

Epistemology and the Wikipedia*

P.D. Magnus

August 8, 2006

This is a draft of a paper to be presented at th North American Computing and Philosophy Conference in Troy, New York
Comments are welcome.
e-mail: [pmagnus\(at\)fecundity.com](mailto:pmagnus@fecundity.com)
web: <http://www.fecundity.com/job>

Abstract

Wikipedia is a free encyclopedia that is written and edited entirely by visitors to its website. I argue that we are misled when we think of it in the same epistemic category with traditional general encyclopedias. An empirical assessment of its reliability reveals that it varies widely from topic to topic. So any particular claim found in it cannot be relied on based on its source. I survey some methods that we use in assessing specific claims and argue that the structure of the Wikipedia frustrates them.

Introduction

For many of us, the Internet has become a reference of first resort. If I want to know what the capital of Mongolia is, for example, I open my laptop and have an answer in moments. As a result, assessing the credibility of claims we find on-line has become a regular part of our epistemic practice.

Of course, the internet connects us to sources more far flung than the ones I am connected to off-line. I read the weblog of Greg Costikyan, a game-designer in Manhattan, and learn things about the game industry that I would not have learned before the internet. Nevertheless, the *kinds* of epistemic connections are the same. To consider that same example, Costikyan's blog at costik.com is written by him and it is of the epistemic kind *testimony* or *commentary*. I

*I want to thank Jon Mandle, Ryan Hickerson, and Bonnie Steinbock for helping with the study.

know how to assess such things. If an anonymous blogger comments on a bit of news that he read on another blog... and so on with the news passed from person to person, it is *gossip*. I am dubious of it, just as I would be dubious of gossip off-line. On-line species are different, perhaps, but the epistemic genera are familiar.

I have argued elsewhere that this holds for claims made on homepages, in forums, and in chatrooms [Mag01]. I want to argue here that it does not hold for Wikipedia.

Wikipedia (<http://en.wikipedia.org>), which bills itself as ‘The Free Encyclopedia,’ is constituted entirely by the contributions of visitors to its website. Participants write, rewrite, and discuss the entries. In the words of a recent *New York Times* article, Wikipedia raises “a single nagging epistemological question: Can an article be judged as credible without knowing its author?” [Str06]. The mere fact that “epistemological” issues are raised in the *Times* should pique our interest. The suggestion, I take it, is that Wikipedia is something new. Is it? Or is it just another on-line variant of a familiar off-line epistemic kind?

First: the genus encyclopedia

It may be tempting to give this quick answer: The Wikipedia is a species of the genus *encyclopedia*, and we know how to deal with encyclopedias. It is an instance of an already familiar kind. Although grade school kids might use the Encyclopædia Britannica as their primary source in writing a report, we hope they outgrow it. For an adult, an encyclopedia may serve as a first source, providing orientation in an unfamiliar field, but only as a last source if the information does not need to be especially accurate. This is not meant to stigmatize encyclopedias, but to put them in perspective. So, the tempting reply concludes, the Wikipedia is no different.

Wikipedia’s own guide contains this sort of reasoning,¹ but I want to argue that pigeon-holing the Wikipedia as an encyclopedia overlooks ways in which it is different than Britannica. Perhaps Wikipedia is not even so reliable as a typical encyclopedia, but set that concern aside for a moment; I will return to it in the next section. Several practical differences make the Wikipedia more likely than Britannica to be a first— and last— source.

First, Wikipedia is more readily accessed. General encyclopedias compete with books; once I am already going to a bookshelf or the library, the incremental cost of checking a weightier source is relatively small. Wikipedia only directly competes with other on-line resources. For example, a friend of mine was preparing a lecture in which she was going to briefly discuss pragmatism. She checked the Stanford Encyclopedia of Philosophy (<http://plato.stanford.edu>) first, but the SEP entry on pragmatism was not written yet. Pressed for time, she consulted the Wikipedia. It was *easy*.²

¹“Remember that *any* encyclopedia is a starting point for research, not an ending point” [Wik06].

²Wikipedia says of itself, “Wikipedia is increasingly used by everyone in the academic

Second, we are often led to the Wikipedia even if we do not start there. For many topics, the Wikipedia entry will be on the first page of web search hits. Even if I avoid visiting the Wikipedia directly, I will still encounter content from it. The content of Wikipedia is under a GNU Free Documentation License, and so may be freely reproduced. The About.com encyclopedia, for example, is copied from Wikipedia verbatim.

