P. D. MAGNUS

ON TRUSTING *WIKIPEDIA*

ABSTRACT

Given the fact that many people use *Wikipedia*, we should ask: Can we trust it? The empirical evidence suggests that *Wikipedia* articles are sometimes quite good but that they vary a great deal. As such, it is wrong to ask for a monolithic verdict on *Wikipedia*. Interacting with *Wikipedia* involves assessing where it is likely to be reliable and where not. I identify five strategies that we use to assess claims from other sources and argue that, to a greater of lesser degree, *Wikipedia* frustrates all of them. Interacting responsibly with something like *Wikipedia* requires new epistemic methods and strategies.

In a conversation with a colleague, I mentioned the work I was doing on *Wikipedia* (which eventually appeared as Magnus 2008). That's fine, she said, but how is that philosophy? On the face of it, the question of whether we should trust *Wikipedia* is like the question of whether I should trust my auto mechanic. The latter question is *epistemic*, because it is about knowledge, and we can pose it in a formal way: Can I know P on the basis of my mechanic M saying P? Of course, epistemologists have things to say about general formulae of this kind. But when it comes down to a particular Tuesday afternoon when my mechanic tells me that I need a new alternator, there is nothing philosophical about the issue. If I trust my mechanic, it will not have anything much to do with my philosophical commitments. By 'trust' here I do not mean anything esoteric, but just that 'M says P' gives me a defeasible reason to believe that P. The question of trusting *Wikipedia* is just whether '*Wikipedia* contains the claim that P' is similarly some reason to believe that P.

The buzz about *Wikipedia* is that it raises new philosophical problems. A *New York Times* article says that *Wikipedia* raises "a single nagging epistemological question: Can an article be judged as credible without knowing its author?" (Stross 2006). (If I read in the *Times* that *Wikipedia* raises epistemological questions, is that enough reason to believe that it does?) When we read *Wikipedia* entries, we read the uncredited, collective work of individuals whose only qualifications for contributing were an internet connection and an interest in doing so.

I can pose the question, Can I know P on the basis of *Wikipedia* saying P? and this is formally like the question I can ask about my mechanic. Yet *Wikipedia*

is rather unlike my mechanic. Jaron Lanier calls it an "online fetish site for foolish collectivism." (2006) He is at least right in that this shambling information aggregate is something new.

In what follows, I begin by reviewing some empirical results about the reliability and stability of *Wikipedia* articles (§1). I then argue that it is unhelpful to pigeonhole *Wikipedia* into our pre-existing category *encyclopedia* (§2). After considering some of the methods by which we determine when to trust more traditional sources (§3), I argue that they are frustrated when applied to *Wikipedia* (§4). Beyond a few scanty suggestions, I do not know what the appropriate strategies are for using *Wikipedia*. Whatever they are, though, they are likely to exploit the features that make *Wikipedia* different than traditional sources (§5).

1 EMPIRICAL QUESTIONS

Whether or not we should trust *Wikipedia* depends, at least in part, on whether doing so will yield true beliefs. This seems to be simply an empirical question. The most natural way to approach it is to examine *Wikipedia* entries and determine the density of true and false claims in them.

A much-discussed study in the journal *Nature* (Giles 2005) 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. (900–1)

The editors of *Britannica* replied to the study (*Encyclopædia Britannica* 2006), arguing over the methodology. The editors of Nature replied to Britannica's reply (*Nature* 2006).[1] 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*.

The *Nature* study reports that "the difference in accuracy was not particularly great" (Giles 2005, 900), but we should not forget that the results *do* favor *Britannica* over *Wikipedia*. The *Wikipedia* articles, considered altogether, contain 32% more errors than the *Britannica* articles. Moreover, 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):

- *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.

P. D. Magnus

To summarize these differences: In addition to having more errors overall, *Wikipedia* entries varied more than *Britannica* entries.[2]

This wide variability should come as no surprise, since *Wikipedia* entries rely on volunteer contributors. Different entries will attract different contributors. Although entries on a specific cluster of topics may be maintained by a core community of contributors, there is not likely to be any overlap of contributors for two entirely unrelated articles. The specific community of contributors matters to reliability. For example, George Bragues (2007) examined the *Wikipedia* entries for seven famous philosophers and found significant omissions, a quirky fixation on biographical detail, but no outright errors.

In a small survey of philosophical topics, I had respondents compare (blinded) entries from *Britannica* and *Wikipedia* (Magnus 2006). The entry on bioethics, I was told, "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 entry on phenomenology, however, was reported to be exemplary; the respondent said, "Now *that's* the way an encyclopedia entry on phenomenology should be organized." The number of entries that I examined in this way was too small to provide meaningful numbers, but the range of responses reinforces the lesson that the quality of entries is highly variable. As such, even if we are asking about the ratio of true to false claims, it is a mistake to ask about all *Wikipedia* entries as one population.

