Reflections following the ASAPbio Peer-Review conference

Peter V. Lidsky

University of California San Francisco

peter.lidsky@ucsf.edu

The ASAPbio initiative makes a major call for renovation in scientific publishing in biological sciences¹. The suggested reforms include open access preprints and a fully transparent review process. If introduced into practice, these changes should result in a dramatic cultural change in scientific publishing, probably making it more predictable and consistent. Innovate ways for evaluation of scientific impact were also discussed during the meeting in February 2018. The crowd-sourced evaluations were suggested as a key element for the new metrics. This implies the establishment of a novel entity that should complement or even replace journal reputations and journal impact factors. The citation-based impact factor that is currently a major metric for the evaluation of scientific quality has been considered as biased and uninformative at least by several speakers. Previously, the San Francisco Declaration on Research Assessment (DORA²) also considered journal impact factors to be highly skewed, non-transparent, and potentially manipulated by editorial policies (see also ³).

However, can we entirely abandon the citation-based metrics? The professional feedback on the papers, proposed by the ASAPbio initiative, can be also biased and artificially exaggerate or diminish positions of certain studies in the scientific landscape. Works containing some emotionally charged conclusions will draw more attention and therefore more feedback. For example, the study of increased alcohol consumption by male fruit flies who were rejected by the females⁴ would have much more impact in scientific and general media because it places our day-to-day experience into the broad evolutionary context. Personal relations and strong opinions can further affect this way of evaluating scientific publications. Furthermore, this proposal requires introducing new ways to motivate scientists to invest time in writing post-publication reviews. To produce a fair, quantitative evaluation of scientific impact, each paper should receive many reviews from many researchers.

On the other hand, citations that represent the logical connections between scientific ideas and experiments seem to be less biased criteria than readers' feedback and, importantly, require less effort to be invested into pre- or post-publication curation. Thus, considering that the collection of post-publication feedback on the papers might be very helpful in many ways, it is nonetheless hard to believe that this can completely replace citation-based metrics. Methods for the improvement of the existing citation indexes are debated in the literature (see e.g. ^{5,6}) and the services for contextualizing citations are developing⁷. This essay aims to discuss possible developmental trajectories and a possible strategy of changing publishing practices to make citation-based indexes more reliable, fair, and informative and thus to complement the ASAPbio proposals.

Phase I. Citations indexes in the existing format are biased towards reviews and methodological papers. Indeed, it is frequently impossible to acknowledge every original discovery by a citation,

making the referencing of reviews convenient. As a result, reviews are collecting scores deserved by the authors of original research papers.

Method papers are on average cited more frequently than original research papers. For example, one of the most cited astrophysics papers of all time is an infrared-emission map of the cosmic dust that is required to be accounted for in most astronomical observations⁸. Publishing papers that describe online tools and database updates might significantly improve the journal's impact factor in biology⁹.

Reviews are an efficient way to help researchers stay up to date with the recent state of the art in science. Developing new methods and tools, assembly and curation of the databases are also indispensable for the scientific community. However, these important activities are different from performing basic research, and the possible way to remove these citation biases is to divide citations (and correspondingly the scientific scores) into classes: for basic research papers, for methods, and reviews. Importantly the reviews, now serving as citation hijackers, will transmit the citation scores to the original research articles. If a review gets cited, all the papers cited in this review are getting some proportion of these scores. The exact numeric factor of how much scores should be transmitted to the cited paper by the review upon citing is to be decided by the scientific community.

This modification of the citation index seems to be feasible just by retrospective reannotation of the existing citation databases. All reviews and most methods/tools/databases articles are unambiguously marked either by the journal they are published in or by the type of publication, which could be easily recognized. Furthermore, many scientific manuscripts contain "Materials and Methods" sections. All references from this section can be considered as "method" type, even if the "original research" paper is cited. The mistakes made by this classification algorithm may be revised on a crowdsourcing basis. By the end of this phase, the citation biases created by method and review papers will be removed, and the community will get a more detailed indexing system to evaluate the professional profile of a given scientist or institution (see Fig 1 for Phases I -III).

