

From: Collabra Editorial Office editorialoffice@collabra.org
Subject: [collabra] Editor Decision - "Open access as a mechanism of revolutionary science and the limitations of data"
Date: 7 January 2016 at 9:08 PM
To: Dr Costa Vakalopoulos hinemoa@bigpond.net.au

CE

Dear Dr Costa Vakalopoulos,

After review, we have reached a decision regarding your submission to Collabra, "Open access as a mechanism of revolutionary science and the limitations of data". We all believe you touch upon an interesting and important topic that would interest the readership of the journal. But all of us also agree that there needs to be a great deal of stronger argumentation, elaboration, justification, coherence, and clarification before a final decision can be reached. Therefore, I am formally rejecting the manuscript, but encouraging a revision and resubmission.

Both reviewers provide many detailed and excellent suggestions that deserve your careful attention. One theme that emerges from the reviews is that many claims need to be backed up by references or arguments. If you can make a compelling case that 1) we need more theory driven research and 2) OA can help achieving this goal, this would make a very strong contribution. But the current version falls short on both counts.

I hope you will undertake a revision, although my suspicion is that the manuscript needs to be rewritten almost from scratch to turn it into a compelling piece. If you are willing to do that, I will send the revision over for re-review.

The full review information should be included at the bottom of this email.

To access your submission account, follow the below instructions:

- 1) login to the journal webpage with username and password
- 2) click on the submission title
- 3) click 'Review' menu option
- 4) download Reviewed file and make revisions based on review feedback
- 5) upload the edited file
- 6) Click the 'notify editor' icon and email the confirmation of re-submission and any relevant comments to the journal.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing; therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please do contact us.

Kind regards,

Prof. wolf vanpaemel
university of leuven
wolf.vanpaemel@kuleuven.be

Reviewer A (Daniel Graziotin):

1) General comments and summary of recommendation

Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:

- Are the methodologies used appropriate?
- Are any methodological weaknesses addressed?
- Is all statistical analysis sound?
- Does the conclusion (if present) reflect the argument, is it supported by data/facts?
- Is the article logically structured, succinct, and does the argument flow coherently?
- Are the references adequate and appropriate?:

This manuscript is an opinion piece that reflects upon how data-driven papers are now the de-facto standard for publishing "real science" at the expense of theoretical works. Furthermore, the authors argues that high-impact factor (HIF) journals are pushing for data-driven publications, while downplaying theoretical works unless if produced by well-known actors

in the various communities. What can hopefully save us from the dictatorship of data-driven research (which the author never attempts to kill as useless, by the way) are open access (OA) journals. As the author states, "The argument in this paper is [...] to make the case for OA as a revolutionary tool for theoretical development of a data driven culture."

The manuscript is beautifully written. The author makes bold claims and accusations (including some lamentations directed at those OA journals which appear to be strongly data-driven). Overall, I enjoyed reading the article, and I agree with most of the author's stance. At first, I was put off by the fact that the article does not seem to be a (state of the art) review as Collabra would consider for its scope. However, that is actually one of the points of the article itself. Yet, I will recommend the editor for revisions. I have got some major concerns that I feel that the author should address in order to support his argumentation "to make the case for OA as a revolutionary tool for theoretical development of a data driven culture".

First, it is not clear to me how the argumentation _as it is currently expressed_ supports the claim that OA journals have the potential to counterbalance the current trend of accepting mainly data-driven publications. While I understand that some OA journals were created, and are supported, by open science advocates that welcome changes and innovation in science, that does not imply that OA journals as such are more "theory friendly". The author states that "PLOS eschews hypothesis manuscripts and PeerJ only accepts them as preprints, all pointing to the misconception of the infallibility of data and the delusion of quality control." (298-300). Yet, PLOS journals and PeerJ are among the biggest OA journals to date. What evidence is there that OA journals will allow more hypothesis manuscripts? Furthermore, the author needs to distinguish between OA as a publishing model to the OA (or open science) "movement". This is because, in my opinion, OA journal simply means "barrier free access to the literature" (pardon my lack of using any of the established definitions for OA), while the OA "movement" is distinguished _also_ by several innovators (e.g., those behind The Winnower, RIO Journal, etc.). One could claim that, shall the fashion and trends change in science, HIF journals will go back accepting theoretical pieces. Please note that I am playing the devil's advocate here, but I think that the paper should reflect upon this to build a stronger argumentation.

Second, while I enjoy opinion pieces, I believe that they should not be exempted from backing up their claims with references and/or explanations, shall the references not be available. Here follow some of the claims that I believe need to be expanded one way or another. Some of the following comments are also born from the fact that Collabra is open access. Therefore, the article would be accessible from the general public and from scientists coming from several fields, and it needs to be as clear and simple as possible.

