



# Theory, Fact or Fake? Global Wikipedia Research Social

Led by: J. Nathan Matias, Sarah Gilbert & Aparajita Bhandari  
August 14, 9:40 UTC

WIKIMANIA



WIKIMANIA



**J. Nathan Matias**  
Assistant Professor at Cornell University & CAT Lab founder



**Sarah Gilbert**  
Research Manager at CAT Lab

**Aparajita Bhandari**  
PhD Candidate at Cornell University



Citizens and  
Technology  
Lab

[citizensandtech.org](http://citizensandtech.org)

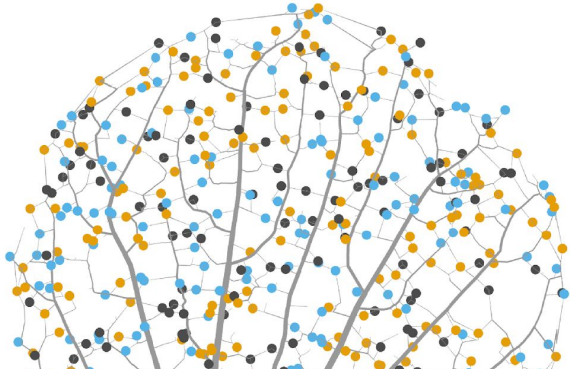


College of Agriculture  
and Life Sciences  

---

Communication

# CAT Lab works alongside communities to test ideas for improving digital life



## Research & Evidence

Organizing communities of millions in **actionable research** that also **grows science**



## Organizing & Education

**Co-designing research** together with affected communities

```
tblch('praw.Reddit', autospec=True)
test_initialize_experiment(mock_reddit):
    r = mock_reddit.return_value
    patch('praw.')

    experiment_name_to_controller = {
        "sticky_comment_0": AMASlickyCommentExperimentController,
        "sticky_comment_frontend_test": FrontPageStickyCommentExperimentController
    }

    for experiment_name in experiment_name_to_controller:
        with open(os.path.join(BASE_DIR, "config", "experiments", experiment_name + ".yaml"), "r") as f:
            experiment_config = yaml.load(f)['test']

        assert(len(db_session.query(Experiment).all()) == 0)

        controller = experiment_name_to_controller[experiment_name]
        controller_instance = controller(experiment_name, db_session, r, log)

        assert(len(db_session.query(Experiment).all()) == 1)
        experiment = db_session.query(Experiment).first()
        assert(experiment.name == experiment_name)
        assert(experiment.controller == experiment_config['controller'])
        assert(pytz.timezone("UTC").localize(experiment.start_time) == parser.parse(experiment_config['start_time']))
        assert(pytz.timezone("UTC").localize(experiment.end_time) == parser.parse(experiment_config['end_time']))

        settings = json.loads(experiment.settings_json)
        for k in ['username', 'subreddit', 'subreddit_id', 'max_eligibility_age',
                'min_eligibility_age', 'start_time', 'end_time', 'controller']:
            assert(settings[k] == experiment_config[k])

    ## NOW TEST THAT AWL OBJECTS ARE ADDED
    for condition_name in experiment_config['conditions']:
        with open(os.path.join(BASE_DIR, "config", "experiments",
            experiment_config['conditions'][condition_name]['randomizations']), "r") as f:
            conditions = []
            for row in Csv.DictReader(f):
                conditions.append(row)
```

## Methods & Software

**Inventing & Maintaining** secure, scalable systems to collect data and intervene with consent online

[citizensandtech.org/research/](https://citizensandtech.org/research/)

# CAT Lab **works alongside communities to** **test ideas for improving digital life**

## **Testing the effects of thanks** **(DE, PL, AR, FA Wikipedias)**

- Does receiving thanks influence how people contribute to Wikipedia
- Does sending thanks influence how editors feel about contributing?

## **Broadening Digital Inclusion** **(WikiLovesAfrica)**

- What is the effect of outreach and recruitment messaging campaigns for WikiLovesAfrica?

## **Welcoming newcomers** **FR & DE Wikipedias**

- What are the effects of welcome messages and mentorship offers on French Wikipedias?
- What are the effects of showing newcomers information about how many people viewed their contributions?

