

University at Albany, State University of New York

## Scholars Archive

---

Philosophy Faculty Scholarship

Philosophy

---

9-2008

### Early response to false claims in Wikipedia

P.D. Magnus

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

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 [Philosophy Commons](#)

---

#### Recommended Citation

Magnus, P.D., "Early response to false claims in Wikipedia" (2008). *Philosophy Faculty Scholarship*. 50.  
[https://scholarsarchive.library.albany.edu/cas\\_philosophy\\_scholar/50](https://scholarsarchive.library.albany.edu/cas_philosophy_scholar/50)

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. For more information, please contact [scholarsarchive@albany.edu](mailto:scholarsarchive@albany.edu).

Early response to false claims in *Wikipedia*

by P.D. Magnus

---

A number of studies have assessed the reliability of entries in *Wikipedia* at specific times. One important difference between *Wikipedia* and traditional media, however, is the dynamic nature of its entries. An entry assessed today might be substantially extended or reworked tomorrow. This study paper assesses the frequency with which small, inaccurate changes are quickly corrected.

## Contents

[Introduction](#)

[Methods](#)

[Results](#)

[Conclusion](#)

---

## Introduction

A number of studies have assessed the reliability of entries in *Wikipedia* at specific times (for example: Giles, 2005; Magnus, 2006; Chesney, 2006). One important difference between *Wikipedia* and traditional media, however, is the dynamic nature of its entries. An entry assessed today might be substantially extended or reworked tomorrow.

Sometimes contributors to *Wikipedia* respond quickly and effectively to inaccurate additions. For example, in 2004 Alex Halavais created a pseudonymous account and inserted 13 false claims in various entries. All of the false claims were deleted within three hours [1].

Halavais' case is just one anecdote, because the 13 changes were not independently corrected. Some other *Wikipedia* user had noticed that Halavais' account was responsible for bad changes and undid them all. Someone even sent a message to the account, discouraging him from any further shenanigans. Call this an "association effect"; all the errors inserted at the same time were corrected, not because they detected independently but because one was detected by a user who checked what other changes had been made by the same person. Some users even describe themselves as being dedicated to "vandal patrol."

Dan Tynan (2008) worries about this process and recounts a similar experiment:

Wikipedians say the encyclopedia ultimately corrects itself, and that might be true. But how long does it take, and what happens meanwhile? As an experiment, I once added a harmless fictional 'fact' to the *Wikipedia* biography of a notable technology executive. Three months and nearly 200 edits later, the bogus sentence was removed.

Of course, Tynan's experiment is just another anecdote. It would be good to have some more systematic sense of early response to erroneous changes in *Wikipedia*. Thus, the present study assesses the frequency with which inaccurate changes are quickly corrected.

Note that the study is not aimed to show that *Wikipedia* is vulnerable to malicious tampering. Deliberate tampering could easily have employed more effective methods: usernames, fabricated citations, and others best left to the reader's imagination. The aim, rather, is to see how effective *Wikipedia* users are at responding quickly to false claims added to otherwise adequate entries. The insertion of false claims is inevitable even without vandalism, because some *Wikipedia* users have false beliefs. They will, in good faith, transcribe these falsities into *Wikipedia*. This study gets at response to such falsities by *Wikipedia* users.

## Methods

One and two sentence fictitious claims (fibs) were introduced into the biographical or factual parts of *Wikipedia* entries about notable, deceased philosophers. For example, of 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." Of 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." [2]

Fibs were inserted three at a time, so as to mitigate association effects. Also, different IP addresses were used for different groups. Each fib was inserted at a plausible point in the entry, but no other changes were made so as to the surrounding sentences. Fibs did not include any hyperlinks. Although some of the fibs mentioned "sources", no citations were provided. Changes were made anonymously and were given default edit summaries.

Each fib was monitored for 48 hours to see whether it was corrected by a *Wikipedia* user. If it was still part of the entry after 48 hours, it was anonymously removed.

Fibs were placed only in a single type of entry (philosophers) because such entries are likely to be maintained by similar communities of users; nevertheless, there is no single group of users maintaining the entries on all philosophers. Fibs were inserted only in entries that were reasonably well-tended, not in entries that were already marked as failing to meet *Wikipedia* quality standards or articles with 'semi-protected' status.

Fibs were about biographical or factual matters, rather than philosophical content or interpretive questions. There are several reasons for this choice: First, *Wikipedia* entries typically have more thorough discussion of philosophers' biographies than about their philosophical views (Bragues, 2007). Second, biographical facts are more clearcut; fibs about philosophical content could be read merely as heterodox interpretations. Third, biographical facts are more likely to mislead even people who have some philosophical acumen.

## Results

Of 36 fibs, 15 were removed within 48 hours. Three others were not removed, but were marked as needing a citation. This is an entirely appropriate response, since changes that are marked as needing citation are typically removed later if no citation is provided.