Moreover, studies like the ones discussed so far only consider *Wikipedia* entries at particular times. This ignores the important fact that *Wikipedia* entries change over time. As new contributors supplement or revise existing articles, entries might change very quickly. The entry praised by the phenomenologist, for example, soon accumulated several thousand words of new material.

Falsehoods may enter into any *Wikipedia* entry and persist for some time. Dan Tynan (2008) worries about this problem and provides an anecdote:

> 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.

Yet sometimes contributors respond quickly and effectively to inaccurate additions. In a 2004 prank that has been used as a stock example of attempts to probe *Wikipedia*, Alex Halavais created a pseudonymous account and inserted 13 false claims in various entries. All of the false claims were deleted within three hours. Some *Wikipedia* user noticed that Halavais's account was responsible for bad changes and undid them all. (Halavais 2004 detailed the episode on his blog; see also Read 2006.)

The Halavais episode illustrates what I'll call an *association effect*; all the errors inserted at the same time were corrected, not because they were detected independently but because one was detected by a user who checked what other

changes had been made by the same person. Association provides a check on rampant vandalism. The probability that a particular falsehood will be corrected grows as the user introduces more falsehoods over a short period of time. However, not all falsehoods will be entered by prolific vandals. Many will be entered by well meaning contributors who transcribe their false beliefs into *Wikipedia*. Others will be entered by more careful and deliberate vandals. For example, I know an eight-year-old boy who uses *Wikipedia* as a place to practice his creative writing. He finds entries and adds fictions to them. In a given evening of writing, he might only revise a few articles. If *Wikipedia* is reliable, then eventually his stories will be discovered and removed. In the meantime, however, unsuspecting users will read the affected entries and encounter his prose.

In an attempt to test *Wikipedia*'s response to isolated falsehoods, I systemically inserted fibs and tracked their persistence. Adjusting for association and other factors, 10 out of 28 fibs (36%) were corrected or flagged as dubious within 48 hours. I removed the fibs after 48 hours if they had not been removed already. (For a more detailed discussion of this method and the results, see my 2008.) I inserted fibs into the biographies of philosophers. For reasons already discussed, there is no reason to think that 36% would be the correction rate for entries on other topics. However, it does moderate between the extreme anecdotes provided by Tynan (the fib uncorrected for three months) and Halavais (every fib corrected in a few hours). The important point is that a static estimate of the truth ratio of claims in *Wikipedia* does not tell the whole story. The dynamic entries of *Wikipedia* will regularly have errors introduced to them and errors removed. When you consult a *Wikipedia* entry at a particular time, you might capture it at a well-tended or fib-ridden moment.

In any case, the density of true claims in *Wikipedia* and the rate at which false claims get corrected are empirical questions. Imagine we had precise answers to those questions. Would there be anything left for an epistemologist to worry about? Of course, because we want to know whether someone consulting *Wikipedia* will wind up having true beliefs. If we make the simplifying assumption that a *Wikipedia* user believes everything in *Wikipedia* indiscriminately, then this would just reduce to the ratio of true to false claims in *Wikipedia*. That simplifying assumption would be absurd, however, because no real user would believe *everything* from *Wikipedia*.

Imagine the entry on Ronald Reagan were edited to say that he was the three-headed dog that guarded the gates to the underworld. The entry would quickly be reverted, but suppose a user consults the entry in the brief time before it is corrected. The user might form beliefs *about Wikipedia* (that it is silly or unreliable, for example) but would probably not form any beliefs about Reagan. This is an obvious example, but it illustrates this general point: Whether users form true beliefs depends on *how Wikipedia* is used.

We could approach this as another empirical issue and study how *Wikipedia* users form beliefs when using *Wikipedia*. For example, we could direct subjects to use *Wikipedia* in researching a specified topic and then quiz them on that topic. The distinctively philosophical question, though, is how *Wikipedia ought* to be used. The

question of whether we should trust *Wikipedia* becomes the question of *how* and *to what extent* we should trust *Wikipedia*. Although our answer to this question should be informed by evidence, it is not entirely an empirical matter.

2 WHAT *WIKIPEDIA* IS NOT

It may be tempting to reason in this way: *Wikipedia* is an encyclopedia. We already know how encyclopedias ought to be used. Therefore, we already know how *Wikipedia* should be used. Although school children might use the *Encyclopædia Britannica* as their primary source in writing a report, they should 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. The tempting reply concludes that *Wikipedia* is no different.

Wikipedia's own guide contains this sort of reasoning; the *Wikipedia* article on using *Wikipedia* advises "that *any* encyclopedia is a starting point for research, not an ending point" (*Wikipedia*, "Wikipedia:Academic use"). When Middlebury College prohibited students from citing *Wikipedia* in early 2007, *Wikipedia* co-founder Jimmy Wales said similarly, "students shouldn't be citing encyclopedias. I would hope they wouldn't be citing Encyclopaedia Britannica, either" (Cohen 2007).

