

University at Albany, State University of New York

## Scholars Archive

---

Philosophy Faculty Scholarship

Philosophy

---

9-4-2023

### Early response to false claims in Wikipedia, 15 years later

P.D. Magnus

*University at Albany, State University of New York, pmagnus@albany.edu*

The University at Albany community has made this article openly available.

**Please share** how this access benefits you.

Follow this and additional works at: [https://scholarsarchive.library.albany.edu/cas\\_philosophy\\_scholar](https://scholarsarchive.library.albany.edu/cas_philosophy_scholar)



Part of the [Communication Technology and New Media Commons](#), [Epistemology Commons](#), and the [Information Literacy Commons](#)

---

#### Recommended Citation

Early response to false claims in Wikipedia, 15 years later by P.D. Magnus. *First Monday*, Volume 28, Number 9 - 4 September 2023 <https://firstmonday.org/ojs/index.php/fm/article/download/12912/11332>  
doi: <https://dx.doi.org/10.5210/fm.v28i9.12912>



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 4.0 International License](#).

#### Rights Statement



#### License

This Article is brought to you for free and open access by the Philosophy at Scholars Archive. It has been accepted for inclusion in Philosophy Faculty Scholarship by an authorized administrator of Scholars Archive. Please see [Terms of Use](#). For more information, please contact [scholarsarchive@albany.edu](mailto:scholarsarchive@albany.edu).

# Early response to false claims in Wikipedia, 15 years later

by P.D. Magnus

---

## Abstract

Fifteen years ago, I conducted a small study testing the error-correction tendency of Wikipedia. Not only is Wikipedia different now than it was then, the community that maintains it is different. Despite the crudity of that study's methods, it is natural to wonder what the result would be now. So I repeated the earlier study and found surprisingly similar results.

## Contents

[Introduction](#)

[Methods](#)

[Results](#)

[Conclusions](#)

---

## Introduction

In 2008, I published a short paper in this journal reporting the results of research that I had conducted the year before (Magnus, 2008). I inserted fibs into Wikipedia entries about dead philosophers. Each was a one or two sentence falsehood about that philosopher's biography, written to be implausible but not incoherent. For each fib, I observed whether it was removed or flagged as needing citation in the 48 hours after I inserted it. If neither of these things happened, then I removed the fib myself.

The raw percentage of fibs which were removed or marked was 50 percent (18/36). However, some of these corrections were due to *association effects* — that is, a Wikipedia user correcting one fib noticed others which I had added from the same IP address at the same time, so they went on to correct the others as well. My method in the study was to use a range of different IP addresses, but I inserted fibs in three different articles from each location. As a result of association effects, one conspicuous fib could carry two others down with it [1]. Correcting for these effects, 36 percent (10/28) of fibs were removed or flagged within 48 hours.

To put it in rough terms: About a third to a half of the fibs were removed or flagged within a 48-hour window.

When I discussed the result, many people found that rate surprisingly high. Although they might have been extrapolating (wrongly) that every 48-hour period would have that kind of correction rate, some just found it to be better than they would have expected.

Of course, I am not the only person to have inserted false claims as a way of testing Wikipedia. In most cases, like the much discussed example of Halavais (2004), the changes have been unsystematic. It is common in discussion threads for someone to comment that they have put a false claim or two of their own into Wikipedia just to see what would happen. In 2015, Gregory Kohs mounted a more systematic campaign. He made each edit from a different IP address, so there were no association effects. His result was that in 63 percent of cases the phony information was not corrected (Kohs, 2015). The flipside, of course, is that they were corrected in 37 percent of cases — almost the same as the (adjusted) result in Magnus (2008). This is an odd coincidence rather than a robust finding. Methodologically, Kohs’ “months-long experiment” was kind of a mess. His phony claims were inserted in all different kinds of articles. The insertions were of highly variable lengths, ranging from just changing a number or a name to adding a whole paragraph-length subsection. He let them stand for an arbitrary, inconsistent amount of time.

---

## Methods

Between July and November 2022, I made 33 changes to Wikipedia: one at a time, anonymously, and from various IP addresses. The changes were not given edit summaries. When an IP address was reused, it was only after considerable delay. Each change consisted of a one or two sentence fib inserted into the Wikipedia entry on a notable, deceased philosopher. The fibs were about biographical or factual matters, rather than philosophical content or interpretive questions. Although some of the fibs mention “sources”, no citations were provided. If the fibs were not corrected within 48 hours, they were removed by the experimenter.

