The bot that makes history

# Introduction: What is Wikipedia time?

# Theoretical Overview: Wikipedia and the primacy of the past

How does Wikipedia portray the past? Scholars typically give three answers. Some argue that Wikipedia produces *history*: it represents the past in literary form. Wikipedia history may be more “colorful,” “anecdotal” and “factualist” than “professional history,” observes Roy Rosenzweig, but history it most certainly is (Rosenzweig 2006, 142). Others argue that Wikipedia articles comprise *collective memories* that evoke shared experiences. From this perspective, Wikipedia’s Talk pages are more important than the articles themselves, and its editors are more important than its readers. As Christian Pentzold argues, Wikipedia’s Talk pages are non-physical “memory places,” where editors meet to “negotiat[e]” the “memorable elements” of their experiences (Pentzold 2009, 264). Numerous scholars have followed in Pentzold’s wake to examine how editors “build” or “form” collective memories in Wikipedia (Ferron and Massa 2011b, 2011a; Porter, Krafft, and Keegan 2020). A third group of scholars argue that Wikipedia is a repository of *facts*. Wikipedia may well publish works of history and store collective memories, but its main role is to produce atomistic facts that are propagated through knowledge graphs (Ford 2020, 2022). Wikipedia may be *memory* to thousands of editors. It may be *history* to millions of readers. But it is mere *fact* to billions of search requests and API calls.

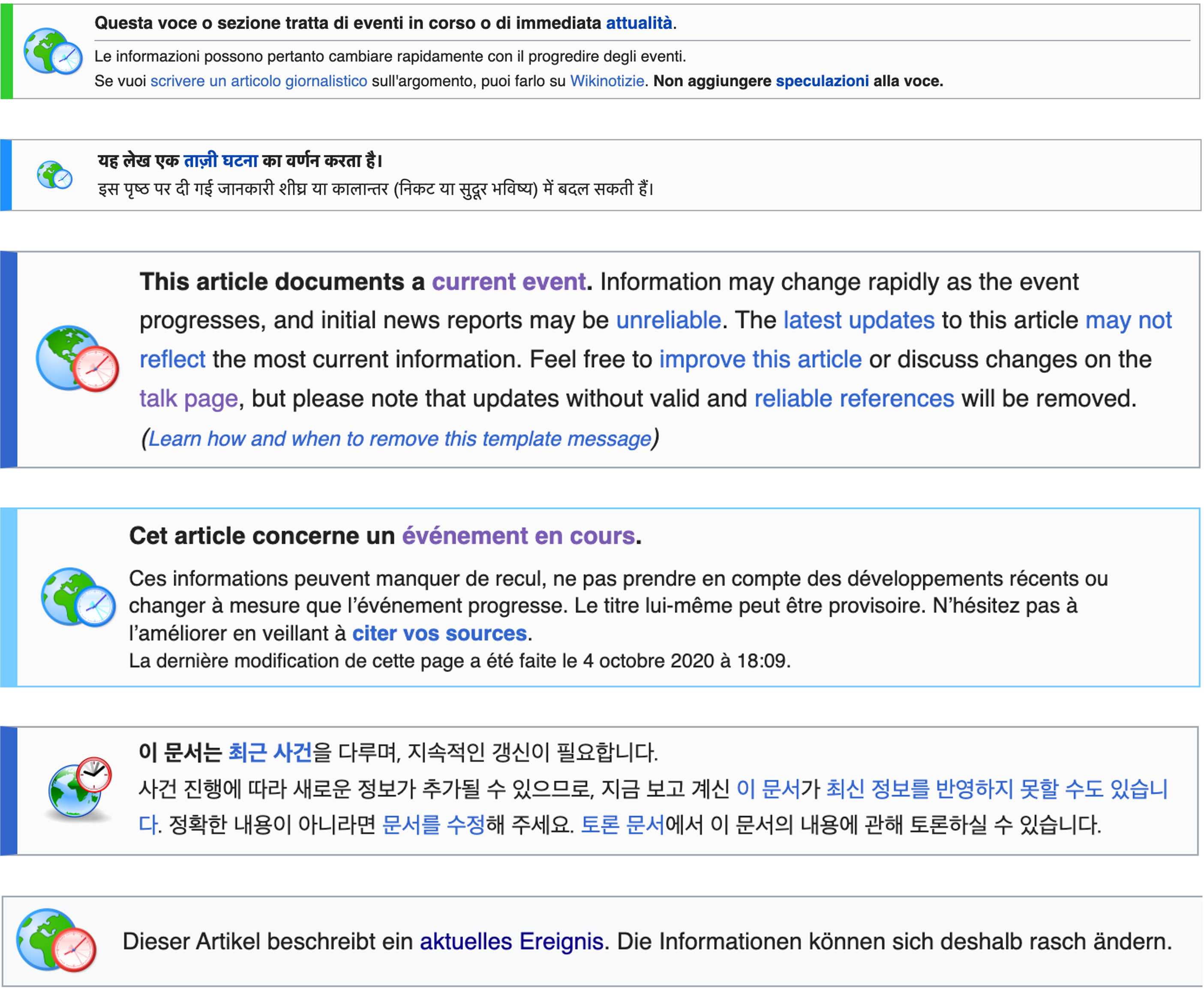
These approaches are not mutually exclusive. Search engines, readers and editors all produce and consume Wikipedia in different ways, and a complete account of the encyclopedia must include them all. In which case, we must ask: how are the historical, memorial and factual aspects of Wikipedia related?