However, pigeonholing *Wikipedia* as an encyclopedia overlooks ways in which it is different than a familiar hardcopy encyclopedia like *Britannica*. Practical differences make *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 effort 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* (SEP) first, but the SEP entry on pragmatism was not written yet. Pressed for time, she consulted *Wikipedia*. It was *easy*. As *Wikipedia* says of itself, "*Wikipedia* is increasingly used by people in the academic community, from first-year students to professors, as an easily accessible… source for information about anything and everything" (*Wikipedia*, "Wikipedia:Academic use"). (Maybe this is true.) To underscore the point, ask yourself how many times you have consulted *Wikipedia* in the last year. And how many times have you consulted *Britannica*?

Second, users are often led to *Wikipedia* even if they do not start there. For many topics, a *Wikipedia* article will be on the first page of internet search hits. Even if users were to avoid visiting *Wikipedia* directly, they would still encounter content from it. The content of *Wikipedia* is under a GNU Free Documentation License, which requires acknowledgement but explicitly allows content to be freely reproduced. Many sites copy from *Wikipedia* verbatim, and many do not even clearly acknowledge that they have done so.

Third, *Wikipedia* has a breadth that general encyclopedias do not. One can consult *Wikipedia* on matters about which traditional encyclopedias are silent. For example, it has an entry for the Polish philosopher Kazimierz Twardowski; *Britannica* has none. And because *Wikipedia* receives new contributions all the time, it has more information about popular culture and current events than a traditional encyclopedia.

Therefore, we do not use *Wikipedia* in the way we use traditional encyclopedias. The point of this argument is not that *Wikipedia* is *worse* than an encyclopedia, since these features arguably make it better. The point is just that these features make it *different*. Saying that we should just treat it as an encyclopedia would naïvely ignore the respects in which it is something new and different. So how should one use *Wikipedia*? The warning that it should not be taken as a final authority on important matters is a start, but does not tell us when or how we should actually form beliefs on the basis of using it.

3 EVALUATING CLAIMS ON-LINE

In trying to decide how we ought to use *Wikipedia*, it may help to consider how we form beliefs on the basis of claims from single-author websites, personal blogs, and internet forum posts. I do so in this section and return to *Wikipedia* in the next. There are many practical guides to dealing with this problem, but they typically take the form of specific questions arranged as loosely organized checklists. (The Virginia Tech library maintains an extensive bibliography of such guides (Auer and Sebek 2007). Frické and Fallis (2003) found that many of the criteria suggested by such guides are not good indicators of website accuracy.) Here I'm interested not just in methods, but in the structure of the methods. I want to specifically consider five criteria which I will call authority, plausible style, plausible content, calibration, and sampling. These exhaust most, if not all, of the ways we assess claims found on the net.

Authority

One may believe a claim if it is made by a reliable source. This shifts the problem from deciding which claims are to be believed to deciding which sources are to be trusted.

One relies on background knowledge to decide who can be trusted. As an example, suppose someone – let's call her Alice – goes to a fan site that describes what happened at a recent concert. Suppose that the fan site is maintained by Bob, who claims to have been at the concert himself. If Alice justifiably believes that he was at the concert and that he is being honest, she is justified in believing that things happened as he describes. Her ability to trust a particular claim (that such and so happened at the concert) relies on her accepting an indefinite list of other background claims (that Bob went to the concert, that Bob is not spreading

malicious rumors about the band, and so on). Call this reliance on *authority*. The problem with it is readily apparent: how can Alice be sure of her background beliefs?

We can imagine ways she might address this question in an off-line context. Bob may be a friend of Alice or at least a friend of a friend; someone besides Bob himself will swear for him. Perhaps Bob can show his ticket stub to Alice or provide other physical evidence. Such checks allow her to judge whether he went to the concert and so on. She can make these checks against a source besides his own testimony.

In an on-line community or chat room, Alice may be able to rely on similar resources. Alice may trust some members of the group who would swear for Bob. This is only possible, though, if Alice has been involved in the community long enough to trust some members of it and if Bob has been involved long enough that those others will swear for him. If Alice only knows Bob from his web site, it may be that all she knows about him is what he has had to say for himself.

Although illustrated in terms of individual people, these issues arise with institutional sources as well. Consider, for example, that Aleksandr Solzhenitsyn is dead. Suppose Alice believes this because she read it in the *New York Times* and because she believes that the *Times* is typically accurate about this sort of thing. It makes little difference whether she read it in the actual paper or on the *Times* web site. Some news sites are extensions of traditional publications – as nytimes.com is an extension of the *New York Times* – and we would trust them to whatever degree we would trust their traditional counterpart.

