Do Experts or Crowd-based Models Produce More Bias? Evidence from Encyclopædia Britannica and Wikipedia*

Shane Greenstein
sgreenstein@hbs.edu

Feng Zhu
fzhu@hbs.edu

Organizations today can use both crowds and experts to produce knowledge. While prior work compares the accuracy of crowd-produced and expert-produced knowledge, we compare bias in these two models in the context of contested knowledge, which involves subjective, unverifiable, or controversial information. Using data from Encyclopædia Britannica, authored by experts, and Wikipedia, an encyclopedia produced by an online community, we compare the slant and bias of pairs of articles on identical topics of US politics. 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 in bias between a pair of articles decreases with more revisions. The bias on a per word basis hardly differs between the sources because Wikipedia articles tend to be longer than Britannica articles. These results highlight the pros and cons of each knowledge production model, help identify the scope of the empirical generalization of prior studies comparing the information quality of the two production models, and offer implications for organizations managing crowd-based knowledge production.

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

* We thank Bin Gu, 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.

INTRODUCTION

Recent technological advances have made it significantly easier for organizations to harness the collective intelligence of online communities (e.g., Afuah and Tucci 2012; Gu et al. 2007; Kane and Fichman 2009; Zhang and Zhu 2011; Gorbatai 2014; Gallus forthcoming). A few studies have addressed the skepticism about the wisdom of the crowd by demonstrating that collective decision making can be more accurate than that of experts (e.g., Antweiler and Frank 2004; Galton 1907; Giles 2005; Rajagopalan et al. 2011; Shankland 2003). But beyond accuracy, little is known about how well collective decision making performs along other dimensions of information quality, such as objectivity, consistency, relevance, and timeliness. We address this gap by examining two broad and related questions in the context of contested knowledge, defined loosely as a debate between viewpoints in which there is no single "right answer": (1) Is knowledge produced by crowd-based organizations more or less biased than the knowledge produced by experts? (2) What key factors make the difference larger or smaller?

We address our two research questions by examining the entries about US politics in Wikipedia, the largest online encyclopedia, and Encyclopædia Britannica, the most popular offline English-language encyclopedia. Wikipedia relies on tens of millions of volunteers to generate its content. 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 each is the most common reference source in its domain. In addition, they resolve disputes with different decision-making processes. Editors at Britannica resolve disputes in consultation with experts and "edit-to-fit" proscribed article lengths, while Wikipedia resolves disputes through decentralized discretion of decision making to contributors.[1]

In addition to being the first paper to compare crowd-based and expert-based models in the context of contested knowledge, we develop an empirical approach to address two issues. First, a central challenge in knowledge production is that disputes cannot be resolved by referencing authoritative information from "outside" the dispute. Researchers on this topic face the same challenge: because they cannot accurately identify "the truth," it is virtually impossible for them to meaningfully quantify biases. Second, some topics are inherently more slanted or biased than others. And as articles from both sources are constantly updated, their bias and slant can be changed by many factors. For example, some random events (i.e., when the phrase "death tax" is created) could make some topics more biased. As a result, statistical identification requires controlling for unobserved factors that shape slants and biases of each topic, especially when those unobserved factors are shared by two sources. We overcome both challenges by developing a matched sample of paired articles appearing *at the same time* in both sources and covering *identical or nearly identical* topics in US politics. For any matched pair of articles, we compare the slant and bias. Direct comparison of matched articles allows us to focus on which source is more biased than the other without identifying the truth for each topic. It also controls for otherwise unobserved factors shared by the two articles of the same topic.

We find that the bias of content sourced from collective intelligence differs from the bias of content sourced from experts. Overall, we find that Wikipedia articles display greater bias and, importantly, substantial revisions of Wikipedia articles reduce the differences in biases and slants to negligible statistical differences (between sources). We also find that, on average, Wikipedia's article on a given topic is longer than Britannica's and is slightly less biased on a per-word basis,

[1] See, respectively, Greenstein (2016) and Greenstein and Zhu (2012, 2016), for descriptions about the production models of Britannica and Wikipedia.

although the statistical difference is barely meaningful.