One way to approach this question is to focus precisely on the *pastness* of history, memory and fact. Pastness is central to Wikipedia’s self-definition. “Wikipedia is not a crystal ball,” reads [a famous policy](https://en.wikipedia.org/wiki/WP:NOT), wherein we also read that Wikipedia is “not a newspaper.” It is the pastness of Wikipedia that allows it to function simultaneously as history, memory and fact. Pastness is obviously a feature of both history and memory: I cannot remember an event nor write its history until it has happened. The pastness of *fact* is less obvious. Wikipedia contains facts about fictional spacecraft, embroidery techniques and the heat death of the universe. In what sense can such facts be said to be “past”?

Wikipedia itself provides an answer in two of its foundational policies. According to [‘No Original Research’](https://en.wikipedia.org/wiki/Wikipedia:No_original_research), no new facts are to be admitted to the encyclopaedia. The only allowable facts are—the *old*. According to [‘Neutral Point of View’](https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view), no controversial facts are to be admitted to the encyclopaedia. The only allowable facts are—the *settled*. Facts are geological. Only time can grind down the seashells of evidence and bring forth the limestone of objectivity. Editors who wish to include new or unsettled facts in the encyclopaedia are advised that [‘There is no deadline’](https://en.wikipedia.org/wiki/Wikipedia:There_is_no_deadline). *Þæs oferēode; þisses swa mæg.* That passed; so may this. Eventually everything is past.

Despite its supposed pastness, Wikipedia is well-known as a source of information on current events. It is ‘An Encylopedia with Breaking News’ (Keegan 2019). Current events dominate Wikipedia, accounting for the lion’s share of user contributions and page views at any given time (Keegan, Gergle, and Contractor 2011). Scholars have analysed Wikipedia’s coverage of current events in detail. We now know how Wikipedia’s editors clash over the nature and definition of current events (Ford 2022; Pentzold 2009), how they link current events into larger thematic structures (Twyman, Keegan, and Shaw 2017), how they adopt newsroom practices to co-ordinate their efforts (Avieson 2019), how they revisit old articles to commemorate traumatic events (Ferron and Massa 2014), and how they shape the interpretation of events using images (Porter, Krafft, and Keegan 2020). One thing we *don’t* know is how Wikipedia’s editors decide what is ‘current’. How does Wikipedia distinguish the past from the present at the very threshold of time? How does it resolve the contradiction between the pastness of the encyclopaedia and the presentness of the current?

Most readers of Wikipedia will have seen what editors do when an article trespasses on the present: mark it with one of the available [Current Event Templates](https://en.wikipedia.org/wiki/Wikipedia:Current_event_templates). The main template is [Template:Current](https://en.wikipedia.org/wiki/Template:Current), which at the time of writing is available on 115 language editions of Wikipedia. When the template is added to an article, a familiar banner appears at the top of the page , and the article is automatically added to [Category:Current Events](https://en.wikipedia.org/wiki/Category:Current_events) or a related category. Each language edition has its own distinct version of Template:Current, and may also sport a range of related Templates. French Wikipedia, for instance, distinguishes ‘Événements en cours’ [ongoing events] from ‘Événements récents’ [recent events] in its [main template](https://fr.wikipedia.org/wiki/Mod%C3%A8le:%C3%89v%C3%A9nement_en_cours), and provides several related templates such as [Modèle:Bataille en cours](https://fr.wikipedia.org/wiki/Mod%C3%A8le:Bataille_en_cours) [Template:Ongoing battle] and [Modèle:Mort récente](https://fr.wikipedia.org/wiki/Mod%C3%A8le:Mort_r%C3%A9cente) [Template:Recent death]. German Wikipedia, by contrast, has only a single [Current Events template](https://de.wikipedia.org/wiki/Vorlage:Laufendes_Ereignis), but it is customisable, so that editors can replace the phrase “aktuelles Ereignis” [current event] in the banner with a more specific description such as “die derzeitige Sportveranstaltung” [the ongoing sporting event].



