remain, and consequently the need for relatively simple (and potentially faulty) prestige hierarchies would also remain.

Reforming prestige differentials. While open science advocates may tend to foreground the importance of flattening prestige differentials, realistically these cannot be flattened. The question rather concerns how prestige differentials can be compatible with the values of inclusiveness or democracy. Much here depends on how prestige is allocated.

Some of the basic questions facing egalitarianism resurface here (Gosepath, 2021): how can prestige differentials be organized in a way that is, in general, egalitarian and, in particular, consistent with open science values? For instance, a Rawlsian open science advocate might require the allocation of prestige to occur under a veil of ignorance, according to rules that every scientist agrees on beforehand.

To develop such a Rawlsian approach to prestige differentials in science, two classes of challenge would need to be met. The first lies in identifying what rules could be both reasonable and specific enough to inform policy. For instance, one reasonable rule would be that prestige is allocated in proportion to the "quality" of the scientific contribution. The problem with this rule is that it is not clear how "quality" can be operationalized: how are researchers to distinguish between *genuine* and merely *apparent* quality when assigning prestige?

A second class of challenge concerns the evident gap between the ideal distribution and the real distribution. For instance, the Matthew effect, as originally described by Merton, refers to how high-prestige individuals receive more "credit" (i.e., additional prestige) than low-prestige individuals for the same contribution (Merton, 1968). However, the Matthew effect is not necessarily *unfair.* For instance, Michael Strevens has argued that the endorsement of a result by a high-prestige individual will allow the result to circulate more rapidly, and hence such an endorsement is more valuable for the scientific community (and more deserving of a reward) than an endorsement by a low-prestige individual (Strevens, 2006). It is clear then that a full discussion of how prestige differentials should be structured in science would need to consider phenomena such as the Matthew effect, which may be only apparently unfair.

These considerations help emphasize the scope of the argument of this paper, which is not to evaluate scientific prestige differentials *in general,* but rather to evaluate the specific proposal of gatekeeper elimination and whether the subsequent restructuring of prestige differentials would automatically be more in line with open science ideals. The next sections examine two alternatives to gatekeeping-influenced prestige hierarchies: prestige differentials based on citation count, and differentials based on search engine rank.

4 Citation metrics

If gatekeepers are to be eliminated, the unavoidable information overload faced by scientists implies that another strategy will be needed to decide what publications and individuals are worth noting. Concretely, if all scientific manuscripts are to be posted to a repository, without the imprimatur of journal and of prepublication review, then some other mechanism would be needed for even basic distinctions between rigorous publications and, to take the other extreme, pseudoscientific blog posts. One proposal among open science advocates has been is to rely on citation counts as measures of scientific quality. This proposal, in effect, redefines "prestige differentials" in terms of citation count. The question for this and next section is whether the incentives and evaluative judgments implied by this proposal are consistent with open science values.

The intuition behind the citation count proposal is that citation counts are the scientific analog of vote counts in democracies. Consider for instance, a recent proposal to eliminate peer review, where Heesen and Bright stipulate that citation count is one of the central ways of capturing "long-run credit" (Heesen & Bright, 2021, p. 646). Following Merton (Merton, 1957), they conceptualize credit as the reward given to scientists for their work from their peers. They also follow others (Kitcher, 1990; Strevens, 2003) in conceiving of the agency of scientists—i.e., their motivational structure for doing scientific work—as directed towards the maximization of credit.[5] In this view, a journal publication is a "short-run credit", since it indicates only the recognition given by a small number of peer-reviewers and an editor. By contrast, the long-term impact on the scientific community is what truly matters, and this is better measured by citation count (Heesen & Bright, 2021, p. 646).[6] In their view, citations are explicitly comparable to "votes" in democracies (e.g., Heesen & Bright, 2021, pp. 646–647).

How justified is it to view citation count as a good indicator of scientific achievement? Consider a recent contribution by Hannah Rubin, who sketches the evidence for a citation gap between male and female researchers, with the former being disproportionately cited (Rubin, 2022). Such citation gaps can have multiple causes, and Rubin mentions the possibility that the citation gap can be partially explained by a number of factors: (1) differences in productivity, with men publishing at a higher rate than women, (2) differences in self-citation, again with men self-citing more than women, and (3) differences in career length, with women's careers, on average, being shorter and thus leaving less scope for the accumulation of a body of work and of reputation. The latter indicates the most structural cause: more males than females occupy central nodes in scientific networks, and hence their work will have a better chance of being informally disseminated and cited. Even if individual scientists cite for idealistic reasons—i.e., because they hold the cited work in high esteem—they may end up citing scientists with many social connections because these scientists are simply more likely to be widely known than less well-connected scientists.

---

[5] For a critique of this conception of scientific agency with regards to trust, see (Desmond 2021).

[6] They also mention prizes and inclusions in anthologies as long-run credit, but presumably the decisions about who receives a prize or an inclusion in an anthology is not done by a small group of gatekeepers (judges), but by citation count.

This one case study shows that citation count is not necessarily a good indicator of scientific achievement, but rather of social connectedness. The extensive research done by bibliometrics scientists on the patterns and underlying causes of citation behavior, gives a good overview of the different variables that citation count can track. In their landmark review of citation behavior, Bornmann and Daniel report on the following list (originally due to Shadish et al. (1995)) of possible reasons to cite an article:

(1) "Exemplar citation" (the cited item is a widely recognized reference);
(2) "Negative citations" (the cited item is a target for criticism or an example of bad science or scholarship);
(3) "Supportive citations" (the cited item adds credibility to a claim);
(4) "Creative citations" (the cited item is unusual or innovative);
(5) "Personally influential citations" (the author is intellectually indebted to the cited item);
(6) "Citations made for social reasons" (the cited item was authored by someone who played a role in the review process; the cited item was published in a prestigious journal).[7]

Only reason (5) is a direct measure of "scientific achievement" or "scientific quality"; whether the others also track achievement is more ambiguous. Reason (1) could potentially indicate scientific quality, but a citation to a canonical reference could also be the result of network effects. The same can be said of reason (3): citing an article in order to add credibility to a claim can simply indicate that the author is using the citation as a convenient claim to authority. Citing a source merely to add credibility to a claim does not necessarily signal an intellectual debt, nor even that the cited source is a genuine authority on the claim. For instance, it could signal that the cited article has a useful title for the citer's bibliography, or it could indicate that the journal in which the cited article appears is sufficiently prestigious for the citer's needs.