Phase II. Another problem is that current topologies of citation networks are not non-informative in the sense that they do not allow us to reconstruct the real pedigree of the ideas as they evolve from paper to paper. Indeed, the majority of citations in scientific publications come from references of secondary importance. Citations for the concepts that were fundamental to nucleate a given study are technically indistinguishable from citations to the papers, the authors do not agree with, discuss or refute. Very few crucial references that set the preceding publications as the key ancestral studies are not highlighted and therefore their recognition is impossible at high throughput scales.

Assigning each reference a **type** (in addition to methods and reviews discussed above) might be the next step towards evolving the citation networks into the structures reflecting the internal logic of scientific advance. First of all, before submitting the manuscript, authors would need to annotate the references by their contribution to the study. The two main types of "basic research" references are required. The first type is for the few papers that immediately contributed to a given study by preparing grounds for the ideas and experiments described in

this manuscript. This can include the crucial methods and approaches, observations that lead to the hypotheses tested in the current study, etc. All the other references are considered ones of secondary importance. These can be further subdivided by subtypes which might include: (i) key discovery in the field, (ii) paper of secondary importance in the context of the particular study, (iii) conflicting data, (iv) potential connections for future research, (iv) resonating results in a neighboring field, etc. Different types of references can have different weights, thus giving different numbers of the “original research” scores. The exact specifications of these types and subtypes, as well as their weights, should be decided by the scientific community.

According to these guidelines, citation scores of the papers and individual scientists might become more informative and fair than they are now. The topology of citation networks will become more detailed and instructive for navigation in the growing body of scientific literature. Annotation of references is already evolving, for example, some journals require the authors of reviews to write short annotations to the most important articles they cite. It seems that the shift to reference lists with citations, classified by types, might be relatively easy and will not require a big effort from the authors.

An important technical problem would be the retrospective reannotation of several hundreds of millions of scientific publications that are already published now. It is clear that such a project can be performed only with the participation of many members of the scientific community and is feasible only on such a large-scale crowdsourcing basis. To accomplish this, annotation of the references in other people’s papers should become a routine activity. This should be a significant change in scientific work, comparable to the routine writing of the post-publication reviews proposed by the ASAPbio initiative. In the next paragraph, the potential ways of transition to this model are discussed.

Phase III. In this step, the exact topology of citation networks should get a post-publication curation from the scientific community. Each scientist should get an individual authorization (e.g. by ORCID¹⁰ or some analogous ID). Every scientist will be encouraged to vote for a type of particular citation. Additionally to the type, citations will be getting a non-integer weight that will specify their importance and allow accounting for the opinions of many different scientists by averaging the weights they proposed. Importantly, readers should be allowed to add references that were intentionally or unintentionally missing by the authors. This should make citing mistakes and personal “not-citing rivalries” less damaging for the integrity of the knowledge networks. The exact rules of this reference editing activity should be carefully discussed and established by the scientific community. Questions like: “Are all researcher’s votes equal or also can have different weights depending on the position, age in science or reputation?”, “Is a biologist allowed to edit references in astronomy?”, “How many positive votes are required to insert a new reference into a paper?”, “Can a researcher add/vote for citations that reference to her/his work?” etc. should be thoroughly investigated.

In case of conflicts or disagreements, the arguments are posted on the post-publication discussion page, which will be attached to every scientific paper, analogous to the “Talk” pages in Wikipedia. Such discussions might become more efficient in providing an accurate evaluation of scientific work than the existing system of peer review and journals reputations. Today the

outcome of the publishing process is highly biased since the number of reviewers is limited and the time they can spend on the paper is restricted. The post-publication analysis proposed by the ASAPbio initiative can recruit many more peers, who can work on the paper without fixed deadlines. However, researchers might get more motivation from spending time in discussions about formalized issues, than if they would be just asked to write general post-publication feedback. The questions like “does this citation correspond to this formally defined reference type?” might nucleate scientific discussions that will also be very helpful for the readers to understand the exact place of a given piece of research in the existing scientific landscape. These topics will be provocative since they will not only affect the numeric citation scores of the papers but will also be crucial for the integrity of scientific literature as a whole. The topics of discussions will be very specific and these constraints will relax potential biases coming from personal relations between scientists. Saying that your friend’s study is “elegant” or “awesome” is much easier than giving an incorrect answer to a formalized logical question.

