HARVARD | BUSINESS | SCHOOL

Do Experts or Collective Intelligence Write with More Bias? Evidence from Encyclopædia Britannica and Wikipedia

Shane Greenstein
Feng Zhu

Working Paper 15-023

Do Experts or Collective Intelligence Write with More Bias? Evidence from Encyclopædia Britannica and Wikipedia

Shane Greenstein
Harvard Business School

Feng Zhu
Harvard Business School

Working Paper 15-023

Do Experts or Collective Intelligence Write with More Bias? Evidence from Encyclopædia Britannica and Wikipedia*

Shane Greenstein
Harvard Business School
Harvard University
sg@hbs.edu

Feng Zhu
Harvard Business School
Harvard University
fzhu@hbs.edu

March 1, 2016

*We thank Gerald Kane, Karim Lakhani, Michael Luca, Abhishek Nagaraj, Michael Norton, seminar participants at Boston University, Northwestern University and University of Pennsylvania, and participants at the Social Media and Digital Innovation Workshop at Boston College, the Conference on Open and User Innovation, the ZEW Conference on the Economics of Information and Communication Technologies, the 2015 DRUID conference, and the Digital Economy Workshop at Collegio Carlo Alberto for their valuable comments. We also thank Alex Risman for his excellent research assistance.

Do Experts or Collective Intelligence Write with More Bias? Evidence from Encyclopædia Britannica and Wikipedia

Organizations today have the option of using crowds or experts for knowledge production. While prior work focuses on comparing the factual accuracy of knowledge from crowd-based and expert-based production models, we compare bias from these two models when knowledge is contested. Contested knowledge is endemic to topics involving subjective, unverifiable, or controversial information, and addressing it successfully lay behind the promise of the production model based on collective intelligence. Using data from Encyclopædia Britannica, an encyclopedia authored by experts, and Wikipedia, an encyclopedia produced by an online community, we compare the slant and bias of pairs of articles on identical political topics. Our slant measure is less (more) than zero when an article leans towards Democratic (Republican) viewpoints, while bias is the absolute value of the slant. We find that Wikipedia articles are more slanted towards Democratic views than are Britannica articles, as well as more biased. The difference for a pair of articles decreases with more revisions of Wikipedia articles. The bias on a per word basis hardly differs between the sources, pointing towards the key role of article length in online communities. We stress a mechanism for resolving disputes in online communities: contributors tend to add text instead of reducing it, taking advantage of the lower costs to acquiring, storing, and revising information, as well as the absence of organizational discipline to restrain additions. These results highlight the pros and cons of each knowledge production model, and have implications for how organizations manage crowd-based knowledge production.

Keywords: online community, collective intelligence, wisdom of crowds, bias, Wikipedia, Britannica, knowledge production

INTRODUCTION

Technological advances in the past few years have made it significantly easier for users to communicate and collaborate with each other in online communities (e.g., Afuah and Tucci 2012; Gu et al. 2007; Kane and Fichman 2009; Ren et al. forthcoming). Many of these communities operate at a scale that exceeds even that of the biggest global organization, bringing in many more individuals to focus on a given task. Organizations seek to harness the collective intelligence from these self-organizing user communities to accomplish a variety of tasks, including developing new products (e.g., Lee and Cole 2003), evaluating product or service quality (e.g., Ba et al. 2014; Bin et al. 2014), funding startups (e.g., Agrawal et al. 2013, 2015; Wei and Lin 2015; Zhang and Liu 2012), innovation tournaments (e.g., Lakhani et al. 2012), prediction markets (e.g., Wu and Brynjolfsson 2013), scientific research (e.g., Lakhani 2009), and knowledge production (e.g., Gulati et al. 2012; Kane 2011; Kummer et al. 2012; Lih 2009; Ren et al. 2007; Ren et al. forthcoming; Xu and Zhang 2014).

When online collective intelligence emerged, scholars initially expressed concerns over the “madness of crowds” (Mackay 1852), arguing that online collective intelligence could be vulnerable to group thinking (e.g., Janis 1982), confirmation bias (e.g., Park et al. 2013), emotional contagion (Barsade 2002), and herding (e.g., Banerjee 1992; Bikhchandani et al. 1992). More recent studies have begun to characterize the properties of online collective decision making in a variety of situations, and have offered counter-examples that highlight positive attributes—for example, that collective decision making can be more accurate than experts’ decision making (e.g., Galton 1907; Shankland 2003; Antweiler and Frank 2004; Lemos 2004; Surowiecki 2004; Giles 2005; Rajagopalan et al. 2011). Focusing on characterizing the properties of online communities, Mollick and Nanda (forthcoming) examine crowd-funding and traditional venture funding, and stress that collective decisions can exhibit tastes or preferences not otherwise present in traditional sources.

We characterize properties of online organizations, and advance the field by examining situations where knowledge is contested. Broadly, we consider what happens to contested knowledge in collective decisions, and consider how the outcomes compare to experts' decisions. Contested knowledge—which we will define loosely for now as a situation in which there is no single 'right answer'— is endemic to topics involving subjective, unverifiable, or controversial information. It presents a challenge to online communities because online communities bring together participants who originate from communities with different traditions for expressing opinions, and have different cultural and historical foundations for those opinions, potentially without common bases of facts.

Contested knowledge also draws our interest because it is a crucial component to how online communities assemble and present information. In political settings, there is an open question as to whether online communities can support the institutions of democratic plurality (Sunstein 2001), and allow exposure to diverse viewpoints in social policy, law, and government decisions (Cohen 2012). The treatment of contested knowledge in online communities also contributes to our understanding of whether the general movement to online reference sources will alter civil discourse, compared to eras in which public disputes involved expert presentations (Tetlock 2005). Understanding how contested knowledge is treated in online communities also helps organizations understand which kinds of problems crowds may be better able to handle than experts.

This study examines political discourse in online communities, an arena we believe offers a good opportunity to test how such organizations treat contested knowledge. A significant portion of political opinion involves contested knowledge. Political slant arises when people try to frame their opinions to suit their political views on a subject, often by omitting opinions that disprove their viewpoints. Questions about slant are endemic to discussions about, for example, governments' taxation choices, or the operation of health care policies, or the biographical details of presidential candidates. They are also pervasive in presentations of scientific knowledge that touch on persistent ideological divides, such as forecasts of climate change, the consequences of diffusing genetically-modified crops, or the funding of stem-cell research. Consistent with this reasoning, Yasseri et al. (2014) find that the subject of politics contains the highest percentage (25%) of controversial articles in Wikipedia's top 100 controversial topics.

We use data from the entries about US politics in Wikipedia, the largest online encyclopedia, and Encyclopædia Britannica, the most popular offline English language encyclopedia. Wikipedia relies on an enormous group of volunteers to generate its content, receiving contributions from tens of millions of individual users. Conflicts in viewpoints are addressed in a highly decentralized process during countless arguments about whether a passage reflects a 'neutral point of view'. In contrast, Britannica sources its material from experts, and fosters a reputation for being an "august repository of serious information" (Melcher 1997), producing its final content after consultations between editors and experts. We choose these two sources because they both aspire to provide comprehensive information, and they are both the most common reference source in their respective online and offline domains. In addition, both sources resolve disputes—especially over contested knowledge—with distinct decision making processes.

As the first study to compare contested knowledge in two settings, we overcome a number of novel statistical challenges. First, some topics are inherently slanted and biased, so we need to take that into account when comparing the two sources. Second, opinion does not remain unchanged: as editors of Britannica or Wikipedia revise an article to improve the writing or to incorporate new information, the bias of its content may change, as Greenstein and Zhu (2012) find in their study of Wikipedia articles. To overcome these challenges, we develop a matched sample that compares paired articles appearing at the *same point in time* in both sources and which cover (nearly) *identical* topics. We are able to identify 3,918 pairs of articles about US politics that appeared in both outlets in 2012 and verify that these articles are a representative selection among the many topics covered within Wikipedia's political content.

To mitigate concerns that we manipulate our statistical procedures, we rely on a modification of an existing method, developed by Gentzkow and Shapiro (2010), for measuring slant and bias in newspapers' political editorials. 'Slant' indicates which way a particular piece of knowledge 'leans' (and is thus positive or negative after we normalize a neutral point of view to 0), and 'bias' is the absolute value of that slant (or 'lean'). This combined definition measures the direction of an article's 'opinion' and how strongly 'opinionated' it is. For example, Gentzkow and Shapiro (2010) find that Democratic representatives are more likely to use phrases such as "war in Iraq," "civil rights," and "trade deficit," while Republican representatives are more likely to use phrases such as "economic growth," "illegal immigration," and "border security." They characterize how newspapers also use such phrases to speak to constituents who lean towards one political approach over another. Several studies have applied their approach to analyze political biases in online and offline content (e.g., Greenstein and Zhu 2012; Jelveh et al. 2014)[1]—in a similar manner, we compute an index for the slant of each article from each source, tracking whether articles employ language that appears to slant towards 'Democrats' or 'Republicans.'

