

A Clockwork Wikipedia: From a Broad Perspective to a Case Study

Omer Benjakob^{*1} and Rona Aviram^{†2}

^{*}*Cohn Institute for the History and Philosophy of Science and Ideas, Tel Aviv University, Tel Aviv, Israel, and*

[†]*Weizmann Institute of Science, Rehovot, Israel*

Abstract While research has shown that scientists use Wikipedia and that scientific content on Wikipedia ramifies back into scientific literature, many questions remain on how the two sides interact and through what paradigm this dynamic may be best understood. Using the circadian clock field as a case study, we discuss this scientific field's representation on Wikipedia. We traced the changes made to the articles for "Circadian clock" and "Circadian rhythm" and reviewed the debates that informed them over a span of a decade, using Wikipedia's native and third-party tools. Specifically, we focused on how groundbreaking research pertaining to the function of biological oscillators was integrated into the articles to reflect a wider paradigmatic shift within the field. We also identified the articles' main editors to detail the dynamic collective editorial process that took place during a time that saw the field undergo a fundamental change. We discuss the different concerns the academic community has with Wikipedia—specifically regarding its content and its contributors—to ask whether the online encyclopedia's open model is inherently at odds with scientific culture or whether the model could reflect science or even expand on its core values and practices such as peer review and the idea of communicating science.

Keywords Wikipedia, circadian clocks, citizen science, sociology of science, open science

WIKIPEDIA AND SCIENCE

A forthcoming study from the MIT Sloan School of Management (Cambridge, MA) attempted to address a question that up until a few years ago may have been unimaginable: Does content from Wikipedia, the online encyclopedia that anyone can edit, find its way into academic works? The researchers commissioned PhD students to write articles on topics that fell under their field of expertise; half of the articles were introduced to

Wikipedia, while the remaining were held as a control group. Using textual analysis, the researchers claimed that "word-usage patterns" from the articles introduced into Wikipedia show up more in peer-reviewed papers than do those from the control group. Less than a week after being uploaded to the Social Science Research Network website, the unreviewed preprint received widespread media coverage, including in *Nature*, which parroted its findings that "1 in every 300 words in a scientific paper was influenced by language

1. To whom all correspondence should be addressed: Omer Benjakob, The Cohn Institute for the History and Philosophy of Science and Ideas, Tel Aviv University, Tel Aviv, Israel; e-mail: omerbj@gmail.com.

2. To whom all correspondence should be addressed: Rona Aviram, Weizmann Institute of Science, 234 Herzl Street, Rehovot, 76100, Israel; e-mail: anorona@gmail.com.

in the Wikipedia article" (Zastrow, 2017). In their conclusion, the study's authors claimed that "Wikipedia doesn't just reflect the state of the scientific literature, it helps shape it" (Thompson and Hanley, 2017).

Despite the stereotype of the lazy student using the easily accessible database of articles to fulfill class assignments, the aforementioned research shows that publishing scientists also use Wikipedia. Indeed, despite long-standing concerns regarding Wikipedia, some academics have been willing to engage with it, using it as an arena for disseminating—and even elaborating—on scientific knowledge. Prof. Erik Herzog from Washington University in St. Louis integrated Wikipedia into his Biological Clocks undergraduate course, tasking students with writing well-researched articles as an integral part of the coursework. Herzog and his students even wrote a paper about their work (Chiang et al., 2012), describing the project's goal as both educational and part of a large attempt "to enhance public access to important discoveries in chronobiology."

Meanwhile, in Wikipedia, academic projects have long played an important role. For example, Gene Wiki (Huss et al., 2008)—a project led by Prof. Andrew Su and Prof. John Hogenesch in which computers automatically open Wikipedia articles for the genes defined through the Human Genome Project—was the first case of automated article opening in Wikipedia. As of 2017, more than 10,000 articles have been created under the initiative, and many have since expanded with the help of human editors—students, professors, and even genetics enthusiasts, creating what Prof. Hogenesch called a "virtuous circle" that starts in the laboratory and ramifies out into Wikipedia.

Examples of scientists like Herzog, Su, and Hogenesch and Wikipedia-related research highlight the growing interplay between the scientific world and the online encyclopedia and how, despite legitimate concerns regarding the encyclopedia, it can no longer be wholly rejected or ignored by academics.

In this review, we discuss the complex ties between Wikipedia and science through a case study on the circadian clock research field and its representation on the site. By examining scientific content, as opposed to politically charged articles (which have usually been the focus of research into Wikipedia), we ask whether Wikipedia contradicts, reflects, or even possibly expands on academic practices.

AUTHORSHIP AND EDITORS: WIKIPEDIA VS. SCIENCE

Wikipedia was established in 2001 by Jimmy Wales and Larry Sanger as a community-based website

aimed at compiling "the sum of all human knowledge" (Mesgari et al., 2015). As of December 2017, Wikipedia has facilitated the collective creation of approximately 5.5 million English-language articles (Wikipedia, 2017). These articles attract roughly 500 million unique monthly users (Alexa.com, 2017), currently making Wikipedia the fifth most popular website in the world and the first go-to source for knowledge for most netizens.

Wikipedia is free in two distinct senses: free to use in terms of cost, and free as in "open." As its own definition stresses, "anyone can edit" its content, with almost absolute editorial power given to anybody visiting the site—no login required. With the exception of a handful of "protected" articles locked to public editing, changes to articles' text are updated online immediately. As a result of this open form, articles in Wikipedia have numerous authors of all backgrounds (Reagle, 2010), engaging millions worldwide in a process called "commons-based peer production" (Benkler and Nissenbaum, 2006).

Research has shown that academics' apprehension regarding Wikipedia seems to stem from concerns about the validity of its content—which is often taken as inherently unreliable—as well as suspicions regarding those authoring it (Jemielniak and Aibar, 2016). The former president of the American Library Association, Michael Gorman, expressed this position clearly when he said that academics who encourage the use of Wikipedia are "the intellectual equivalent of a dietitian who recommends a steady diet of Big Macs" (Reagle, 2010). Larry Sanger voiced similar concerns after he left the project to set up a more rigid online encyclopedia: "This arguably dysfunctional community is extremely off-putting to . . . academics," Sanger wrote, saying it seems Wikipedia is "committed to amateurism" (Sanger, 2006).