By the end of this phase, the community will acquire an accurately annotated landscape of scientific knowledge, with the input of each paper accurately characterized. It will help not only in evaluating the achievements of a given scientist, but also make literature more integrated and accessible. The evolutionary trajectories of scientific ideas will become more apparent from topologies of the citation networks, now overcrowded by dozens of citations to the papers of secondary importance.

It will also help to prevent predatory publishing, as publications without recognition from the scientific community would have much lower scores. It will also exclude the retracted papers from the body and will solve the current problem of the post-retraction life of these articles¹¹.

The transformation trajectory described above goes in a stepwise, evolutionary way. Each phase requires only minor changes in publication processes, and even if some part of the community (a university, a country, or a publishing company) will not support the change at a given phase, reannotation can be done later or externally. It is also important that the three phases can largely overlap or coexist as different metrics, allowing the community to adopt the changes and to develop rational and transparent rules in a timely and comfortable way. It should be emphasized that only Phases III requires a moderate change in common practice and that the reference annotation activity is not very time-consuming: typically, the reader would need to identify several relevant references in the swarm of secondary ones and to verify if citations for reviews and methods are recognized properly.

The next part of this essay is a foreseeing attempt at the possible evolutionary paths of a scientific paper as the minimal publishable unit. Indeed it would be surprising if the publication practices largely coming from the mid of 17th century¹² will not change in the digital era. It is feasible to expect that scientific publications will become better integrated into the knowledge databases and more accessible for machine learning algorithms.

Phase IV. To further expand and formalize the information incorporated in the connections between scientific papers, the individual article should become separated into logical pieces or sub-divisions or nano-publications¹³, such as “hypotheses”, “experimental results/datasets”, “methods”, “analyses”, “conclusions”. All these elements would become connected by the

“bonds” of different citation types (see Fig 2). For example, the path from a hypothesis to an experiment to test this hypothesis is one class of “bond”, and the step from raw dataset to its analysis is a second class, etc. These new bond types are different from the types of references between articles introduced in phase II (Fig 1). The bond types reflect logical relations between subdivisions within a single manuscript or between different articles. In addition, each bond may have a “weight” reflecting, for example, the importance of different experiments or analyses, supporting the same conclusion; or the importance of different conclusions for nucleation of a novel hypothesis. Each classical reference might now be represented as a single or several (a bundle) bonds that connect not only the articles as a whole but also the specific sub-divisions. Thus, the topology of bonds will provide an improved representation of the evolution of scientific ideas from hypothesis to experiments and conclusions.

Each sub-division will be also supplemented with a list of authors. These changes will assist to understand better the individual input of each researcher and will make scientific literature more easily classifiable for machine algorithms. But most important, these changes will prepare grounds for the next evolutionary stage of scientific publishing (see Fig 2 for Phases IV-VI).

Phase V. The next suggested change consists of the possibility to publish smaller pieces of data, e.g. a single dataset, an analysis, or a hypothesis, rather than a fully accomplished study. The readers will be endorsed to edit the topology of interactions between different logical elements, assign weights and integrate the results within the knowledge databases. Such transformation might significantly accelerate the scientific advance, improve the quality of research, and make it more cooperative (see ¹⁴ for review on current skews and biases in science). Some advantages of such changes are discussed below.