The findings show that the slant and bias of content sourced from collective intelligence differs from an expert-based source. Overall, we find that Wikipedia articles are slanted more towards Democratic views and display greater bias. Two factors explain these differences. First, substantial revisions of Wikipedia articles reduce the differences in biases and slants to negligible statistical differences. In other words, the greatest biases and slants arise in Wikipedia articles that are based on fewer contributions. The rate of convergence between the

[1] Budak et al. (2014) use alternative approaches to measure ideological positions of news outlets and their results are consistent with Gentzkow and Shapiro (2010).

two sources due to revision is also comparatively slow, so there are many more Wikipedia articles with bias and slant than without. While revised articles in the upper quartile (of our sample) based on the number of revisions have received enough revision to achieve no difference in slant and bias from their Britannica counterparts, the median article (and lower quartile) does not receive enough revision and thus, show considerable difference in slant and bias from their Britannia counterparts.

Article length also plays an important role in sharing the difference. We find evidence that, on any given topic, Wikipedia's articles are longer than their Britannica counterparts, as well as slightly less biased on a per word basis, although the statistical difference is barely meaningful. Our results show that longer articles in Wikipedia are more likely to include more slanted phrases. We interpret this as evidence for a novel and simple explanation for how disputes about contested knowledge are resolved in online communities. Online collective intelligence faces comparatively lower acquisition, storage, and distribution costs, enabling Wikipedia contributors to add a new fact or viewpoint to existing articles without the constraints Britannica faces when it wants to lengthen an article. Crucially, disputes about achieving a 'neutral point of view' on Wikipedia are often resolved through text additions rather than reductions. Moreover, Wikipedia lacks the organizational discipline that Britannica possesses, which keeps article length lower in Britannica. As far as we know, this is the first study to examine how this mechanism for addressing disputes in online communities affects the knowledge produced.

The paper proceeds as follows. In the next section, we provide a brief description of each organization, and then describe our theory development and hypotheses. After presenting our dataset and making general comparisons between the two sets of articles from the two production models, we present our regression results and a number of robustness tests for those results. We conclude with a discussion of the theoretical and practical implications of our findings.

BACKGROUND

Brief History of Britannica's Production

This study examines Britannica in 2012—its 244th year and the last time it was published in book form—when the substantial bulk of its revenue came from CD sales and online content. Online licensing accounted for 15% of the organization revenues, and educational curricula accounted for most of the other 85%. Book sales accounted for less than 1%. At its peak two decades earlier, Britannica had approximately half the market share for household encyclopedias, and was widely regarded as the most authoritative and widely consulted compendium of expert knowledge (Evans and Wurster 2000; Greenstein and Devereux 2009). While its initial decline in the 1990s was due to the rise of Encarta (a digital multimedia encyclopedia published by Microsoft Corporation from 1993 to 2009), many observers attributed the last post-millennial decline in Britannica's print sales to the rise of the Wikipedia online encyclopedia (Bosman 2012).

Throughout its long and august history, Britannica entries have been written by experts in every field imaginable, and edited by Britannica staff. Britannica's world famous sales force sold the product at high margins, a substantial fraction of which went to covering the fixed costs of employing its editorial staff to produce and organize its content into book format. The organization was a privately held company owned by the Benton Family: after the death of William Benton in 1973, ownership of the organization passed to a foundation that donated all its profits to the University of Chicago (Greenstein and Devereux 2009).

A large fraction of Britannica's content came from experts at little or no expense. Experts jumped at the chance to write entries in order to enhance their professional reputations. At the same time, the organization devoted considerable resources to maintaining its reputation as a comprehensive source of information. To prevent customers from perceiving Britannica as outdated, it issued annual reviews of newsworthy events of the prior year, and also operated a program guaranteeing customers the answer to any question not addressed in its volumes, which necessitated a large staff of researchers as well. Both programs fed into annual updates of Britannica's content in its book form, and more frequent updates of its online entries.

The organization also employed a large number of editorial staff. There was no set rule for the length of Britannica articles, but concerns about the overall length of the volumes played a significant role in the duties of editors. The sales department regarded additional length as a negative attribute, arguing that customers resisted buying books that took up too much shelf space. Management had fixed the total physical length of the encyclopedia for decades, and these concerns were enforced by an editing department that followed strict rules —no pages were ever added without an equal number being subtracted. Editors were hired and promoted on the basis of their ability to 'edit to fit,' i.e., to make content fit a prescribed length (Greenstein and Devereux 2009).

Even during its decline under competition from Encarta and Wikipedia, Britannica never turned away from relying on experts for sourcing its content, so the 2012 edition used in our study is still an excellent representative of 'expert-based content.'

Brief History of Wikipedia's Production

Wikipedia was founded in 2001, and after some initial challenges, positioned itself as 'the free encyclopedia that anyone can edit'—that is, as an online encyclopedia entirely written and edited via user contributions. Users could select any page to revise—expertise played no role in such revisions. By November 2011, it was the world's largest Wiki, supporting 4.8 million articles in English and over 36 million articles in all languages: it had become the world's largest 'collective intelligence' experiment, and one of the largest human projects to ever bring information into one source.[2]

Wikipedia is the largest application of wiki technology, which was developed in 1995. Wiki technology allows anyone to contribute content without special training, and enables the creation of hypertexts with nonlinear navigation structures: each page contains a series of cross-links to other pages, so readers can decide for themselves how to navigate through the site. Since 2003, Wikipedia has been owned and administered by the Wikimedia Foundation, a not-for-profit group established to manage the operations behind the Wikipedia website and related efforts.

At no time in its history has Wikipedia ever paid for content. Contributions come from tens of millions of dedicated contributors, who are not under any central control from the Wikimedia Foundation (Kane and Fichman 2009; Te'eni 2009; Zhang and Zhu 2011). All these voluntary contributors are considered editors on Wikipedia. The organization relies on contributors to discover and fix passages that do not meet the site's content tenets, but no central authority tells contributors how to allocate editorial time and attention. Available evidence on conflicts suggests that contributors who frequently work together do not get into as many conflicts, nor do their conflicts last as long (Piskorski and Gorbatai 2013). Additional evidence suggests a taste for prosocial and reciprocal behavior among contributors also plays an important role in fostering long-lasting cooperation among the participants (Algan et al. 2013).

A key aspiration for all Wikipedia articles is a 'neutral point of view' (Majchrzak 2009). To achieve this, "conflicting opinions are presented next to one another, with all significant points of view represented" (Greenstein and Zhu 2012). In practice, when multiple contributors make inconsistent contributions, other contributors devote considerable time and energy

[2] Source: https://en.wikipedia.org/?title=Wikipedia:Size_of_Wikipedia, accessed June 2015. Wikipedia is not yet the largest collection of knowledge in human history. The British Library has over 150 million items, and the Library of Congress has over 155 million, with 12 million searchable. The online version of Encyclopædia Britannica, which we use in this study, has 120,000 articles and 55 million words.

debating whether the article's text portrays a topic from a neutral point of view. As Wikipedia articles face no limits to their number or size—due to the absence of any significant storage costs—conflicts are often addressed by adding more points of view to articles, rather than eliminating them. Like all matters at Wikipedia, contributors have discretion to settle disputes on their own—no discipline or restraint comes from the center of the organization.

The Wikimedia Foundation makes no claims that the end product is ever 'finished.' It advises users to check the recent history of revisions, and not to treat any passage as definitive. In practice, hundreds of millions of readers do treat Wikipedia as comprehensive and definitive. For many searches, Google, the largest online search engine, displays the Wikipedia answer in a formatted box at the top of search results, and other question-answer sites also source from Wikipedia.

Over time a *de facto* norm has developed that keeps articles under 6-8 thousand words. As articles grow, contributors tend to either reduce their length, or split them into sub-topics to maintain the length norm. As we show below, the average Wikipedia article in our sample is shorter than this norm (just over 4,000 words), but the sample does include a few longer articles (the longest is over 20,000 words).

In summary, Wikipedia is a widely used reference site that is written by an online community. There is minimal organizational hierarchy, and all contributions are voluntary. Articles do not follow a fixed design, and quality control is allocated to contributors.

THEORY DEVELOPMENT AND HYPOTHESES

A substantial amount of work has considered problems associated with devising methods for sampling user-generated contributions (e.g., Luca 2015). Much of this work begins with an optimistic premise, presuming it is possible to tap the wisdom of the crowd (Surowiecki 2004), and the only questions are practical ones about methods, i.e., how to discover the best answer from a community of potential contributors (Afuah and Tucci 2012; Budescu and Chen 2014; Larrick and Soll 2012; Ray 2006). Much of this work concentrates on settings when there is a single 'right answer' hidden amongst a large group of contributors (e.g., Hasty et al. 2014). There is either one person in the crowd who has the answer, and the methods seek to identify this individual in a contest, or they compute averages (or some other statistic) to aggregate participant's estimates in a useful way (Page 2007), or aggregate many minor contributions into a whole text (such as in Wikipedia).

Existing thinking largely does not address online communities in which knowledge is contested—where content is controversial, unverifiable, or subjective—and so there is no single 'right answer.' In this section we develop hypotheses about how online collective intelligence performs in contested fields of knowledge, and, more specifically, whether and how well online communities present political information in comparison to expert presentations.

