

Wikipedia is significantly amplifying the impact of Open Access publications.

 blogs.lse.ac.uk/impactofsocialsciences/2015/09/08/wikipedia-amplifying-impact-of-open-access/

9/8/2015

When you edit Wikipedia to include a claim, you are required to substantiate that edit by referencing a reliable source. According to a recent study, the single biggest predictor of a journal's appearance in Wikipedia is its impact factor. One of the exciting findings, writes [Eamon Duede](#), is that it appears Wikipedia editors are putting a premium on open access content. When given a choice between journals of similar impact factors, editors are significantly more likely to select the "open access" option.



How does the public learn about science? How do you? Most basic science literacy comes from primary and secondary school education. Of course, college can offer more formal scientific training, and, perhaps you got some of that, too. Yet, even if you are a practicing researcher, even if you are a world leading expert in this or that field, even still, science as a whole is bigger and mercilessly more diverse than your expertise will ever be. It turns out that, in many ways, you are just like everybody else and you read the science section of your favorite newspaper or magazine, maybe watch a science show, or listen to a science podcast.

Yet, in those more serious or desperate moments, when the stakes are really high, perhaps challenged over coffee as to when the [Cambrian Explosion](#) started, you turn to Wikipedia. But why? Teachers say you can't trust it. You certainly can't cite it in your senior thesis. Don't even mention it in a scholarly, scientific, article submitted for peer review. So when you think you know about the Cambrian explosion and egos are at stake, can you turn to Wikipedia?

You can. In a [recent study](#) (currently under review), my colleagues Misha Teplitsky, Grace Lu, and I looked at the world's 50 largest Wikipedias (the English language Wikipedia is just one of hundreds) to learn about the sources of the scientific information contained in them and whether those sources are reputable.



Image credit: Wikipedia User:Oblivious editing wikipedia (Wikimedia Commons CC BY-SA)

First, what would count as reputable? Well, satisfying answers to this question are notoriously nuanced. Here is a far too simplistic answer: a source is more reputable than another source if the former is relied on by scientists and scholars more than the latter. Journals can, in fact, be ranked in this way by an imperfect metric known as “impact factor” which is, put simply, the average number of times articles from that journal are cited in the literature. So, journals with high impact factors are relied on more than journals with low impact factors.

Now, when you edit Wikipedia to include a claim (as opposed to correcting grammar or spelling), you are required by Wikipedia’s guidelines to substantiate that edit by referencing a reliable source. For a pop or indie music claim, reliable sources might be Billboard or Pitchfork respectively. For a science related claim, a reliable source would simply be one that a practicing researcher is likely to cite in a scholarly paper, and that citation is likely to come from a scholarly, scientific journal.

Nevertheless, when it comes to science edits to Wikipedia, there has been some debate about whether the sources that are referenced are, in fact, reliable. And, this debate turns on a question about access to reliability.

Access to scholarly, scientific journals is either open or closed. Open access journals make all of their published research freely available to anyone who cares to read it. Closed access journals sit behind paywalls and require extremely expensive subscriptions in order to read them. These subscriptions are so expensive that really only institutions of higher education and large companies with research and development arms have them. The intuition that motivates the debate about the reliability of referenced sources on Wikipedia is that the journals that are most heavily cited (that is, the journals with the highest impact factors) are almost uniformly “closed access”. So, the intuition continues, if you are a member of the general public making a science edit to Wikipedia, chances are you do not have access to the most heavily relied on sources and, in order to substantiate your claim, you will need to turn to something else.

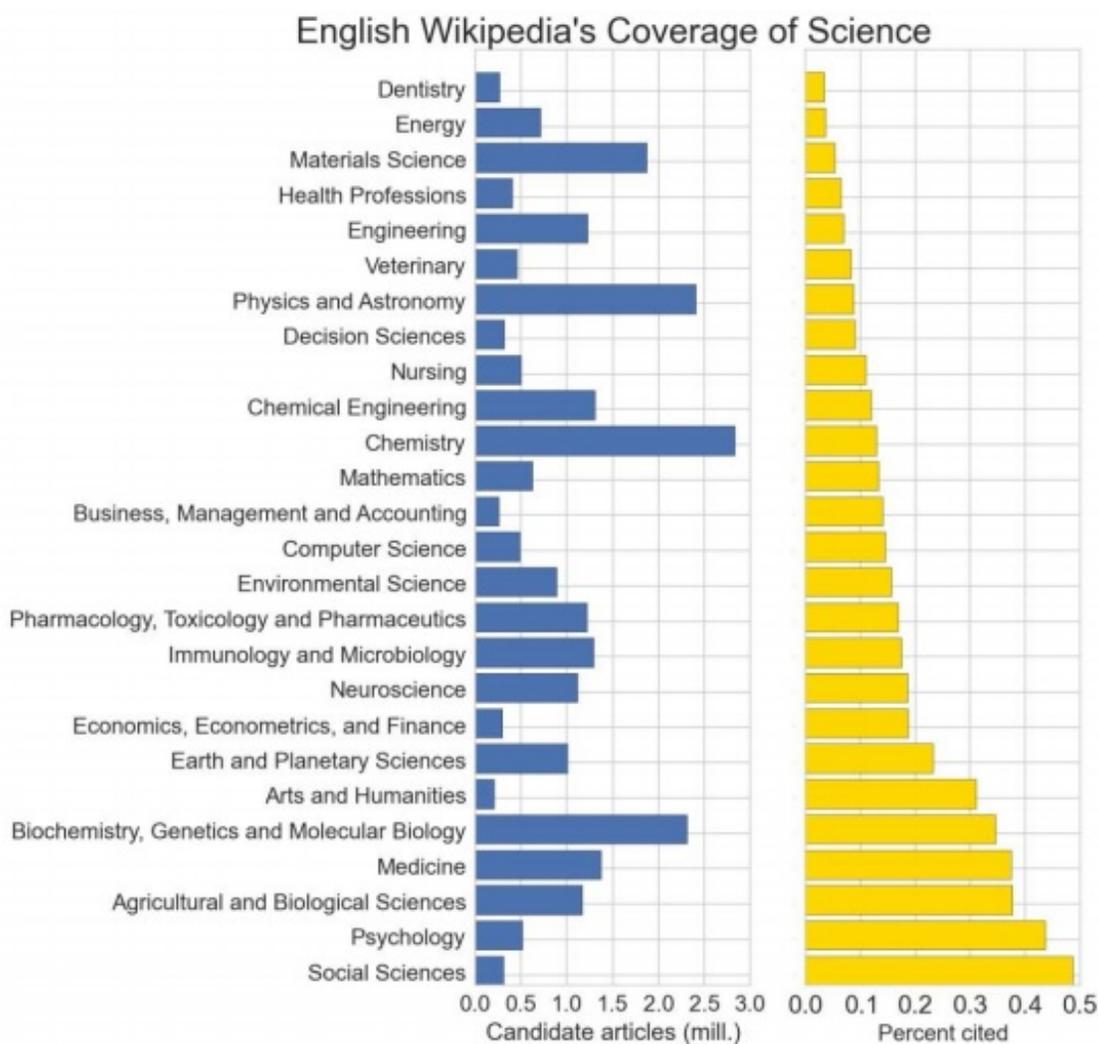
To figure out whether this intuition hold for actual Wikipedia edits, we first turned to Elsevier’s Scopus database. This database indexes more than 20,000 peer reviewed scientific and scholastic journals. These are the journals that practicing researchers turn to. Of the journals indexed by Scopus about 15% are classified as “open access”.