Third, Wikipedia has a breadth that general encyclopedias do not. For example, it has a brief entry for the Polish philosopher Kazimierz Twardowski; Britannica has none. Because Wikipedia receives new contributions all the time, it also has more information about popular culture and current events than a traditional encyclopedia.³

Second: data and anecdotes

Suppose that I am correct and that Wikipedia is not, in use, a member of the genus encyclopedia. Is it even so reliable?

This is not a question that can be answered in the abstract. The mechanism underlying the Wikipedia, in and of itself, has no intrinsic reliability. If an entry is maintained by informed contributors, without pervasive bias or sabotage, then it will probably have things right. If an entry is written by clueless rubes or hacked out by partisans, then it probably has things wrong.⁴

So the Wikipedia's reliability is a contingent, empirical matter. A recent study in the journal *Nature* [Gil05] compared Wikipedia and Encyclopædia Britannica articles on a great range of scientific topics. The study concludes that the two are not so different. It is hard to perform significance tests on results like these, but this is how the result was summarized:

Only eight serious errors, such as misinterpretations of important concepts, were detected in the pairs of articles reviewed, four from each encyclopaedia. But reviewers also found many factual errors, omissions or misleading statements: 162 and 123 in Wikipedia and Britannica, respectively. [Gil05, p. 900–901]

The editors of Britannica replied to the study [Inc06], arguing over the methodology. The editors of *Nature* replied to Britannica's reply [Nat06].⁵ As they note, there is no reason to think either that the shortcomings of the study are 'fatal' or that they favor Wikipedia over Britannica.

Although the "the difference in accuracy was not particularly great" [Gil05, p. 900], we should not forget that the results *do* favor Britannica over Wikipedia.

community, from first-year students to professors, as the easiest source of information about anything and everything" [Wik06].

³There are some exceptions. As of this writing, the Wikipedia entry for 'abortion' has no substantive discussion of the ethical arguments for and against.

⁴There are also trivial possibilities: If an entry is copied verbatim from another encyclopedia, then it will be exactly as accurate as the source encyclopedia.

⁵Nature has links to all of the relevant articles at <http://www.nature.com/nature/britannica/>.

	Wikipedia			Britannica		
	Inf.	Min.Err.	Maj.Err	Inf.	Min.Err.	Maj.Err
Rawls	6	5	2	7	2	0
Husserl	7	3	6	12	16	4
Bioethics	1	1	0	1	0	0
total	14	9	8	20	18	4

Table 1: Wikipedia versus Britannica on philosophical topics.

The Wikipedia articles, considered altogether, contain 32% more errors than the Britannica articles. There are other differences which are not mentioned in the Nature article but can be calculated from the raw data (provided on the Nature website). In addition to having more errors overall, Wikipedia entries varied more than Britannica entries. Britannica had a mean error per article of 3.0, with a standard deviation of 2.4; Wikipedia had a mean error per article of 3.9, with a standard deviation of 3.5. Wikipedia contained more entries than Britannica with zero errors, but two Wikipedia articles were worse than the worst of Britannica’s.

This wide variability should come as no surprise, since Wikipedia entries rely on contributors. Different entries will attract contributors. Given this wide variability in among Wikipedia’s science entries, it is not clear whether the results of the Nature study generalize to entries about non-scientific topics. I have made a few tests in this regard.

I prepared material from Britannica and Wikipedia so that they could be blindly reviewed by three obliging experts. The topics covered included Rawls’ theory of justice, Husserl and phenomenology, and bioethics. The topics were chosen because Britannica and Wikipedia provide similar depth, and because expert readers were readily available.⁶ The readers were asked to identify infelicities, minor errors, and major errors in the articles. The definitions of each category are provided in an appendix to this paper.

I introduced the category of ‘infelicity’ because it is easy to make philosophical claims that are misleading but which are not, strictly speaking, incorrect—especially when interpreting specific philosophers. Nevertheless, quantifying infelicities proved difficult. The reader in bioethics complains that the article from Wikipedia “doesn’t exactly have errors. It’s just bizarre. . . . The stuff on utilitarian bioethics is pure axe-grinding. It doesn’t seem to have been written by someone in the field.”

The numerical results are summarized in table 1, but they are something of a muddle. Britannica has fewer major errors overall, but more minor errors and infelicities.

Nevertheless, some of the reader comments are informative: The reader on Rawls identified (unbeknownst to him) the Wikipedia entry as “clearly worse”

⁶Often Wikipedia provides depth than Britannica. However, this is not always the case. I meant to include the ethics of abortion in the study, but the Wikipedia article—which was locked because of controversy—had no substantive discussion of philosophical issues.

than material from Britannica. The reader on Husserl wrote that one of the Wikipedia articles was his “favorite one, hands down.” He added, “Now *that’s* the way an encyclopedia entry on phenomenology should be organized.”