Number of Opinions

Experts have long held privileged positions in proffering knowledge, acting as arbiters of factual truth, and matching fact to theory. They play an important role in contested fields of knowledge, such as political discourse, in the same way that they do in many markets as shapers of opinion, and framers of public discourse (Abbott 1988; Reinstein and Snyder 2005). Encyclopedias such as Britannica have acted to compile such expertise and present it in digestible summaries.

An online compilation of contested knowledge can offer very different summaries of the same topics. For example, Wikipedia gets its articles from self-selected voluntary contributors. Rather than relying on a limited set of experts to summarize a topic, it draws its facts and opinions from many sources. And rather than relying on professional editors for text revisions, it synthesizes multiple revisions from many sources. A high membership turnover rate also allows new contributors to offer insights and knowledge the community previously did not possess (Ransbotham and Kane 2011). Because of a larger number of contributors to each article, the revisions occur more often in Wikipedia than in traditional encyclopedias, so online articles may also contain a wider range of opinions. Consistent with this view, Kane et al. (2014) find that adding new content—such as new perspectives on solutions to problems—is one of three core behaviors in the online knowledge production process.

The different methods used for settling disputes in online communities and the expert-based model may also contribute to the difference in the number of opinions sampled. As noted earlier, in traditional encyclopedia publishing, revisions follow ‘edit-to-fit’ norms because of space constraints, which can oblige editors to remove sections of text, potentially reducing the discussion about one point of view, when settling disputes. Britannica also possesses tools to readily enforce this norm across its organization, such as monitoring total page numbers in a volume, and assessing personnel’s ability to realize targets for length.

Wikipedia’s articles do not face the same pressures to reduce article length, as the cost of storing and displaying online text online is quite low, and makes longer articles more costly to produce. Wikipedia editors also lack any organizational mandate to reduce length. Therefore, in an online community one potential resolution to a dispute about contested knowledge is to add more discussion, because it costs very little to do so. This mechanism can result in longer articles with more contrasting opinions, so long as disagreements do not descend into reversion wars (Brown 2011; Yasseri et al. 2014). Hence we posit that:

Hypothesis 1 (H1): Online communities will present longer discussions of contested knowledge, and sample a greater number of opinions than expert-sourced knowledge.

Content Bias

Would the difference between these two kinds of organizations lead to their articles having different biases? We see several reasons why they might. An expert-based source gets its articles from authoritative contributors, and its staff selects those contributors on the basis of their authority and reputation. They work together to establish facts, and decide which opinions deserve attention, but these sources may reflect many biases in the expert consensus about a topic.

In contrast, by sampling facts and opinions from many sources, an online community produces content that no single individual would have produced otherwise. The diversity in contributors offer many advantages—for example, Ren et al. (forthcoming) show that both variety in tenure and interests between group members may increase group productivity; Ransbotham and Kane (2011) show that contributions from a mixture of new and experienced participants increase the quality of Wikipedia articles; Arazy et al. (2011) find that the creative abrasion generated when cognitively diverse members engage in task-related conflict leads to higher-quality Wikipedia articles; Østergaard et al. (2011) show that diversity of education and gender among employees is associated with better innovative performance in organizations; and Malhotra and Majchrzak (2014) point out that diversity is important for knowledge integration in crowds. In our setting, a diverse set of potential contributors to an article can help increase its likelihood of including facts and opinions that experts dismiss, and may present a rather different discussion of competing viewpoints (Page 2007). Benefitting from the efforts of many contributors, an article is also more likely to present controversial content in an unbiased way: thus diversity may help reduce content bias.

On the other hand, even when contributors to an online community aspire to write and edit entries that reflect a neutral point of view, they may differ in their interpretations of that goal and in their interpretations of what constitutes 'unbiased' content. Such communities bring together participants with "socially disembodied ideas" (Faraj et al. 2011) who may originate from communities with different traditions for expressing opinions, and different cultural and historical foundations for those opinions. In the absence of shared social context or work history, contributors can have difficulty developing mutual understanding (Hinds and Bailey 2003), integrating knowledge (e.g., Robert et al. 2008), or achieving convergence on solutions (Majchrzak et al. 2015). Conflicts may arise when contributors disagree as to whether knowledge should be changed or kept, and their intensity may grow with the number of contributors (e.g., Kittur and Kraut 2010). Although editing processes in online communities are often guided by norms and rules for effective collaboration (Butler et al. 2008), a high turnover rate and inability to hold anyone accountable may weaken their effectiveness: achieving consensus on neutral content among a disparate group of collaborators is likely to be a daunting task.

Although anyone can contribute to crowd-based knowledge, the contributions in online communities are often skewed, with relatively few contributors providing a disproportionate amount of the content (e.g., Ba and Wang 2013; Swartz 2006). Kane (2011) examined the development of the Wikipedia article on the 2007 Virginia Tech massacre and found that the top 10% of contributors contributed more than 60% of the content, and that most contributors (69%) contributed only once or twice. Hence, the article content may simply represent the viewpoint(s) of its most diligent and persistent contributors.

Group dynamics could also lead to content bias. In a less optimistic assessment, some theories suggest that crowds may be swayed through "group-cognition," tending towards the unique positions of the group that examines a topic, so that articles tend to reflect unique positions of those groups that examine a topic. Thus Wikipedia articles—especially those on narrow topics—could become swayed by relatively small groups (Barsade 2002; Frith and Frith 2012; Janis 1982).

Finally, self-selection among contributors in online communities may give rise to bias. Studies have shown that, in offline settings, consumers segregate their allegiances between sources, preferring to consume content that confirms their pre-existing views (see, e.g., Mullainathan and Shleifer 2005). As the Internet has made it easier for consumers to filter content according to their ideological preference, some analysts forecast an extreme form of self-selection among online readers. Sunstein (2001) frames this issue provocatively: "Our communications market is rapidly moving [toward a situation where] people restrict themselves to their own points of view—liberals watching and reading mostly or only liberals; moderates, moderates; conservatives, conservatives; Neo-Nazis, Neo-Nazis." Consistent with these studies, the psychology literature on confirmation bias suggests that individuals may selectively seek information that is consistent with their prior beliefs, or interpret ambiguous information in a manner that enhances their own beliefs, because confirmatory information reduces their psychological discomfort (e.g., Nickerson 1998; Oswald and Grosjean 2004; Park et al. 2013). The same concern applies to online knowledge production—if online communities with specific slants only attract contributors with similar ideologies, we expect knowledge generated by such organizations to exhibit strong ideological biases relative to that sourced from expert-based models.

These arguments suggest that it is unclear *ex ante* whether online communities perform better or worse than expert-based models in reducing bias. We therefore hypothesize:

Hypothesis 2a (H2a): Crowd-sourced knowledge contains more biased summaries of contested knowledge than does expert-sourced knowledge.

Hypothesis 2b (H2b): Crowd-sourced knowledge contains less biased summaries of contested knowledge than does expert-sourced knowledge.

The Role of Revisions in Crowd-Based Production

The discussion above also suggests that revisions could play an important role in shaping content bias in an online community. The revision process can help achieve a more neutral point of view for each article as the article can source from diverse contributions. But conflicts may arise when contributors with different ideological preferences start to collaborate with each other. For example, when the *Los Angeles Times* used a social media platform to capture opinions about the involvement of the U.S. military in Iraq, participants on one side of the debate deleted and replaced contributions from the other side (Ransbotham and Kane 2011; Wagner and Majchrzak 2006). Moreover, if contributors self-aggregate based on their ideological preferences to contribute to content, then revisions are likely to make the content even more biased. The revision process may also influence contributors' own viewpoints through a process known as group polarization, which describes the tendency of people to become more extreme in their viewpoints following interactions with other members of the group (e.g., Sia et al. 2002). Group polarization may happen when people try to 'outdo' each other by changing their positions more towards the direction valued by their group after being exposed to positions during such interactions (Isenberg 1986). Siegel et al. (1986) find that groups with members located in geographically dispersed areas tend to become more polarized than groups in which people communicate face to face.

We therefore do not know—without empirical investigation—whether revisions to online community articles make those articles more or less ideologically biased than those produced by expert-based models. We therefore hypothesize:

> *Hypothesis 3a (H3a): The greater the number of revisions to an article generated by an online community, the more different that article's political leanings will be from its counterpart produced by an expert-based source.*
>
> *Hypothesis 3b (H3b): The greater the number of revisions to an article generated by an online community, the less different that article's political leanings will be from its counterpart produced by an expert-based source.*

DATA

We examine a sample of articles from Wikipedia focused on broad and inclusive definitions of US political topics, including all Wikipedia articles that included the keywords 'Republican' or 'Democrat.' We gather a list of 111,216 relevant entries from the online edition of Wikipedia on June 8, 2012. Many of these articles concern events in countries other than the United

States, necessitating further assessment for relevance,[3] which reduces the list to 70,668 articles focused on US politics. This sample covers an enormous array of topics, including many controversial ones—such as on abortion, gun control, civil rights, taxation, and foreign policy—as well as many articles that lacked anything controversial, such as undisputed historical accounts of minor historical political events and biographies of comparatively obscure regional politicians. We compare this list of Wikipedia articles to all (120,000+) articles in the Britannica's online edition (also obtained on June 8, 2012)[4] and are able to identify 3,918 pairs of 'matching' articles. In 73% of the pairs the titles are identical: in the remainder they are nearly identical, and we check manually that the pairs covered similar topics. As we show below, these 3,918 articles cover a representative sample of topics from Wikipedia articles on US politics.