Template:Current in Italian, Hindi, English, French, Korean and German (as of 14 March 2023)

On the surface, Template:Current might seem like a simple phenomenon. Editors mark an article when it is “current,” then remove the template when its currency is ended. Avieson (2019) likens Template:Current to the “live” icon on a newspaper blog or television news. While Template:Current is present, the article functions as live coverage; when the template is removed, the article gradually becomes encyclopaedic.

But the use of Template:Current is not simple, and has vexed Wikipedia’s editors for years. Avieson (2019) herself grapples with the problem. Although she argues for a distinction between “news” and “encyclopaedias,” she also observes that Wikipedia’s coverage of current events “blurs the boundaries of both news and temporality.” These blurred boundaries are a problem for Wikipedia’s editors, and editors in different langauges have clarified the distinction between past, present and future in different ways. In this context, English Wikipedia is extremist. Unlike other language editions, English Wikipedia strictly polices the use of Template:Current with a bot: [Yapperbot/uncurrenter](https://en.wikipedia.org/wiki/User:Yapperbot). Yapperbot/uncurrenter scans English Wikipedia hourly, examining every article that includes Template:Current and deleting the template if the article has not been edited in the last five hours. English Wikipedia is also one of the few larger language editions without a [Template:Future](https://wikipedia.org/wiki/Template:Future) to mark events that have yet to occur. English Wikipedia deleted Template:Future in 2009 after an official process, and several attempts to resurrect the template have foundered. Meanwhile French, Italian, Bengali, Chinese and 51 other-language Wikipedias maintain a Template:Future.

Why does practice vary across the different language editions? What led to the extremely strict approach of English Wikipedia, in which Template:Current is ruthlessly policed by an artificial agent, and the future is not explicitly marked? What can this tell us about Wikipedia’s “temporal regime” (Assmann 2020)?

We try to answer these questions by focussing on Yapperbot/uncurrenter. We describe the history of Template:Current and recount the debates that led to the bot’s creation. We then examine Yapperbot/uncurrenter’s contributions to English Wikipedia, comparing its practice with the practice of human editors on English Wikipedia and other-language Wikipedias. As many scholars have observed, bots are powerful actors in Wikipedia’s “sociotechnical system”, and account for a large share of contributions (Niederer and Dijck 2010; Dijck 2013, 137–40; R. Stuart Geiger and Ribes 2010; R. Stuart Geiger and Halfaker 2013, 2017; Halfaker and Riedl 2012; Livingstone 2016). Bots are also culturally significant. In a series of pathbreaking papers, Stuart Geiger has demonstrated how bots enact or incarnate Wikipedia’s culture (R. Stuart Geiger 2009, 2011, 2013; R. Stuart Geiger 2017; see also Kennedy 2010). As he explains, it is not sufficient to read a bot’s source code, although there may well be important policies, procedures or ideals “encoded” in the source (R. Stuart Geiger 2017, 9). To understand bots, it is essential to observe how they act in the wild, and to observe how human users and other bots interact with *them* (R. Stuart Geiger 2011; R. Stuart Geiger 2017). In that spirit, we pursue Yapperbot/uncurrenter through Wikipedia, to see how and when it consigns articles to history. Does it solve the problems identified by the editors who summoned it into existence? And what were those problems anyway?