LITERATURE REVIEW

Our paper is related to several streams of literature. First, it contributes to the literature on collective intelligence in the crowd (e.g., Mackay 1852, Janis 1982, Park et al. 2013, Barsade 2002). Recent studies have begun to characterize the properties of online collective decision making in a variety of situations. Some studies show that collective decision making can generate high-quality output and can sometimes be more accurate than experts' decision making (e.g., Antweiler and Frank 2004; Chesney 2006; Galton 1907; Giles 2005; Lemos 2004; Rajagopalan et al. 2011; Shankland 2003; Surowiecki 2004). On the other hand, several studies find the opposite (e.g., Rector 2008). Still others take a middle ground. For example, Mollick and Nanda (forthcoming) examine crowdfunding and traditional venture funding and stress that collective decisions can exhibit tastes or preferences that traditional sources do not.

Much work has considered the problems of devising methods for sampling user-generated contributions (e.g., Luca 2015), often optimistically assuming that it is possible to tap the wisdom of the crowd (Surowiecki 2004) and seeking the best methods (Afuah and Tucci 2012; Budescu and Chen 2014; Larrick and Soll 2012; Ray 2006). Underlying this optimism is a presumption that there is a *single* "correct answer" amongst the many contributions (e.g., Hasty et al. 2014). While the quality of crowd-produced knowledge has been demonstrated to be comparable to expert-produced knowledge in cases of uncontested knowledge, no paper has identified the scope of the empirical generalization in a setting in which knowledge is contested (Barwise 1995).

Second, our paper is related to the literature on information quality, for which scholars have defined multiple important dimensions, such as accuracy, objectivity, consistency, relevance, and timeliness (e.g., Eppler and Wittig 2000; Madnick and Wang 1992; Miller 1996; Stvilia et al. 2008; Wang and Strong 1996; Xu and Zhang 2014). Objectivity is difficult to evaluate without expert input (Klein 2001; Naumann and Rolker 2000), which often limits the amount of information that can be evaluated. Similar to Jelveh et al. (2014) and Greenstein and Zhu (2016), we use an approach that allows research to systematically assess bias in a large number of articles without expert input. These papers differ in several key dimensions. First, the two papers focus on different research questions. Greenstein and Zhu (2016) examine the evolution of political articles at Wikipedia over ten years to answer questions such as whether the Internet has a tendency to increase ideological segregation. As a result, it does not address questions related to differences between Wikipedia and Britannica specifically, or expert and crowd-source production broadly, as this paper does. This paper builds on the literature that makes direct comparison between crowd-generated content and experts-generated content, such as Mollick and Nanda (forthcoming). The comparison yields strong inferences, but comes with extraordinary research challenges: It requires us to develop data and hypotheses that compare both production models and account for their salient features. The empirical approaches and the datasets also are different – Greenstein and Zhu (2016) identifies inferences from panel regressions of all Wikipedia's political articles, while this paper generates identification from cross-sectional analysis of matched pairs of articles. In sum, the research objectives, the datasets, the theory development and narrative, and the core empirical approaches and the extended tests of inferences are all different.

Third, our paper is related to the literature on mechanisms of self-organization in online communities, such as coordination mechanisms (Kittur and Kraut 2008; Kittur et al. 2007a, 2007b; Kittur et al. 2008; Schroeder and Wagner 2012), social interactions (e.g., Forte et al. 2012; Halfaker et al. 2011), an extensive set of rules, norms, and policies (Forte et al. 2009; Jemielniak

2014; Schroeder et al. 2012), quality assurance procedures (Stvilia et al. 2008), conflict resolution mechanisms (Arazy et al. 2011), and a comprehensive scheme of access privileges that formally defines organizational roles (Arazy et al. 2015; Burke et al. 2008; Collier et al. 2008; Forte et al. 2012). Our study complements them by comparing the output of an online community to that of an expert-based community, treating these self-organization mechanisms as a given feature of the setting.

HYPOTHESIS DEVELOPMENT

We focus on the differences between the expert- and crowd-based production models in two areas: the amount of content bias and the role of revisions in changing the difference between biases from the two product models.

Content Bias

Sampling a large number of opinions from the online community could introduce more biases to content when compared to the expert-based approach for several reasons. First, the contributors to an expert-based source are selected for their authority, objectivity, and reputation. While this selection process could introduce biases, it will also limit them. Indeed, reducing the number of contributors may reduce biases in the output. Bias may also be kept in check if the experts operate with norms for presenting biases in debates where knowledge is contested.