15-16 "OA is not in danger of lowering the standards of science as its critics claim" Why? And what do the critics claim? I think that several critics of OA are available nowadays (both as a publishing model and as a "political movement", e.g., <http://triplec.at/index.php/tripleC/article/view/525/514>). Although I do not agree with the majority of those claims and critiques, they are there and they should be addressed.

18-19 "OA is a self-regulatory response to the consequences of the distorted and inhibitory contemporary practice of science." Please explain why.

36-37 "Open access is more than a debate about models of access to the research literature and the burden of cost. It is driven by a broader agenda that challenges the control of ideas in science" Where is this agenda? And why can it be claimed that OA is part of this agenda? Please note that I actually agree with this statement, but I find it lacking of support. Also, the sentence could be softened by stating that open access and open science in general are supported by many individuals and organizations that support disruptive ideas in science. A similar argumentation could be made for lines 39-41.

78-79 "However, the emphasis on data can be restrictive because they assume a normative rather than revolutionary course for science" I agree that data-driven studies are framed by the data itself, thus constraining reasoning and intellectual production of ideas and theories. However, there can be revolutionary discoveries from data, as well. I wonder if the author could reflect on this.

88-90 "Peer review generally discourages innovative ideas" Why is that (or who says so)?

103-104 "OA appears, at least currently, to show a greater tolerance for hypothesis papers from lesser status sources although, several OA venues exclude hypothesis based submissions." I think that this sentence is in contradiction with what is stated at lines 298-300, as the biggest OA journals appear to not be theory-friendly. It would help if the author could point to some OA journals that allow theoretical contributions. The Winnower, RIO journal, and Science Open come to mind.

145-146 "This level of explanation will not come from the data sets themselves though." ..but also from? Please explicate

162-154 "Most likely the perceived quality is based on personalities and institutions behind the work as a default measure of likely impact and desirability in the context of a culture of too few highly significant papers." Does this come from (Osborne 2015) or is it the author's opinion?

186-187 "It is about the efficiency of the data, that is the targeting of new heuristic principles of research" And how does help the authors in building his argumentation that we need more theoretical work?

191-192 "Most often data drives hypotheses rather than the reverse case, where good theory creatively spawns experimental design" Here the author could cite some work on Grounded Theory, as it might help the argumentation that even the most famous theory building techniques (also the most widely accepted in journals as studies) are data-driven.

264-265 "The response generally, then is to hide or ignore serious concerns about a study to protect the reputation of the journal rather than to illuminate methodological concerns" I wonder if comparing this thought to the work of Lakatos would help to build up the reasoning.

280-281 Whenever a discussion in PubPeer is mentioned, I would appreciate (as a reader) to have a reference to it.

298 Also to help to build the argumentation, bioRxiv appears to exclude theoretical pieces in its aims and scopes: <http://biorxiv.org/about-biorxiv>

306-307 "Frontiers does publish hypothesis-based pieces and appears to be a covert reason for recent affronts by establishment tools countering what perhaps is perceived as theoretical vandalism of cherished ideas ingrained by the status of a collaborative artifact." I would love to see an example of this here, or a reference.

404-408 "Peer review as currently practiced outside OA is often seen with some justification as a self-referential opportunity to dictate the course of scientific exposition." Citation, please. "The advantage of OA peer review to scientific dissemination of ideas and the bane of OA critics is its regulatory imposition of a non-discriminatory process. Reviewers can suggest improvements, but not arbitrarily block the publication of papers." While I agree that editors have the ultimate choice on the fate of papers in the ordinary peer review processes, I do not see how reviewers could not block the publication of papers. Could the author elaborate on this?

423 Please indicate the page number (2) in the reference for the direct quotation.

466 "of the open access movement such green, gold and platinum." Either explain the various open access publishing models, or avoid mentioning them. Also, I do not consider those as movements.

Finally, shall the author need some further support in building up the argumentation for more theoretical works, these are some classic pieces that I have enjoyed in the last years:

Cannella, A. A., & Paetzold, R. L. (1994). Pfeffer 'S Barriers To the Advance of Organizational Science : a Rejoinder. *Academy of Management Review*, 19(2), 331–341.

Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, 50(6), 1346–1352.

Helfat, C. E. (2007). Stylized facts, empirical research and theory development in management. *Strategic Organization*, 5(2), 185–192.