We measure slant and bias using methods developed by Gentzkow and Shapiro (2010 - hereafter G&S). An article's slant is a cardinal number, which we normalize to zero, so negative (positive) numbers represent a slant towards a Democratic (Republican) 'view.' The degree of bias is the absolute value of the slant; larger numbers indicate more bias than smaller numbers. Bias represents whether an article is "opinionated."

Following G&S, we investigated whether Wikipedia or Britannica articles used phrases favored more by Republican or by Democratic members of Congress. G&S select such phrases based on the number of times they appear in the text of the 2005 *Congressional Record*, applying statistical methods to identify those that separate Democratic from Republican representatives. This approach rests on the notion that each group uses a distinct 'coded'

---

[3] The words "Democrat" and "Republican" do not appear exclusively in entries about U.S. politics. If a country name shows up in the title or category names, we then check whether the phrase "United States" or "America" shows up in the title or category names. If yes, we keep this article. Otherwise, we search the text for "United States" or "America." We retain articles in which these phrases show up more than three times. This process allows us to keep articles on issues such as "Iraq War," but drop articles related to political parties in non-US countries.

[4] We checked the online edition of Britannica to ensure that, just like Wikipedia, it is constantly updated to incorporate the latest information.

language to speak to its respective constituents.[5] Each phrase is associated with a cardinal value representing how 'slanted' the phrase is. After offering considerable supporting evidence, G&S estimate the relationship between the use of each phrase and the ideology of each newspaper, using 1,000 phrases to identify whether those newspapers' views tend to be more aligned with Democrat or Republican ideologies. We label the phrases from the G&S lexicon as 'code words.'

G&S's approach has several key strengths: it has passed many internal validity tests and it avoids many subjective elements. It provides a general yardstick for measuring the bias of newspaper articles—Jelveh et al. (2014) demonstrate its effectiveness when examining political bias in articles in economic journals—which we believe can be transferred to the context of Internet articles. Wikipedia's contributors are unlikely to have used this yardstick to target these words for editing, though they might have included or excluded these phrases to try to represent or exclude a view. The method adds up numbers to get a total slant for an article. It considers an article 'unslanted' or 'unbiased' when it includes no code words, and also when it uses an equal numbers of Republican/Democrat code words with the same cardinal values.[6]

In general, just as there is no definitive way to measure the 'true bias' of a newspaper article in G&S, there is no definitive way to measure the 'true bias' of an online encyclopedia article. In this paper, however, every online article is paired with its own offline match, so we can net out any effects of mismeasurement that are common to the pair. It will thus be possible to say which article—from either Wikipedia or Britannica—is more slanted or biased. In addition, by comparing articles on the same topics from the two sources at the same point in time, we in effect control for any unobservable topic-specific or time-specific effects that might

[5] See Table I in Gentzkow and Shapiro (2010) for more examples.

[6] Greenstein and Zhu (2013) looked for—and found no evidence—that these two types of unslanted articles differ in their underlying traits. Hence, in this paper we treat them as identical.

influence their bias or slant.

Comparing Slants and Biases

Table 1: Comparing Slants in Wikipedia and Britannica Articles

	No. of Obs.	Wikipedia		Britannica		Mean
		Mean	Std. Dev.	Mean	Std. Dev.	Difference
Abortion	13	-0.14	0.23	-0.06	0.18	-0.07
American Politicians	438	-0.05	0.20	-0.05	0.19	0.00
Budgets	249	-0.02	0.16	-0.01	0.16	-0.02
Civil Rights	263	-0.15	0.26	-0.11	0.23	-0.03**
Corporations	28	-0.09	0.21	0.02	0.18	-0.11*
Crime	244	-0.04	0.19	-0.03	0.18	-0.01
Drugs	39	-0.02	0.23	-0.02	0.14	0.00
Education	311	-0.05	0.22	-0.01	0.15	-0.04***
Energy	52	-0.03	0.14	-0.02	0.13	-0.01
Family	126	-0.03	0.19	-0.03	0.13	0.00
Foreign Policy	524	0.01	0.17	0.01	0.13	-0.00
Trade	104	0.03	0.17	0.04	0.13	-0.01
Government	1183	-0.14	0.24	-0.05	0.17	-0.09***
Gun	9	-0.07	0.12	-0.13	0.16	0.07
Health Care	120	-0.03	0.24	-0.05	0.19	0.02
Homeland Security	132	-0.03	0.17	-0.04	0.19	0.01
Immigration	99	0.01	0.16	-0.02	0.14	0.04*
Infrastructure & Technology	277	-0.03	0.21	-0.02	0.13	-0.01
Employment	256	-0.03	0.19	-0.01	0.15	-0.01
Value	165	-0.05	0.22	-0.03	0.16	-0.03
Taxation	21	-0.15	0.22	-0.21	0.27	0.06
War & Peace	578	-0.01	0.17	-0.01	0.15	0.00
Welfare & Poverty	109	-0.03	0.19	-0.02	0.17	-0.01

*Note: We report both means and standard deviations for Wikipedia and Britannica articles, respectively: the last column shows the difference in means. We also test whether the difference is significantly different from zero. * significant at 10%; ** significant at 5%; *** significant at 1%.*

Table 2: Comparing Biases in Wikipedia and Britannica Articles

	No. of Obs.	Wikipedia		Britannica		Mean
		Mean	Std. Dev.	Mean	Std. Dev.	Difference
Abortion	13	0.19	0.19	0.08	0.18	0.11*
American Politicians	438	0.14	0.15	0.10	0.17	0.04***
Budgets	249	0.11	0.12	0.08	0.13	0.03**

Civil Rights	263	0.23	0.19	0.15	0.21	0.08***
Corporations	28	0.15	0.18	0.10	0.15	0.05
Crime	244	0.13	0.14	0.09	0.16	0.04***
Drugs	39	0.15	0.17	0.07	0.12	0.08**
Education	311	0.15	0.16	0.07	0.13	0.08***
Energy	52	0.10	0.09	0.09	0.1	0.02
Family	126	0.12	0.15	0.06	0.12	0.07***
Foreign Policy	524	0.12	0.12	0.08	0.11	0.04***
Trade	104	0.13	0.11	0.09	0.10	0.04***
Government	1,183	0.20	0.20	0.09	0.16	0.11***
Gun	9	0.10	0.09	0.15	0.15	-0.05
Health Care	120	0.16	0.18	0.10	0.17	0.06***
Homeland Security	132	0.13	0.12	0.12	0.16	0.01
Immigration	99	0.10	0.13	0.06	0.12	0.05***
Infrastructure & Technology	277	0.14	0.15	0.06	0.12	0.08***
Employment	256	0.13	0.14	0.07	0.14	0.06***
Value	165	0.17	0.16	0.08	0.15	0.09***
Taxation	21	0.20	0.17	0.23	0.26	-0.02
War & Peace	578	0.12	0.13	0.08	0.13	0.04***
Welfare & Poverty	109	0.14	0.14	0.10	0.13	0.04***

*Note: We report both means and standard deviations for Wikipedia and Britannica articles, respectively: the last column shows the difference in means. We also test whether the difference is significantly different from zero. * significant at 10%; ** significant at 5%; *** significant at 1%.*

Tables 1 and 2 break down these articles based on their topics, using category information defined in Wikipedia. As an article may be affiliated with more than one topic, the categories are not mutually exclusive. The most common topic is ‘Government,’ followed by ‘War and Peace,’ ‘Foreign Policy,’ and ‘American Politicians.’

The tables also show the slants and biases of articles in our sample, computing the mean and standard deviations for the average bias and slant for all articles in each category. Both Britannica and Wikipedia’s articles display considerable variance in the levels of bias and slant across topics. The two sources also track one another: the differences in slant between the two sources is insignificant in 19 of the 23 categories, but are quite pronounced in the other four categories. For example, Wikipedia entries about civil rights, corporations, and government have a more Democratic slant than those in Britannica, but entries on immigration have a more

Republican slant. Overall, Wikipedia articles appear to be mildly more slanted towards the Democratic 'view' than those published in Britannica. The findings for slant show that the articles from Wikipedia are often more biased than those from Britannica. In only five topic categories are these differences insignificant—in many topics they are considerable, with Wikipedia articles displaying more bias in every instance.