## Instructions

1. We will assign you to a room
2. Get to know each other: "If someone visited your region, what food would you serve them?"
3. **Game Time:** The google form will show you 10 series of 3 statements:
  - a. 2 statements about Wikipedia established through research
  - b. 1 statement about Wikipedia that research has not yet uncovered
  - c. **If someone in your group is vision impaired, please read aloud**
4. You will have 25 minutes for your group to guess which is "unknown"
5. At the end we will review answers
6. Everyone wins by meeting people, learning, and imagining research!

## How to play:

### Two Knowns & an Unknown

#### Form:

[bit.ly/wikimania-trivia-2022](https://bit.ly/wikimania-trivia-2022)

#### PDF:

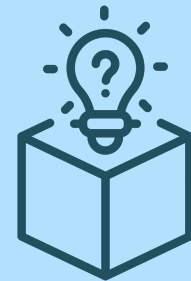
[bit.ly/wikimania-trivia-2022-pdf](https://bit.ly/wikimania-trivia-2022-pdf)

## An Example

Receiving thanks from editors on multiple language Wikipedias leads volunteers to express more thanks to other Wikipedians as well.

Wikipedia contributors who spend more time removing vandalism also report spending more time mentoring newcomers, across multiple language Wikipedias.

Adding an image to Wikipedia biography will increase the views that the article receives.



**See you in 25 minutes!**

**Form: [bit.ly/wikimania-trivia-2022](https://bit.ly/wikimania-trivia-2022)**

**PDF: [bit.ly/wikimania-trivia-2022-pdf](https://bit.ly/wikimania-trivia-2022-pdf)**

**WIKIMANIA**

1.

We can reliably identify the gender of registered Wiki users based on profile information. (Minguillón et al., 2021)

The reduction of mobility during COVID increased the volume of people seeking information on Wikipedia. However, once mobility returned to normal, the volume also returned to normal, but the kind of information people looked for did not. (Ribeiro et al., 2021).

Contributions from Tor users that slip through detection are similar to contributions from unregistered and new editors (Tran, 2019).





2.

Wikipedia editors who spend more time monitoring Wikipedia for damaging content feel more emotionally drained but also more positive about their contributions compared to those who do less monitoring work. (Matias et al., 2020)

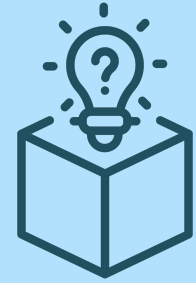
Requiring account registration to edit Wikia wikis reduced both the number of low-quality edits and the number of high quality edits, causing an overall decrease in quality, across 136 communities. (Hill & Shaw, 2021)

Data scientists have created a single reliable measure of information inequality on Wikipedia that works across languages and cultures. This measure can be used to monitor and improve how well Wiki Projects represent the world's knowledge. (Beytía, 2020; Zia et al 2019)



3.

As of 2021, the Wikimedia movement has found effective ways to increase the retention of editors from Asia, Africa, and Latin America year on year (Community Insights, 2021)



Edit-a-thon organizers have motivations beyond closing gaps in Wikipedia. For example, they also hope Edit-a-thons can build information literacy and foster community outside of Wikipedia. (March & Dasgupta, 2020).

While female music artists are underrepresented in the music industry, males are underrepresented on Wikipedia. (Wang et al., 2021)

4.

While volunteer contributions to Wikipedia increased during COVID-19, larger Wikipedias grew more than small Wikipedias, which largely stayed at the same levels of participation. (Ruprechter et al., 2021)

Financial support intended to incentivize contribution to Wikipedia does not lead to active participation across all Wikipedias. (Khatri et al., 2022)

Over the last decade, the coverage gap on Wikipedia between Europe and Africa, has reduced by roughly five times. (Dittus & Graham, 2022)



5.

Sociocultural norms impact Indian women's ability to contribute to Wikipedia in ways that have not been identified in other language Wikipedias. (Chakraborty & Hussain, 2022)

Recommender algorithms that suggest priorities to editors based on WikiData, are an effective way to fill knowledge gaps across Wikipedia. (Redi et al 2021)

A person's biography is more likely to be available in languages common to the person's nationality, ethnicity, and background. (Field et al., 2022)



6.

Wikipedia is used to satisfy a variety of motivations, from looking up a topic that was referenced in media or conversation to wanting to learn something, with no dominant individual motivation. (Singer et al., 2017)

Biography pages of transgender women and non-binary people tend to be longer and available in more languages than comparison articles, indicating a possible glass ceiling effect in which there is a higher bar for transgender women and non-binary people to have a Wikipedia article. (Field & March, 2022).