Given this multitude of reasons for citing, a safe conclusion is that citation count does not track scientific achievement in any straightforward one-to-one fashion. Given a high citation count, one cannot simply deduce that an article is a valuable scientific achievement. Any combination of the five reasons above could be responsible for the high citation count. The inference from citation count to scientific quality is only possible as a complex abductive inference, when other possible explanations have been discounted through other sources of evidence.

Even if the inference from citation count to scientific quality is invalid and problematic for any single individual case, could one not still put forward a heuristic based on citation count as a generally reliable strategy? This is an undecided issue. Bornmann and Daniel document how the most common intended targets of citations are articles that review prior work, articles that are cited because they are representative of a field, or authoritative articles that help establish legitimacy (Bornmann & Daniel, 2008, p. 62). As Bornmann and Daniel conclude, "citations can therefore be viewed as a complex, multidimensional and not a unidimensional phenomenon. Why authors cite can vary from scientist to scientist." (Bornmann & Daniel, 2008, p. 69).

A preliminary lesson here is that it is misleading to take citation count as a measure of "credit", if "credit" implies that it tracks genuine scientific achievement. A more

---

[7] Bornmann and Daniel 2008, p. 61: the terms are theirs, but explanations in parentheses are mine.

accurate conclusion would be that scientists cite what they find *useful*, and that what they find "useful" depends on the context. Only sometimes do they find it useful to cite genuine scientific achievements. Often enough it is quite useful to cite lesser-quality articles, for instance in order to fulfill the function of negative or supportive citation. Citations track, first and foremost, those articles that scientists find useful to cite: this may seem disappointingly uninformative, but it is closer to the truth than claiming that citations track achievement or quality.

5 Gatekeeper elimination as market deregulation?

The preceding discussion suggests a different metaphor for citation count. A citation seems less like the vote of an informed citizen in a healthy democracy, and more like that the cash a consumer buying a product in a market. A citation is not a vote of confidence; rather, it indicates that the author wanted to use the publication for their own purposes. A high citation count for an article then indicates a high degree of usefulness, and a relatively high price placed on that item. The aggregation of citations leads not so much to the *democratization* of evaluations of scientific quality as to their *marketization.*

This metaphor suggests a very different narrative: what gatekeeper elimination does is to "deregulate" judgments of scientific quality and submit them to the judgment of the "market" that will set the "price" of the publication. The mercantile metaphor also implies that the most successful articles are "commoditized", packaged as useful for citation purposes. It also implies that bringing an article to the market is only a part of the work, and must be accompanied by "advertising it" or even "selling it" (on, e.g., social media or the conference circuit) to "potential users" (i.e., scientific colleagues).

Why would such prestige hierarchies—where prestige is understood as being in demand by users—be inconsistent with open science values? Could one not still claim that, in general, the "marketization" of scientific research ultimately promotes its "democratization"? Markets, after all, offer equality of opportunity. However, to continue the metaphor, this only is true when the competition is *fair* and when market agents do not act to undermine fair competition (e.g., by forming monopolies, or by engaging in unfair practices). Reorganizing prestige differentials according to citation counts alone implies the endorsement of practices that are at odds with open science ideals; moreover, to avoid such practices, one needs other factors besides citation count (such as those of gatekeepers to regulate competition). To add empirical detail to this general argument, consider the causes of the current trend in increasing citation inequality.

5.1 The emergence of extreme citation inequality

According to a recent study, the top 1% of the most cited authors capture almost a quarter of all citations, translating to a Gini coefficient of 0.70 (Nielsen & Andersen,

2021).[8] This proportion grew from 15 to 25% between 2000 and 2015, with no sign of deceleration (Nielsen & Andersen, 2021, Fig. 1A).

What are the causes of rising citation inequality? Nielsen and Andersen note that it is not that the top 1%-authors have become a lot more productive. If one assumes that a publication with n co-authors gives each co-author 1/n publication credit,[9] then the total number of publication credits for the top 1%-authors has declined from 3 per year in 2000 to 2 per year in 2015 (Nielsen & Andersen, 2021, Fig. 2B). However, the number of co-authored publications has increased significantly: increased collaborations have been driving citation inequality.

Further, the increased collaborations are not due to more collaborations between scientists of equal rank in the academic hierarchy.[10] The dominant driver seems to be the greater incidence of collaborations between senior scientists (PIs) and junior or transient members of academia (PhDs and postdocs). This is the "lab model" of research. And this is how senior scientists are accruing citation credits at a much higher rate than previous generations of senior scientists, all the while without accruing a higher number of publication credits. Nielsen and Andersen describe a feedback dynamic where "Highly cited authors have better odds of winning grants, allowing them to expand their research laboratories and collaborative networks, ultimately leading to increased (full-count) publication and citation rates" (Nielsen & Andersen, 2021, p. 6).

In the resulting picture, a very high citation rate indicates an author's central position in a scientific network (this corroborates the view of Rubin, 2022), and more specifically, the author's institutional position. Institutional privilege leads to funding, which leads to hiring (junior) scientific collaborators, which leads to publications, citations, and, in turn, more funding. In this way, citation counts (and derivative metrics such as the h-index) do not just measure the usefulness of scientific outputs; rather, the "usefulness" of the outputs of the most highly cited authors seems to proceed from their *social status,* which in turn primarily indicates institutional privilege and access to research funding, and only secondarily (and in principle, optionally) scientific capacity or scientific achievement.

5.2 Academic rentiership

This is a Matthew effect in the sense that success begets success, but it has little in common with the Matthew effect originally described by Merton. As previously discussed, the original Matthew effect could be argued to be compatible with open science values (via Strevens, 2006). However, this does not seem to be possible with the type of Matthew effect just described. Instead, the element producing an upwards

---

[8] For context, citation inequality is higher than even the highest income inequality on a national level. According to the World Bank, that exists in South Africa, with a Gini coefficient of 0.63 (World Bank, 2021).

[9] This assumption is not unreasonable. In some cases, a co-authored paper with 2 authors would not reduce the work by half. However, in other cases, a co-authored paper with many other authors, in which one has contributed but a small part, might entail less work than the corresponding fractional credit.

[10] However, notable exceptions here are the large-scale collaborative projects in astronomy and high-energy physics, which have had a such a large influence on publication and citation numbers in physics that averages across science as a whole have shifted to a surprising extent. See Nielsen and Andersen (2021, Fig. 1).

spiral is an institutional position of privilege. The causal chain is thus quite different than in the classic Matthew effect. It goes from institutional position, to increased access to research funding, to the ability to hire (more) junior academics, to increased number of publications, to increased number of citations, to higher scientific prestige, and then back again to increased research funding.