Second, several prior studies find that crowds may draw more diverse contributors, which can offer advantages such as greater productivity, more innovation, and better knowledge integration in a contested debate (e.g., Østergaard et al. 2011; Sun and Zhu 2013; Malhotra and Majchrzak 2014; Ren et al. forthcoming). However, crowd-sourced content can also produce a large sample with a great variety of biased opinions.

Norms and processes for dispute resolution will also play a role in online communities.

Even when contributors aspire to write and edit entries that reflect a neutral point of view, their assessments of what constitutes unbiased content may differ. Such communities bring together participants with "socially disembodied ideas" (Faraj et al. 2011) and with different traditions for expressing opinions and different cultural and historical foundations for those opinions. In the absence of a shared social context or work history, it can be difficult for contributors to develop mutual understanding (Hinds and Bailey 2003), integrate knowledge (e.g., Robert et al. 2008), or achieve convergence on solutions to unresolved conflicts in opinion (Majchrzak et al. 2015). Conflicts may arise when contributors disagree as to whether knowledge should be changed or kept and may become more intense as the number of contributors increases (e.g., Kittur and Kraut 2010). Although editing processes in online communities are often guided by norms and rules (Butler et al. 2008), a high turnover rate and an inability to hold anyone accountable may weaken their effectiveness. Thus, the difficulty of achieving consensus on neutral content among a disparate group of collaborators may allow many biases to remain embedded in the content.

The composition of the contributors may also shape biases. Although anyone can contribute to crowd-based knowledge, the contributions are often skewed, with relatively few contributors providing a disproportionate amount of the content (e.g., Ba and Wang 2013; Swartz 2006). For example, Kane (2011) examined the development of the Wikipedia article on the 2007 Virginia Tech massacre and found that the top 10 percent of contributors contributed more than 60 percent of the content and that 69 percent of the contributors only contributed once or twice. The article's content thus contained the viewpoint(s) and bias(es) of its most diligent and persistent contributors. An implication is that the self-selection of the most significant contributors could make the content more biased than in an expert-based production model.

Group dynamics could also lead to content bias. Crowd-produced content, especially on narrow topics, could become biased by the influence of relatively small groups (Barsade 2002; Frith and Frith 2012; Janis 1982; Sun et al. 2017); bias is more likely to occur in such settings than in settings involving a few experts.

Additional aspects of self-selection may also give rise to bias. As the Internet has made it easier for consumers to filter content according to their ideological preferences, some analysts forecast an extreme form of self-selection among online readers. For example, Kitchin (1998) finds that people often sample a large number of communities and migrate to those that confirm their own viewpoints. As a result, an online community could, over time, become less diverse and more dysfunctional (Arazy 2011). Consistent with these studies, the psychology literature on confirmation bias suggests that individuals may seek information that is consistent with their beliefs or interpret ambiguous information so as to enhance their beliefs because confirmatory information reduces their psychological discomfort (e.g., Nickerson 1998; Oswald and Grosjean 2004; Park et al. 2013). If online communities with specific slants only attract contributors with similar ideologies, we expect knowledge generated by such organizations to exhibit strong ideological biases compared to that sourced from expert-based models.

We therefore hypothesize:

Hypothesis 1 (H1): On a given topic, crowd-sourced knowledge contains more biased summaries of contested knowledge than does expert-sourced knowledge.

The Role of Revision in Crowd-based Production

The revision process plays an important role in the development of online content and differs substantially from the production of knowledge using experts. In expert-based production, experts negotiate with editors and revise accordingly. Such editing occurs before publication and is typically not repeated. In online communities, however, revision is frequent. With repeated revision, the content is more likely to obtain input from groups with different viewpoints. As long as each group is of sufficient size, the negotiation process could help achieve a more neutral point of view by including different viewpoints. For example, Kane and Fichman (2009) show an example of two camps of opposing editors confronting one another over gun control in Wikipedia. The advocates reached a compromise on how to define the message. Neither side persuaded the other of its points; rather, the opposing camps settled on expressions for each point of view.