Edit-a-thon organizers across language Wikipedias have motivations beyond closing gaps in Wikipedia. For example, they also hope Edit-a-thons can build information literacy and foster community outside of Wikipedia. (March & Dasgupta, 2020)



7.

Globally, women are underrepresented as readers of Wikipedia. (Johnson et al., 2021).

On the weekends and late at night, Wikipedia readers are more likely to be led to Wikipedia by media coverage; on Fridays and Saturdays, they are more likely led by conversations. (Singer et al., 2017)

Data on behavior from Wikipedia can provide a reliable, indicator across languages of conflict and edit wars, including conflicts between bots. (Geiger & Halfaker, 2017)

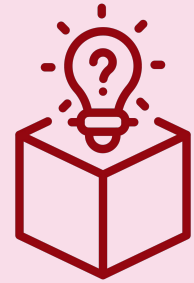


8.

Wikipedians and social scientists have found reliable ways to increase the visibility of notable women in Google results and the Wikipedia's link structure by adding new biographies. (Langrock & González-Bailón, 2022)

Biographies about women who meet Wikipedia's criteria for inclusion are more frequently considered non-notable and nominated for deletion compared to men's biographies. (Tripodi, 2021)

Newcomers feelings of empowerment have increased since 2019. Especially among women and newcomers in East Asia. (Community Insights, 2021)

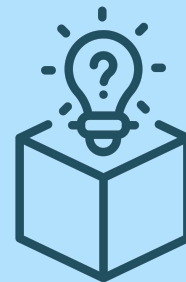


9.

Most reported harassment on Wikipedia is perpetuated by multiple people over time (vs harassment perpetuated by one person once).  
(Wikimedia Harassment Report, 2015)

Wikipedia use among urban and rural users is similar.(Redi et al, 2021)

Most Wikidata items propagate to only a few language editions.  
(Valentim et al., 2021)





10.

There is a link between users' circumstances and how they use Wikipedia. For example, people who use Wikipedia for work or school tend to access articles directly and spend more time reading, while those satisfying boredom use internal links and spend less time reading. (Singer et al 2017)

Showing newcomers view counts on articles they edited causes them to see the value of their edits and contribute more to Wikipedia. (Growth Team 2022 - contact them if you want to help!)

Adding a trust meter to Wikipedia pages influences to trust individual articles in ways that match their actual reliability— lower trust for unreliable articles and higher trust for reliable ones. (Kuznetsov et al 2022)



# THANK YOU!

# WIKIMANIA

- Nominate a team VIP in the chat
- Expect a blog post soon!
- Add errors or research we missed to the form
- Connecting with researchers:
  - [wiki-research-l@lists.wikimedia.org](mailto:wiki-research-l@lists.wikimedia.org)
  - [www.mediawiki.org/wiki/Wiki media Research/Showcase](http://www.mediawiki.org/wiki/Wiki_media_Research/Showcase)
- Many thanks to everyone who submitted ideas for quiz answers!



[citizensandtech.org](http://citizensandtech.org)