To kickstart and maintain this causal chain, scientific merit or capacity is optional. Of course, some scientific merit or quality is present *before* this reinforcing dynamic becomes active: i.e., only high-quality scientists are given such positions. Moreover, one could perhaps also assume that considerable scientific capacity raises the probability that resulting collaborations with junior academics will generate high citation counts. However, scientific merit or quality becomes much less important once the institutional position is sufficiently established, and good junior faculty can be attracted; the rewards then sustain themselves in a way that is structurally similar to buying property in a prime location.

This type of Matthew effect could be termed "academic rentiership", since it clearly conveys key elements of the dynamic, and since it further develops the market metaphor (rather than the democracy metaphor). Academic "capital" can be accumulated to the point that a scientist (or academic) can further increase this capital by collecting "rents" (in the form of citations from collaborative authorships). According as citation count becomes increasingly institutionalized as the primary indicator of scientific quality, so the incentive grows to pursue any successful strategy to maximize citation count, including "rentiership" strategies. Not just individual scientists are incentivized here; institutions may also be incentivized to provide the best conditions for their scientists to be efficient rentiers.

Could one reply that "rentiership" is potentially consistent with the ideals of open science, and that it *can be* democratic and inclusive? In response, to return to the property metaphor, it is important to underline the difference between a rentier—an absentee landlord who collects rent without actively providing any kind of service—and a "good" landlord, who tends to tenants' needs and invests in the upkeep of a property. A similar distinction is relevant here. As long as a scientist is able to win a position of institutional privilege in order to guide junior collaborators towards doing better work than they would otherwise, it would seem fair that this type of inequality exists in the scientific community. Senior scientists bring skills, insight, or competence to the table, and this kind of collaboration is a type of the "division of cognitive labor" many maintain is crucial for scientific progress (e.g., Kitcher, 1990). Thus, the inequality could be viewed as fair in the Rawlsian sense, namely as an arrangement that would be endorsed by all researchers behind the veil of ignorance. By contrast, the ideal type of "academic rentiers" would refer to scientists who seek to extract the maximum of "rent" (citations) while investing the minimum of work to ensure quality. Whether they obtained their position of privilege through some past demonstration of superior scientific capacity would be irrelevant here: in rentiership, the continued exercise of superior scientific capacity is strictly optional, and in this sense would not seem to be an arrangement that would be endorsed by all researchers.

Could one still insist that rentiership, while perhaps unfair, is consistent with democracy and inclusiveness? For instance, one could perhaps argue that equality of opportunity—one of the hallmarks of (liberal) democracies—is not endangered by

rentiership, since every scientist, in principle, can become a rentier?[11] As a first, formal response, the argumentative onus here would seem to be on open science advocates: rentiership describes a state of affairs that is *prima facie* closer to an oligarchy than a democracy, and one that is in any case a far cry from "the culture of sharing" envisaged by some advocates (Persic et al., 2021). Advocates of open science have typically used "democracy" and "inclusiveness" in only loose senses, assuming that eliminating gatekeepers will automatically make science more democratic and inclusive, and the phenomenon of rentiership problematizes any such assumption. The onus is on open science advocates to engage with the relevant literatures on the nature of democracy (Christiano & Bajaj, 2022) in order to show exactly how rentiership is consistent with the open science vision.

However, a second, more substantive response is possible too: there are reasons to believe rentiership is genuinely at odds with democracy and inclusiveness. Rentiership may indeed be, in principle, open to any scientist, but the prior decisions about the nature of the arrangement (i.e., rentiership) and about who may be granted institutional privilege are not made in a democratic way (i.e., the whole scientific community does not participate in this decision-making). Instead, they are made by a small group of individuals linked to the institution in question, e.g., hiring committees, funding agency committees, or science policy-makers. This small group plays an indirect and de facto gatekeeping role by deciding what institutional privilege will be obtained by which individuals. Hence the question becomes: do the decisions of this small group represent the whole scientific community? Deciding on rentiership allocation seems most straightforwardly analogous to an election where powerful donors decide who the candidates will be. While this may not entirely eliminate the democratic character of an election, insofar the donors themselves have not been democratically elected, it is at least a negative factor. It also erodes genuine inclusiveness: since rentiership is a negative factor for science regardless of how such positions are assigned (a rentier per definition need not exercise superior scientific capacity to accrue citations), it can be assumed to be a form of unfair inequality (since presumably scientists in the original position would not endorse it). The decisions on the allocation of rentiership therefore cannot be assumed to be made by all scientists in the scientific community: at least some scientists are excluded from this process.

5.3 The reappearance of gatekeepers?

Another objection open to advocates of gatekeeper elimination would be to argue that promoting citation-based prestige differentials does not entail an *endorsement* of rentiership. After all, the primary motivation to eliminate gatekeepers was make the evaluation of scientific work "a more democratic process" (Heesen & Bright, 2021, p. 647), and this entails a rejection of rentiership as described here. However, in response, the real problem is that nothing in the concept of citation-based prestige differentials involves a *rejection* of rentiership. Citation-based prestige incentivizes a variety of citation-seeking research strategies, some of which, like rentiership, are incompatible with open science values. Other evaluative principles, in addition to

---

[11] Thanks to an anonymous referee for pushing me on this point.

citation metrics, would be needed to support a rejection of rentiership. There is no conceptual resource present in the proposal to eliminate gatekeepers that can help address the problem of rentiership.

This is a more fundamental problem than may seem at first. The existence of citation rentiership strategies indicates that there is a difference between "good citations" (accrued through strategies that respect open science values) and "bad citations" (accrued through other strategies that do not respect open science values). This means that citations are *neutral* with regards to open science values; whether or not citations "tend to be good for science" depends on *other* factors (independent of citation count) that structure scientists' motivation and scientific prestige hierarchies. There are all sorts of strategies to increase citation count, including the overtly unethical ones: exaggerating results, refusing to share data even when appropriate, or manipulating citation patterns and general "citation hacking" (Van Noorden, 2020). Taking citation count as the only indicator of prestige is entirely dependent on the assumption that most scientists will not engage in various forms of citation hacking—and that assumption only makes sense if scientists have *other* motivations for doing good science, for instance, intrinsic motivations, or a sense of professionalism (Desmond, 2020).

Given that citation-based prestige only promotes open science values when prestige and motivation are *not solely* defined by citation count, one can then also enquire as to *who* will make the distinction between "good" and "bad" citations. Who will determine whether this specific set of citations is consistent or not with the ideals of open science? This judgment cannot be based on citation count alone. This constitutes in effect a rationale for gatekeepers: analogous to market regulators, gatekeepers can step in and dampen the pernicious effects of, for instance, rentiership, by making decisions about quality that override citation count (and with it, the "price" put on an article by the market). They can also ensure a basic trustworthiness (as regulators do when mandating basic product standards), and the process by which such trustworthiness is assured inevitably takes on many of the properties of the peer-review process.