Journal-based metrics push scientists to “oversell” their research. As the criteria for high-profile publications include both the importance and the quality, the one is sometimes sacrificed for the gain in the other. Both the papers with big but insufficiently supported claims as well as experiments made with cutting-edge techniques but lacking a major conceptual outcome are routinely being published by high-profile journals. During the few decades, we have encountered numerous examples of sensational discoveries that were not confirmed or entirely refuted shortly after the publication, and that would probably not happen if the scientific impact of researchers would be scored differently. One can mention here the bacterium, able to use arsenic instead of phosphorus¹⁵, reprogramming of mouse fibroblasts into pluripotent cells with acidic media treatment¹⁶, pictures of bacteria remains from Mars¹⁷, reparation of point DNA mutations with use of the postulated RNA cache¹⁸, etc. The existing system pushes researchers to publish high, and they are sometimes therefore forced to invoke big claims even if they are not sufficiently supported by reliable data. Importantly, similar things might occur at the level of less ambitious, minor results that, while being less apparent, might be even more damaging for the scientific progress, as they stay uncovered for a longer period of time. Once published, the reported results are rarely verified by independent groups as such verification would hardly give high benefits with either positive or negative outcomes. Thus a large part of published research findings in certain fields may be false¹⁹. This problem is especially harmful with methodological papers. The authors of the methods often tend to exaggerate the robustness of their tools by selecting favorable demonstration models or conditions. At the same time, the negative

experience of researchers who tried to use these tools is rarely coming into publications and is largely invisible. The key problem here is the way of scientific data evaluation: the main factor is the level of a journal the research is going to be published in, and the decision is made by few reviewers within a short time.

Currently, researchers often feel paranoid to disclose the primary data and even the conceptual ideas before the whole publication is completed and submitted. By revealing the intermediate results, researchers are exposing themselves to the risk of getting scooped and losing months or even years of work. Thus, the publication of individual experiments is strongly disfavored by the existing publishing practices. Conferences, where scientists usually disclose their fresh and often unpublished research for a limited audience, became points of exchange and stimulation of new ideas.

At the same time research is rarely performed by a single individual, but rather by collaborating groups. As the complexity of scientific approaches increases, more people are getting involved in performing and analyzing the experiments. The analysis is, more and more often, made by a different person than the one who has collected the primary data. Finding a good collaboration, therefore, becomes more important, while a wrong collaboration often ruins the project.

Taking together, today's publishing system introduces some disconnections between the goals of individual carrier development and goals of scientific advance. By breaking down publications into smaller logical units with a specified set of coauthors, the scientific community might resolve most of these conflicts. By adjusting the scores that an individual researcher is gaining for each activity, one can accurately move the priorities in a way that the individual's interests are matching the overall aims of scientific advancement. The scores from connections between logical pieces can be adjusted in a way that scientists benefit not only from the big claims in their papers but also from the speed of the progress in their field. Disclosing smaller pieces of data might become more advantageous: for example if the analysis performed on a published dataset will give its creators the scores comparable to the co-authorship, opening such data will become a good strategy and will make searches for good collaborations redundant. Instead, the interactions and exchange between scientists that now take place mainly at the conferences will become constantly running online. In general, science might become much more cooperative, and individual researchers might find their niches more easily. Different researchers have different expertise, different approaches, ways of thinking, strengths, and weaknesses. Some of them are better at performing experiments, some in the analysis, some in theoretical thinking and critical evaluations. The described system might help them to complement more productively.

The new system might also resolve the problem of reproducibility: if the experimental result is corroborated by the later research, the authors of this second confirmatory work would also receive a percentage of the scores from references to the original study and its conclusions. The exact proportion of this benefit has to be decided by the community to reach a perfect balance between the speed of research and the reliability of published data.

The proposed system of flexible and continuous post-publication data annotation will make exaggerations and data overselling a bad strategy, as it will open an opportunity for other

researchers to refine and correct the conclusions, thus sharing or partially taking the citation scores away from the authors of the original study.