# Methods

## Rows: 1648 Columns: 10  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (3): user, title, comment  
## dbl (5): userid, pageid, revid, parentid, ns  
## lgl (1): texthidden  
## dttm (1): timestamp  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# Results

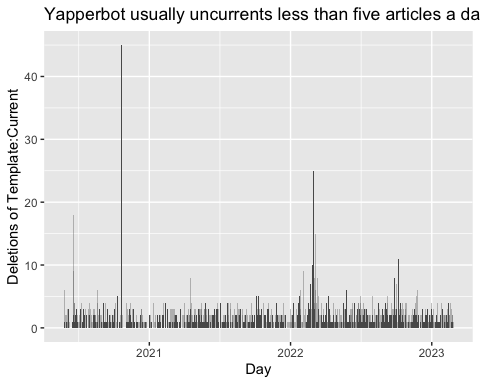
## Document analysis

*How did `Yapperbot/uncurrenter’ come about? What were the debates and discussions of the editors? What was the perceived problem the bot was supposed to fix?*

## Quantitative analysis

*What does the bot actually do? How does that compare with what it is supposed to do?*

How many pages is yapperbot uncurrenting per day?



What was happening on that day it uncurrented 45 pages?

## # A tibble: 45 × 3  
## pageid title timestamp   
## <dbl> <chr> <dttm>   
## 1 63362621 COVID-19 pandemic in Ethiopia 2020-10-21 23:00:28  
## 2 63431783 COVID-19 pandemic in Northern Ireland 2020-10-21 22:00:48  
## 3 63181042 COVID-19 pandemic in Europe 2020-10-21 22:00:41  
## 4 63178596 COVID-19 pandemic in Hong Kong 2020-10-21 22:00:38  
## 5 63313047 COVID-19 pandemic in Moldova 2020-10-21 21:00:43  
## 6 63239190 COVID-19 pandemic in Sweden 2020-10-21 21:00:37  
## 7 64307024 Timeline of the COVID-19 pandemic in October 20… 2020-10-21 20:00:49  
## 8 63395521 COVID-19 pandemic in Sarawak 2020-10-21 20:00:45  
## 9 63391509 COVID-19 pandemic in Quebec 2020-10-21 20:00:39  
## 10 63389195 COVID-19 pandemic in Sabah 2020-10-21 20:00:36  
## # … with 35 more rows

How often does Yapperbot have to remove the template more than once?

## # A tibble: 7 × 2  
## count num\_pages  
## <int> <int>  
## 1 1 1236  
## 2 2 133  
## 3 3 23  
## 4 4 9  
## 5 5 4  
## 6 6 2  
## 7 9 1

Which page had to be uncurrented nine times?

## `summarise()` has grouped output by 'pageid'. You can override using the  
## `.groups` argument.

## # A tibble: 7 × 3  
## # Groups: pageid [7]  
## pageid title count  
## <dbl> <chr> <int>  
## 1 70168267 Siege of Mariupol 9  
## 2 65760352 Tigray War 6  
## 3 70161957 Siege of Chernihiv 6  
## 4 64399515 2020–2021 Belarusian protests 5  
## 5 70157964 Timeline of the 2022 Russian invasion of Ukraine 5  
## 6 70160923 Battle of Kharkiv (2022) 5  
## 7 70809573 2022–2023 monkeypox outbreak 5

# Discussion: What is distinctive about Wikipedia time?

*Compared to other temporal regimes*

Assmann, Aleida. 2020. *Is Time Out of Joint?: On the Rise and Fall of the Modern Time Regime*. Translated by Sarah Clift. Ithaca, UNITED STATES: Cornell University Press. <http://ebookcentral.proquest.com/lib/uts/detail.action?docID=5964894>.

Avieson, Bunty. 2019. “Breaking News on Wikipedia: Collaborating, Collating and Competing.” *First Monday*, April. <https://doi.org/10.5210/fm.v24i5.9530>.

Dijck, José van. 2013. *The Culture of Connectivity: A Critical History of Social Media*. Oxford ; New York: Oxford University Press.

Ferron, Michela, and Paolo Massa. 2011a. “The Arab Spring Wikirevolutions: Wikipedia as a Lens for Studying the Real-Time Formation of Collective Memories of Revolutions.” *International Journal of Communication* 5: 20.

———. 2011b. “Collective Memory Building in Wikipedia: The Case of North African Uprisings.” In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, 114–23. WikiSym ’11. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2038558.2038578>.

———. 2014. “Beyond the Encyclopedia: Collective Memories in Wikipedia.” *Memory Studies* 7 (1): 22–45.

Ford, Heather. 2020. “Rise of the Underdog.” In *Wikipedia@20: Stories of an Incomplete Revolution*, edited by Joseph Reagle and Jackie Koerner, 189–201. Cambridge, Mass: MIT Press. <https://direct.mit.edu/books/book/4956/chapter/1879827/Rise-of-the-Underdog>.