In sum: on closer inspection, the proposal of citation-based prestige loses plausibility. One cannot claim that citation-based prestige is an *improvement* with regards to inclusiveness and democracy over current structures of prestige allocation.[12] Not only do citation counts not necessarily track scientific achievement, but there is a great variety of citation-maximizing behaviors, and some of these are at odds with open science values. Eliminating gatekeepers and promoting citation counts would not seem to be the policy that would, by itself, promote open science values.

6 Search algorithms

Search algorithms rank scientific publications according to the search terms used, and thus allow scientists to make judgments about what publications are worth learning from. In this sense, search rank instead of citation count could be considered as

---

[12] In what is called "citation-based prestige", prestige is allocated solely in light of citation count or some other citation metric. This is an ideal type, and in current structures of prestige allocation, many factors contribute, including gatekeeping (i.e., whether or not a manuscript has passed through peer and editorial review).

a contender for shaping prestige differentials in gatekeeper-less science. Using search algorithms can aid scientists to avoid some of the pitfalls associated with relying on citation-based heuristics, since some valuable articles may be under-cited. To return to the market metaphor: search algorithms help scientists find undervalued gems that have been overlooked by the market. In a gatekeeper-less science, search algorithms would be even more crucial in navigating vast academic repositories such as arXiv, PubMed, or Scopus. Even though search algorithms are not promoted quite as visibly as an alternative to gatekeeper elimination as citation metrics are, we should consider whether there are good grounds for believing that relying on algorithms would help realize the ideal of democratic, inclusive, and transparent science.

6.1 The rise of machine-learning algorithms

Let us start with PubMed's search algorithm as a case study, since information regarding the design choices for PubMed's search algorithms are in the public domain (see the report in Fiorini et al., 2018). The case of PubMed illustrates how machine-learning search algorithms are both very effective aids for users to deal with information overload, and how they constitute a de facto form of gatekeeping.

Prior to 2017, users could only navigate PubMed by searching the meta-data of articles, such as author, title, or journal, for certain terms. Initially, PubMed returned all search hits in reverse chronological order, with the newest research items first. Then, in 2013, a new sorting algorithm was introduced, where search hits were ranked according to "relevance", which was an estimate for how close the content of the document corresponded to the search terms.[13]

Then, in 2017, the "Best Match algorithm", a machine learning algorithm, was launched. The reason for this was the perception that users, accustomed to using Google, had developed an expectation that searches had to be fast and simple. Specifically, the relevant information had to be found on the first page of results (only 20% of clicks are for the results on the second and further pages). In response, PubMed introduced a machine learning algorithm ("learning-to-rank": Fiorini et al., 2018, p. 17), where the algorithms "learn" how to respond to queries with particular results based on feedback they receive from the user's actual clicking behavior. This feedback is contained in the PubMed search logs.

Machine learning algorithms are very powerful tools for quickly navigating very large databases, but they come at the cost of introducing novel biases. All algorithms inherit the biases of the computer scientists who design them. For instance, the relevance sorting algorithm requires certain values to be assigned to parameters, and the assigning of these values can be biased (Kiester & Turp, 2022). However, machine

---

[13] More precisely, "relevance" was operationalized through the inverse document frequency model (see Fiorini et al., 2018, p. 17). Here the content of the documents was scanned for the search terms, and the rank of hits was determined by (1) how often one of the search terms appeared in the document, and (2) how often the search term appeared in the whole database. So, if a search term appeared frequently in a particular document ("term frequency"), but turned out to be common across all documents, this was judged to be less relevant than the non-common search term being found frequently (hence "inverse document frequency").

learning algorithms are influenced by the biases of *users* as well. If users, upon searching for some combination of keywords "X, Y, Z", consistently click on search result #5, then the algorithm will "learn" that this search result needs to be ranked higher.

Machine learning algorithms proceed on two fundamental assumptions: clicks reveal user judgments of what they believe is relevant, and users are good judges of what is in fact relevant. Both assumptions are problematic. Like citation behavior, click behavior does not necessarily reveal what users believe to be important or relevant—a user may click on sensational content while despising it.[14] The second assumption is flawed in a more interesting way for purposes of this paper. Users may click on what they *believe* to be relevant, but their judgments in that regard can be influenced by various "societal biases" (Kiester & Turp, 2022).

To the best of my knowledge, detailed studies on just how such biases could affect click behavior on PubMed are lacking. However, the literature on how machine learning algorithms perpetuate existing biases in society—whether sex, race, age, or socio-economic status—is burgeoning (Mehrabi et al., 2021), and there seems to be no principled reason why searches in scientific repositories would escape the reach of at least some such biases. Given the general role that heuristics and biases play in human cognition (Haselton et al., 2016), and given that machine learning algorithms seek to imitate statistical patterns of human decision-making (i.e., patterns of clicks), it would not be surprising should prestige and conformist biases show up in the results returned by searches through scientific repositories.

This discussion gives a first indication of how an increasing reliance on search algorithms would simply throw up *new* challenges to the ideals of democracy and inclusiveness, and cannot be considered as a means of avoiding all the pitfalls associated with gatekeeping. Search algorithms—especially machine learning algorithms trained on large datasets—do not neutrally highlight which research items can be considered worthy of the attention of the searcher, and which cannot. They reflect statistical patterns of clicking behavior, and just as citations do not necessarily indicate scientific excellence, neither do clicks. In effect, *new barriers* are erected, namely the relative position in search results. A lowly position (i.e., not appearing on the first page) is tantamount to being ignored.

6.2 Quality raters

The latest iterations of search algorithms are not only informed by user click behavior. Consider the most comprehensive academic search engine, with more records and better citation data than that of its rivals: Google Scholar (Gusenbauer, 2019; Martín-Martín et al., 2021). How precisely Google Scholar ranks search results is not public knowledge, but there have been attempts to reverse engineer the Google Scholar algorithm. One such study indicated that the most weight was given to citation count in determining the rank of research items (Beel & Gipp, 2009). It seems to be unknown whether this weight has changed over the years, but a later study sampling 64,000

---

[14] There is much more to be said about this point. There is an analysis to be made here about how click behavior may follow first-order desires without necessarily following second-order desires.

documents found a strong correlation between the citation count of an item and the item's position in the ranking (Martin-Martin et al., 2017).