Revision could also reduce bias if online content receives early edits from passionate advocates who are drawn to starting articles and revising newer articles. Over time, as the content receives revision from contributors with less extreme viewpoints, it may become more neutral. Consistent with this view, Kittur et al. (2007a) study Wikipedia and del.icio.us, a social bookmarking website, and find that, in both cases, most contributions initially come from a small set of passionate users. Gradually, however, there are more edits by another population of contributors, often those who only seldom edit.

In addition, many online communities share a norm known as Linus' Law: "given enough eyeballs, all bugs are shallow" (Greenstein and Zhu 2016; Kittur and Kraut 2008). While Linus' Law emerged in the domain of crowd-developed code, many assume that it applies other online domains. This belief fosters a tendency to allow the revision process to offset the negative effects of self-selection and group cognition and make the content less biased.

We therefore hypothesize:

Hypothesis 2 (H2): When crowd-generated content receives a greater number of revisions, its bias will differ less from that of comparable content produced by experts.

METHODOLOGY

To compare contested knowledge in two settings, we face novel statistical challenges. First, some topics are inherently slanted and biased. Second, an article's bias can change as editors revise it to improve the writing or to incorporate new information. To overcome these challenges, we develop a matched sample that compares paired articles which (a) appear at the same time in both sources and (b) cover identical topics.

We examine a sample of Wikipedia articles on broad US political topics, including all Wikipedia articles that included the keywords "Republican" or "Democrat." We first gather a list of 111,216 entries from Wikipedia on June 8, 2012. Many of these articles concern events outside the United States. We apply the procedure developed in Greenstein and Zhu (2016) to drop such articles and obtain a sample of 70,668 articles focused on US politics. We compare this list of Wikipedia articles to all 120,000-plus articles in Britannica's online edition (also obtained on June 8, 2012)[2] and identify 3,918 pairs of matching articles. We check manually that the pairs covered similar topics. In 73 percent of the pairs, the titles are identical; in the remainder, they are nearly identical. These 3,918 articles cover a representative sample of topics on US politics.[3] We then measure slant and bias of these articles, adapting the methods of Gentzkow and Shapiro (2010). Tables 1 and 2 break down these articles according to topic categories defined in Wikipedia. An article may be in more than one category. The most common is "Government," followed by "War and Peace," "Foreign Policy," and "American Politicians." The tables show the slants and biases of articles in our sample, computing the mean and standard deviation for the average slant and bias

[2] We checked the online edition of Britannica to ensure that, just like Wikipedia, it has been updated periodically to incorporate the latest information.

[3] We also consider the representativeness of our sample by comparing the matched sample against the much larger initial set of 70,668 Wikipedia articles (see Appendix Table 1). We observe no significant troubling features in our matched sample. Overall, the matched sample and the original sample are roughly in proportion with each other, with a few obvious exceptions. The most common topic is "Government," followed by "War and Peace," "Foreign Policy," and "American Politicians." The matched sample has a large representation of American political biographies, which we think is a byproduct of how easy it is to exactly match biographies of individuals across the two samples. The matched sample also comparatively over-represents entries about "Government," although that seems innocuous since this category label is used so frequently in Wikipedia. Underrepresented areas, such as "Education," "Foreign Policy," "War and Peace," and "Infrastructure and Technology," do not reveal any obvious selection issues.

for all articles in each category. Both Britannica's and Wikipedia's articles display considerable variance in slant and bias across topics. The two sources also track one another: the difference in slant between the two sources is insignificant for 19 of the 23 categories, but quite pronounced in the other four. Wikipedia entries about civil rights, corporations, and government have a more Democratic slant than those in Britannica, while entries on immigration have a more Republican slant. Overall, Wikipedia articles appear to be mildly more Democratic-slanted than Britannica articles. We also find that Wikipedia articles are often more biased than their Britannica counterparts. 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 provides descriptive statistics for the entire matched sample dataset. At first sight, the finding about bias and slant in Tables 1 and 2 reflects the different frequencies of code phrases across the two sources. On average, Wikipedia articles contain more code phrases than Britannica articles: a much higher percentage of Wikipedia articles (73%) have at least one more code phrase than those published in Britannica (34%). Although both sources are slanted towards Democratic viewpoints, Wikipedia articles are more slanted and more biased. We also find that Wikipedia articles are longer, measured by number of words, than their Britannica matches, unsurprising given Wikipedia's cheaper storage costs and Britannica's "edit-to-fit" editorial process. Although Britannica has the longest single article in our dataset, the average lengths for Wikipedia and Britannica articles are 4,113 and 1,778, respectively. Wikipedia articles are more likely to include code phrases because of their greater length. Normalizing the slant, and the bias by article length, we find that, on a per-word basis, 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. Due to this concern, it is important to control for article length in regression analysis, and focus on average effects.