# “Knowns” References

- Chakraborty, A., & Hussain, N. (2022). Documenting the gender gap in Indian Wikipedia communities: Findings from a qualitative pilot study. First Monday.
- Community Insights 2021 Report. [https://meta.wikimedia.org/wiki/Community\\_Insights/Community\\_Insights\\_2021\\_Report](https://meta.wikimedia.org/wiki/Community_Insights/Community_Insights_2021_Report)
- Dittus, M., Graham, M (2022) The language geography of Wikipedia. Oxford Internet Institute and Whose Knowledge?
- Field, A., Park, C. Y., Lin, K. Z., & Tsvetkov, Y. (2022, April). Controlled analyses of social biases in wikipedia bios. In Proceedings of the ACM Web Conference 2022 (pp. 2624-2635).
- Hill, B. M., & Shaw, A. (2021). The Hidden Costs of Requiring Accounts: Quasi-Experimental Evidence From Peer Production. *Communication Research*, 48(6), 771-795.
- Johnson, I., Lemmerich, F., Sáez-Trumper, D., West, R., Strohmaier, M., & Zia, L. (2021, May). Global gender differences in Wikipedia readership. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 15, pp. 254-265).
- Kuznetsov, A., Novotny, M., Klein, J., Saez-Trumper, D., & Kittur, A. (2022, April). Templates and Trust-o-meters: Towards a widely deployable indicator of trust in Wikipedia. In CHI Conference on Human Factors in Computing Systems (pp. 1-17).
- March, L., & Dasgupta, S. (2020). Wikipedia Edit-a-thons as Sites of Public Pedagogy. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 1-26.
- Matias, J. N., Al-Kashif, R., Kamin, J., Klein, M., & Pennington, E. Volunteers Thanked Thousands of Wikipedia Editors to Learn the Effects of Receiving Thanks.
- Ribeiro, M. H., Gligoric, K., Peyrard, M., Lemmerich, F., Strohmaier, M., & West, R. (2021, April). Sudden Attention Shifts on Wikipedia During the COVID-19 Crisis. In ICWSM (pp. 208-219)
- Rodolfo Valentim, Giovanni Comarela, Sounel Park, Diego Saez-Trumper. 2021. Tracking Knowledge Propagation Across Wikipedia Languages. Proceedings of the Fifteenth International AAAI Conference on Web and Social Media (ICWSM '21). Dataset.
- Ruprecht, T., Horta Ribeiro, M., Santos, T., Lemmerich, F., Strohmaier, M., West, R., & Helic, D. (2021). [Volunteer contributions to Wikipedia increased during COVID-19 mobility restrictions](#). *Scientific reports*, 11(1), 1-12.
- Singer, P., Lemmerich, F., West, R., Zia, L., Wulczyn, E., Strohmaier, M., & Leskovec, J. (2017, April). Why we read Wikipedia. In Proceedings of the 26th international conference on world wide web (pp. 1591-1600).
- Support and Safety Team. Harassment Survey 2015 Results Report. [https://upload.wikimedia.org/wikipedia/commons/5/52/Harassment\\_Survey\\_2015\\_-\\_Results\\_Report.pdf](https://upload.wikimedia.org/wikipedia/commons/5/52/Harassment_Survey_2015_-_Results_Report.pdf)
- Tran, C., Champion, K., Forte, A., Hill, B. M., & Greenstadt, R. (2019). Tor users contributing to Wikipedia: Just like everybody else. arXiv preprint arXiv:1904.04324.
- Tripodi, F. (2021). Ms. Categorized: Gender, notability, and inequality on Wikipedia. *New Media & Society*, 14614448211023772.
- Wang, A., Pappu, A., & Cramer, H. (2021, May). Representation of Music Creators on Wikipedia, Differences in Gender and Genre. In ICWSM (pp. 764-775).

# “Unknowns” References

- Beytfa, P. (2020, April). The positioning matters: Estimating geographical bias in the multilingual record of biographies on wikipedia. In Companion Proceedings of the Web Conference 2020 (pp. 806-810).
- Community Insights 2021 Report.  
[https://meta.wikimedia.org/wiki/Community\\_Insights/Community\\_Insights\\_2021\\_Report](https://meta.wikimedia.org/wiki/Community_Insights/Community_Insights_2021_Report)
- Geiger, R. S., & Halfaker, A. (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(CSCW), 1-33.
- Khatri, S., Shaw, A., Dasgupta, S., & Hill, B. M. (2022, April). The social embeddedness of peer production: A comparative qualitative analysis of three Indian language Wikipedia editions. In CHI Conference on Human Factors in Computing Systems (pp. 1-18).
- Langrock, I., & González-Bailón, S. (2022). The Gender Divide in Wikipedia: Quantifying and Assessing the Impact of Two Feminist Interventions. Journal of Communication, 72(3), 297-321.
- March, L., & Dasgupta, S. (2020). Wikipedia Edit-a-thons as Sites of Public Pedagogy. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2), 1-26.
- MediaWiki. Growth/Growth Team Updates. [https://www.mediawiki.org/wiki/Growth/Growth\\_team\\_updates](https://www.mediawiki.org/wiki/Growth/Growth_team_updates)
- Minguillón, J., Meneses, J., Aibar, E., Ferran-Ferrer, N., & Fàbregues, S. (2021). Exploring the gender gap in the Spanish Wikipedia: Differences in engagement and editing practices. PLoS one, 16(2), e0246702.
- Redi, M., Gerlach, M., Johnson, I., Morgan, J., & Zia, L. (2020). A taxonomy of knowledge gaps for wikimedia projects (second draft). arXiv preprint arXiv:2008.12314.