The fibs were all, *verbatim*, ones that I used in Magnus (2008). Two of the original fibs had to be excluded because the target articles were, at the time of the present study, *protected* or *semi-protected*; anonymous changes are not allowed in protected articles [2]. Another was excluded because the target article now contains details which explicitly contradict the fib; as such, the fib could not be sensibly inserted in the current article. From the 36 fibs used in the original study, this left 33 — all of which I used this time (See [Table 2](#)).

---

## Results

Thirty-six percent (12/33) of changes were corrected within 48 hours. Rounded to the nearest percentage point, this is the same as the adjusted result in Magnus (2008) [3]. See [Table 1](#).

**Table 1: Results for fibs inserted as part of the study, sorted alphabetically by philosopher. Results from 2007 and 2022 are given as time to correction, in HH:MM format; “—” indicates that the fib was not corrected within 48 hours. The IP address column indicates when a particular IP address was reused for this study;**

<b>where no value is given, the IP address was used just once.</b>			
<b>Wikipedia article</b>	<b>2007 results</b>	<b>2022 results</b>	<b>IP address</b>
Thomas Aquinas	0:43	0:01	
Corpus Aristotelicum	1:30	—	IP-A
Jeremy Bentham	—	—	
George Berkeley	—	—	IP-B
Boethius	21:09	36:07	
F.H. Bradley	—	—	IP-C
Rudolf Carnap	—	—	
René Descartes	—	—	
T.H. Green	5:01	—	
Norwood Russell Hanson	—	—	
G.W.F. Hegel	7:45	1:17	
Martin Heidegger	5:30	—	
Carl Gustav Hempel	0:44	—	
Heraclitus	0:05	0:59	
David Hume	—	0:21	IP-A
Immanuel Kant	—	0:07	
Søren Kierkegaard	0:02	—	
Gottfried Wilhelm Leibniz	—	—	
Norman Malcolm	5:05	0:01	IP-B
Nicolas Malebranche	45:27	—	
J.M.E. McTaggart	—	5:43	
John Stuart Mill	0:09	2:24	
Michel de Montaigne	0:43	24:58	
G.E. Moore	—	—	IP-B
Friedrich Nietzsche	—	—	
Karl Popper	—	0:01	
A.N. Prior	—	—	IP-C
Thomas Reid	—	—	
Bertrand Russell	—	—	

Gilbert Ryle	—	—	IP-B
George Santayana	5:02	—	
Baruch Spinoza	8:32	3:02	
Ludwig Wittgenstein	2:36	—	

The community of users responding to the fibs is surely different between the two studies. And even if a user from 15 years ago were still contributing Wikipedia, they would be much older now. So one may naturally worry that the result is more about the method than about Wikipedia. To simplify the worry: If about a third of the fibs used in the study were preposterous howlers which anyone could easily detect while the remainder were cunning lies that would slip by, then a correction rate of about a third would be explained more by the structure of the dataset than by features of the community.

<b>Table 2: Fibs used for both Magnus (2008) and for this study. Note, crucially, that none of these are true.</b>	
<b>Wikipedia article</b>	<b>Fib</b>
Thomas Aquinas	In order to highlight the contrast between Christian living and pre-Christian Greek thought, Aquinas encouraged the eating of beans.
Corpus Aristotelicum	There are no surviving editions of Aristotle's <i>Theophrastian ethics</i> , which considered issues in the ethics of animal care. Records indicate that a copy existed as late as the tenth century, in the city of Cordoba.
Jeremy Bentham	As a child, he wrote a series of imaginative dialogues between an unnamed boy and wisdom incarnate in the form of a tiger. These were never published, but reflected the author's early interest in writing and philosophy.
George Berkeley	The <i>Principles</i> consisted of three parts, elaborating consequences for metaphysics, ethics, and aesthetics respectively. Of these, only the first was ever published, and Berkeley's drafts of the second and third parts have not survived.
Boethius	It is known that he lost two fingers on his left hand in a childhood

	accident, although there is no record of how exactly it occurred.
F.H. Bradley	In 1900, Bradley was nearly blinded in a sporting accident. He continued to be philosophically active, but his subsequent works were dictated to an assistant.
Rudolf Carnap	The Vienna Circle was also a tightly-knit social group. They regularly met to play cards, including a bridge-like game of their own devising called <i>Whistenschaft</i> .
René Descartes	While there, Descartes first encountered hermetic mysticism. Although he was briefly a Free Mason, he later abandoned mysticism in favor of reasoned inquiry.
T.H. Green	Green's correspondence, published in 1912, also gives insight into his philosophy. In a letter to Victoria Regina, he suggests that moral perfectibility will allow humans to transcend their limitations within the next century.
Norwood Russell Hanson	Hanson was skilled at sleight-of-hand, and would often entertain dinner party guests with card tricks and other feats of legerdemain.
G.W.F. Hegel	Hegel found the work isolating and drank heavily when not working. While drunk, Hegel ran naked through the foyer of the house while chanting the Lord's Prayer in Latin.
Martin Heidegger	Some of the faculty at Freiburg called him 'Edmund II', a moniker that Heidegger found demeaning.
Carl Gustav Hempel	Hempel was renowned for whittling at departmental colloquia. If he liked the talk, he would give the resulting figure to the guest speaker.
Heraclitus	According to some ancient sources, Heraclitus was mildly hydrophobic and refused to travel by boat. This is connected with the probably apocryphal story that he died by drowning.
	Hume had begun wrestling with local sportsmen in Bristol, and