Plausibility of style

Suppose Alice does a search for high-energy physics research and finds the web page of someone – call him George – who claims to have a PhD in physics and who describes several cutting-edge projects. Alice may evaluate what George has to say by trying to determine if George writes like a physicist. If George writes like a high school student, Alice should not give his site much credence. Note that Alice can make these judgments about George regardless of what particular claims he is making. Call this an appeal to *plausibility of style*.

By *style*, I have something more subtle in mind than the markers tested by Frické and Fallis (2003). They tested markers that anyone can readily inspect, such as where the site is hosted, whether it has advertising, or whether it has a copyright notice. One might expect university websites to be more reliable than commercial sites, for example, but for the questions they investigated, this was not so. *Writing like a physicist*, although it is something we can recognize, is not something we can so readily codify.

Of course, the fact that someone can successfully write like a physicist does not logically entail that they are actually an expert in physics. Having interactional expertise is not the same as being a reliable source; cf. Collins et al. (2006). Nevertheless, interactional expertise is frequently correlated with substantive

knowledge of the subject matter. If someone *can* write in the physicists' idiom, that is some reason to believe claims that person makes about physics. Conversely, if someone *cannot* write like a physicist, that is some reason to disbelieve.

Plausibility of content

One can also assess the plausibility of a particular claim, independently of whomever the source might have been. Suppose that Alice reads on a web site that Princess Diana's death was arranged by a conspiracy of Idaho potato farmers. This is such a wild claim that she ought not believe it, even if, by some quirk, it appears on a website she otherwise trusts. Call this an appeal to *plausibility of content*.

Admittedly, the content of a claim may be plausible even though the claim is false. Creative speculation can generate plausible explanations for any event. As such, the plausibility of a claim provides no positive reason to believe it. Nevertheless, the implausibility of a claim provides a reason *not* to believe it. Alice has *prima facie* reason not to believe that a consortium of Idahoans masterminded Diana's demise because of her background knowledge about Diana and Idaho. Her background beliefs might be wrong, but that worry will arise for any criterion. Judging whether or not to believe claims found on the internet always relies on some background knowledge.

If we consider all of the claims one encounters on the internet as candidate beliefs, then negative criteria like appeals to plausibility serve to vet the candidates. They help sort the claims more likely to be true from the ones more likely to be false. If Alice's background beliefs are mostly true, then resisting implausible content will help her avoid adopting false new beliefs. (As Mark Wunderlich 2008 argues, error avoidance rather than truth ratio is the appropriate reliabilist measure of adequacy for vetting procedures.)

Calibration

Of course, one does not encounter each claim made on the internet in isolation. One might evaluate the quality of a website based on its discussion of matters one knows something about, and rely on that evaluation when deciding whether to believe claims made on the website about which one has no prior belief.

Consider Alice again, as she reads George's description of physics research. If she knows something about physics, she can compare the partial knowledge she has with the claims George makes. If he is correct on those points, then she might expect him to be correct on other points as well. Call this *calibration*: Alice can judge the page reliable if, on the things she can check, it gets things right. If the website that says $P_1, P_2, \ldots P_n$, and Q gets P_1–P_n right, then that is some reason for her to think that it is also right about Q.[3]

Perhaps calibration is an inference of this form: 'This source is right about several points I have checked. By induction, this source is probably right about

most things. This establishes the source's authority.' If so, then it is really just an appeal to authority, where the claim of authority is underwritten by an inductive inference. Regardless, since calibration does not require any prior belief about the authority of the source, I will treat it as a separate method.

Although calibration does not depend on prior knowledge about the source, it does depend on prior knowledge of the subject matter. If one already knows almost all of what is said by a source, then one might be more comfortable believing the remainder than if one can only check a small fraction of what they say. However, the crucial issue is *not* the mere volume of claims that one can check. Suppose you had an evil twin with exactly the same background knowledge but an intent to deceive you. The evil twin would be correct on all the matters you could check, because the evil twin would tell you what you already knew, so calibration would lead you to trust the twin – but the twin might lie about *everything else*. The trick would work because the twin already knows the extent of your background knowledge. A less perfectly matched deceiver would either tell you some falsehoods that you could catch or accidentally tell you something true that you did not already know.

To consider a less abstract example, imagine that George's knowledge of physics comes entirely from some recent magazine articles and that the rest of his website is the product of his fevered imagination. If Alice calibrates by reading the same articles and checking whether he gets the facts in them right, then she will be suckered into believing his fantasies. (Perhaps his style of writing would change when he shifts from the facts to his ravings. If Alice relies on such cues, however, she is appealing to plausibility of style and not to calibration.) In general, calibration goes wrong if what one can check is not representative of the claims made by the source; that is, when the claims one can check are a biased sample of the population of claims made by the source.