Looking into Wikipedia's content, *Nature* published in 2005 a now infamous report comparing randomly selected articles on Wikipedia to those from *Encyclopaedia Britannica*, the preeminent expert-written encyclopedia of yesteryear. The news article, titled "Internet Encyclopedias Go Head to Head," claimed that "Wikipedia comes close to Britannica in terms of accuracy of its science entries" (Giles, 2005). The article sparked a fierce debate between the two, with Britannica blasting the study as "fatally flawed," promoting *Nature* to respond both in an editorial and in official rebuttal to the media. A follow-up study conducted by the University of Oxford in 2012 together with Wikimedia Foundation, the nonprofit that manages Wikipedia, reached similar results in respect to accuracy, references, style, and readability, specifically regarding articles on the natural sciences (Casebourne et al., 2012). As content on Wikipedia is

never stable and can always be re-edited, these studies may do little to quell skeptics' concerns.

Personal expertise and accountability are cornerstones of academic culture, and Wikipedia seems to fall short on both accounts. In Wikipedia, academic credentials do not necessarily confer status, and editors are even instructed, "Share your expertise, but don't argue from authority." Moreover, users can easily lie about their bona fides or even open numerous accounts, so-called "sock puppets," that allow a biased editor to cover his or her tracks. Some users abuse the site's anonymous editing function for nefarious purposes, engaging in what is called "disruptive editing" or "vandalism," which can entail expletives, unwarranted deletions, or the addition of irrelevant content.

For example, a user called Brian Phosphorus made the following edit to the article for "Circadian rhythm" (CR) on May 17, 2008:

Circadian rhythms is a rapper from Peabody, Ma. He has yet to release an album, but he performs many live shows from his house/car. Born in 1985, CR has an older sister and a younger brother, and two parents.

The example of the aspiring rapper shows how people try to use the encyclopedia for self-promotion. Tellingly, only 7 minutes later a user called Hordaland deleted the rapper's unremarkable biography from the "Circadian rhythm" article, which more readily defines circadian rhythms as "any biological process that displays an endogenous, entrainable oscillation of about 24 hours."

This interaction demonstrates how the open format is used not only to contribute content but also to regulate content added by others. As discussed in the next section, this collective editorial process can be considered similar to that of the academic practice of review. Taken as such, it raises the opportunity to rethink some of the academic world's assumptions regarding Wikipedia.

CAN WIKIPEDIA REFLECT SCIENTIFIC PROCESSES?

Knowledge from what can be loosely called the circadian clock field is represented in a number of articles in Wikipedia, for example, "Jet lag," "Melatonin," and even the articles of Nobel laureates Jeffrey C. Hall, Michael W. Young, and Michael Rosbash (whose Wikipedia page, incidentally, was created as part of Herzog's class). To elaborate on the relationship between Wikipedia and circadian research activity, we focused on the articles that explicitly address the field itself, namely "Circadian

clock" (CC) (2017) and "Circadian rhythm" (CR) (2017). Reviewing each of these articles, we moved along two axes: content (both the article's present formulation and its evolution over time) and the article's editors.

Wikipedia offers built-in tools that are accessible through the "page information" sidebar button and through the "view history" tab of any Wikipedia article. For example, the chronological "view history" tool, also known as the *changelog* is an archive of the different versions of the article and a database of all its edits. The talk pages are a forum-like arena where editors can discuss potential changes to the article and reach a consensus on future formulations. Auxiliary data analysis tools have been developed by the Wikipedia community, supported by Wikimedia Foundation. The third-party XTools suite mines Wikipedia and its Wikimedia database, not for content but for data—tracking, for example, the number of edits made to any given article. Ordinarily, these tools are used to facilitate the oversight and regulation of the site. In recent years, such auxiliary tools have also been spurred on by the emergence of a "Wiki research" community dedicated to researching Wikipedia's data, for both academic and internal purposes. We suggest an expansive approach and use these tools in a manner that can shed light on how Wikipedia might reflect science.

The CC and CR Articles: A Snapshot

As of December 2017, each article had a slightly different focus: While CR addresses the wider context of "24-hour rhythms . . . driven by a circadian clock," the article for CC focuses on the "biochemical oscillator" driving those rhythms. The table of contents for each article gives form to the differences: For example, CR has sections dedicated to "History" and "Origins," which describe the field's history and the evolution of clocks, and a section on "Human health" depicting physical disorders linked to clocks. Meanwhile, CC posits a more mechanistic outlook, with sections dedicated to "Vertebrate anatomy," "Post-transcriptional modification" and "Systems biology approaches to elucidate oscillating mechanisms."

Wikipedia's native "what links here" tool shows how the CC and CR articles maintained hypertextual ties to a large body of articles: CC was hyperlinked in about 90 different articles and CR in more than 700, among them articles about scientists (e.g., Jürgen Aschoff, Ueli Schibler), genes/proteins (e.g., Rev-ErbA, NONO), and scientific concepts (e.g., zeitgeber, chronotype).

Much like content in academic papers, all claims in Wikipedia "must include an inline citation that directly supports the material" and all content in

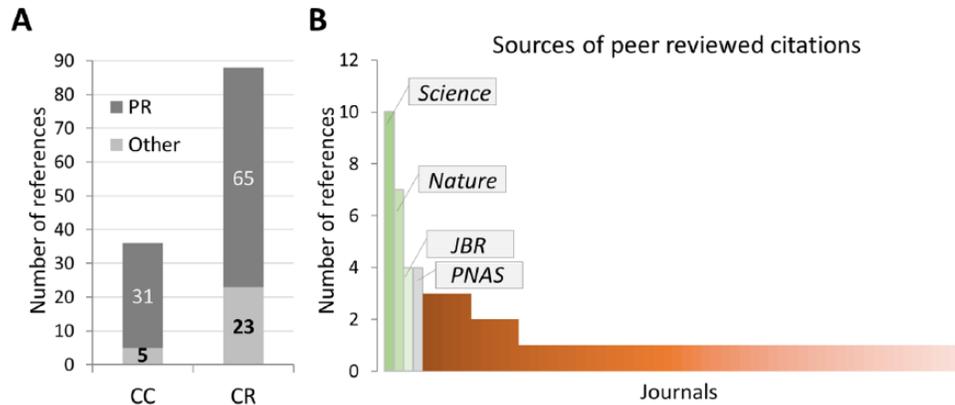


Figure 1. Distribution of references from the Wikipedia entries for “Circadian clock” (CC) and “Circadian rhythm” (CR). (A) The number of references from peer reviewed (PR) journals (darker shade) in comparison with other sources (lighter shade) in the articles for CC and CR as of December 2017. (B) The peer-reviewed journals cited in both articles, sorted by the number of references per source: a total of 60 different peer-reviewed journals were referenced. The most cited journals in CC and CR are indicated in lighter shade and include *Science*, *Nature*, *Journal of Biological Rhythms*, and *Proceedings of the National Academy of Sciences of the United States of America*.

articles “should be backed up by reliable sources,” its policy pages suggest (“Wikipedia: You don’t need to cite that the sky is blue,” 2017). Wikipedia’s guidelines instruct editors that to substantiate a factual claim, editors should use “academic and peer-reviewed publications” (“Wikipedia: Verifiability,” 2017). In this regard, Wikipedia stands out from Britannica and most printed encyclopedias, which usually only include internal references to other entries. Research has even shown that the higher the impact factor an academic paper has, the more likely it will be cited in Wikipedia (Teplitskiy et al., 2017).