Fibs were inserted in groups of three. This mitigated but did not entirely remove the influence of association effects. Sometimes a group of fibs would be removed in rapid succession by a single *Wikipedia* user. In such a case, the first fib clearly alerted the user to a problem; the second and third were not noticed in themselves, but were removed merely because they were changes made from the same IP address as the first. To be cautious, then, we might count such a trio only as a single instance of diligence.

Some *Wikipedia* articles are "Featured", meaning that they have been recognized as being of high quality. Featured articles are more closely watched by users on vandal patrol, and *Wikipedia* policy specifically

directs users to be more aggressive in removing dubious claims from them. Unfortunately, two of the fibs were inserted in Featured articles. Since "Featured" articles are importantly different, it would be prudent to not count these two instances.

These adjustments leave 28 fibs, 10 (36 percent) of which were fixed within 48 hours.

Most of the corrections were made within just a few hours. The median is skewed low because the changes were only watched for 48 hours before being fixed regardless, as a matter of method. However, it is interesting to note that only one item was corrected after 24 but before 48 hours.

<b>Table 1: Fibs corrected.</b>		
	raw	adjusted
total items	36	28
total fixed	50% (18)	36% (10)
median response time	2h 5m	5h 13m

Although the study was aimed to mitigate association effects, rather than examine them, the results give some indication about them. Of the 12 groups of fibs, three were corrected one after another by a single user. One of the two groups containing a Featured article was subject to association effects; the other two articles were promptly corrected by the user who corrected the Featured article. The other was not. Removing the two groups that included fibs in Featured articles leaves 10 groups, two of which experienced association effects.

<b>Table 2: Association effects</b>		
	raw	adjusted
total groups	12	10
fixed as a group	25% (3)	20% (2)

## Conclusion

In short: About one third to one half of the fibs were corrected within 48 hours. One fifth to one quarter of the fib groups experienced association effects [3].

There would be little point in trying to refine these results with a larger sample. If the effort became large enough to draw attention from *Wikipedia* users, then the sample as a whole might suffer from association effects. Moreover, different topics and areas of *Wikipedia* are maintained by different portions of the user community. And the very same entries will be maintained by different partially overlapping communities over time. An effort expanded to many more entries would inevitably test the diligence of different subcommunities who would not form a homogenous reference class.

Nevertheless, these results provide something more than anecdotes and can serve as a compliment to assessments of *Wikipedia* entries at-a-time (such as Giles, 2005 and Chesney, 2006) and indirect measures of reliability (such as Nielson, 2007).

## About the author

P.D. Magnus is an Assistant Professor in the Department of Philosophy at the University at Albany, State University of New York.

Web: <http://www.fecundity.com/pmagnus>

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

## Notes

1. Halavais detailed the episode on his blog (Halavais, 2004), and it has been offered as a stock example of attempts to probe *Wikipedia* (Read, 2006).

2. A complete list of the fibs and outcomes is available at <http://hdl.handle.net/1951/43003>.

3. Without unrealistic assumptions about the population of possible fibs, there is no natural way to put error bars around these numbers. Rather, the higher and lower proportions represent the raw and adjusted values, respectively.

## References

George Bragues, 2007. "Wiki-Philosophizing in a Marketplace of Ideas: Evaluating Wikipedia's Entries on Seven Great Minds" (April), available at <http://ssrn.com/>, accessed 27 August 2008.

Thomas Chesney, 2006. "An empirical examination of Wikipedia's credibility," *First Monday*, volume 11, number 11 (November), at [http://firstmonday.org/issues/issue11\\_11/chesney/](http://firstmonday.org/issues/issue11_11/chesney/), accessed 27 August 2008.

Jim Giles, 2005. "Internet encyclopedias go head to head," *Nature*, volume 438, number 7070 (15 December), pp. 900-901, and at <http://www.nature.com/nature/journal/v438/n7070/full/438900a.html>, accessed 27 August 2008.

Alex Halavais, 2004. "The Isuzu Experiment," blog entry at *A Thaumaturgical Compendium* (29 August), at <http://alex.halavais.net/index.php?p=794>, accessed 8 December 2007.

P.D. Magnus, 2006. "Epistemology and the Wikipedia," paper presented at the North American Computing and Philosophy Conference in Troy, New York, at <http://www.fecundity.com/job/wikipedia.pdf>, accessed 27 August 2008.

Brock Read, 2006. "Can Wikipedia Ever Make the Grade?" *Chronicle of Higher Education*, volume 53, number 10 (27 October), p. A31, and at <http://chronicle.com/free/v53/i10/10a03101.htm>, accessed 8 December 2007.

Dan Tynan, 2008. "The truth is out there ... somewhere," *US Airways Magazine* (January), p. 42.

---

## Editorial history

Paper received 7 February 2008; accepted 5 August 2008.

---



"Early response to false claims in Wikipedia" by P.D. Magnus is licensed under a [Creative Commons Attribution 3.0 United States License](#).

Early response to false claims in Wikipedia

by P.D. Magnus

*First Monday*, Volume 13 Number 9 - 1 September 2008

<http://journals.uic.edu/ojs/index.php/fm/rt/printerFriendly/2115/2027>