Data like this tell us about Wikipedia entries at a particular time, but there have been many changes to the articles that I sent to readers since I collected them in May. Most of these are minor, but some of are substantive; for example, there is a new discussion of Rawl’s *The Law of Peoples* and over 2500 words of new material on Husserl’s connection to Frege, his criticisms of psychologism, and his philosophy of logic. Scoring the old articles does not tell us what to expect from this new material. We need a way of evaluating how Wikipedia changes over time.

Fortunately, Wikipedia itself provides information about the history of its entries. Any visitor to the site can view any previous versions of the entry or compare different versions side-by-side. Note that many users do not use this feature. When quickly consulting Wikipedia, one typically just views the current entry— if the current entry is misleading, then one is misled. However, the articles histories do tell us about the vigilance of Wikipedia’s contributors. When an article is vandalized or when outrageous content is inserted, someone is quick to revert the article to its prior state. This often happens within a minute or two. Less outrageous errors, ones which slip past the first responders, may persist for much longer periods. As an illustration, imagine that Alice introduces an error into a section of the article on Zeno of Elea. Sometime later, the vandal Bob erases the section and replaces it with obscenity. The vigilant Charles detects the vandalism and reverts the entry to its prior state— which includes Alice’s error. Later still, visitor Dana fixes a formatting problem in another section of the article. It is not too long before Alice’s error is no longer near the top of the list of recent changes. It is corrected at some point, perhaps, but during the interim people will read it when they consult the Wikipedia.⁷ Moreover, duplicates of the entry will preserve Alice’s error on other sites even after it has been corrected in Wikipedia.

I want to draw a modest lesson from all this: While some articles in the Wikipedia are rather good and others are abysmal, this is not merely a matter of extraordinary outliers. Wikipedia entries vary widely in quality. Mistakes and lies can persist long enough that they might be present in any entry you decide to consult.

Third: some general considerations

Because of its spotty quality, using Wikipedia requires assessing the legitimacy of specific claims. If I read something in the Wikipedia, how can I tell whether or not to believe it? Let’s consider four general ways that we might assess

⁷For more discussion on this point, see <http://www.fecundity.info/foe/comments.php?y=06&m=07&entry=entry060727-031457>.

claims.⁸

Appeal to Reliability involves relying on the reputation of the source. If I read an article on the New York Times website, I give it roughly the same weight as I would an article in the print edition. Note that this is not a question of knowing who the author is. If the Wikipedia were robustly reliable, then I would be justified in trusting claims found in it. Yet the Wikipedia is not so reliable. The quality of its entries varies widely.

Appeal to Plausibility involves assessing whether the claims sound like they are even in the right ball park. This can be done either in terms of style or in terms of content.

Consider style first: Imagine I read a website maintained by a single author, and it makes some claims about astrophysics. If it is written in sentence fragments and marred by misspellings, that alone makes me disinclined to trust it. Similarly, if it reads like a potboiler romance, then I distrust it regardless of the specific claims made. If the author at least sounds like a physicist, then the claims pass a preliminary test. The test is not infallible, but it does help to sort out those that merit no further attention.

Appeal to plausible style, as a method, is frustrated by the collective nature of the Wikipedia. As we can tell by looking at articles' histories, contributors will often edit Wikipedia articles and merely correct the formatting, add hyperlinks, or fix usage. These contributors are prettying up any mistakes that might be in the article. The style will make the claims seem more plausible without making any changes to the content.

One might object: Surely these contributors who prune entries would also weed out any wild, ridiculous falsities. I concede this, but it is not as much of a virtue as it might at first appear. More on this in a moment.

Appeal to plausible content involves considering whether the claim is even *prima facie* plausible. If a claim is obviously false, then I should not believe it when I read it in the Wikipedia.

Consider again the pruning contributors to Wikipedia. By removing claims that are obviously false, they assure that only plausible claims remain—yet these might be either true claims or plausible falsehoods.

To make this worry precise, let's simplify and suppose that intuitions of plausibility are an oracle that answers 'yes' or 'no' when asked whether it is wise to believe a claim. Suppose that they are incredibly reliable, such that if a claim is true then intuition necessarily says 'yes' and if a claim is false then there is a 95% chance that intuition says 'no.' It does not immediately follow from this assumption that we should believe a claim if intuition says 'yes' about it. Rather, it is crucial to know about the population of claims from which this claim was drawn. If there are many false claims and only a few true ones, then

⁸In the paper where these are introduced [Mag01], I argued that combinations and variants of these methods exhausted the possibilities. I am not relying on that claim here.

a ‘yes’ answer is more likely to be a false positive than it is to be an actual mark of truth.