David Hume	continued the activity in France until a shoulder injury forced him to stop.
Immanuel Kant	Kant's poetry was much admired, and handwritten manuscripts circulated among his friends and associates.
Søren Kierkegaard	As a young boy Kierkegaard was mauled by a wild dog. Although he recovered, some have suggested that the episode prefigures later themes of anxiety and dread.
Gottfried Wilhelm Leibniz	Many of his manuscripts are written in a shorthand of his own invention which uses binary numbers to encode sequences of characters.
Norman Malcolm	He built a greenhouse at his home in Ithaca. He raised orchids, producing several new hybrids including one that bears his name.
Nicolas Malebranche	Malenbranche's tutor, Pierre Gassendi, was himself a notable philosopher, but there is no indication that philosophy was part of the curriculum.
J.M.E. McTaggart	Among McTaggart's many interests was antique collecting. His collection boasted the sword that removed Tycho Brahe's nose.
John Stuart Mill	Following the death of his wife, Mill had a series of mistresses who helped him prepare manuscripts as well as sharing his bed.
Michel de Montaigne	Montaigne had been an avid duellist at Guyenne. During this period of isolation, he carried a rapier with him and would challenge anyone who disrupted his work.
G.E. Moore	His influence outside philosophy includes a reference to him in the signature line of the musical <i>Oliver</i> .
Friedrich Nietzsche	In a letter to Victoria Regina, Nietzsche even entertained the possibility of burning the remaining copies to collect on insurance.
Karl Popper	While there, he lived on a cooperative farm. He later claimed that nothing prepares the mind for philosophy like milking a cow.

A.N. Prior	While at Oxford, Prior wrote a draft of a book on the formal structure of interpersonal awareness. Although he showed parts of the draft to various colleagues, it appears to have been lost.
Thomas Reid	Reid and Hume met once when both were in London, and the former indicated a fireplace poker as an example of a material object which certainly exists.
Bertrand Russell	In the same year, Russell published a volume of poetry under the pseudonym Christian Bellows. The poems primarily addressed humanistic concerns that he later revisited in works such as “Why I am not a Christian.”
Gilbert Ryle	After retiring, Ryle bought a small farm. He tinkered with automated processes to care for livestock, although they never proved to be commercially viable.
George Santayana	He was an avid cyclist and, in 1923, he went on a cycling tour of Italy with the novelist Taylor Caldwell.
Baruch Spinoza	He supplemented his income by selling stolen jewelry that had been smuggled into Holland from France.
Ludwig Wittgenstein	He was twice forced to pay fines for misuse of strychnine, which he used to control squirrels around the garden.

It is possible to address this worry, since the same fibs were used in both studies. If corrections occurred only because of the content of the fibs, then precisely the same fibs should have been corrected in both studies. So we can compare the outcomes for particular fibs, noting whether the two studies got the same outcome (that either the fib was corrected in both studies or not corrected in either study). The number of fibs which led to the same outcome in the two experiments was only 64 percent (21/33). This is within the range of what one would expect if it were simply due to chance [4]. So it looks as if the outcome is not merely due to the plausibility and implausibility of particular fibs themselves.



## Conclusions

Despite the recurrence of an error correction rate of about 36 percent, that specific value should be taken

with a grain of salt. Nevertheless, the original study and this one give a rebuttal to the worry raised by Brian Keegan that “it might be trivial to persistently embed disinformation into provincial articles about distant historical events, specialized scientific topics, or marginal trivia about national anthems that lack sustained editorial oversight” [5]. Adding disinformation is not trivially easy. A few sentences of disinformation added to Wikipedia may persist, but they might be caught relatively quickly more than a third of the time.