Table 3: Summary Statistics

Wikipedia Articles					
Variable	Obs.	Mean	Std. Dev.	Min	Max
Num of Code Words	3,918	6.12	12.30	0	239
Slant	3,918	-0.06	0.21	-0.61	0.62
Bias	3,918	0.14	0.17	0.00	0.62
Length	3,918	4,113.20	3,536.17	3.00	23,218
Num of Code Words/Length	3,918	0.0015	0.0025	0	0.040
Slant/Length	3,918	-0.00003	0.00019	-0.0042	0.0013
Bias/Length	3,918	0.00007	0.00018	0	0.0042
Contributors	3,918	839.50	1,077.40	1.00	14,160
Revisions	3,918	1,924.23	2,826.28	1	44,880

Britannica Articles					
Variable	Obs.	Mean	Std. Dev.	Min	Max
Num of Code Words	3,918	2.02	9.75	0	342
Slant	3,918	-0.02	0.15	-0.61	0.62
Bias	3,918	0.07	0.14	0.00	0.62
Length	3,918	1,778.28	8,179.78	7.00	155,874
Num of Code Words/Length	3,918	0.0015	0.0036	0	0.056
Slant/Length	3,918	-0.00006	0.00050	-0.0085	0.0063
Bias/Length	3,918	0.00015	0.00048	0	0.0085

Table 3 provides descriptive statistics for the entire matched sample dataset. At first sight, our finding about bias reflects the different frequencies of code words across the two sources. On average, Wikipedia articles contain more code words than Britannica articles: a much higher percentage of Wikipedia articles (73%) have at least one more code word than those published in Britannica (34%). It shows again that Wikipedia articles are more slanted towards Democratic viewpoints than are Britannica articles (although both are slanted towards

Democratic viewpoints), and they are also more biased. We also find that Wikipedia articles are longer than their Britannica matches (measured by the number of words in each article), as can be expected given Wikipedia's cheaper storage costs and different editorial processes. Although Britannica has the longest single article in our dataset, Wikipedia articles contain 4,113 words on average, and Britannica articles only 1,778 (43% as long). So Wikipedia articles are more likely to include code words because of their greater length. We also normalize the number of code words, slant, and bias by the article length. We find that, on a per-word basis, the number of code words is similar across the two sources, and Wikipedia articles lean less left and are less biased than Britannica articles. These results suggest that the difference in slant and bias may be associated with the length of the articles.

For Wikipedia articles, we are able to observe the number of contributors and the number of revisions for each article. We find wide variance for both measures, with the average Wikipedia article in this sample having 839 contributors (s.d. = 1,077) and 1,924 revisions (s.d. = 2,826). Because revisions are skewed, the summary statistics suggest that some articles may receive enough revisions to yield big changes in their slant and bias, and many will not.

REGRESSION RESULTS

We next examine the differences in slant and bias via a regression framework. Several factors may shape article-by-article comparisons simultaneously, so multivariate regression analysis can help yield additional insights about the causes.

Our dependent variables are the total number of code words, the slant or bias of each article. We create a dummy variable, *Wikipedia*, measured as 1 if the article is from Wikipedia and 0 if it is from Britannica. We use *Log (Length)*—the logarithm of article length —as a control variable: we log it because it is a positive and skewed variable. We use fixed-effects specifications at the matched article level to control for unobserved underlying slant or bias in the articles.

Table 4: Fixed-Effects Regressions Comparing Slant and Bias of Wikipedia and Britannica Articles

Panel A

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Num of Code Words	Num of Code Words	Slant	Slant	Bias	Bias
Wikipedia	1.502***	0.350***	-0.036***	-0.013***	0.074***	0.023***
	[0.030]	[0.023]	[0.004]	[0.005]	[0.003]	[0.004]
Log(Length)		0.789***		-0.013***		0.030***
		[0.014]		[0.002]		[0.002]
Observations	7,836	7,836	7,836	7,836	7,836	7,836
Adjusted R-squared			0.027	0.033	0.128	0.166
Number of Articles	3,918	3,918	3,918	3,918	3,918	3,918
Article Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Specification	Negative Binomial	Negative Binomial	OLS	OLS	OLS	OLS

Panel B

Model	(7)	(8)	(9)
Dependent Variable	Num of Code Words/Length	Slant/Length	Bias/Length
Wikipedia	-0.00002	0.00002***	-0.00008***
	[0.00005]	[0.00001]	[0.00001]
Observations	7,836	7,836	7,836
Adjusted R-squared	0.000	0.002	0.026
Number of Articles	3,918	3,918	3,918
Article Fixed Effects	Yes	Yes	Yes
Specification	OLS	OLS	OLS

*Note: Heteroskedasticity-adjusted standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.*

Models (1) and (2) in Panel A of Table 4 use *Num of Code Words* as the dependent variable, and use a negative binomial model. In both models, we find that (consistent with H1) Wikipedia articles tends to have more code words, suggesting that crowd-sourced knowledge tends to sample a greater number of opinions. Controlling for article length matters for the result, suggesting longer articles are more likely to have more code words. Models (3) and (4) use *slant* as the dependent variable. We find that, overall, Wikipedia articles are more Democratic-slanted than are Britannica articles. Once we control for length in Model (4), we also find that longer articles are more Democratic. The estimated coefficient of the length variable is of moderate size: a doubling in length (i.e., adding approximately 4,000 words on average to each article) generates a change towards the Democratic direction of approximately -0.01 in Wikipedia. Even with this control, the Wikipedia articles are still overall more Democratic (-0.01) than their Britannica counterparts.

We repeat the analysis using *bias* as the dependent variable in Models (3) and (4), and find Wikipedia articles to be more biased than Britannica articles. We thus find support for H2a. Again, article length is responsible for a substantial part of this difference—doubling the length generates an increase in the bias of approximately 0.3 for Wikipedia articles, which accounts for a major part of the difference between the average biases found in Wikipedia and Britannica articles.

The columns in Panel A try to account for the skewed distribution of article length by adding it as a control variable. As an alternative approach, we normalize our measures by the length of the article to capture number of code words, slant, and bias per word, and use them as our dependent variables in Models (7)-(9) in Panel B. In Model (7), we find that there is no significant difference between the number of code words at the per word level between the two sources. In Model (8), we find that the sign of the *Wikipedia* variable reverses—so Wikipedia articles are now more *right*-leaning than their Britannica counterparts at the per word level. But, since both Wikipedia and Britannica articles exhibit overall Democratic slants at the per word level (Table 3), this result suggests that Wikipedia articles are closer to neutral than their Britannica counterparts at the per-word level. Similarly, results from Model (9) confirm that Wikipedia articles are less biased than Britannica articles at the per word level.

Table 5: OLS Regressions to Examine the Impact of Revisions on Biases in Wikipedia Articles

Model	(1)	(2)	(3)	(4)
Dependent Variable	Wikipedia Bias	Wikipedia Bias	Wikipedia Bias	Wikipedia Bias
Britannica Bias	0.237***	0.245***	0.262***	0.224***

	[0.022]	[0.022]	[0.022]	[0.021]
Log(Length)	0.020***	0.033***	0.030***	0.025***
	[0.002]	[0.003]	[0.004]	[0.003]
Log(Revisions)		-0.011***	-0.018***	-0.014***
		[0.003]	[0.004]	[0.004]
Average Revisions Per Contributor		-0.002	0.007	0.004
		[0.004]	[0.004]	[0.004]
Year Created = 2002			0.065***	0.049***
			[0.007]	[0.007]
Year Created = 2003			0.006	0.006
			[0.008]	[0.008]
Year Created = 2004			0.009	0.005
			[0.010]	[0.010]
Year Created = 2005			-0.023*	-0.021*
			[0.013]	[0.013]
Year Created = 2006			-0.033*	-0.027
			[0.017]	[0.017]
Year Created = 2007			-0.038*	-0.028
			[0.019]	[0.019]
Year Created = 2008			-0.063***	-0.047**
			[0.020]	[0.020]
Year Created = 2009			-0.096***	-0.085***
			[0.015]	[0.017]
Year Created = 2010			-0.110***	-0.086***
			[0.017]	[0.017]
Year Created = 2011			-0.191***	-0.166***
			[0.016]	[0.019]
Dummies for Categories	No	No	No	Yes
Observations	3,918	3,918	3,918	3,918
Adjusted R-squared	0.067	0.071	0.109	0.185

*Note: Heteroskedasticity-adjusted standard errors in brackets. In Models (3) and (4), Year Created = 2001 is used as the benchmark group. * significant at 10%; ** significant at 5%; *** significant at 1%.*

We next examine how the Wikipedia revision process might change the bias of an article: in particular, we are interested in discovering whether articles become less biased as the number of revisions increase. To address this question, we use the bias of each Wikipedia article as the dependent variable, and the bias of its Britannica counterpart as a control. This model is valid under the assumption that Britannica's content is statistically exogenous, i.e., Britannica's writers did not alter their content in reaction to Wikipedia's content (we test this in the robustness check section).

We include two explanatory variables related to the revision process at Wikipedia. The first is *Log(Revisions),* the logarithm of the total number of revisions the article has already received, and since each contributor to a Wikipedia article can revise that article multiple times, we also include a measure of the average revisions per contributor for each article, *Average Revisions per Contributor*. We retain the logarithm of the length of the Wikipedia article as a control, and (in some specifications) add year dummies to indicate when the Wikipedia articles were created, as well as dummies for the articles' categories.