<http://doi.org/10.1177/1476127007077559>

Pfeffer, J. (1999). Barriers To The Advance Of Organizational Science:

Krieger, J. (1993). Barriers to the Advance of Organizational Science. Paradigm Development As A Dependable Variable. Academy of Management Review, 18(4), 599–620. <http://doi.org/10.5465/AMR.1993.9402210152>
 Van de, A. H. (1989). Nothing Is Quite So Practical as a Good Theory. Academy of Management Review, 14(4), 486–489. <http://doi.org/10.5465/AMR.1989.4308370>

I wish the author best of luck for the endeavor of this manuscript.

2) Figures/tables/data availability:

Please comment on the author's use of tables, charts, figures, if relevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here).:

The paper does not make use of figures and data

3) Ethical approval:

If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared.

If not, please detail where you think a further ethics approval/statement/follow-up is required.:

Ethical approval not required

4) Language:

Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.:

The paper is beautifully written.

As a minor concern, I think that the paper would improve in terms of readability if the author does not assume a readership from the biomedical and behavioral science. I think it would be sufficient to explain certain acronyms, e.g., "DSM V" -> the Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association.

Some examples of needed clarity are:

325 "ASD"

328 "GABA"

Also, not related to the previous fields:

238 "IF" is used as acronym here first. I would specify "Impact Factor" even if the acronym HIF is specified before.

247 "of of"

 Reviewer B (Jonas Kubilius):

1) General comments and summary of recommendation

Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:

- Are the methodologies used appropriate?
 - Are any methodological weaknesses addressed?
 - Is all statistical analysis sound?
 - Does the conclusion (if present) reflect the argument, is it supported by data/facts?
 - Is the article logically structured, succinct, and does the argument flow coherently?
 - Are the references adequate and appropriate?:
- # General

This manuscript describes how open access publishing could improve the quality of published research. In particular, it brings forward two key ideas: (i) the importance (and the current downfall) of theory-driven research and (ii) open access publishing as a means to decentralize power distribution in publishing.

However, while the ideas are important, the manuscript severely suffers from

a poor presentation of the issue. In fact, I am afraid the manuscript reads as a second draft rather than a finished and well-thought-through argument. The essence of the argument is muddled by the lack of structure, multiple seemingly unrelated and often unfamiliar examples that the author fails to explain, and the overall incoherence left me wondering if the paper was not generated by a machine (even the title of the paper is somewhat bizarre!).

Even more importantly, overall the manuscript sounds as a rant suitable for a blog post rather than a scholarly work suitable for a publication in a peer-reviewed journal. In many instances, the other side of the argument is not represented, let alone addressed, and claims made by the author are often not backed by appropriate references. While in some cases such format is understandable, I think in this instance it is problematic as I believe the author is not representing the argument fairly, as I explain below.

Overall, I want to give this manuscript a chance as it clearly taps at important issues. However, I could not recommend accepting it at its present form. With significant restructuring and expansion of the arguments, I would be happy to reconsider this manuscript again.

Theory-driven approach

The author is working on the premise that theory-driven research is somehow superior to data-driven approach. To support his claim, the author often refers to the "traditional" way of carrying out research in physics, where theory often precedes supporting data. But so what? Mail also used to work pretty well until this new idea of email came across. Should email be dismissed? Why couldn't there be a paradigm shift where data-driven research becomes more important? Perhaps theory-driven approach is only viable in certain scenarios but once complex systems are considered, such purely mathematical approach is insufficient. One example that comes to mind is deep learning. For years the technique was lying around and not much was happening. At some point machines became fast and data abundant to finally train these deep networks. The result is that object recognition has made an incredible progress in the past few years. Moreover, advances in deep networks are helping to understand neural data. And yet, there is no theory behind it; nobody can explain how precisely deep nets learn and why they work so well. Of course, researchers want to know this and are trying to derive theories, but this is a successful example where a data-driven approach resulted in advances in vision sciences.

Or how about the Human Genome Project? It was also just a data driven approach, but it provided us with new tools and abilities to develop drugs faster. Is it not already important?

So while I also sympathize with the idea that more theory is needed, the author needs to provide more convincing arguments than this.

One other way the author tried to provide support for this claim was by pointing to multiple areas of research and claiming that little progress has been achieved. I am certainly not intimately familiar with mental disorders, GABA receptors and Alzheimer's disease to evaluate the merit of these claims, but I am well familiar with neuroimaging and here comparisons to phrenology reveal author's poor knowledge of the state of the art techniques and their importance to the field. The field, at least in visual sciences, has long moved beyond a mere cortical localization of a function to studying the underlying mechanisms in these areas by applying pattern classifiers and directly comparing predictions of theories. Moreover, it is the best technique currently available to at least confirm neurophysiology data from primates in humans.