In this way, Google Scholar combines multiple ranking mechanisms, including machine learning (giving a higher rank to items that other users seem to prefer) and citation heuristics (giving a higher rank to items that are highly cited). However, even though there are, to the best of my knowledge, no detailed empirical studies about the biases of Google Scholar, it would not be surprising if Google Scholar also manifested the status biases present in citation behavior.

Moreover, the current iterations of Google's search algorithm reveal yet another dimension in the relation between search algorithms and gatekeeping. Today Google employs many of what it calls "search quality raters" to tweak the results, or weighting given to different factors. These Raters are given extensive guidelines (Google, 2021) on how to check and correct the results of the Google algorithm. In particular, Google looks for their input with regard to searches on topics where the cost of error is high, such as "Your Money or Your Life" webpages, i.e., webpages about personal finance and healthcare (Google, 2021, p. 10). Here Google emphasizes the importance of getting the rank of these webpages right.

What are the guidelines? In particular, search quality raters have to look for E-A-T: Expertise, Authority, and Trustworthiness (Google, 2021, pp. 19–20). How are these properties discerned? By traditional markers of prestige and trustworthiness: professional qualifications (J.D., M.D., etc.), awards by accredited bodies (Pulitzer Prizes, etc.), signs that a webpage has been edited and reviewed, and websites with rigorous editorial and review policies (Google, 2021, p. 20). Search quality raters must even be able to judge information pages on whether they have been written by "people or organizations with appropriate scientific expertise" and whether the pages "represent well-established scientific consensus on issues where such consensus exists" (Google, 2021, p. 20). This means that the latest iteration of Google search depends on the judgments of two relatively small groups of individuals: the quality raters who decide which results to promote and which results to suppress, and the traditional gatekeepers (peer-reviewers, professionals, editors) on whom the raters base their judgments.

It is most likely *not* the case that such quality raters are used for the algorithm of Google Scholar (for who could "rate the quality" of a technical scientific article other than expert scientists themselves?). In fact, the need for quality raters is not crucial, precisely because of the prevalence of peer-review, which is, in effect, a different type of quality rating. However, the fact that the general Google search algorithm needs a large number of human quality raters does hold general lessons for what type of algorithm management is needed when gatekeeping is eliminated. In the scenario where gatekeepers are eliminated, Google Scholar or similar algorithms *would* need such quality raters, because the requirements for publishing in a scientific repository would no longer be any more stringent than publishing a blog.

From the perspective of the open science rationale for eliminating gatekeepers, this is a highly ironic. Machine-learning algorithms were originally intended to learn automatically from the behavior of crowds of users (a wisdom-of-the-crowds rationale). However, today, even the most widely used search engine, with privileged access to the data of user behavior, must *nonetheless* be redirected by small groups of specially

trained employees in order to make more trustworthy judgments about which search results are "truly" relevant.

6.3 The reappearance of gatekeeping

Search algorithms may seem to be *prima facie* allies of open science. Scientists can use them in an individualized way to find valuable articles even though the articles in question may not have been approved by peer-reviewers, even though they may not have been included for publication in a prestigious journal, and even though they may be under-cited. However, once we delve into the details of how search algorithms are actually engineered, a very different picture emerges. As the importance of search algorithms grows, the most accurate search algorithms learn from the aggregate patterns of user behavior. These aggregate patterns perpetuate harmful biases that are not consistent with the type of ideals that open science professes.

This is the same type of problem that arose with relying on citation behavior (as opposed to click behavior) as a proxy for quality; however, it is a problem that has been around for longer, and it is interesting that corporations have evidently decided that the best solution is to reintroduce new types of gatekeepers in order to avoid relying too much on "bad" click behavior. These quality raters in effect feed machine learning algorithms "good" click behavior (based on the guidelines). Moreover, these new gatekeepers to a surprising extent defer to the judgments of the traditional gatekeepers that were originally believed to be superfluous. This sequence of events has played out on the world-wide web, but not on scientific repositories precisely because of the presence of prepublication peer-review. If it were to be eliminated, and scientific publication were to become like blog posting, the search algorithms of scientific repositories would need new gatekeeping "quality raters" once sufficient untrustworthy content would be posted.

In sum, our increased reliance on search algorithms is either at odds with open science values of inclusiveness, or else implies the need for a new form of gatekeeping that, moreover, relies on traditional gatekeeping (peer-review). Hence, far from providing a plausible path towards a gatekeeper-less future, search algorithms simply involve novel forms of gatekeeping.

7 Conserving gatekeeping in open science

The main thrust of this paper is negative: to undermine a certain widespread line of reasoning where gatekeeping elimination is unproblematically viewed as a step towards the promotion of open science values, such as democracy and inclusiveness. Such lines of reasoning explicitly assume that it is more democratic and inclusive to judge scientific merit by a type of wisdom-of-the-crowd process rather than a peer-review process where a small number of gatekeepers judge on merit.

In response, this paper draws attention to the recalcitrant nature of prestige biases in science. Given information overload, one should not expect prestige biases to be

eliminated, only to be transformed. Redefining prestige in terms of, for instance, citation counts will not flatten prestige differentials but simply introduce novel prestige differentials. The question is whether these novel prestige differentials are more inclusive and democratic, and this paper argues that at least two of the proposed prestige differentials are not more inclusive or democratic. Differentials based on citation count imply that scientists such as academic rentiers would be accorded high status, but that does seem consistent with the open science values of inclusiveness and democracy. Differentials based on search engine rank not only rely heavily on citation metrics but also on teams of gatekeepers to adjust the results of such algorithms behind the scenes. These scenarios are not acknowledged by the idealized arguments for eliminating gatekeeping.

Could gatekeeping be replaced by yet some other reform of prestige differentials that *would* guarantee the realization of open science ideals? This question is not explicitly considered in this paper; however, if one were to generalize from the failure of citation-based or rank-based prestige, a more general impossibility result could be suggested. Consider why citation metrics are so popular: they allow both non-specialists and non-scientists (administrators, managers) to judge the quality of specialist researchers. In fact, citation count has long become institutionalized as an evaluation metric for hiring, promotion, and even in the awarding of grants and prizes (Safer & Tang, 2009).