Although the numerical result is about the same as for the earlier study, the significance is not. Fifteen years ago, concerns about the reliability of Wikipedia were widespread. It was common to see Wikipedia as posing an epistemological problem: How could we trust a source which might have been written by anyone? [6] The original study was thus a small contribution to answering questions of broader epistemology [7]. In the years that followed, these concerns seem to have faded. Wikipedia has become more central to online life, and whining about it has become less common [8]. Search engines offer information from Wikipedia as a sidebar in search results. When we ask questions to the digital assistants on our phones, often the results come from Wikipedia. We seem to have decided, as Don Fallis noted, that “Wikipedia is not all that bad” [9]. So the current study is just a snapshot of what *not all that bad* means. 

## About the author

P.D. Magnus is a professor in the Department of Philosophy at the University at Albany, State University of New York.

E-mail: pmagnus [at] fecundity [dot] com

## Notes

1. I also made the mistake of including two Featured articles. Because these were carefully tended by editors, they are not really comparable to the rest of the data set.

2. See [https://en.wikipedia.org/wiki/Wikipedia:Protection\\_policy](https://en.wikipedia.org/wiki/Wikipedia:Protection_policy).

3. In Magnus (2008), some of the corrected fibs were marked citation needed rather than being removed. In the current study, all the corrected fibs were removed entirely. This may indicate a change in Wikipedia culture to err on the side of removing a dubious passage rather than merely marking it as dubious.

4. Supposing each fib had an identical and independent probability of being corrected, the expected rate of agreement would be  $54 \pm 17$  percent. This calculation overlooks the fact that some of the results of the earlier study were due to association effects, but there is no way to correct for that.

5. Keegan, 2020, p. 86.

6. Frost-Arnold (2018) provides a nice summary of these debates.

7. I situated the earlier study in relation to the broader epistemic issues in Magnus (2009).

8. There is ongoing criticism of Wikipedia, of course, but the focus has shifted. Rather than wondering how Wikipedia could provide knowledge at all, recent critics have more often asked what knowledge — and whose knowledge — it provides. See *inter alia* Vrana, *et al.* (2020) and McDowell and Vetter (2022).

9. Fallis, 2011, p. 302.

## References

Don Fallis, 2011. “Wikipistemology,” In: Alvin I. Goldman and Dennis Whitcomb (editors). *Social epistemology: Essential readings*. Oxford: Oxford University Press, pp. 297–313.

Karen Frost-Arnold, 2018. “Wikipedia,” In: David Coady and James Chase (editors). *Routledge handbook of applied epistemology*. London: Routledge. pp. 28–40.  
doi: <https://doi.org/10.4324/9781315679099>, accessed 3 September 2023.

Alex Halavais, 2004. “The Isuzu experiment,” *A Thaumaturgical Compendium* (29 August), at <http://alex.halavais.net/index.php?p=794>, accessed 28 December 2022.

Brian Keegan, 2020. “An encyclopedia with breaking news,” In: Joseph Reagle and Jackie Koerner (editors). *Wikipedia @ 20: Stories of an incomplete revolution*. Cambridge, Mass.: MIT Press, pp. 55–70.  
doi: <https://doi.org/10.7551/mitpress/12366.003.0007>, accessed 3 September 2023.

Gregory Kohs, 2015. “Experiment concludes: Most misinformation inserted into Wikipedia may persist,” *Wikipediocracy* (13 April), at <https://wikipediocracy.com/2015/04/13/experiment-concludes-most-misinformation-inserted-into-wikipedia-may-persist/>, accessed 28 December 2022.

P.D. Magnus, 2009. “On trusting *Wikipedia*,” *Episteme*, volume 6, number 1, pp. 74–91.  
doi: <https://doi.org/10.3366/E1742360008000555>, accessed 3 September 2023.

P.D. Magnus, 2008. “Early response to false claims in Wikipedia,” *First Monday*, volume 13, number 9.  
doi: <https://doi.org/10.5210/fm.v13i9.2115>, accessed 3 September 2023.

Zachary J. McDowell and Matthew A. Vetter, 2022. *Wikipedia and the representation of reality*. New York: Routledge.  
doi: <https://doi.org/10.4324/9781003094081>, accessed 3 September 2023.

Adele Godoy Vrana, Anasuya Sengupta, and Siko Bouterse, 2020. “Toward a Wikipedia for and from us all,” In: Joseph Reagle and Jackie Koerner (editors). *Wikipedia @ 20: Stories of an incomplete revolution*. Cambridge, Mass.: MIT Press, pp. 239–257.  
doi: <https://doi.org/10.7551/mitpress/12366.003.0021>, accessed 3 September 2023.

---

## Editorial history

Received 13 April 2023; revised 22 May 2023; accepted 3 September 2023.



This paper is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Early response to false claims in Wikipedia, 15 years later  
by P.D. Magnus.

*First Monday*, volume 28, number 9 (September 2023).

doi: <https://dx.doi.org/10.5210/fm.v28i9.12912>