Consider the population of claims in the Wikipedia. It begins as a mixture of true claims and false claims. Now the pruning contributors go to work and correct the style and formatting of the entries. If they find an implausible claim, one to which intuition says ‘no’, then they remove it. The final population of claims contains only ones to which intuition says ‘yes’, but it contains false claims as well as true ones. The exact density of truth will depend on the density of truth in the starting population, but it has no upper or lower bound.⁹ For example, if 9 in 10 claims in the original pool are false, then 1 in 3 claims judged to be plausible will nonetheless be false.¹⁰

Our intuitions of plausibility are not a reliable oracle. *Prima facie* plausibility is hit-or-miss. Nevertheless, the lesson generalizes. When many of the implausible claims are summarily removed, the probability that an arbitrary plausible claim is true depends on the density of truth among input claims—but we never see the great mass of implausible claims that were excised. Appeal to plausibility will lead us astray with Wikipedia more than with a single authored source, *just because* other people have gone ahead ahead of us and weeded out the flatly implausible claims.¹¹

Calibration involves checking the facts where you can and extrapolating: If the source gets things right on matters you can check, then that is some reason to believe that it gets things right on matters you cannot check.

Again, the collaborative nature of the Wikipedia frustrates this method. If the things that you can check independently are the things that other people could check, then those things will probably be correct— someone will have corrected any mistakes. The correctness on those points will fail to be evidence for the correctness on the remainder, if the background knowledge of honest and conscientious contributors runs out where yours does.

Sampling involves checking multiple sources and comparing them against one another. Of course, if I just look at Wikipedia, then I have not bothered to sample. Even if I do sample, however, I am frustrated: Contributors may copy from known sources *to* Wikipedia and from Wikipedia to their own webpages, so sampling is apt to overcount any claim that persisted in a Wikipedia entry.

Conclusion

I have argued that we should not assimilate Wikipedia to the kind *encyclopedia*. For one thing, we use it differently. Moreover, it frustrates the methods by

⁹For more about base rates— especially base rates of truth— see [MC04].

¹⁰If 9 in 10 false seems like too grim a ratio, recall that there are a great many changes made to Wikipedia articles that are immediately undone.

¹¹There is a further, related problem: Contributors may also remove true claims that seem counterintuitive.

which we judge the claims of traditional information sources like encyclopedias. This does not mean that Wikipedia is worthless or that we ought not use it at all. Yet it does mean that we should be wary of it and that we should try to develop methods which are suitable to it.

Appendix: Methodology

The experts whose reports are summarized in table 1 were given these instructions:

Attached are a number of articles taken from different sources. You are asked to read each item and judge its accuracy. Keep a tally of the number of mistakes in each, organizing mistakes into the following categories:

Infelicities are points at which the article presents an eccentric or peculiar interpretation of some point; it is not strictly-speaking incorrect, but it fails to represent the dominant philosophical opinion.

Minor errors are mistakes that would not seriously effect the readers understanding of the issue.

Major errors seriously misrepresent the issue or facts.

You should decide what counts as a discrete mistake. I appreciate that it is difficult to count mistakes in this way, but please try to be consistent between entries.

References

- [Gil05] Jim Giles. Internet encyclopedias go head to head. *Nature*, 438:900–901, December 15 2005.
- [Inc06] Encyclopædia Britannica, Inc. Fatally flawed, March 2006. http://corporate.britannica.com/britannica_nature_response.pdf.
- [Mag01] P.D. Magnus. Reliability on the crowded net: Finding the truth in a web of deceit. In *MacHack 16 (proceedings)*. MacHack: the Conference for Leading Edge Developers, 2001. <http://hdl.handle.net/1951/25991>.
- [MC04] P.D. Magnus and Craig Callender. Realist ennui and the base rate fallacy. *Philosophy of Science*, 71(3):320–338, July 2004.
- [Nat06] Britannica attacks. . . and we respond. *Nature*, 440:582, March 30 2006.
- [Str06] Randall Stross. Anonymous source is not the same as open source. *New York Times*, 2006. March 12. <http://www.nytimes.com/2006/03/12/business/yourmoney/12digi.html>.

[Wik06] Wikipedia. Academic use — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Wikipedia:Academic_use&oldid=%63044815, 2006. [Online; accessed 15vii2006].