Citation metrics provide a *rule* for making evaluative judgments that does not require expert knowledge or even much individual deliberation. Anyone can go to a database, look up a virologist's citation count (or some derivative metric such as their h-index) and make a judgment about their prestige. Such rules for distinguishing between good and bad research, or more and less capable scientists are bound to fail because they can be *gamed*. In general, any metric can lead to a "looping effect" (Hacking, 1996), where the measurement methodology ends up changing that very feature. Citation metrics were intended to capture the quality of individual scientists, but institutionalizing those metrics as direct proxies for quality has led to new dynamics where scientists pursue citation-maximizing strategies that do not necessarily involve producing quality work.[15]

In this way, if yet another evaluative rule that does not involve any gatekeeping *were* to be proposed, it would need to include a rule for making evaluative judgements *without* any expert knowledge being required (because otherwise it would amount to a form of gatekeeping by experts). However, as experts would tailor their work to maximize this rule, one would obtain the situation where the rule would no longer be able to distinguish between good and bad research. Then there would be a need for human beings with their capacity for deliberation and judgment to produce evaluative judgments. Such humans would simply be de facto gatekeepers.

This suggests a closing line of thought, where the close links between gatekeeping, scientific agency, and scientific professionalism are intimated. We can expect that gatekeeping practices will continue to evolve, just as they did in the wake of past changes in communication technology (i.e., the advent of the typewriter or photocopier). More positively, since gatekeeping can be viewed as a form of scientific

---

[15] To rephrase the issue in terms of Hacking's original context: bibliometrics such as citation count try to capture "excellent science" as if it were a natural kind, but in so doing change the phenomena of research that bibliometrics try to classify.

self-regulation (Redman, 2023), its implicit norms of professionalism can help align gatekeeping with open science values, and help inform new gatekeeping practices. To situate this positive proposal, it is helpful to first consider the broader sociological-historical context of professionalism, and the way in which gatekeeping is a part of the social structures of scientific professionalism.

Reviewer professionalism. The distrust shown towards gatekeeping inherits many of the motifs of the distrust once shown towards ideals of professionalism, starting in the 1980s. Under pressure from new ideals such as New Public Management (which sought to impose market and bureaucratic forms of organization, see e.g. Carvalho & Correia, 2018), old structures of professionalism were dismantled or "hybridized" in virtue of professionals' unjustified power and privilege, self-interestedness, and tendency to error. Output or performance metrics were introduced to judge professionals. Performance metrics are very expedient because they allow administrators and managers to judge the work of professionals. For this reason, some sociologists view such metrics as tools of managerial and bureaucratic models of organization (see e.g. Vican et al., 2020).

This sociological-historical context casts yet a different light on the proposal of gatekeeper elimination. Eliminating gatekeepers and increasing the importance of metrics would be one of the textbook strategies to undermine the professional autonomy of the scientific community (for a *locus classicus* on this issue, see Freidson, 2001, chapter 8). Demonstrable shortcomings of forms of professional expertise—as is clearly the case in gatekeeping (Lee et al., 2013)—typically lead to renewed calls for more bureaucratic control and diminished professional autonomy. Here, open science proposals such as gatekeeper elimination inadvertently align with proposals to bring scientific research under expanded bureaucratic forms of surveillance.

To what extent science *should* be organized according to the model of professionalism (where scientists maintain considerable autonomy but are expected to adhere to relatively demanding codes of ethics), or the model of bureaucracy (where managers and other non-scientists make decisions on how scientific activity is organized) is a separate question. Elsewhere it is argued that the logic of professionalism and other organizational logics, such as bureaucratic or market control, are always present in mixed form in science, but that the logic of professionalism is, nonetheless, the single most appropriate one for scientific research (Desmond, 2020).

Gatekeeping fits the model of a professional activity quite well: scientific gatekeepers are given considerable autonomy in forming their judgments, and in return are expected to uphold standards of impartiality, diligence, and competence.[16] By contrast, there are no such expected standards for casual readers. Even if they are experts in the field, they are free to read as deeply or superficially as they see fit, or to cite in a merely perfunctory matter (Bornmann & Daniel, 2008). Casual readers may ignore a paper because they do not like the conclusion. By contrast, a peer-reviewer would be considered to be failing basic norms of reviewer integrity if they rejected a paper because they happened to not like the conclusion, or because they did not read

---

[16] See for instance *The Ethical Guidelines for Peer Reviewers* published by the Committee on Publication ethics: reviewers are expected to review only within areas of "expertise", and to remain "unbiased by considerations related to […] characteristics of the atuhors, origins of a manuscript or by commercial considerations. (COPE Council 2017, pp. 4–5).

the paper carefully, or because they lacked the requisite knowledge to understand the paper. This points to the normative dimension of gatekeeping: de facto gatekeeping may of course be partial, lazy, and incompetent with disappointing frequency, but the fact remains that the *disappointment* points to the presence of norms and expectations that govern peer-review.

It is in this specific sense that peer-review can be categorized as a professional activity. Moreover, whether peer-review occurs before publication or after, or whether referee reports are published or not, are not really relevant. If post-publication review is carried out with rigorous standards, and if prestige is attached to receiving positive post-publication review, then post-publication review is also a form of gatekeeping, because publications that fail to receive positive post-publication reviews will be most likely ignored or discounted.

Professionalism mandates continual reform. Gatekeeping professionalism is flexible enough to inform how gatekeeping practices can and must change in response to a changed scientific environment. One component of professionalism is diligence, which entails taking *reasonable precautions* (such as one could expect of peers of similar background or training) to avoid even negligent errors in judgment. What types of behavior can be considered "diligent" depends on the circumstances. In this sense, as the scientific environment changes with online publication and open science practices (e.g. preprints, or pre-registration of protocols), new precautionary measures become reasonable. Hence, one could even claim that professionalism mandates continual reform of practices, since as the scientific environment changes, new types of precautionary measure become reasonable to expect.

Consider, for instance, the pre-registration of protocols. This allows reviewers to compare the submitted paper to the pre-registered protocol and make more informed judgments about scientific quality: e.g., whether the research was hypothesis-driven, or whether some form of P-hacking occurred, and so on. As pre-registration becomes widespread, it becomes reasonable to expect of reviewers that they take the pre-registered protocol into consideration. Similarly, as the sharing of data through online supplements becomes widespread, it becomes reasonable to expect reviewers to consult this supplement in evaluating a submission, since it allows in principle for a more informed judgment. These expectations—checking pre-registered protocols, examining online supplements—cannot be reasonably held of most causal scientific readers, even if they are experts in the relevant field.

In this way, gatekeeping should be thought of as part and parcel of scientific professionalism, where individual, considered, and autonomous judgment of quality is given special weight and where the process of judging is expected to adhere to norms of integrity. What "considered" judgment means is dependent on what information is readily available, and here open science practices can be integrated to strengthen gatekeeping professionalism.