Next, we looked at every reference on every page of each of the world’s 50 largest Wikipedias (the English language Wiki has about 5 million pages, alone) to determine whether the reference cites a scholarly journal. When they did,

we tried to match those journals to a journal represented in our pool of reputable sources indexed by Scopus. It turns out that, of the citations in Wikipedia that used a journal as a reference, we were able to match the majority to a journal indexed by Scopus. So, when Wikipedia editors make contributions to science topics, they tend to cite the same journals that practicing scientists cite.

Things get more interesting when you start to look at what Wikipedia editors are citing.

Figure: English-language Wikipedia’s coverage of academic research.



Source: Misha Teplitskiy, Grace Lu, Eamon Duede (2015) “Amplifying the Impact of Open Access: Wikipedia and the Diffusion of Science” Arxiv preprint.

In the figure above, the left panel shows our pool of reputable sources; the number of articles published in the 26 major subfields of science and scholarship. If you are making a science edit to Wikipedia and you are including a citation to a reputable source, then your citation is most likely one of the articles represented in the left panel. The uneven distribution of candidate articles is rather remarkable. For instance, the Arts and Humanities and the Social Sciences do not publish nearly as frequently as, say, Chemistry or Physics. However, it would probably be a mistake to assume that the Humanities or Social Sciences are somehow slower, or that they make less progress than Chemistry or Physics. What is reflected here are merely publication conventions. For one thing, papers in the Social

Sciences tend to be quite long and Chemistry papers tend to be quite short.

Yet, look at the right panel of the figure. This panel shows the percentage of the papers represented in the left panel that are actually distilled into content and referenced on Wikipedia. A considerably higher proportion of the reputable Social Sciences and Humanities sources make it into Wikipedia. This could be an indication of many things. One might be that there is significantly more demand for citations from the Social Sciences than from Chemistry. Perhaps Wikipedia editors require that claims to social science articles be substantiated with more citations than, say, a claim about the start of the Cambrian Explosion. This asymmetry opens up a whole space of intriguing questions, some of which my collaborators and I are looking into.

Primarily, the single biggest predictor of a journal's appearance in Wikipedia is its impact factor –the higher the better. Yet, a really exciting finding to pop out of the data is that, for any given journal, those that are designated as “open access” are 47% more likely to appear in Wikipedia than comparable “closed access” journals. It looks like Wikipedia editors are putting a premium on open access. It is important to emphasize that this does not mean that Wikipedia editors are citing “open access” journals more often than closed access journals. What seems to really matter most to Wikipedia editors is impact factor. Nevertheless, when given a choice between journals of highly similar impact factors, Wikipedia editors are significantly more likely to select the “open access” option.

There are so many possible reasons for this. Perhaps the open nature of Wikipedia itself inspires a kind of preference for similarly open resources. Maybe “closed access” journals actually are more likely to appear in Wikipedia than comparable “open access” journals at some point in time but that these citations are later removed when a suitable “open access” alternative is found.

One thing is clear, Wikipedia is serving to significantly amplify the impact that “open access” publications are making beyond the scientific community – an impact on society as a whole. [Previous research](#) by James Evans has shown that, while “open access” policies have a very limited impact on scientific communities in developed countries, they serve to make findings more widely available, particularly to scientific communities in developing countries. Our study shows, when scientific findings are coupled with open knowledge sharing platforms such as Wikipedia, that this widening effect is, perhaps, even more pronounced.

Note: This article gives the views of the authors, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

About the Author

Eamon Duede is Executive Director of [Knowledge Lab](#), a computational science of science research center at the University of Chicago's [Computation Institute](#). His academic interests are focused on the constellation of problems fixed around the concepts ‘perception’, ‘belief’, and ‘knowledge’, as well as the philosophy of biology (particularly evolution), logics, and mathematics.

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.