Table 5 reports the OLS regression results. We find that the correlation of bias between Wikipedia and Britannica is about 25% and is significant, and that Wikipedia articles that have received more revisions tend to be more neutral. In addition to the article length, the number of revisions contribute to the slant difference between Wikipedia and Britannica articles. The impact of revisions is not as strong as article length: doubling the number of revisions reduces bias by -0.01, but doubling article length increases bias by 0.03. However, the average number of revisions per contributor has no significant effect on the bias. The variable *Revisions* is skewed, so the articles receiving the most attention are much less biased, even when they are longer. However, most articles receive a mean number of 1,924 revisions or lower, and that is insufficient to erase the bias.

We also find that further controls add some nuance to the results. Articles created in early years tend to have more bias. The differences between 2002 and 2011 are the greatest, and the pattern is monotonic across all years in Model (4), which suggests that the greatest differences between Britannica and Wikipedia appear in the oldest articles. In summary, the biases from the two sources converge when articles have been heavily revised, even when they come from vintages with large biases: this is consistent with H3b.

To summarize, this econometric evidence is consistent with our conclusions from summary statistics. Wikipedia articles are likely to sample more opinions, and are longer than their Britannica counterparts. Longer articles and more revised Wikipedia articles are more likely to include more slanted phrases, which is consistent with a simple explanation for the mechanism for resolving disputes with contested knowledge: lower production and storage costs online enable contributors to add new facts or viewpoints to an existing article without the constraints faced offline, and thus online communities are more likely to *add* and *compare* new points of view in the same text in the face of contested knowledge.

Robustness Checks[7]

We conduct several checks to ensure our results are not driven by alternative explanations. Our first concern is that article lengths exhibit significant variations (as Table 3 shows), and longer articles are more likely to include code words, so it is theoretically possible that our results are mainly driven by outlying long articles. As a robustness check, we exclude all matched articles if either the Wikipedia or Britannica versions are longer than two standard deviations above the mean. 105 pairs fit these criteria. We obtain similar results when they are excluded, and so conclude that articles with outlying article lengths do not drive our results.

Our second concern is with a potential unintended consequence of our methods. Because our approach examines article slant conditional on the topic of the article, we are concerned that articles whose titles contain code words might exhibit more slant merely because those words are likely to be used many times in their texts. To ensure such examples were not driving our results, we identify all articles whose titles contain code words (50 pairs – 1.3% of the total), and exclude them from the analysis. Again, we obtain similar results.

Our next concern is with a subtle property of the G&S approach. It identifies two factors that shape slant and bias: (1) the choice of phrasing when there are multiple possible ways of

[7] We include all the results for robustness checks in the appendices.

describing the same concept (e.g., using ‘death tax’ or ‘estate tax’), and (2) the choice of topics (e.g., some newspapers may choose to run more articles about illegal immigration than others). By design, our study focuses much more on the former than the latter, i.e., the choice of phrasing conditional on the topic. Some phrases in G&S (e.g., ‘Saddam Hussein,’ ‘World Trade Organization,’ and ‘Endangered Species Act’), however, do not have natural variants that exhibit no or opposite slants. When such phrases are used, it is unclear whether they present actual slant or choice of content. To ensure that these special phrases do not drive an article’s slant, we recruit an experienced copy-editor with both an academic and legal background to go through the 1,000 code words to identify instances of variations in phrasing for the same concept, and check all the variations she identifies. This exercise reveals that 638 of the 1,000 code words have substitutes. We then repeat our analysis using these 638 words as our code words to measure slants and biases (essentially, ignoring any slant and bias arising from the remaining code words). Our results continued to hold.

The last of our robustness checks test the assumption of exogeneity of Britannica articles. While we can identify the dates when Wikipedia articles are created, we do not know when the matched Britannica articles are created, so it is possible that biases in Britannica articles might have arisen because some of them may have been altered by the experts in reaction to Wikipedia content. To address this concern, we obtain a copy of the Britannica edition for 2001 (the year when Wikipedia is founded) which must be exogenous, by design. Of the 3,918 Britannica articles in our dataset, 2,855 existed in the 2001 edition. When we repeat the analysis using only these 2,855 articles and their matched Wikipedia versions, we obtain similar results, supporting our assumption that biases in Britannica articles are exogenous to the processes that create bias in Wikipedia articles.

DISCUSSION AND CONCLUSION

The Wikipedia community and website represent a remarkable experiment in collective intelligence, and have carried the ideals of the notion further into practice than any other reference material on the Internet. In the ideal of collective intelligence, it should be possible to aggregate disparate ideas into a cohesive and presentable whole—but this would surely be difficult to accomplish even if all such ideas were uncontroversial, objective, and verifiable. This study has sought to examine the output of collective intelligence in a context where knowledge is contested.

We focus on contested knowledge, and specifically the factors that generate bias and slant in text, especially in settings where disagreements are likely. Our research objective leads us to ask a novel question about whether contributors with different ideologies engage in fruitful conversations with each other online and whether their output captures their respective points of view. Relatedly, we are also the first to suggest that the lower costs of producing, storing, and distributing knowledge, and the absence of organizational discipline to restrain additions, may lead to increased—rather than fixed or decreased—article length and thus shape different biases and slants in online communities,.

Our results suggest that, in comparison to expert-based knowledge, collective intelligence does not aggravate the bias of online content when articles are substantially revised. This is consistent with a best-case scenario in which contributors with different ideologies appear to engage in fruitful online conversations with each other, in contrast to findings from offline settings (e.g., Mullainathan and Shleifer 2005). In that light, we think this is an important and novel finding.

But, of course, collective intelligence does not always achieve its ideals. The absolute level of bias of Wikipedia articles remains higher than that of Britannica content, and varies considerably across content categories. On one level, this is not surprising, as Wikipedia contains an enormous corpus of text, and does not receive enough editorial or participant contributions to revise all of it, particularly in niche content categories. On the other hand, it is surprising because the average Wikipedia article receives over 1,900 revisions—but that is still not enough for eliminating bias. So, Wikipedia falls short of its ideal because it takes a lot of contributions to reduce considerable bias and slant. In other words, because it takes a lot of contributions to make any changes to bias and slant, there is not enough contribution and editorial attention to cover the full breadth of all contested articles.

Managerial Implications

The main reason many organizations still resist crowds is that managers do not clearly understand the pros and cons of the crowd compared to internal production (Boudreau and Lakhani 2013). Our results show that, indeed, crowds do not necessarily perform better than experts in every dimension. Our results also suggest that the allocation of editorial time and user contributions is central to the minimization of differences in bias and slant between organizational models. If editorial time and attention tends to go to the articles with the most readers, such an allocation minimizes the differences in readers' experiences of biases and slants in the two models. We note that the Wikimedia Foundation allocates discretion to a large community, and eschews central authority. It uses a large set of principles and norms for etiquette, but then asks its participants to decide how to implement them. As a result, it would be heroic to assume such an optimal allocation in such a highly decentralized organization as Wikipedia. There is some evidence that allocation of editorial attention is only weakly correlated with reader interest in Wikipedia in general, and for numerous reasons (Gorbatai 2014). Hence, our finding frames an open question about how organizations that depend on collective intelligence should conceptualize the optimal allocation of editorial time and user contributions.

We see no reason to be sanguine. Concerns about contested knowledge arise in discourse in a wide array of current events and scientific topics. As Wikipedia increasingly becomes many online readers' primary source of comprehensive information, there may be strong incentive for those with strong opinions to manipulate the site's content to foster their specific points of view. As the world moves from reliance on expert-based sources to collectively-produced intelligence, it seems unwise to blindly trust the properties of widely used information sources. Their slants and biases are not widely appreciated, nor are the properties of these organizational forms fully understood.

While this study focuses on a setting where we can implement a viable empirical strategy, the same dilemma arises in many communities beyond Wikipedia. For example, the largest for-profit Wiki, Wikia[8], hosts a wide set of topics for many communities in which the information is subjective, controversial, and unverifiable. The site was founded by Wikipedia alumni who were interested in topics that Wikipedia considered inappropriate, such as cooking, celebrity gossip, popular music, movies, gaming, and hobbies. Wikia uses principles and norms similar to those at Wikipedia to guide its communities of contributors and to attract readers (Greenstein et al. 2009). Knowledge communities are also formed around other technologies such as online bulletin boards and review systems (e.g., Ba et al. 2014; Bin and Ye 2014; Wasko and Faraj 2005). Our results imply two normative pieces of advice for such community sites: representing both sides of an issue typically takes a lot of contributions and considerable revisions; and length by itself is not usually sufficient to guarantee a balanced view, unless considerable revision is also involved.

A similar piece of managerial advice goes for aggregation efforts inside closed communities within private firms (Majchrzak et al. 2009; Surowiecki 2004; Wagner 2005; Wagner and Majchrzak 2006). Many private firms today use Wikis or other knowledge management technologies to organize their internal knowledge management efforts (e.g.,

[8] Source: http://www.wikia.com/, accessed March 2015.