———. 2022. *Writing the Revolution: Wikipedia and the Survival of Facts in the Digital Age*. Cambridge, Massachusetts: The MIT Press.

Geiger, R Stuart. 2017. “Beyond Opening up the Black Box: Investigating the Role of Algorithmic Systems in Wikipedian Organizational Culture.” *Big Data & Society* 4 (2): 205395171773073. <https://doi.org/10.1177/2053951717730735>.

Geiger, R. Stuart. 2009. “The Social Roles of Bots and Assisted Editing Programs.” In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, 1–2.

———. 2011. “The Lives of Bots.” In *Critical Point of View: A Wikipedia Reader*, edited by Geert Lovink and Nathaniel Tkacz, 78–93. Institute of Network Cultures: Amsterdam.

———. 2013. “Are Computers Merely" Supporting" Cooperative Work: Towards an Ethnography of Bot Development.” In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, 51–56.

Geiger, R. Stuart, and Aaron Halfaker. 2013. “When the Levee Breaks: Without Bots, What Happens to Wikipedia’s Quality Control Processes?” In *Proceedings of the 9th International Symposium on Open Collaboration*, 1–6.

———. 2017. “Operationalizing Conflict and Cooperation Between Automated Software Agents in Wikipedia: A Replication and Expansion of’even Good Bots Fight’.” *Proceedings of the ACM on Human-Computer Interaction* 1: 1–33.

Geiger, R. Stuart, and David Ribes. 2010. “The Work of Sustaining Order in Wikipedia: The Banning of a Vandal.” In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, 117–26.

Halfaker, Aaron, and John Riedl. 2012. “Bots and Cyborgs: Wikipedia’s Immune System.” *Computer* 45 (3): 79–82. <https://doi.org/10.1109/MC.2012.82>.

Keegan, Brian. 2019. “An Encyclopedia with Breaking News.” In *Wikipedia@ 20: Stories of an Incomplete Revolution*, edited by Joseph Reagle and Jackie Koerner, 55–70. Cambridge, Mass: MIT Press. <https://doi.org/10.7551/mitpress/12366.003.0007>.

Keegan, Brian, Darren Gergle, and Noshir Contractor. 2011. “Hot Off the Wiki: Dynamics, Practices, and Structures in Wikipedia’s Coverage of the Tōhoku Catastrophes.” In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, 105–13. Mountain View California: ACM. <https://doi.org/10.1145/2038558.2038577>.

Kennedy, Krista. 2010. “Textual Machinery: Authorial Agency and Bot-Written Texts in Wikipedia.” In *The Responsibilities of Rhetoric*, edited by Michelle Smith and Barbara Warnick, 303–9. Long Grove, Illinois: Waveland. <https://surface.syr.edu/wp/1>.

Livingstone, Randall M. 2016. “Population Automation: An Interview with Wikipedia Bot Pioneer Ram-Man.” *First Monday*, January. <https://doi.org/10.5210/fm.v21i1.6027>.

Niederer, Sabine, and José van Dijck. 2010. “Wisdom of the Crowd or Technicity of Content? Wikipedia as a Sociotechnical System.” *New Media & Society* 12 (8): 1368–87. <https://doi.org/10.1177/1461444810365297>.

Pentzold, Christian. 2009. “Fixing the Floating Gap: The Online Encyclopaedia Wikipedia as a Global Memory Place.” *Memory Studies* 2 (2): 255–72. <https://doi.org/10.1177/1750698008102055>.

Porter, Emily, P. M. Krafft, and Brian Keegan. 2020. “Visual Narratives and Collective Memory Across Peer-Produced Accounts of Contested Sociopolitical Events.” *ACM Transactions on Social Computing* 3 (1): 4:1–20. <https://doi.org/10.1145/3373147>.

Rosenzweig, Roy. 2006. “Can History Be Open Source? Wikipedia and the Future of the Past.” *The Journal of American History* 93 (1): 117–46. <https://doi.org/10.2307/4486062>.

Twyman, Marlon, Brian C. Keegan, and Aaron Shaw. 2017. “Black Lives Matter in Wikipedia: Collaboration and Collective Memory Around Online Social Movements.” In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1400–1412. <https://doi.org/10.1145/2998181.2998232>.