In the two articles, most references were from well-respected scientific journals, with the bibliographies serving as a possible indication of a correspondence with the scientific outlook that the articles aim to represent. In CC, 31 of its 36 citations were from peer-reviewed journals, and in CR, 65 of 88 (Fig. 1A). Other sources included books and websites. In the two articles combined, 60 different journals were cited, most of them only once (Fig. 1B). The most cited journals were *Science* (10 references), *Nature* (7 references), *Journal of Biological Rhythms* (4 references), and *Proceedings of the National Academy of Sciences of the United States of America* (4 references).

In line with the academic distinction between primary and secondary sources, and following other encyclopedias and academic textbooks, the article for CR also offered a “Further reading” section that included key texts from the field’s history, like Aschoff’s *Circadian Clocks* (1965). Overall, based on their content and the standing of their bibliographic sources, these articles appear to be grounded in the world of science they purport to represent, both internally (i.e., hypertext) and externally (i.e., academic references).

The CC and CR Articles: A History

Because Wikipedia is constantly changing, it is not enough to examine the content of articles (as discussed above) to understand how they reflect their respective scientific fields; such understanding also requires a historical analysis of their growth over time.

CC and CR were added to Wikipedia over a decade ago (opened in 2005 and 2002, respectively) and were merged in 2006 before splitting up again in 2012. Until December 2017, CR was edited 1652 times whereas CC was edited 127 times, the latter a relatively low number, possibly due to that article’s more focused scope. On average, CC had 18 edits every year, while CR had 103. Throughout the articles’ history, there was no significant correlation between the number of edits and the overall size of the text (Fig. 2A), suggesting their limited value in terms of locating content-dependent editorial trends. Therefore, we used the *changelog* and Wikipedia’s native “compare” tool (Fig. 2, B and C) to manually review the history of edits and attempted to locate key changes to both content and the references since the articles opened and until December 2017.

Some edits—like the addition of well-sourced research—were deemed beneficial by Wikipedia editors and were therefore conserved in the articles’ texts. Other edits were rejected, being labeled either as irrelevant or as a form of vandalism: from “happy birthday” (23 May 2005; deleted after 2.5 h) to racist expletives (e.g., 5 June 2009; deleted after 1 min) to full-blown deletion of the entire article’s text (e.g., 19 April 2005; reverted after 7 min). Overall, 3 of CC’s and 108 of CR’s edits were reverts, taken by some researchers to be a possible metric for vandalism that