Kankanhalli et al. 2005). Such tools are viewed as well suited for aggregating information from many unverifiable sources, but our results imply that this strength is also a potential weakness in the absence of close managerial oversight. There is considerable debate among practitioners about how closely to moderate such activities, because strong views can take over text if only a few employees participate regularly, and there are few revisions, so that the knowledge involved can easily become biased and slanted. Our findings favor the view that managers must do more than just offer guidelines—they must work towards a balanced view to ensure that intervention alleviates disputes and the right kind of principles for governing participation around contested areas of knowledge are generated.

Limitations and Future Research

Our study has several limitations. First, we focus on a large online community for knowledge production—it is not clear whether our results are generalizable to small online communities. In such communities, especially those within private firms, contributors may know each other or share similar social contexts, so it may be easier for the communities to develop mutual understandings of neutral content and also enforce norms. On the other hand, smaller communities are more likely to be prone to group thinking, and the lack of anonymity may reduce contributors' psychological safety (e.g., Edmondson 1999; Kang et al. 2013) and hence prevent them from expressing their own perspectives freely. It is thus not clear whether content from small communities will be more or less biased than that from large communities. Applying our approach to studying content produced by small communities would be an interesting area of future research.

Second, our empirical methods rely on phrases used by contributors to detect the ideological bias of each article. A drawback of this approach is that these phrases are only coarse proxies of contributors' complex underlying beliefs, and do not capture other dimensions such as positive or negative sentiments, which could influence readers' perceptions and acceptance of biased content. Future research can extend our empirical methodology to capture additional dimensions of knowledge produced.

Finally, we focus on ideological bias in our paper. Bias, however, can come in many forms, and other forms of bias, such as ethnic, racial, and gender bias (Hinnosaar 2015; Reagle and Rhue 2011) can also be consequential to the culture of organizations and our society. It is well known that multiple forms of biases can co-exist in online communities—for example, Wikipedia does not present knowledge traditionally associated with women with the same depth and attention paid as that associated with men (Knibbs 2014). Future research could aim to develop empirical methods to analyze different types of bias, and identify factors that minimize them.

References

Abbott A. 1988. *The System of Professions: An Essay on the Division of Expert Labor*, Chicago, IL: University of Chicago Press.

Afuah A., and Tucci C. 2012. "Crowdsourcing as a Solution to Distant Search," *Academy of Management Review* (37:3), pp. 355–375.

Agrawal A., Catalini C., and Goldfarb A. 2013. "Some Simple Economics of Crowdfunding," *Innovation Policy and the Economy* (14:1), pp. 1–46.

Agrawal A., Catalini C., and Goldfarb A. 2015. "Crowdfunding: Geography, Social Networks, and the Timing of Investment Decisions*," Journal of Economics & Management Strategy* (24:2), pp. 253–274.

Algan Y., Benkler Y., Morell M.F., and Hergueux J. 2013. "Cooperation in a Peer Production Economy: Experimental Evidence from Wikipedia," Working paper, NBER Workshop on the Economics of IT and Digitization.

Antweiler W., and Frank M.Z. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *Journal of Finance* (59:3), pp. 1259–1294.

Arazy, O., Nov, O., Patterson, R., and Yeo, L. 2011. "Community-Based Collaboration in Wikipedia: The Effects of Group Composition and Task Conflict on Information Quality," *Journal of Management Information Systems* (21:4), pp. 71–98.

Ba, S., and Wang, L. 2013. "Digital Health Communities: The Effect of their Motivation Mechanisms," *Decision Support Systems* (55:4), pp. 941–947.

Ba, S., Li, X., and Lu, X. 2014. "One Size Does Not Fit All: The Differential Impact of Online Reviews," Working Paper.

Banerjee, A.V.1992. "A Simple Model of Herd Behavior," *Quarterly Journal of Economics* (107:3), pp. 797–817.

Barsade, S.G. 2002. "The Ripple Effect: Emotional Contagion and its Influence on Group Behavior," *Administrative Science Quarterly* (47:4), pp. 644–675.

Bikhchandani S., Hirshleifer D., and Welch I. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy* (100:5), pp. 992–1026.

Bosman, J. 2012. "After 244 years Encyclopaedia Britannica Stops the Presses*," The New York Times*, March 13 (http://mediadecoder.blogs.nytimes.com/2012/03/13/after-244-years-encyclopaedia-britannica-stops-the-presses/?_r=0).

Boudreau, K.J., and Lakhani, K.R. 2013. "Using the Crowd as an Innovation Partner, *Harvard Business Review* (91:4), pp. 61–69.

Brown, A.R. 2011. "Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage," *Political Science and Politics* (44:2), pp. 339–343.

Budak, C., Goel, S., and Rao, J.M. 2014. "Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis," Working paper. Available at SSRN: http://ssrn.com/abstract=2526461.

Budescu, D.V., and Chen, E. 2014. "Identifying Expertise to Extract the Wisdom of Crowds," *Management Science* (61:2), pp. 267–280.

Butler, B.S., Joyce, and E., Pike, J. 2008. "Don't Look Now, But We've Created a Bureaucracy: The Nature and Roles of Policies and Rules in Wikipedia," *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, pp. 1101–1110.

Cohen, G.L. 2012. "Identity, Belief, and Bias." in *Ideology, Psychology, and Law,* J. Hanson and J. Jost (eds), Oxford, UK: Oxford University Press, pp. 385–404.

Edmondson, A. 1999. "Psychological Safety and Learning Behavior in Work Teams," *Administrative Science Quarterly* (44:4), pp. 350–383.

Evans, P., and Wurster, T.S. 2000. *Blown to Bits: How the New Economics of Information*

Transforms Strategy, Boston, MA: Harvard Business School Press.
Faraj, S., Jarvenpaa, S.L., and Majchrzak, A. 2011. "Knowledge Collaboration in Online Communities," *Organization Science* (22:5), pp. 1224–1239.
Frith, C.D., and Frith, U. 2012. "Mechanisms of Social Cognition," *Annual Review of Psychology* (63), pp. 287–313.
Galton, F. 1907. "Vox Populi," *Nature* (75:1949), pp. 450–451.
Gentzkow, M., and Shapiro, J.M. 2010. "What Drives Media Bias? Evidence from US Daily Newspapers," *Econometrica* (78:1), pp. 35–71.
Gentzkow, M., and Shapiro, J.M. 2011. "Ideological Segregation Online and Offline," *Quarterly Journal of Economics* (126:4), pp. 1799–1839.
Giles, J. 2005. "Internet Encyclopaedias Go Head to Head," *Nature* (438:7070), pp. 900–901.
Gorbatai, A. 2014. "The Paradox of Novice Contributions to Collective Production: Evidence from Wikipedia," Working Paper, University of California, Berkeley, CA.
Greenstein, S., and Devereux, M. 2009. "Wikipedia in the Spotlight," Kellogg School of Management Case No. 5-306-507, Northwestern University, Chicago.
Greenstein, S., Frazzano, R., and Meagher, E. 2009. "The Triumph of the Commons: Wikia and the Commercialization of Open Source Communities in 2009," Kellogg School of Management Case No. 5-309-509, Northwestern University, Chicago, IL.
Greenstein, S., and Zhu, F. 2012. "Is Wikipedia biased?" *American Economic Review: Papers and Proceedings* (102:3), pp. 343–348.
Greenstein, S., and Zhu, F. 2013. "Collective Intelligence and Neutral Point of View: The Case of Wikipedia," NBER Working Paper 18167.
Gu, B., Konana, P., Rajagopalan, B., and Chen, H-W. M. 2007. "Competition among Virtual Communities and User Valuation: The Case of Investing-Related Communities," *Information Systems Research* (18:1), pp. 68–85.
Gu, B., and Ye, Q. 2014. "First Step in Social Media - Measuring the Influence of Online Management Responses on Customer Satisfaction," *Production and Operations Management* (23:4), pp. 570–582.
Gulati, R., Puranam, P., and Tushman, M. 2012. "Meta-organization Design: Rethinking Design in Interorganizational and Community Contexts," *Strategic Management Journal* (33:6), pp. 571–586.
Hasty, R.T., Garbalosa, R.C., Barbato, V.A., Valdes, P.J., Powers, D.W., Hernandez, E., John, J.S., Suciu, G., Qureshi, F., Popa-Radu, M., San Jose, S., Drexler, N., Patankar, R., Paz, J.R., King, C.W., Gerber, H.N., Valladares, M.G., and Somji, A.A. 2014. "Wikipedia vs. Peer-reviewed Medical Literature for Information about the 10 Most Costly Medical Conditions," *Journal of the American Osteopathic Association* (114:5), pp. 368–373.
Hinds, P.J., and Bailey, D.E. 2003. "Out of Sight, Out of Sync: Understanding Conflict in Distributed Teams," *Organization Science* (14:6), pp. 615–632.
Hinnosaar, M. 2015. "Gender Inequality in New Media: Evidence from Wikipedia," Working paper, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2617021.
Isenberg, D.J. 1986. "Group Polarization: A Critical Review and Meta Analysis," *Journal of Personality and Social Psychology* (50:6), pp. 1141–1151.
Janis, I.L. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*, University of California: Houghton Mifflin.
Jelveh, Z., Kogut, B., and Naidu, S. 2014. "Political Language in Economics," Columbia Business School Research Paper No. 14-57.
Kane, G.C. 2011. "A Multimethod Study of Information Quality in Wiki Collaboration," *ACM Transactions on Management Information Systems* (2:1), Article 4.
Kane, G.C., and Fichman, R.G. 2009. "The shoemaker's children: Using wikis for information systems teaching, research, and publication," *MIS Quarterly* (33:1), 1–17.