It will bring to the light a huge layer of “orphan” experiments that, while being consistent, are never published. For example, the hypothesis predicts the experimental outcome of a white circle if it is true and a black circle if it is wrong. However, the real experiment reproducibly gives the green triangle, clearly indicating the involvement of additional factors, which are not in the scope of the current study. Such experiments are usually not pursued further, as the effort of writing it up as a paper might be too much of a time- and energy investment, while the significance of the occasional results for the “triangle” field might be not well understood by the “circles” specialists. The opportunity to publish smaller pieces of evidence, without investing too much effort into composing a long manuscript, would bring to light these data, preventing the loss of resources invested. The small experiments, challenging or refining previously published papers as well as negative results will get more visibility. Frequently scientists are testing if the effect published by someone else is working in their system of interest (e.g. different organism or cell type). If the results entirely or partially contradict the published data, the experiment is usually discarded, as again it would take too much effort to bring this minor piece to the publication level. The methodological caveats will come to light much faster than they do now, preventing scientists from wasting their time with error-prone methods. It should be mentioned that the growing importance of publishing confirmatory and contradictory results is already considered by the scientific community, e.g. bioRxiv specifically has ‘Confirmatory Results’ and ‘Contradictory Results’ sections to increase reproducibility in science²⁰.

The platform for releasing minor pieces of reliable data, which would be otherwise not included in any publications, seems to be a good entry point to establish the rules of the novel publication format. Conceptually it should be a short and synthetic description of a single series of experiments. Importantly, it should be exhaustively annotated, and these annotations should be highly standardized. The scientific community would need to elaborate the rules to avoid potential caveats of the novel system, such as the potential influence of the large institutions or “spam” publications. Protocols of collective decision-making and voting should also be established. This should go in parallel with the standardization of experimental techniques that will improve reproducibility²¹. As the data quality control and annotation procedures will be established, the scientific community will become more familiar with this format, more regular papers will get annotations within the novel system. In the end, the whole moiety of scientific publications should get annotated. Thus, starting from citation indexes and GitHub-like portal, the system should integrate the functions of scientific journals, preprint services, scientific-oriented online forums (e.g. Research Gate), cross-referencing, and annotation platforms.

It should be also emphasized that the proposed system will not reduce competition in science. The individual researcher will still decide to publish or not to publish a hypothesis or a dataset. The conflict of interest here would be similar to the one we experience now when starting a collaboration with another group: we share the credit for our discovery by the gain in speed and quality of the outcome. Currently, the protection from indecent behavior (e.g. data hijacking) occurs due to a limited number of collaborators and their ethics. The suggested system allows (i) to claim priority on smaller pieces of data or scientific knowledge and (ii) accurately and fairly

weight each individual's contribution to a given scientific discovery. The system will still prize the big discoveries, but it also will create a strategy, when many researchers progress faster by working together and sharing the benefits. Within such groups, the competition will be shifted from the timeframe of years to the months or weeks that are required to perform a single series of experiments. The failure in such competition will be less painful as not that much time would be lost and also the reproduction of the published experiment would provide some citation scores for the researcher.

Phase VI. In the next stage, the scientific publishing system should become integrated with knowledge databases, such as GeneBank, SwissProt, FlyBase, etc. Currently, database curation is an activity separate from publication and is performed by moderators or by volunteering researchers. The importance of databases is constantly growing, especially in biology, studying the systems composed of hundreds of thousands of elements: genes, different biological molecules, species of animals, etc. Dividing research papers into the smaller logical pieces in Phase V will ease the curation process, ideally making it fully automated. The main content of the publication unit should find its formalized designation in the knowledge databases as a list of interacting entities (e.g. proteins, genes, or developmental processes), types of their interactions (e.g. protein-protein binding or involvement of a particular gene in a given developmental process) and the types of evidence that demonstrated it. Thus, the scientific community should obtain an integrated database of knowledge itself integrated with the database that provides a detailed description of the advance of this knowledge accumulation. The topologies of these networks potentially could be analyzed by machine-based algorithms that might identify some overlooked connections and suggest experiments. It might be especially helpful in the extended fields, such as biology or astronomy, where the human brain is clearly unable to encompass the whole variety of proteins or galaxies. Such algorithms, trained with prior topologies of scientific research publications might become valuable tools at least for brainstorming the scientific concepts. Further, these approaches might mediate our first steps towards technological singularity. Indeed the creation of machines capable to advance science without human supervision would require at least learning them how to read the papers and how to classify the obtained knowledge. An accurate annotation and classification produced by the scientists within the system suggested above can allow analyzing data with machine algorithms, bypassing building the sophisticated robots, capable to read the scientific papers themselves.