Sampling

The methods considered so far apply to a particular report or web page in isolation, but one can assess the credibility of a claim by comparing related claims made by several sources. If all sources disagree, then the choice of which, if any, to trust must be made on the basis of some other method. If all or most agree, then those attract more credibility than they would if they were considered separately. Call this method *sampling*. It will correct for individual bias and error but will fail to increase reliability if distorting biases are shared by all or most of the sample.

A mundane example of sampling is the practice of asking for a second opinion. If Alice is diagnosed with a dread disease and her doctor recommends surgery, she may insist on asking another doctor. If the diagnosis was the result of some uncommon mistake or if the first doctor was unusually keen on surgery, then the second doctor will most likely offer a different opinion. If the test used to detect the disease systematically returns false positives or if both doctors rely on the same

protocol, then the second opinion will accord with the first and lead Alice to place potentially unwarranted trust in the diagnosis and recommendation.

Sampling has similar strengths and weaknesses when used to evaluate claims found on-line. If a web site contains errors that result from the author being careless, confused, or idiosyncratically deluded, other sites are unlikely to contain those same errors; so, a comparison of multiple sites can separate those errors from other claims. Conversely, where many authors have copied information from a common source or from each other, sampling will not increase reliability. This may occur where the many presentations of a news item or rumor are actually unattributed second- or third-hand repetitions of a claim made on one influential site. And even if authors do not share sources, sampling will fail to help if they share some distorting bias. Sampling applied to claims made on the internet automatically excludes people who do not use the internet, and for some topics this may introduce a significant bias.

Unlike the other methods considered so far, sampling does not require knowledge about the subject matter *per se*. Nevertheless, it requires at least the implicit assumption that errors will be randomly distributed rather than systematic. Thus, it also implicitly relies on some background knowledge.

Other methods?

Of course, these strategies can be used in conjunction with one another. Sources that calibrate positively may count for more in sampling; claims made by one reliable source may be used as a basis against which to calibrate another source; and so on. Furthermore, there may be ways of assessing claims that are entirely distinct from the methods discussed here. Nevertheless, the list serves as a helpful rubric in considering how to use *Wikipedia*. My argument in the next section requires only that we do often rely on these methods, not that the list be exhaustive.

4 STANDARD METHODS UNDONE

Because of its spotty quality, using *Wikipedia* requires assessing the legitimacy of specific claims. If I read something in a *Wikipedia* entry, how can I tell whether or not to believe it? In this section, I consider each of the methods for evaluating claims on-line that I discussed above and argue that *Wikipedia* frustrates each and every one of them to some extent – and it does so because of the features that distinguish it from single-author websites.

Authority

We should acknowledge at the outset that the question of whether "an article [can] be judged as credible without knowing its author" (Stross 2006, cited above) is a red herring. Stories in the *New York Times* typically carry bylines, but our believing what they say does not typically depend on what we know about the specific reporter

credited. The article has the authority of something printed in the *Times*. Knowing who wrote it does not usually matter. So, too, for *Wikipedia* articles. (Fallis (2008, 1667) makes a similar point.)

Wikipedia policies encourage contributors to cite sources. One might suppose that an article with citations to public sources has more authority than an article that does not. However, there is nothing to stop contributors who fabricate claims from making up sources as well or to stop contributors who misunderstand a topic from misattributing a claim to a source. Where there are references, we might follow them to the published source and check the claims that are made. For sources which are not themselves readily available on the internet, however, we are unlikely to do this; even if we did, it would undo any convenience gained by using *Wikipedia* in the first place.

If *Wikipedia* were robustly reliable, then one could safely trust claims made in *Wikipedia* articles. Yet (as we saw in §1) the reliability of *Wikipedia* articles varies widely. The lack of any centralized control means that the class of claims made in any *Wikipedia* entry is just too broad. It is more like 'claims made in New York' than 'claims made in the *New York Times*.' As such, it would be a mistake to appeal to the authority of *Wikipedia tout court*.

Don Fallis suggests that further research should aim to "compare the reliability of *Wikipedia* with respect to different subject areas" (2008, 1672). He suggests, for example, that *Wikipedia* might be better on science and current events than on philosophy. I am less sanguine. Even if present *Wikipedia* entries about scientific matters were more accurate than entries about philosophical matters, this does not tell us what those entries will be like in a year. This problem is especially acute for *current events*, since the topics that count as current will also change in that time. Moreover, there is no reason to think that the communities maintaining the articles on (for example) philosophical topics are a homogenous group. The contributors writing about phenomenology seem to be different than the ones writing about bioethics.