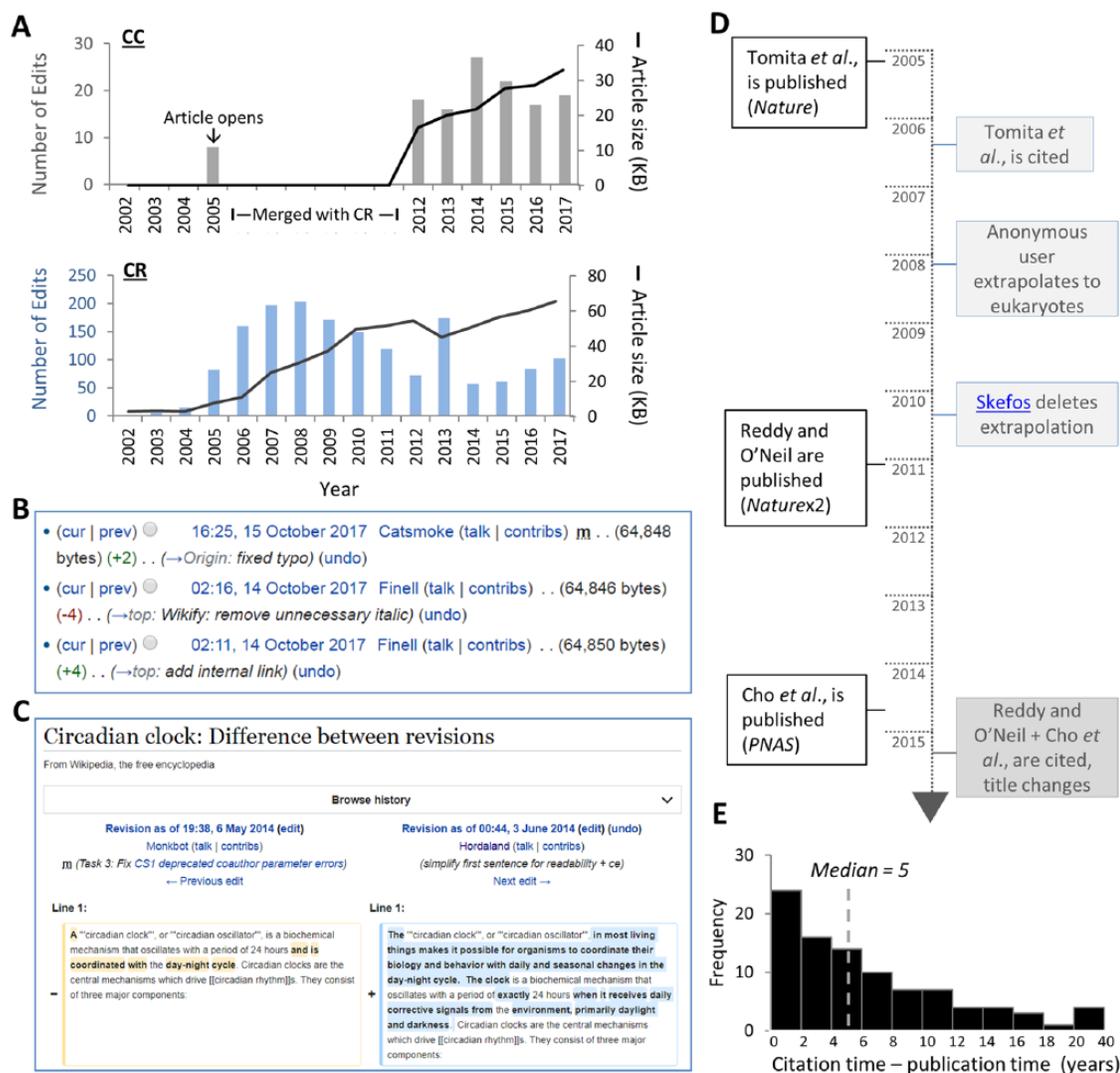


Figure 2. Mining the edit history of Wikipedia entries for “Circadian clock” (CC) and “Circadian rhythm” (CR). (A) Timeline of CC (top) and CR (bottom) articles in respect to number of edits per year (bars in light shade) and the size of the articles (line, dark shade). (B) An image of Wikipedia’s *changelog* (history tool), as seen in the link for “view history.” Each line indicates a saved edit, in this case an example of 3 edits that were made to CR in October 2017. (C) The compare tool is a native textual comparison function that allows a side-by-side comparison of different versions of the article’s text, with differences to specific lines or words highlighted (old text on the left and new text on the right). It is accessible through the “view history” tab of each article, in this case taken from the CC *changelog*. (D) A timeline of selected edits pertaining to transcriptional regulation of circadian clocks, from 2005 until 2015. Listed on the left of the axis are selected scientific journal publications (black frames); on the right of the axis are key events in the history of CC (darker gray fill) and CR (brighter gray fill), vis-à-vis publication of the aforementioned articles. The edits, as well as the publication list, are partial and have been manually chosen by the authors. (E) The delay between an article’s publication in a peer-reviewed journal and its integration into the CC and CR Wikipedia articles. The median time for any given article to be cited in either Wikipedia entries was 5 years (the combined references are presented; 93 different citations were examined).

is content independent (Yasseri et al., 2012b; Vuong et al., 2008).

Unexpectedly, one of the most contentious areas in the articles was the translation of the term *circadian*. A November 2003 version of the article stated that the term “comes from the Latin ‘circa,’ meaning ‘about’ and ‘dia,’ ‘day.’” On 11 May 2004, the word *about* was changed to *approximately* and then was changed to *around* just 4 days later. Additional versions were

modified (e.g., *around* was again changed to *approximately* on 29 March 2008) until the term evolved into the somewhat awkward yet just as accurate phrasing of “necessarily almost exactly 24 hours” (16 March 2017, CC). These edits are not a form of vandalism, but neither do they reflect any changes in the scientific understanding of clocks. Nonetheless, other edits pertaining to the body of the text did seem to follow scientific discoveries in the field.

Wiki KaiA

Since the discovery of cell autonomous clocks, one fundamental question in the field has regarded the cellular clocks' capability to generate daily rhythms. According to research, an integral and universal feature of circadian oscillators relies on generation of transcription-translation feedback loops (TTFL). However, accumulating evidence suggests that non-transcriptional processes are in some cases sufficient to sustain approximately 24-h rhythms. In 2005, the laboratory of Takao Kondo at Nagoya University published studies in which circadian rhythms of protein phosphorylation were reconstituted *in vitro*, using only purified cyanobacterial proteins and ATP (Nakajima et al., 2005; Tomita et al., 2005). Hence, oscillations were shown to persist independent of cellular context and, therefore, transcription. A number of years later, circadian cycles were discovered in human red blood cells, which lack a nucleus and are therefore not transcriptionally active (O'Neill and Reddy, 2011). In 2014, Cho et al. observed circadian rhythms in red blood cells of mice and hence confirmed the results of O'Neill and Reddy. These studies, among others, contributed to the formulation of a new generalized paradigm of the function of biological oscillators, namely, by revising the necessity of gene transcription for all aspects of circadian function. These studies were published after the CC and CR articles were created and can thus serve as a test case example for examining how new knowledge is integrated into Wikipedia as it accumulates. By working through the *changelog* for CC and CR, we focused on how the different stages of this paradigmatic reformulation integrated into the two articles and how they reframed the "story" of circadian clocks, one step at a time (Fig. 2D).

On 23 May 2006, the 2005 findings from the Kondo laboratory are first referenced *prima facie* under the "Origin" subsection, as a local phenomenon limited to cyanobacteria. The location and the phrasing implied that the findings were attributed a relatively localized and even marginal role, more as an exception to the TTFL rule than the case, with the contributing editor noting in the text that "transcription/translation feedback mechanism [was] still believed to hold true for eukaryotic organisms."

On 26 February 2008, an anonymous user (identified only through an IP number) deleted the line about the mechanism "still" being held true for eukaryotes and extrapolated from cyanobacteria to suggest that it is an "outstanding question whether circadian clocks in eukaryotic organisms require translation/transcription-derived oscillations." On 14 March 2009, the phrase "outstanding question" was changed to "unanswered question." However,

on 25 April 2010 this extrapolation was edited out, with a user called Skefos explaining it was "infeasible" to make such an inference.

The first reference to O'Neill and Reddy (2011) was made on 16 May 2015, four years after their research was published in *Nature*. It was added to Wikipedia as part of a major revision of the CC article's text, which was edited to state, "In 2011, a major breakthrough in understanding came from the Reddy laboratory. . . . Therefore, the model of the clock has to be considered as a product of an interaction between both transcriptional circuits and non-transcriptional elements." This revision also included the first reference to Cho et al. (2014) as well as another study by Reddy and O'Neill (O'Neill et al., 2011) on transcriptionally inhibited algae.

Interestingly, the initial significance attributed to Kondo's finding was reassessed in accordance with the new research, with the citation changing locations in the text to now support the claim that "Studies in cyanobacteria, however, changed our view of the clock mechanism, since it was found by Kondo and colleagues."

The broad (evolutionary conserved) view of post-transcriptional rhythms also affected the table of contents, with the relevant section's title changing from "Transcriptional and translational control" to "Transcriptional and non-transcriptional control."

These changes to the text and headline, as well as the addition of the group of references in tandem, reflect how in this case Wikipedia managed to successfully bear testimony to the scientific outlook regarding the circadian clock mechanism, even as it evolved. At different periods in the history of the articles, their text and references seemingly reflected the significance that the relevant scientific community attributed to the aforementioned studies. In Wikipedia, as in the scientific world, it seems that a generalization of Kondo's claim regarding prokaryotes could be fully accepted only after research about eukaryotes was published and accepted by the scientific community first.

The dynamics of this process suggest viewing the integration of knowledge as a nonlinear process, with certain discoveries being attributed different significance at different points in time: from the initial discovery of a localized and primitive mechanism, whose wider ramifications regarding the article's fundamental question on "how the clock works" were still unclear, to a wider reformulation of the overall paradigm of the field. This could also be seen to be represented in the delay of the citations' integration into the articles. The citation for Tomita et al. (2005) appeared a year after its publication but only within a minimalistic interpretation. In contrast, the study by O'Neill and Reddy (2011) took 4 years to

enter Wikipedia. This work appeared only after the findings could be generalized, and the generalization accepted independently, as represented by the publication of Cho et al. (2014), which took just shy of a year to enter the article.