I understand that these techniques are not beyond criticism and one could legitimately ask whether we learned anything new, but a simple dismissal of the whole field as "phrenology" is unsubstantiated and further casts doubt on author's willingness to faithfully consider the arguments for and against in his other examples where I lack the expertise to judge.

Also, the author seems to think that nobody even cares about theories, and tries to convince us otherwise: "This does not mean that these technologies are not useful, but their full potential cannot be achieved without better understanding of what they are looking at." But who is arguing with that? I think many researchers are interested in ultimately building theories that would account for their data. Maybe that interest is not as strong as the author wants but that's a different issue from whether people care about theories at all. I believe most people do and their approach is that you cannot formulate a theory without at least some data. Even in the case of

Watson and Crick, which the author brings up as a prime example of a paper without data, their hypotheses relied heavily on R. Franklin's experimental data, even if unacknowledged clearly in the original paper.

Finally, the author brings up an interesting argument that is hardly noticeable among other things but should be emphasized more: "It is about the efficiency of the data, that is the targeting of new heuristic principles of research." I think this is one of the stronger points of his argument and should be expanded! On the other hand, similar criticism can be brought about theories. For instance, string theory has been around for a long time now yet hasn't resulted in any testable predictions. So how is theoretical work necessarily more efficient?

Open Access

I had to read to the end of the manuscript to understand what OA had to do with theory-driven approach. This understanding was also disappointing, as it could be summarized in a single paragraph or even a single sentence along the lines of "if you want to publish theory-driven research and cannot do so in HIF journals, you may succeed in some OA journals". This is a very weak link that, by itself, does not deserve to be in the title of the paper.

A more important message, however, is that OA enables the decentralization of power in publishing. This is a far more interesting argument but it is not specific to theory-driven approach. Any novel research as well as credit assignment and research impact is affected by this approach. So the two big arguments in the paper do not connect well, in my view, and I'm afraid I cannot offer a good solution to this issue. Perhaps the need for the theory-driven approach could be provided as an example of the power of OA. Note that in any case the discussion on OA should be better structured and more comprehensive (explaining in full what decentralization implies) than it is currently.

Finally, I think the author is missing the point that HIF journals are not only controlling ideas but also providing them visibility. The author briefly touches on this issue (and dismisses it), but the value of filtering cannot be understated. You can go ahead and publish anything you want anywhere, but if nobody is going to read it, publishing all these great theories will play no role in forming scientific thought. OA by itself will not address this issue. New mechanisms need to be constructed and I think the author should at least acknowledge that.

References to other works

The manuscript is filled with references that at least to me were unfamiliar and made no sense, starting already with the first one in the opening sentence of the manuscript: Debating Spellman's position is only meaningful if her argument is at least summarized first. Below is a list of such references where more background should be provided:

Spellman
Body-mind and DSM V
Connectome (why is it the new "Tower of Babel"?)
The famous Max Planck's statement
Paper from Imperial College London and Dijkstra's response
"sine qua non"
"well known Kuhnian perspective"

In addition, I found several other references hard to comprehend and would suggest to rewrite those sections:

Criticism of EEG
Bayesian statistics gets mixed with romantic priming -- no idea why

Finally, the discussion of mental disorders and lack of progress in understanding them lacks citations, and the discussion of Alzheimer's disease cites Teplow in a weird way: "A β protein and its role in dementia, is now considered rather simplistic (Kirkitadze, Bitan & Teplow, 2002)" -- how is this paper representative of the current views 2002?

2) Figures/tables/data availability:

Please comment on the author's use of tables, charts, figures, if relevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Colledge here) :

reproducibility (see open data policies per discipline of Collabra here)..

N/A

3) Ethical approval:

If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared.

If not, please detail where you think a further ethics approval/statement/follow-up is required.:

N/A

4) Language:

Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.:

The author should carefully revise grammar and punctuation, which to a non-native English speaker like me appear incorrect (though I might be wrong here), such as "None the less" and "Worse it desperately attempts to show support or otherwise for often outmoded and simplistic models of reputed significance." and "HIF journals to some degree have brought the problem on themselves by cultivating a coveted reputation of data reproduction above any potential or post hoc criticism, after all they are the arbiters of significance."

Moreover, as a non-native speaker, I found the language to be rather complicated when there is no need for that, e.g., "The problem with the general type of proselytizing to the cause of more and ever expansive data sets is that there ultimately must be a confluence in mechanistic aetiology between genes, relevant pathophysiology and studied behavior."

Collabra

<http://www.collabra.org/>

@collabraoa