Gatekeeping as ally of open science values. Another crucial normative aspect of gatekeeping lies in how it *professes* impartiality: it aspires to evaluate scientific manuscripts regardless of the prestige of the author or domain. In this sense, peer-review is *designed* to be a corrective to the types of Matthew effect (like rentiership) described earlier that are antithetical to open science. The practice of blind peer-review, for instance, is an attempt to promote impartiality in scientific evaluation. Of course,

de facto peer-review may fall short of the ideals of impartiality, but such shortcomings do not by themselves constitute an argument to eliminate it. This paper documents how alternatives to gatekeeping reinstate some of these exact same shortcomings but without the professional ideals that govern gatekeeping.

Gatekeeping also lessens the importance of author prestige for the scientific community in deciding about the value of a scientific publication. Gatekeeping is a source of prestige differentials, whether through conferring the prestige of a journal on a publication, or by giving manuscripts the label of "peer-reviewed". However, because these prestige variables are relatively decoupled from the prestige of the publication's author(s), gatekeeping dilutes the importance of author prestige in the overall prestige of a publication. Moreover, it lowers the importance of citation count in order to receive attention from scientific peers. In this process, it also lowers the incentive to single-mindedly maximize citation account (for instance through rentiership strategies). By providing a (relatively) independent source of prestige, gatekeeping dampens some of the pernicious runaway effects that could occur from solely relying on citation count.

Here good gatekeeping presents itself as an ally of open science. Even though gatekeeping can lead to tribal dynamics, where insiders are privileged over outsiders, this represents a type of tribal gatekeeping and a failure of scientific gatekeeping. Good gatekeeping offers protection against the herding dynamics that arise when current status and prestige are given too much importance in evaluative judgments of scientific quality.

Conclusion. Traditional peer-review often succumbs to prestige biases, and the proposal to eliminate gatekeeping seems to be in keeping with the desire to bring about a more egalitarian science: more democratic, more inclusive, and more transparent. While this desire is commendable, it is not realistic to hope that status differentials can be engineered away through an ingenious reform of science policy. Due to information overload, scientists need help from others to make distinctions regarding the quality of a work. Eliminating gatekeeping would not only *not* flatten prestige hierarchies, but alternatives (relying on citation counts or search algorithms) would incentivize behaviors that run counter to open science ideals and in fact call for new types of gatekeepers in order to avoid the unintended consequences from gatekeeper elimination. Instead, good, integrous gatekeeping should be conserved as a part of scientific professionalism, and as an ally of the goals of open science. Open science advocates should not promote the elimination of gatekeeping but rather its conservation and reform.

Declarations

Conflict of interest The author has no conflicts of interest to declare.

References

Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological Bulletin, 141*(3), 574–601. https://doi.org/10.1037/a0038781

Atkisson, C., O'Brien, M. J., & Mesoudi, A. (2012). Adult learners in a novel environment use prestige-biased social learning. *Evolutionary Psychology, 10*(3), 147470491201000320. https://doi.org/10.1177/147470491201000309

Bartling, S., & Friesike, S. (Eds.). (2014). *Opening science*. Springer. https://doi.org/10.1007/978-3-319-00026-8

Beel, J., & Gipp, B. (2009). Google Scholar's ranking algorithm: An introductory overview. In *Proceedings of the 12th international conference on scientometrics and informetrics* (Vol. 1, pp. 230–241).

Bergson, H. [1932] 1963. *The two sources of morality and religion*. University of Notre Dame Press.

Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation, 64*(1), 45–80. https://doi.org/10.1108/00220410810844150

Boyd, R., & Richerson, P. (1985). *Culture and the evolutionary process*. The University of Chicago Press.

Brezzi, M., González, S., Nguyen, D., & Prats, M. (2021). In *An updated OECD framework on drivers of trust in public institutions to meet current and future challenges*. OECD working papers on public governance 48. OECD Publishing. https://doi.org/10.1787/b6c5478c-en.

Carvalho, T., & Correia, T. (2018). Editorial: professions and professionalism in market-driven societies. *Professions and Professionalism, 8*(3), e3052. https://doi.org/10.7577/pp.3052

Christiano, T., & Bajaj, S. (2022). Democracy. In E.N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. https://plato.stanford.edu/entries/democracy/.

Chudek, M., Heller, S., Birch, S., & Henrich, J. (2012). Prestige-biased cultural learning: Bystander's differential attention to potential models influences children's learning. *Evolution and Human Behavior, 33*(1), 46–56. https://doi.org/10.1016/j.evolhumbehav.2011.05.005

COPE Council. (2017). COPE ethical guidelines for peer reviewers. Committee on Publication Ethics. https://doi.org/10.24318/cope.2019.1.9.

Crocker, J., & Lynne Cooper, M. (2011). Addressing scientific fraud. *Science, 334*(6060), 1182–1182. https://doi.org/10.1126/science.1216775

Desmond, H. (2020). Professionalism in science: Competence, autonomy, and service. *Science and Engineering Ethics, 26*(3), 1287–1313. https://doi.org/10.1007/s11948-019-00143-x

Desmond, H. (2021). Incentivizing replication is insufficient to safeguard default trust. *Philosophy of Science, 88*(5), 906–917. https://doi.org/10.1086/715657

Fiorini, N., Canese, K., Starchenko, G., Kireev, E., Kim, W., Miller, V., Osipov, M., et al. (2018). Best match: New relevance search for PubMed. *PLoS Biology, 16*(8), e2005343. https://doi.org/10.1371/journal.pbio.2005343

Freidson, E. (2001). *Professionalism, the third logic: On the practice of knowledge*. University of Chicago Press.

Gewin, V. (2016). Data sharing: An open mind on open data. *Nature, 529*(7584), 117–119. https://doi.org/10.1038/nj7584-117a

Google. (2021). General guidelines. Retrieved from https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf

Gosepath, S. (2021). Equality. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Summer 2021. Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2021/entries/equality/

Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics, 118*(1), 177–214. https://doi.org/10.1007/s11192-018-2958-5

Habgood-Coote, J. (2021). What's the point of authors? *The British Journal for the Philosophy of Science.* https://doi.org/10.1086/715539

Hacking, I. (1996). The looping effects of human kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524021.003.0012

Haselton, M. G, Nettle, D., Murray, D. R., & Buss, D. M. (2016). The evolution of cognitive bias. In *Handbook of evolutionary psychology* (pp. 968–987).