Examining the latency between an article's publication date and its appearance in the reference lists for CC and CR, we found that the median time for any given article to be cited was 5 years (Fig. 2E). If we judge by this baseline, Tomita et al. (2005) and Cho et al. (2014) were included relatively quickly, while the study by O'Neill and Reddy (2011) was rather closer to the average integration time.

Interestingly, the overall shift in understanding of the TTFL also manifested across other articles in Wikipedia: On 28 January 2014, Kondo received his own personal article, and from 27 April 2017 his entry stated that "Kondo's seminal 2005 discovery . . . disproved the universal necessity of the transcription-translation autoregulatory feedback loop."

The CC and CR Articles: Contributors

To fully understand what informed this process, we looked at the articles' editors. According to the data supplied by the third-party XTools, the CC and CR articles had a large number of contributors (855 and 56 in CR and CC, respectively), but only a small cadre of participants were committed to maintaining the articles over time. In line with the "long tail" model usually used to describe editors of Wikipedia articles (Benkler, 2006), the top 10 editors of CC contributed the majority of its content (~90%), with Gorton K providing 44% of the text. In CR, 60% of the text was added by the top 10 editors, and of these, a user called Hordaland was responsible for one third. Interestingly, textual contributions were not directly tied to editorial contributions, as editors who added the majority of the text were not necessarily those who edited most often (e.g., Gorton K in CC added the most text but had only 7 editing events) (Fig. 3A). Bots—software programs created to perform mundane editorial tasks—also contributed edits (Fig. 3B). Overall, 47 different bots contributed 108 edits in CR, and 10 bots contributed 14 edits in CC.

Whereas a bot's function must be clearly described by its creator and verified by the Wikipedia community before it becomes operational, the identity of human contributors is limited to the information they choose to divulge, with many using pseudonyms or claiming expertise that cannot be verified.

For example, Gorton K identifies himself on his user page as a "budding biological scientist of some sort." We failed to confirm this, highlighting how full accountability may always be lacking in Wikipedia. However, others we succeeded in identifying.

The user called Looie496 identified himself as William Skaggs, a neuroscientist and author of peer-reviewed studies published in prominent academic journals. On his blog (Skaggs, 2014) and in an email correspondence, Skaggs confirmed he was the aforementioned Wikipedia editor. On Wikipedia, Skaggs edited various scientific articles, like "Brain" and "Dopamine," as well as the Wikipedia project for neuroscience. His contributions, including those to CC and CR, seem limited to the scope of his academic expertise (as indicated from his publication list on PubMed). Thus, he serves as a testimony to the presence of experts on Wikipedia editing topics pertaining directly to their research field.

The most prominent editor on the CR and CC pages was a user called Hordaland, ranking first in overall edits to these articles (244 in CR and 14 in CC), (Fig. 3A). We successfully identified Hordaland as Beth MacDonald, an American based in Norway who had delayed sleep phase disorder (DSPS). MacDonald, who died in 2017, maintained a blog in collaboration with James Fadden, a biochemist working in private industry who confirmed her identity to us. Together, the two founded an international nonprofit called the Circadian Sleep Disorder Network. In the sleep disorder network's blog, MacDonald wrote that her "mission is to inform [people] about DSPS based on what I've learned since diagnosis" (MacDonald and Fadden, 2017). Describing MacDonald as someone with both intimate and scientific knowledge of the topic to which she contributed vastly online, Fadden wrote, "Despite having no formal scientific training, [MacDonald] steeped herself in the circadian science literature."

Because every edit is logged, one could use editing activity as a form of action metric (Yasserli et al., 2012a). To get to know our main characters better, we created actographs for their editing patterns in Wikipedia: While Looie496 (Skaggs) edited in a highly rhythmic daily cycle, Hordaland (MacDonald), who had a sleeping disorder, edited around the clock. Gorton K, the purported "budding biologist," edited rather sporadically, mostly on Thursdays (Fig. 3C).

CAN WIKIPEDIA EXPAND ON THE SCIENTIFIC TRADITION?

Although limited to a few articles pertaining to the circadian clock, our case study exemplifies how experts and laypeople can join forces to maintain scientific entries in a dynamic manner over time to reflect shifts in scientific understanding. Moreover, it seems that no small part of this process was facilitated by tools and mechanisms that far from being