Still, one might have reasons to trust specific *Wikipedia* articles. For example, David Morgan Mar is a PhD astrophysicist who often discusses technical topics on his website. In lieu of filling in all the details, he sometimes links to *Wikipedia* articles.[4] I have known him both from his web presence and from some on-line chat, and I consider him to be a competent authority on the subjects he discusses. Suppose he has looked at the *Wikipedia* page to check that it is a good place to direct people. In terms of the methods I've described: By the criterion of authority, I believe his claim that the *Wikipedia* article is OK. Appeal to this vouched-for authority of the article gives me some reason to believe claims made in it.

However, the article to which he links will typically have changed between the time he looked at it and the time when I visit his site and follow the link. Even if the article is maintained by a core community of competent contributors, users from outside the core community will occasionally edit the carefully tended pages and introduce errors. When I follow the link from Morgan Mar's site, it is unclear

how much weight I should give to the fact that he gave his nod to some previous version of the article. The point here is not to excoriate Morgan Mar. The same problem arises whenever experts link to *Wikipedia* articles, as has become common practice among bloggers.

Plausibility of style

Appeal to plausible style, as a method, is frustrated by the collective nature of *Wikipedia*. As we can tell by looking at articles' histories, contributors will often edit *Wikipedia* articles merely to correct the formatting, add hyperlinks, or fix usage. Claims that are made carelessly may get prettied up in this way, and mistakes can be disguised in a facade of apparatus. Stylistic flags of implausibility might be removed without any changes to the content.

One might hope that a user who cleans up an entry will have the relevant expertise to correct errors in it. Someone who can write like a physicist is likely to know something about physics, for example, but *Wikipedia* articles have a style and tone of their own. Is someone who can edit an article to give it the tone and style appropriate for a *Wikipedia* entry someone who is likely to know about physics? A user may be good at fixing grammatical and structural mistakes, inserting the appropriate Wiki-markup, and so on without having any special knowledge of the topic. Even if a user knows something about the topic, that does not mean they will take the time to determine whether there are false claims in the article; fixing usage and formatting is quick.

Plausibility of content

Surely contributors who prune *Wikipedia* entries would also weed out any wild, ridiculous falsities. However, this is not as much of a virtue as it might at first appear. By removing claims that are obviously false, these pruning contributors assure that only plausible claims remain – yet the claims might still just be plausible falsehoods.

To make this worry precise, let's simplify and suppose that intuition of plausibility is an oracle that answers 'yes' or 'no' when asked whether it is wise to believe a claim. Suppose for illustration that it is 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 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 *Wikipedia* articles. 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. 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. If 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.

Of course, *prima facie* plausibility is hit-or-miss, and intuitions of plausibility are not a reliable oracle. 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. 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 of us and weeded out the flatly implausible claims. (As a further, related problem: Contributors may also remove true claims that seem counterintuitive.)

There is a flip side to this. For users who would not be able to tell a wild, ridiculous falsity from an insidious, plausible one, the vetting of *Wikipedia* is a good thing: They will have fewer false beliefs because of it. Don Fallis reflects on this tradeoff and suggests that "we have to weigh the epistemic cost of a loss of verifiability for some people against the epistemic benefit of removing information that will be misleading to other people." (2008, 1667) However, this sort of weighing would only answer the question of whether *Wikipedia* is on balance a good or a bad thing. It does not help us with the problem that concerns us here, which is deciding how we ought to use it.

Calibration

Recall that calibration is only effective if the claims one can check are representative of the claims made by a source. Imagine Alice looking at the *Wikipedia* entry on black holes. She knows a bit about black holes, but many *Wikipedia* users know roughly what she does about the topic. As such, the claims she can check independently are the things that other people have checked. Some other user will have corrected any mistakes. However, 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 hers does, then the rest could be written by crackpots.

When applying calibration to a *Wikipedia* article, one must either already know something about the topic or have a separate source that one can consult. In order for calibration to be effective, however, the range of claims one can check must not just be the range of claims that many other *Wikipedia* users can check. Even if this is true about the population of contributing users, one is rarely, if ever, in a position to know that it is true. Therefore, calibration fails when applied to *Wikipedia* articles.

Sampling

If one looks only at a *Wikipedia* article, then one has not bothered to sample. So suppose one widens one's search and looks elsewhere on the web. Many web pages are automatically harvested from *Wikipedia* without explicit acknowledgment (as noted in §2) but perhaps these are not so worrisome. As Don Fallis notes, "it is fairly easy for people to recognize that a Web site containing a word-for-word copy of a *Wikipedia* entry … does not provide much corroboration." (2008, 1668) Since *Wikipedia* is dynamic, however, a page made a year ago by copying *Wikipedia* might agree with many of the claims made in the current *Wikipedia* article without being an obvious copy. Even disregarding word-for-word copies, users of *Wikipedia* may relay claims from known sources to *Wikipedia* and from *Wikipedia* to their own webpages. As a result, when sampling, it is easy to overcount any claim that has persisted in a *Wikipedia* entry.