Heesen, R., & Bright, L. K. (2021). Is peer review a good idea? *The British Journal for the Philosophy of Science, 72*(3), 635–663. https://doi.org/10.1093/bjps/axz029

Henrich, J. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior, 22*(3), 165–196. https://doi.org/10.1016/S1090-5138(00)00071-4

Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*(5843), 1360–1366. https://doi.org/10.1126/science.1146282

Jung, R. G., Di Santo, P., Clifford, C., Prosperi-Porta, G., Skanes, S., Hung, A., Parlow, S., et al. (2021). Methodological quality of COVID-19 clinical research. *Nature Communications, 12*(1), 943. https://doi.org/10.1038/s41467-021-21220-5

Kiester, L., & Turp, C. (2022). Artificial intelligence behind the scenes: PubMed's best match algorithm. *Journal of the Medical Library Association, 110*(1), 15–22. https://doi.org/10.5195/jmla.2022.1236

Kitcher, P. (1990). The division of cognitive labor. *The Journal of Philosophy, 87*(1), 5–22. https://doi.org/10.2307/2026796

Knoepfler, P. (2015). Reviewing post-publication peer review. *Trends in Genetics, 31*(5), 221–223. https://doi.org/10.1016/j.tig.2015.03.006

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2–17. https://doi.org/10.1002/asi.22784

Martin-Martin, A., Orduna-Malea, E., Harzing, A.-W., & López-Cózar, E. D. (2017). Can we use google scholar to identify highly-cited documents? *Journal of Informetrics, 11*(1), 152–163. https://doi.org/10.1016/j.joi.2016.11.008

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & López-Cózar, E. D. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multi-disciplinary comparison of coverage via citations. *Scientometrics, 126*(1), 871–906. https://doi.org/10.1007/s11192-020-03690-4

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 115:1-115:35. https://doi.org/10.1145/3457607

Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review, 22*(6), 635–59. https://www.jstor.org/stable/2089193.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie, N., du Sert, U., Simonsohn, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

NAS, National Academies of Sciences, Engineering, and Medicine. (2018). *Open science by design: Realizing a vision for 21st century research*. National Academies Press.

Nature. (2018). Science benefits from diversity. *Nature, 558*(7708), 5–5. https://doi.org/10.1038/d41586-018-05326-3

Nielsen, M. W., & Andersen, J. P. (2021). Global citation inequality is on the rise. *Proceedings of the National Academy of Sciences, 118*(7), e2012208118. https://doi.org/10.1073/pnas.2012208118

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*(3), 137. https://doi.org/10.1027/1864-9335/a000192

Nuremberg Code. (1947). The nuremberg code. Retrieved from https://history.nih.gov/research/downloads/nuremberg.pdf

OECD. (2021). *Government at a glance 2021. Government at a glance*. OECD Publishing. https://doi.org/10.1787/1c258f55-en

Persic, A., Beigel, F., Hodson, S., & Oti-Boateng, P. (2021). The time for open science is now. In S. Schneegans, J. Lewis, & T. Straza (Eds.), *UNESCO science report: The race against time for smarter development.* UNESCO Publishing.

Popper, K. R. [1945] 2011. *The open society and its enemies*. Routledge.

Redman, B. (2023). *Reconstructing research integrity: Beyond denial*. Springer. https://doi.org/10.1007/978-3-031-27111-3

Rubin, H. (2022). Structural causes of citation gaps. *Philosophical Studies*. https://doi.org/10.1007/s11098-021-01765-3

Safer, M. A., & Tang, R. (2009). The psychology of referencing in psychology journal articles. *Perspectives on Psychological Science, 4*(1), 51–53. https://doi.org/10.1111/j.1745-6924.2009.01104.x

Shadish, W. R., Tolliver, D., Gray, M., & Sen Gupta, S. K. (1995). Author judgements about works they cite: Three studies from psychology journals. *Social Studies of Science, 25*(3), 477–498. https://doi.org/10.1177/030631295025003003

Sheldon, T. (2018). Preprints could promote confusion and distortion. *Nature, 559*(7714), 445–446. Retrieved from https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=00280836&v=2.1&it=r&id=GALE%7CA572720506&sid=googleScholar&linkaccess=abs

Spier, R. (2002). The history of the peer-review process. *Trends in Biotechnology, 20*(8), 357–358. https://doi.org/10.1016/S0167-7799(02)01985-6

Stengers, I. (2018). *Another science is possible: A manifesto for slow science*. Wiley.

Strevens, M. (2003). The role of the priority rule in science. *Journal of Philosophy, 100*(2), 55–79. https://doi.org/10.5840/jphil2003100224

Strevens, M. (2006). The role of the Matthew effect in science. *Studies in History and Philosophy of Science Part A, 37*(2), 159–170. https://doi.org/10.1016/j.shpsa.2005.07.009

Taubes, G. (1993). Publication by electronic mail takes physics by storm. *Science, 259*(5099), 1246–1248. https://doi.org/10.1126/science.259.5099.1246

Van Noorden, R. (2020). Signs of 'Citation Hacking' flagged in scientific papers. *Nature, 584*(7822), 508–508. https://doi.org/10.1038/d41586-020-02378-2

van Schalkwyk, M. C. I., Hird, T. R., Maani, N., Petticrew, M., & Gilmore, A. B. (2020). The perils of preprints. *BMJ, 370*(August), m3111. https://doi.org/10.1136/bmj.m3111

Vasilevsky, N. A., Hosseini, M., Teplitzky, S., Ilik, V., Mohammadi, E., Schneider, J., Kern, B., et al. (2021). Is authorship sufficient for today's collaborative research? A call for contributor roles. *Accountability in Research, 28*(1), 23–43. https://doi.org/10.1080/08989621.2020.1779591

Vican, S., Friedman, A., & Andreasen, R. (2020). Metrics, money, and managerialism: Faculty experiences of competing logics in higher education. *The Journal of Higher Education, 91*(1), 139–164. https://doi.org/10.1080/00221546.2019.1615332

Ware, M., & Mabe., M. (2015). "The STM Report: An Overview of Scientific and Scholarly Journal Publishing." Technical Report. The Hague: International Association of Scientific, Technical and Medical Publishers. http://digitalcommons.unl.edu/scholcom/9.

WMA, (The World Medical Association). (2013). Declaration of Helsinki: Ethical principles for medical research involving human subjects. Retrieved from https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

World Bank. (2021). Gini Index. The World Bank—Data. Retrieved from https://data.worldbank.org/indicator/SI.POV.GINI?most_recent_value_desc=true&type=shaded&view=map

Ziemann, S., Paetzolt, I., Grüßer, L., Coburn, M., Rossaint, R., & Kowark, A. (2022). Poor reporting quality of observational clinical studies comparing treatments of COVID-19—A retrospective cross-sectional study. *BMC Medical Research Methodology, 22*(1), 23. https://doi.org/10.1186/s12874-021-01501-9

Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva, 9*(1), 66–100. https://doi.org/10.1007/BF01553188