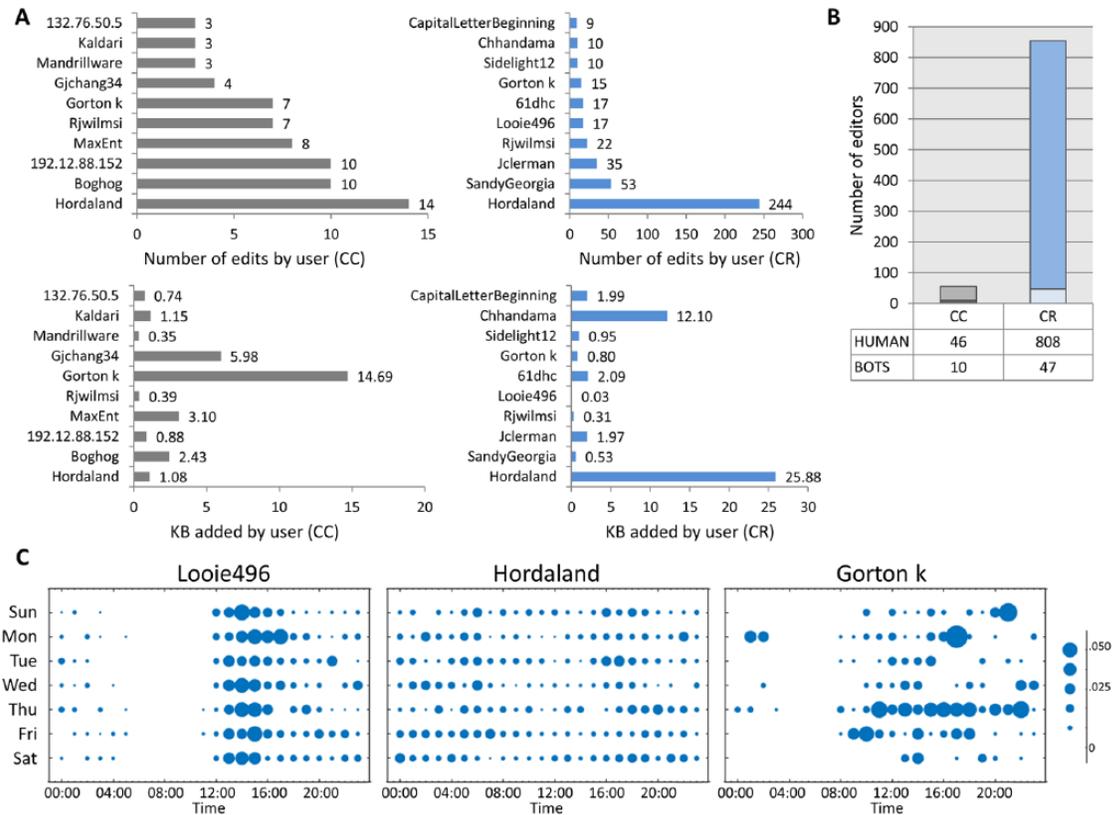


Figure 3. The editors of Wikipedia entries for “Circadian clock” (CC) and “Circadian rhythm” (CR). (A) The number of edits (top) and the amount text added in kilobytes (bottom) of the top 10 editors of CC (left) and CR (right). (B) The number of human editors compared with “bot” editors in CC and CR. (C) Indicated users’ overall editing activity in Wikipedia, referring to 1000 latest edits from 31 December 2017 and earlier (Gorton K had only 500 altogether). Bubble size indicates fraction of overall edits, normalized to each user separately. Time of day is according to UTC.

antagonistic to academic culture may actually expand on it. We will now discuss how Wikipedia (1) embraces and encourages revision as part of a “mob” review process, (2) expands the quantitative and qualitative pool of potential reviewers as what we will call “citizen encyclopedists,” and (3) can be seen as facilitating access to scientific knowledge for free at a widespread level, thus joining the long tradition of communicating science. Taken in this light, Wikipedia’s digital—as opposed to print—format can be viewed as expanding on the understanding that knowledge is always accumulating, never finite.

Scientific Gatekeepers and Citizen Reviewers

On 8 July 2008, a user called Dov Henis, an octogenarian with a purported PhD in biochemistry, edited the CR article to include a theory regarding the origins of circadian rhythms. Allegedly extrapolating from legitimate scientific sources, Henis had in fact just fleshed out his own original theory, in what can be described as a form of academic vandalism. However,

on 13 August 2008, Henis was reminded that Wikipedia bars original research, preferring “well accepted facts,” with one user informing him on his talk page that “the articles [on Wikipedia] are for well sourced mainstream information, NOT for original research.” Henis’ example shows how even vandalism that may seem scientific in wording, style, or origin may fail to pass Wikipedia’s review process.

Wikipedia engrains the revision process into its workings and offers a slew of tools to that end. One main mechanism for encouraging long-term commitment to articles allows editors to “watch” a page they have edited to receive an email alert every time additional changes are made (there were 50 and 480 “page watchers” in CC and CR, respectively). Additionally, so-called “hatnotes” allow users to highlight claims in need of verification, and these serve as a de facto call to either find sourcing or edit the claims out. According to the Wikipedia policy page, “write neutrally,” “no original research,” and “verifiability” are Wikipedia’s three core content guidelines, with the latter meaning “that other people using the encyclopedia can check that the

information comes from a reliable source”—and it is the different editors who are expected to check that articles are properly sourced.

To test this process, on 9 November 2016 we added a line to the CC article based on recent findings from Adamovich et al. (2017). However, we only added the information and neglected to properly cite the published paper. After 24 hours, just before noon on 10 November 2016, Hordaland added a “citation needed” comment on the claim and wrote us (through Wikipedia) requesting that we add a citation to verify its accuracy. And so we did. All in all, it took some 27 hours from the time the claim entered Wikipedia to the time it had been properly cited.

In another case of what we term a beneficial “citizen review,” a purported researcher added a paragraph to the CC article praising his laboratory’s work on clocks, prompting a user called Boghog to re-edit the text to “focus on the results of the research, not the researchers” (3 June 2014).

These examples show how Wikipedia’s open form, the source of much of its vandalism and shortcomings, is also used for maintaining academic standards through a type of review process that allows users to vet one another’s edits. However, unlike academic journals, Wikipedia’s review can be massive in scope, in terms of both the number of possible reviewers and their potential background.

Citizen Encyclopedists, Scientific Gatekeepers, and Bots

Citizen science is becoming increasingly prominent and may offer a framework for understanding some of the ties between Wikipedia and science. The current working definition for citizen science entails including the public in the processes of scientific knowledge production and fostering dialogue between experts and laypeople. Implicit in the idea of citizen science is that nonscientists may hold knowledge that is pertinent to research despite their lack of training. Moreover, the concerns that citizen science inspires among the scientific community (Golumbic et al., 2017) seem similar to those inspired by Wikipedia. Classic examples of citizen science include ornithologists and birdwatchers—dedicated “amateurs” involved in what sociologists term “serious leisure” (Stebbins, 1982).

In contrast to citizen scientists, Wikipedia’s editors do not attempt to produce knowledge. Like all encyclopedias, Wikipedia deals with what the sociologist of science Bruno Latour called “ready-made science,” as opposed to “science in the making” (Latour, 1987). While the latter describes the creation of new and original research regarding the world through

experimentation, the former has to do with assemblage of preexisting scientific findings. Therefore, Wikipedia may be seen as an example of citizen science in the sense that the process of knowledge curation that it hosts is no longer as rigidly bound by the expert-layperson divide. Wikipedia involves both content creators and content regulators, with experts, laypeople, and bots assuming different roles at different times.

Among the CC and CR contributors, we identified those whom we call “citizen encyclopedists,” like Hordaland, who played a role that has historically been reserved for academics. Bots—which played a regulative role on Wikipedia—are also an interesting example of “serious amateur” involvement in Wikipedia as they are predominantly developed by volunteers from the Wikipedia community for internal purposes (Halfaker and Riedl, 2012).

As well, real-world scientists active in Wikipedia also contributed, functioning as a specific kind of reviewer who could be labeled a “scientific gatekeeper.” For instance, Prof. Su, the cofounder of the Gene Wiki project, deleted in CR an entire section called “Music” dedicated to yet another band called “Circadian Rhythms” (5 March 2009, less than an hour after the addition) and, on a different occasion, edited out a line about teens not being able to develop normal circadian rhythms (26 July 2008).