Considering the number of contributors and the number of revisions for each Wikipedia article, we find wide variance in both, with the average article in our sample having 839 contributors (s.d. = 1,077) and 1,924 revisions (s.d. = 2,826). Because the number of revisions is skewed, the summary statistics suggest that only some articles may receive enough revisions to significantly change their slant and bias.

REGRESSION RESULTS

We next examine the differences in slant and bias via a regression framework that controls for shared unobservable factors. Our dependent variables are the slant or bias of each article. We create a dummy variable, *Wikipedia*, equal to 1 or 0 if the article is from Wikipedia or from Britannica, respectively. We use *Log(Length)*—the logarithm of article length—as a control variable; we log it because, according to Table 3, 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.

Models (1) and (2) of Table 4 use *Slant* as the dependent variable. We find that Wikipedia articles are more Democratic-slanted than Britannica articles. Once we control for length in Model (2), we also find that longer articles are more Democratic. The estimated coefficient of the length variable is of moderate size: doubling article length (adding, on average, approximately 4,000 words) increases Democratic slant by approximately -0.01. Even with this control, Wikipedia articles still tend to be 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 H1. Again, article length is responsible for a substantial part of this difference; doubling the length increases the bias by approximately 0.02 for Wikipedia articles, which accounts for a major part of the difference between the average biases found in Wikipedia and Britannica articles.

Models (2) and (4) 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 the number of code phrases per word, the slant per word, and the bias per word, which we use as our dependent variables in Models (5)-(6). In Model (5), we find that there is no significant difference between the number of code phrases per word in the two sources. In Model (6), we find *Wikipedia*'s sign reversed; that is, Wikipedia articles become more right-leaning than their Britannica counterparts at the per-word level rather than more left-leaning. 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. Indeed, results from Model (6) confirm that Wikipedia articles are less biased than Britannica articles at the per-word level.

We next examine how Wikipedia's revision process might change an article's bias; in particular, we are interested in whether articles become less biased the more they are revised. We therefore 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; that is, that Britannica's writers did not alter their content in reaction to Wikipedia's content (we test this assumption in the robustness check section).

We include several explanatory variables related to Wikipedia's revision process. *Log(Revisions)* is the logarithm of the number of an article's revisions. In a few models, we replace this variable with *Log(Contributors)*, the logarithm of the number of unique contributors who have edited the article. Since a contributor can revise an article multiple times, we also include each article's *Average Revisions per Contributor*. We retain the logarithm of the article's length as a control and, in some specifications, add category dummies and, to control for vintage effects, year dummies indicating when an article was created. Because *Revisions*, *Contributors*, and *Length* are highly skewed (see Table 3), we take their logarithms to minimize the influence of outliers.

Table 5 reports the OLS regression results. We find that the correlation of bias between Wikipedia and Britannica is about 25 percent 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 contributes to the slant difference between Wikipedia and Britannica. The impact of the number of revisions or of contributors is not as strong as that of article length: doubling the number of revisions or contributors reduces bias by 0.01, but doubling article length increases bias by 0.02. The average number of revisions per contributor has no significant effect on the bias. The variables *Revisions* and *Contributors* are skewed, so the articles receiving the most attention are much less biased, even when they are longer. However, the majority of articles do not receive such attention. The mean number of revisions is 1,924, which is insufficient to erase their bias.

We also find that further controls add some nuance to the results. Articles created in Wikipedia's early years tend to have more bias. The differences between 2002 and 2011 are the greatest of any two years and the pattern is monotonic across all years in Models (3) and (6), which suggests that the greatest differences between Britannica and Wikipedia appear in the oldest articles. Consistent with H2, the biases of the two sources converge when articles have been heavily revised, even when they come from vintages with large biases.