5 NORMATIVE RECOMMENDATIONS

I have argued so far that we should not naïvely suppose that knowing how to use an encyclopedia means knowing how to use *Wikipedia* (§2) and that *Wikipedia*, to a greater or lesser degree, frustrates all of the methods we use to evaluate single-author web pages and forum posts (§§3–4). These lessons should inspire a self-conscious scepticism about *Wikipedia* entries. The same arguments might be given *mutatis mutandis* about other collaborative websites, and so we should be sceptical about those, too.

However, *contra* Descartes, reasons to be sceptical about an information source are not enough reason to abandon it entirely. Most of us will still use *Wikipedia* because it is convenient and timely. So how should we use it?

One natural and conservative strategy is to use *Wikipedia* as a pointer to other resources. *Wikipedia* entries often include a collection of links to other sources on the web. These were selected by *Wikipedia* users as especially appropriate, and spawning links from *Wikipedia* is often faster than dredging through pages of irrelevant search engine results. One need not actually believe any claims made in *Wikipedia* articles to exploit such links.[5]

Yet we typically go to *Wikipedia* expecting to find answers, rather than just pointers to answers. Cory Doctorow has suggested that puzzling out the right answer requires looking not just at *Wikipedia* articles, but also at the histories and discussion pages for the articles. He writes:

> *Wikipedia entries* are nothing but the emergent effect of all the angry thrashing going on below the surface. No, if you want to really navigate the truth via *Wikipedia*, you have to dig into those "history" and "discuss" pages hanging off of every entry. That's where the real action is, the tidily organized palimpsest of the flamewar that lurks beneath any definition of "truth." (2006)

P. D. Magnus

As a result, Doctorow insists, "reading *Wikipedia* is a media literacy exercise. You need to acquire new skill-sets to parse out the palimpsest." (2006) I agree that reading *Wikipedia* requires a new skill set, and sifting through histories might be part of that. Where some small part of an article is an ephemeral addition, where it is a fib living out its brief lifespan between entry and deletion, then looking at the article history will debunk it. Even so, not every false claim is fleeting, and not every false claim that persists incites a flame war. An urban myth is likely to persist in *Wikipedia* articles, for example, and the users who believe it will change the article to reflect the version of the myth with which they are familiar. Over time, the *Wikipedia* entry may contain the most plausible and nuanced version of the myth. The series of changes that led to that version will look pretty much like the series of changes involved in refinement of a genuine article, so looking at the history will not help identify the myth as such.

The ready availability of *Wikipedia*'s history can be exploited in other ways. If one looks at a *Wikipedia* article on a topic one knows well and decides to point someone else to it, one can point to the dated version which one examined. In the interest of also linking to the most recent version, I suggest that links to *Wikipedia* (e.g., from blogs) should point to *both* to the current article and to the dated article that the linking author inspected. Someone who follows the link but is dubious of the current article has the option of using the history function to compare it to the previous, vouched-for article.

One might worry that looking at article histories is impractical in just those cases where *Wikipedia* is most tempting: when we are too busy or do not care enough to find another source. In such cases, we do little more than rely on plausibility. As I argued above, *Wikipedia* to some degree frustrates appeals to plausibility—but admittedly the fact that a claim is not implausible still helps some. If the claim is a matter of trivia, we might be justified in accepting it without digging any further. If it is so trivial that it does not make any difference what we believe, then it probably does not make any difference how we decide what to believe either. But trivia comes in degrees, and something that seems like trivia today may have some consequence tomorrow. It is incumbent on us to learn how to use *Wikipedia* responsibly; this will mean both discovering new methods that it makes possible and learning to mitigate the ways that it frustrates our familiar methods.

Regarding students' use of *Wikipedia*, some have worried that "the reliance by students on *Wikipedia* for finding information, and acceptance of the practice by teachers and academics, was 'crowding out' valuable knowledge and creating a generation unable to source 'credible expert' views even if desired." (Gedda 2008) Certainly, students should not use *Wikipedia* as a substantive source in research papers. Yet it would be naïve to tell students that they ought never to use *Wikipedia*, as if saying so would settle the matter. They will use it, and so will we. I regret that I do not have anything more substantive to say about what the appropriate strategies are for reckoning with it, but those strategies may well involve features of *Wikipedia* (like history records and discussion pages) that it does not share with traditional

information sources. Users who do not know about such features cannot possibly use them. So teaching people to engage *Wikipedia* responsibly will require getting them to cultivate a healthy scepticism, to think of it differently than they think of traditional sources, and to learn to look beyond the current articles – and it will require learning to engage with it more responsibly ourselves.