Reading through Wikipedia’s talk pages, we also located interactions between these two different types of authors (lay and expert). For example, Hordaland periodically edited another article called “Bacterial circadian rhythms,” which was opened by a user identifying as Carl Hirschie Johnson, whom we independently confirmed to be the clock scientist from Vanderbilt University. Hordaland exemplifies how despite Wikipedia’s lax admission standards, it can attract a different kind of expert—a “lay expert” (Prior, 2003)—with a personal and vested interest who may or may not have formal academic training but who maintains an ongoing dialogue with the relevant scientific community. Indeed, Hordaland even thanked Prof. Johnson for writing “a very interesting article.”

In a telling example, after Hordaland changed the opening section of the CR article to state that “Circadian rhythms are endogenous and can be entrained by external cues,” Looie496 wrote to her on her talk page to say that the word “and” in the sentence should be replaced with “but,” as the former “seems confusing to me. . . . Given that circadian rhythms are generated internally, it will be unexpected that external cues control them, so the word ‘but’, is needed for clarity” (1744 h, 7 February 2010). Hordaland responded cordially, offering an alternative formulation, “hoping this reword[ing] satisfies

us both: ‘Although circadian rhythms are endogenous, they are adjusted (entrained) to the environment by external cues called zeitgebers, the primary one of which is daylight’” (1831 h, 8 February 2010). Looie496 agreed: “That works fine for me, thanks” (2120 h, 8 February 2010).

These interchanges bring to light an encyclopedic effort in which laypeople and experts not only write different parts of the article but also edit it collaboratively. While nonacademics like Hordaland may make textual contributions to the article, experts like Looie496 may also be there to review their changes. This way, new types of actors with new types of motivation can become a positive force in disseminating knowledge: “The goal is to say things in the simplest terminology that makes the statements fully correct,” Looie496 claims during one such debate (19 March 2010, CR talk page), in a characterization that seems to fit the articles we researched as much as a description of Wikipedia writ large.

Digital Encyclopedia, Print Culture

In 1660, a group of learned British gentlemen established the Royal Society of London for Improving Natural Knowledge. Its motto was “take nobody’s word for it” and in a few years’ time they founded *Philosophical Transactions*, the world’s first scientific journal and the first to introduce scientific priority and peer review into the academic world (Kronick, 1976).

Roughly a century later, Denis Diderot’s *Encyclopédie* (1751-1766), considered by many the first modern encyclopedia, was published with the aim of “chang[ing] the way people think” (Diderot, quoted in Hunt et al., 2007) by allowing them easy access to the most up-to-date knowledge of the Enlightenment, bringing the best minds in France to write its entries. The *Encyclopédie* followed up on the ethos first articulated by the Royal Society (Winger, 1980) and paraphrased by one of its most famous members, Charles Darwin, who wrote that “general and popular treatises are almost as important for the progress of science as original work” (Darwin, 2017).

This short historical interlude is meant to show that the origins of the academic world as we know it today have always been inextricably tied to the idea of communicating science outward to nonexperts (Burke, 2013). Wikipedia pushes this ethos even further thanks to its digital form.

In addition to breaking down the economic barrier to knowledge by offering its content for free to anyone with an internet connection, Wikipedia makes use of its digital format to push forward the encyclopedic tradition of communicating science. For example, much like the encyclopedias of the past, articles in Wikipedia are part of a web (or

“cycle”) of knowledge. However, Wikipedia’s hypertext “wikilinks” to other articles and its “See also” sections allow Wikipedia articles to serve as a gateway to the wealth of knowledge available online, academic or otherwise. In this way, the advent of hypertext allows readers of Wikipedia to “take nobody’s word for it.”

Wikipedia’s technological aspect expands on the very concept of revision and review: The constantly updated yet catalogued and accessible database reflects a position that views knowledge production as an endless process. “Perfection is not required: Wikipedia is a work in progress,” Wikipedia’s editorial guidelines explain (“Wikipedia: Wikipedia is a work in progress,” 2017), exemplifying an ethos that is as much academic as it is digital. “Wiki articles can serve as constantly updated open access review articles” a *Nature* news story recently claimed (Zastrow, 2017).

Since 2012, the Wikipedia project for Genetics has collaborated with the journals *PLOS Computational Biology* and *PLOS Genetics* on an initiative called “Topic Pages,” which aims to bridge the “journal-Wikipedia gap” by creating review articles. Once a paper on one of the covered topics is accepted for publication by either of the aforementioned PLOS journals, another copy is uploaded to Wikipedia in the wiki format, where it can be edited like any other article on the site. “Topic Pages expand on earlier attempts to add a dynamic component to scholarly publishing. They provide the English Wikipedia with expert-written and expert-reviewed content, and allow authors to get credit for their work,” PLOS wrote (PLOS Blog, 2012) about the project with its launch, indicating the type of cooperation the future may hold.

Wikipedia’s model appears to derive many of its practices from the print-based academic tradition as well as to permit new forms to find a home. Taken in this light and following the examples noted here, there seems to be a need to reassess the scientific community’s apprehension regarding the open online encyclopedia. Here we have addressed some of the fundamental concerns regarding Wikipedia, and we raise the possibility that far from being a threat to science and academics, as currently understood by many, Wikipedia may also be viewed as a reinvigorating force that harnesses technology to increase diversity and open knowledge production to new actors.

ACKNOWLEDGMENTS

The authors thank Gad Asher and members of his research group, notably Gal Manella, and others at the Weizmann Institute of Science, namely Julie Laffy. We also thank

Prof. Yossi Schwartz and Dr. Moshe Elhanati from Tel Aviv University's Cohn Institute for the History and Philosophy of Science and Ideas. We thank all the brilliant researchers whose involvement with Wikipedia and whose professional insight were invaluable to this project, as well as the scientists, clock enthusiasts, and citizen encyclopedists who edited the Wikipedia articles, with a special acknowledgment to Beth MacDonald, who died in 2017. Rona Aviram received funding from the Azrieli Foundation.