Kane, G.C., Johnson, J., and Majchrzak, A. 2014. "Emergent Life Cycle: The Tension Between Knowledge Change and Knowledge Retention in Open Online Co-production Communities," *Management Science* (60:12), pp. 3026–3048.

Kankanhalli, A., Tan, B. C-Y., and Wei, K-K. 2005. "Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation," *MIS Quarterly* (29:1), pp. 113–143.

Kang, R., Brown, S., and Kiesler, S. 2013. "Why Do People Seek Anonymity on the Internet? Informing Policy and Design," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2657–2666.

Kittur, A., Kraut, E.R. 2010. "Beyond Wikipedia: Coordination and Conflict in Online Production Groups," *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, pp. 215–224.

Knibbs, K. 2014. "Chipping Away at Wikipedia's Gender Bias, One Article at a Time," *The Daily Dot*, February 10 (http://bit.ly/LQ9gDt).

Kummer, M.E., Saam, M., Halatchliyski, I., and Giorgidze, G. 2012. "Centrality and Content Creation in Networks – The Case of German Wikipedia," ZEW Discussion Paper No. 12-053.

Lakhani, K.R. 2009. "InnoCentive.com (A)," Harvard Business School Case 608-170, Boston.

Lakhani, K.R., Garvin, D.A., and Lonstein, E. 2012. "TopCoder (A): Developing Software Through Crowdsourcing," Harvard Business School Case 610-032, Boston.

Larrick, R., and Soll, J. 2012. "Intuitions about Combining Opinions: Misappreciation of the Averaging Principle." *Management Science* (52:1), pp. 111–127.

Lee, G.K., and Cole, R.E. 2003. "From a Firm-based to a Community-based Model of Knowledge Creation: The Case of the Linux Kernel Development," *Organization Science* (14:6), pp. 633–649.

Lemos, R. 2004. "Security Research Suggests Linux has Fewer Flaws," *CNET News*, December 13 (http://news.cnet.com/Security-research-suggests-Linux-has-fewer-flaws/2100-1002_3-5489804.html).

Lih, A. 2009. *The Wikipedia Evolution: How a Bunch of Nobodies Created the World's Greatest Encyclopedia*, New York, NY: Hyperion.

Luca, M. 2015. "User Generated Content and Social Media," Working Paper, Harvard Business School.

Mackay, C. 1852. *Extraordinary Popular Delusions and the Madness of Crowds*, London, UK: Richard Bentley.

Malhotra, A., and Majchrzak, A. 2014. "Managing Crowds in Innovation Challenges," *California Management Review* (56:4), pp. 103–123.

Majchrzak, A. 2009. "Comment: Where Is the Theory in Wikis?," *MIS Quarterly* (33:1), pp. 18–20.

Majchrzak, A., Malhotra, A., and Mertens, A. 2015. "Greater Innovation by the Crowd in Crowdsourcing: The Sequencing of Knowledge Types that Balance Divergence and Convergence," Working Paper.

Majchrzak, A., Cherbakov, L., and Ives, B. 2009. "Harnessing the Power of the Crowds with Corporate Social Networking Tools: How IBM Does It," *MIS Quarterly Executive* (8:2), pp. 103–108.

Melcher, R.A. 1997. "Dusting off the Britannica." *BusinessWeek*, October 20 (http://www.businessweek.com/1997/42/b3549124.htm).

Mollick, E.R., Nanda, R. Forthcoming. "Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts," *Management Science.*

Mullainathan, S., and Shleifer A. 2005. "The Market for News," *American Economic Review* (95:4), pp. 1031–1053.

Nickerson, R.S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology* (2:2), pp. 175–220.

Østergaard, C.R., Timmermans, B., and Kristinsson, K. 2011. "Does a Different View Create Something New? The Effect of Employee Diversity on Innovation," *Research Policy* (40:3), pp. 500–509.

Oswald, M.E., and Grosjean, S. 2004. "Confirmation Bias," in *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory,* R.F. Pohl (ed), Hove, UK: Psychology Press, pp. 79–96.

Page, S.E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies,* Princeton, NJ: Princeton University Press.

Park J., Konana, P.C., Gu, B., Kumar, A., and Raghunathan, R. 2013. "Information Valuation and Confirmation Bias in Virtual Communities: Evidence from Stock Message Boards," *Information Systems Research* (24:4), pp. 1050–1067.

Piskorski, M.J., and Gorbatai, A. 2013. "Testing Coleman's Social-norm Enforcement Mechanism: Evidence from Wikipedia," Working Paper, Harvard Business School, Boston.

Rajagopalan, M.S., Khanna, V.K., Leiter, Y., Stott, M., Showalter, T.N., Dicker, A.P., and Lawrence, Y.R. 2011. "Patient-oriented Cancer Information on the Internet: A Comparison of Wikipedia and a Professionally Maintained Database," *Journal of Oncology Practice* (7:5), pp. 319–323.

Ransbotham, S., and Kane, G.C. 2011. "Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia," *MIS Quarterly* (35:3), pp. 613–627.

Ray, R. 2006. "Prediction Markets and the Financial Wisdom of Crowds," *Journal of Behavioral Finance* (7:1), pp. 2–4.

Reagle, J., and Rhue L. 2011. "Gender Bias in Wikipedia and Britannica," *International Journal of Communication* (5), pp. 1138–1158.

Reinstein, D.A., and Snyder, C.M. 2005. "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics," *Journal of Industrial Economics* (53:1), pp. 27–51.

Ren, Y., Chen, J., and Riedl, J. Forthcoming. "The Impact and Evolution of Group Diversity on Online Collaboration," *Management Science*.

Ren, Y., Kraut, R., and Kiesler, S. 2007. "Applying Common Identity and Bond Theory to Design of Online Communities," *Organization Studies* (28:3), pp. 377–408.

Robert, L., A. R. Dennis, A., and M. Ahuja. 2008. "Social Capital and Knowledge Integration in Digitally Enabled Teams," *Information Systems Research* (19:3), pp. 314–334.

Shankland, S. 2003. "Study Lauds Open-source Code Quality," *CNET News*, February 19 (http://news.cnet.com/Study-lauds-open-source-code-quality/2100-1001_3-985221.html).

Sia, C.L., Tan, B., and Wei, K.K. 2002. "Group Polarization and Computer-mediated Communication: Effects of Communication Cues, Social Presence, and Anonymity," *Information Systems Research* (13:1), pp. 70–90.

Siegel, J., Dubrovsky, V., Kiesler, S., and McGuire, T.W. 1986. "Group Processes in Computer-mediated Communication," *Organizational Behavior and Human Decision Processes* (37:2), pp. 157–187.

Sunstein, C.R. 2001. *Republic.com,* Princeton, NJ: Princeton University Press.

Surowiecki, J. 2004. *The Wisdom of Crowds,* New York, NY: Random House.

Swartz, A. 2006. "Who Writes Wikipedia?," September 4 (http://www.aaronsw.com/weblog/whowriteswikipedia).

Te'eni, D. 2009. "Comment: The Wiki Way in a Hurry—The ICIS Anecdote," *MIS Quarterly* (33:1), pp. 20–22.

Tetlock, P. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?,* Princeton, NJ: Princeton University Press.

Wagner, C. 2005. "Supporting Knowledge Management in Organizations with Conversational Technologies: Discussion Forums, Weblogs, and Wikis," *Journal of Database Management* (16:2), pp. i–viii.

Wagner, C., and Majchrzak, A. 2006. "Enabling Customer-centricity Using Wikis and the Wiki Way," *Journal of Management Information Systems* (23:3), pp. 17–43.

Wasko, M. M., and Faraj, S. 2005. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS Quarterly* (29:1), pp. 35–57.

Wei, Z., and Lin M. 2015. "Market Mechanisms in Online Crowdfunding," Working Paper.

Wu, L., and Brynjolfsson, E. 2013. "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales," Working Paper.

Xu, S.X., and Zhang, X. M. 2014. "Impact of Wikipedia on Market Information Environment: Evidence on Management Disclosure and Investor Reaction," *MIS Quarterly* (37: 4), pp.1043–1068.

Yasseri, T., Spoerri, A., Graham, M., and Kertesz, J. 2014. "The Most Controversial Topics in Wikipedia: A Multilingual and Geographical Analysis," in *Global Wikipedia: International and Cross-Cultural Issues in Online Collaboration*, P. Fichman and N. Hara (eds), Lanham, MD: Rowman & Littlefield, pp. 25–48.

Zhang, J., and Liu, P. 2012. "Rational Herding in Microloan Markets," *Management Science* (58:5), pp. 892–912.

Zhang, X. M., and Zhu, F. 2011. "Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia," *American Economic Review* (101:4), pp. 1601–1615.