ACKNOWLEDGMENTS

REFERENCES

Auer, Nicole and Robert Sebek. 2007. "Bibliography on Evaluating Web Information." Retrieved November 2, 2008, from http://www.lib.vt.edu/help/instruct/evaluate/evalbiblio.html

Bragues, George. 2007. "Wiki-Philosophizing in a Marketplace of Ideas: Evaluating *Wikipedia*'s Entries on Seven Great Minds." Working paper. http://ssrn.com/abstract=978177

Cohen, Noam. 2007. "A History Department Bans Citing *Wikipedia* as a Research Source." *New York Times*, February 21. http://www.nytimes.com/2007/02/21/education/21wikipedia.html

Collins, Harry, Rob Evans, Rodrigo Ribeiro, and Martin Hall. 2006. "Experiments with Interactional Expertise." *Studies In History and Philosophy of Science Part A* 37(4): 656–74.

Doctorow, Cory. 2006. "On 'Digital Maoism: The Hazards of the New Online Collectivism' by Jaron Lanier." *Edge*. http://www.edge.org/discourse/digital_maoism.html

Encyclopædia Britannica. 2006. "Fatally Flawed: Refuting the Recent Study on Encyclopedic Accuracy by the Journal *Nature*." http://corporate.britannica.com/britannica_nature_response.pdf

Fallis, Don. 2008. "Toward an Epistemology of *Wikipedia*." *Journal of the American Society for Information Science and Technology* 59(10): 1662–74.

Frické, Martin and Don Fallis. 2003. "Indicators of Accuracy for Answers to Ready Reference Questions on the Internet." *Journal of the American Society for Information Science and Technology* 55(3): 238–45.

Gedda, Rodney. 2008. "*Wikipedia* Breeds 'Unwitting Trust' Says IT Professor." *Computerworld*, April 14. http://www.computerworld.com.au/index.php/id;1828979092

Giles, Jim. 2005. "Internet Encyclopaedias Go Head to Head." *Nature* 438: 900–1.

Halavais, Alex. 2004. "The Isuzu Experiment." Blog entry at *A Thaumaturgical Compendium*. http://alex.halavais.net/index.php?p=794

Lanier, Jaron. 2006. "Digital Maoism: The Hazards of the New Online Collectivism." *Edge*. http://www.edge.org/3rd_culture/lanier06/lanier06_index.html

Magnus, P. D. 2001. "Reliability on the Crowded Net: Finding the Truth in a Web of Deceit." *MacHack 16 (Proceedings)*. MacHack: The Conference for Leading Edge Developers. http://hdl.handle.net/1951/25991

Magnus, P. D. 2006. "Epistemology and the *Wikipedia*." Paper presented at the North American Computing and Philosophy Conference, Troy, New York. http://hdl.handle.net/1951/42589

Magnus, P. D. 2008. "Early Response to False Claims in *Wikipedia*." *First Monday* 13(9). http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2115/2027

Nature. 2006. "Britannica Attacks… and We Respond." 440: 582.

Read, Brock. 2006. "Can *Wikipedia* Ever Make the Grade?" *The Chronicle of Higher Education* 53(10): A31. http://chronicle.com/free/v53/i10/10a03101.htm

Stross, Randall. 2006. "Anonymous Source Is Not the Same as Open Source." *New York Times*, March 12. http://www.nytimes.com/2006/03/12/business/yourmoney/12digi.html

Tynan, Dan. 2008. "The Truth Is Out There…Somewhere." *US Airways Magazine*, January: 42.

***Wikipedia* contributors**. "Wikipedia:Academic use." *Wikipedia, The Free Encyclopedia*. Retrieved July 19, 2008, from http://en.Wikipedia.org/w/index.php?title=Wikipedia:Academic_use&oldid=226679146

Wunderlich, Mark E. 2008. "Error-Avoidance Reliabilism." Unpublished draft.

NOTES

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

2 One might worry that the judgement as to what counted as an error was made by experts, since these experts might themselves not merit trust. Yet one cannot press scepticism about expertise too far without being a sceptic about everything beyond one's own immediate observations.

3 This differs from relying on plausibility, because she considers the truth of P_1–P_n in deciding whether or not to believe Q. She can assess whether Q is plausible without considering other claims made on the same site.

4 E.g. http://www.irregularwebcomic.net/comic.php?current=1960

5 Don Fallis commented to me that, for all I have said, the selection of links in *Wikipedia* might be peculiar. They might point to poor sources or a biased selection of good sources, so following such links might lead to false or skewed beliefs. Although this problem reaches beyond the scope of this paper, it is not clear why it is more likely to arise when using *Wikipedia* than when using a search engine.

P. D. Magnus is an assistant professor at the University at Albany, State University of New York. He works primarily in the epistemology of science, and much of his work has addressed forms of the underdetermination of theory by data.