CONFLICT OF INTEREST STATEMENT

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

- Adamovich Y, Ladeuix B, Golik M, Koeners MP, and Asher G (2017) Rhythmic oxygen levels reset circadian clocks through HIF1alpha. *Cell Metab* 25:93-101.
- Alexa.com. (2017) The top 500 sites on the web [cited 6 November 2017]. Available from: <https://www.alexa.com/topsites>
- Benkler Y (2006) *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press.
- Benkler Y and Nissenbaum H (2006) Commons-based peer production and virtue. *J Polit Philos* 14:394-419.
- Burke P (2013) *Social History of Knowledge: From Gutenberg to Diderot*. Hoboken, NJ: John Wiley & Sons.
- Casebourne I, Davies C, Fernandes M, and Norman N (2012) Assessing the accuracy and quality of Wikipedia entries compared to popular online encyclopaedias: a comparative preliminary study across disciplines in English, Spanish and Arabic. Brighton, UK: Epic (cited December 31, 2017). Available from: http://commons.wikimedia.org/wiki/File:EPIC_Oxford_report.pdf (cited December 31, 2017)
- Chiang CD, Lewis CL, Wright MD, Agapova S, Akers B, Azad TD, Banerjee K, Carrera P, Chen A, Chen J, et al. (2012) Learning chronobiology by improving Wikipedia. *J Biol Rhythms* 27:333-336.
- Cho CS, Yoon HJ, Kim JY, Woo HA, and Rhee SG (2014) Circadian rhythm of hyperoxidized peroxiredoxin II is determined by hemoglobin autoxidation and the 20S proteasome in red blood cells. *Proc Natl Acad Sci U S A* 111:12043-12048.
- Circadian clock. (2017) Wikipedia [cited 31 December 2017]. Available from: https://en.wikipedia.org/w/index.php?title=Circadian_clock&oldid=803545164
- Circadian rhythm. (2017) Wikipedia [cited 31 December 2017]. Available from: https://en.wikipedia.org/w/index.php?title=Circadian_rhythm&oldid=808746033
- Darwin C (2017) *Life and Letters of Charles Darwin*. Litres. Originally published in 1959 by Basic Books, Inc., New York.
- Giles J (2005) Internet encyclopaedias go head to head. *Nature* 438:900-901.
- Golumbic YN, Orr D, Baram-Tsabari A, and Fishbain B (2017) Between vision and reality: a study of scientists' views on citizen science. *Citizen Science: Theory and Practice* 2(1):6.
- Halfaker A and Riedl J (2012) Bots and Cyborgs: Wikipedia's immune system. *Computer* 45:79-82.
- Hunt L, Martin TR, Rosenwein BH, and Smith BG (2007) *The Making of the West: Peoples and Cultures, A Concise History, Volume II: Since 1340*. Boston, MA: Bedford/St. Martins.
- Huss JW 3rd, Orozco C, Goodale J, Wu C, Batalov S, Vickers TJ, Valafar F, and Su AI (2008) A gene wiki for community annotation of gene function. *PLoS Biol* 6:e175.
- Jemielniak D and Aibar E (2016) Bridging the gap between Wikipedia and academia. *J Assoc Inf Sci Technol* 67:1773-1776.
- Kronick DA (1976) A history of scientific and technical periodicals: the origins and development of the scientific and technological press, 1665-1790. *The Library Quarterly* 32(3):238-239.
- Latour B (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
- MacDonald B and Fadden J (2017) DSPTS, a sleep disorder blog [cited 9 September 2017]. Available from: <https://delayed2sleep.wordpress.com>
- Mesgari M, Okoli C, Mehdi M, Nielsen FÅ, and Lanamäki A (2015) "The sum of all human knowledge": a systematic review of scholarly research on the content of Wikipedia. *J Assoc Inf Sci Technol* 66:219-245.
- Nakajima M, Imai K, Ito H, Nishiwaki T, Murayama Y, Iwasaki H, Oyama T, and Kondo T (2005). Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* 308(5720):414-415.
- O'Neill JS and Reddy AB (2011) Circadian clocks in human red blood cells. *Nature* 469:498-503.
- O'Neill JS, van Ooijen G, Dixon LE, Troein C, Corellou F, Bouget FY, Reddy AB, and Millar AJ (2011) Circadian rhythms persist without transcription in a eukaryote. *Nature* 469:554-558.
- PLOS_Blog (2012) Bridging the journal-Wikipedia gap [cited 31 December 2017]. Available from: <http://blogs.plos.org/plos/2012/04/bridging-the-journal-wikipedia-gap/>
- Prior L (2003) Belief, knowledge and expertise: the emergence of the lay expert in medical sociology. *Sociol Health Illn* 25:41-57.

- Reagle JM (2010) *Good Faith Collaboration: The Culture of Wikipedia*. Cambridge, MA: MIT Press.
- Sanger L (2006) Toward a new compendium of knowledge. Citizendium.org [cited 31 December 2017]. Available from: <http://larrysanger.org/2006/09/toward-a-new-compendium-of-knowledge-longer-version/>
- Skaggs W (2014) Brains and more [cited 5 October 2014]. Available from: https://web.archive.org/web/20141018031723/http://weskaggs.net:80/?page_id=4
- Stebbins RA (1982) Serious leisure: a conceptual statement. *Pac Sociol Rev* 25:251-272.
- Teplitskiy M, Lu G, and Duede E (2017) Amplifying the impact of open access: Wikipedia and the diffusion of science. *J Assoc Inf Sci Technol* 68:2116-2127.
- Thompson N and Hanley D (2017) Science is shaped by Wikipedia: evidence from a randomized control trial. Unreviewed preprint article on Social Science Research Network (SSRN) online archive/database (cited December 31, 2017). Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039505
- Tomita J, Nakajima M, Kondo T, and Iwasaki H (2005) No transcription-translation feedback in circadian rhythm of KaiC phosphorylation. *Science* 307:251-254.
- Vuong BQ, Lim EP, Sun A, Le MT, Lau HW, and Chang K (2008) On ranking controversies in Wikipedia: models and evaluation. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pp 171-182. Palo Alto, CA: Association for Computing Machinery (ACM).
- Wikipedia is a work in progress. (2017) Wikipedia [cited 31 December 2017]. Available from: <https://en.wikipedia.org/wiki/Wikipedia:WIP>
- Wikipedia: Verifiability. (2017) Wikipedia [cited 31 December 2017]. Available from: <https://en.wikipedia.org/wiki/Wikipedia:Verifiability>
- Wikipedia: You don't need to cite that the sky is blue. (2017) Wikipedia [cited 31 December 2017]. Available from: https://en.wikipedia.org/wiki/Wikipedia:You_don%27t_need_to_cite_that_the_sky_is_blue
- Winger HW (1980) The business of enlightenment: a publishing history of the "Encyclopédie," 1775-1800, by Robert Darnton. *Libr Q* 50:251-252.
- Yasseri T, Sumi R, and Kertesz J (2012a) Circadian patterns of Wikipedia editorial activity: a demographic analysis. *PloS One* 7:e30091.
- Yasseri T, Sumi R, Rung A, Kornai A, and Kertesz J (2012b) Dynamics of conflicts in Wikipedia. *PloS One* 7:e38869.
- Zastrow M (2017) Wikipedia shapes language in science papers. *Nature*. Available from: <https://www.nature.com/news/wikipedia-shapes-language-in-science-papers-1.22656>