Importantly, while attempting to introduce major cultural changes in science and the scientific publication, the suggested mechanisms do not require centralized organizational amendments. This novel publishing system might peacefully coexist with the traditional publishing system until its very last phases, persisting as a citation metric and annotation platform. By getting more and more adjusted, it will recruit more and more scientists for data curation, gradually increasing its importance in scientific publishing.

Acknowledgements

I would like to thank Dr. Miguel Garcia Knight, Dr. Michel Tassetto, Dr. Leonid Gitlin, and Dr. Samantha Hindle for reading and discussing this essay.

1. ASAPbio – Accelerating Science and Publication in Biology. *ASAPbio* Available at: <http://asapbio.org/>. (Accessed: 25th July 2018)
2. DORA – San Francisco Declaration on Research Assessment (DORA). Available at: <https://sfdora.org/>. (Accessed: 27th May 2018)
3. Waltman, L. A review of the literature on citation impact indicators. *J. Informetr.* **10**, 365–391 (2016).
4. Shohat-Ophir, G., Kaun, K. R., Azanchi, R., Mohammed, H. & Heberlein, U. Sexual deprivation increases ethanol intake in *Drosophila*. *Science* **335**, 1351–1355 (2012).
5. Hutchins, B. I., Yuan, X., Anderson, J. M. & Santangelo, G. M. Relative Citation Ratio (RCR): A new metric that uses citation rates to measure influence at the article level. *PLoS Biol.* **14**, e1002541 (2016).
6. Janssens, A. C. J., Goodman, M., Powell, K. R. & Gwinn, M. A critical evaluation of the algorithm behind the Relative Citation Ratio (RCR). *PLoS Biol.* **15**, e2002536 (2017).
7. scite_. Available at: <https://scite.ai>. (Accessed: 23rd October 2018)
8. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525 (1998).
9. Junca, H. The “real” impact factor of Nucleic Acids Research is ~5.6. *Bergman Lab* (2010).
10. ORCID. Available at: <https://orcid.org/>. (Accessed: 30th May 2018)
11. Bar-Ilan, J. & Halevi, G. Post retraction citations in context: a case study. *Scientometrics* **113**, 547–565 (2017).
12. Kronick, D. A. A history of scientific and technical periodicals: the origins and development of the scientific and technological press, 1665-1790. (1976).
13. Mons, B. & Velterop, J. Nano-publication in the e-science era. in 14–15 (2009).
14. Azoulay, P. *et al.* Toward a more scientific science. *Science* **361**, 1194–1197 (2018).

15. Wolfe-Simon, F. *et al.* A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus. *Science* **332**, 1163 (2011).
16. Obokata, H. *et al.* Stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature* **505**, 641 (2014).
17. McKay, D. S. *et al.* Search for Past Life on Mars: Possible Relic Biogenic Activity in Martian Meteorite ALH84001. *Science* **273**, 924 (1996).
18. Lolle, S. J., Victor, J. L., Young, J. M. & Pruitt, R. E. Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis. *Nature* **434**, 505 (2005).
19. Ioannidis, J. P. A. Why Most Published Research Findings Are False. *PLOS Med.* **2**, e124 (2005).
20. bioRxiv.org - the preprint server for Biology. Available at: <https://www.biorxiv.org/>.
(Accessed: 14th October 2018)
21. protocols.io - Life Sciences Protocol Repository. Available at: <https://www.protocols.io/>.
(Accessed: 14th October 2018)

Now

+1
integrated
citation
score

Types of references

+1 original research score
+1 method score
+1 review score

Phase I

Types of "original research" references

key ancestral paper
secondary
supporting data
conflicting data
etc.

Phase II

Post-publication crowd-sourced annotation of references

readers added a reference article
— refining types
— adding references
— voting for weights

Phase III

Phase IV

Phase V

Phase VI

Neural network analysis.
Artificial PI's, Singularity?