Robustness Checks

We conduct several checks to ensure that our results are not driven by alternative explanations. First, longer articles may include more code phrases, so it is theoretically possible that our results are mainly driven by outlying long articles. As a robustness check, we exclude all matched pairs if the length of either article is more than two standard deviations above the mean; this excludes 105 pairs. We obtain similar results.

Our second concern is with a potential unintended consequence of migrating the methods to this application: articles whose titles contain code phrases might exhibit more slant merely because those words are likely to be used many times in their texts. To ensure that such examples are not driving our results, we exclude all pairs with articles whose titles contain code phrases—50 pairs, or 1.3 percent of the total—from the analysis. Again, we obtain similar results.

Our third concern is with a subtle property of Gentzkow and Shapiro (2010)'s approach; namely, that it identifies two factors that shape slant and bias: (1) the choice of phrasing when there are multiple ways of describing the same concept (for example, "death tax" versus "estate tax") and (2) the choice of topics (for example, some newspapers may choose to run more articles about illegal immigration than others). By design, our study focuses on the former—the choice of phrasing conditional on the topic. Some phrases in Gentzkow and Shapiro (2010) (such as "Saddam Hussein," "World Trade Organization," and "Endangered Species Act"), however, do not have natural variants that exhibit an opposite slant. To ensure that these special phrases do not drive an article's slant, we recruited an experienced copy editor with both academic and legal backgrounds to go through the 1,000 code phrases, identify variations in phrasing for the same concept, and check all the variations. This exercise reveals that 638 of the 1,000 code phrases have substitutes. We repeat our analysis using only these 638 code phrases—essentially, ignoring any slant and bias arising from the remaining code phrases. Our results continue to hold.

Our fourth robustness check tests our assumption of the 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 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. Because Wikipedia was founded in 2001, Britannica's content must be exogenous. Of the 3,918 Britannica articles in our dataset, 2,855 are in the 2001 edition, which is a much smaller sample than analyzed above. When we repeat the analysis using only these 2,855 articles, 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

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 even if all such ideas were uncontroversial, objective, and verifiable. Our study examines the output of collective intelligence in a context in which aggregation is most difficult, when knowledge is contested.

Our study finds that crowd-based knowledge production does not result in articles with more biased than articles produced by experts when the crowd-based articles are substantially revised. This is consistent with a best-case scenario. Contributors with different ideologies engage in fruitful online conversations and do not segregate into communities with others who share similar views (e.g., Mullainathan and Shleifer 2005, Gentzkow and Shapiro, 2011). We think this is an important and novel finding.

Our findings also suggest that in online communities, the lower costs of producing, storing, and distributing knowledge may lead to increased—rather than unchanging or decreased—article length and thus result in different biases and slants. This is a new observation, and raises the question of whether the pattern arises in other comparisons of crowds and expert sources.

Theoretical Implications

Our findings include a strong note of caution against simplistic theories of collective decision making in the face of contested knowledge. Appropriate inferences require measured comparisons while interpretation of measured differences requires deep understanding of norms and processes. That combination requires embedding theory in a specific institutional setting.

We find that the level of bias of Wikipedia articles remains higher than that of Britannica content and varies across content categories. On one level, this is not surprising, as Wikipedia contains an enormous corpus of text and does not receive enough editorial contributions to revise all of it, particularly in niche content categories. On the other hand, the average Wikipedia article receives over 1,900 revisions, and our evidence indicates that this time and effort only mildly reduces bias rather than eliminating it. Wikipedia falls short of its goals because many contributions are needed to reduce considerable bias and slant to something close to neutral.

Our study therefore suggests that earlier empirical findings on the quality of crowd-sourced knowledge in the context of uncontested knowledge may not be immediately generalizable to settings with contested knowledge. We thus help identify the scope of the empirical generalization (Barwise 1995). It is important to evaluate whether other quality dimensions, such as relevancy and consistency, differ amongst types of knowledge.

Our results also suggest that the phrase, “Given enough eyeballs, all bugs are shallow” (Raymond 1998), can apply to bias and slant as well as to bugs. However, our study also suggests that theoretical speculation needs to account for the resources available and the speed with which goals are reached.

Managerial Implications

The main reason many organizations still resist crowds is that managers do not clearly understand the pros and cons of crowd production compared to those of internal production (Boudreau and Lakhani 2013). Our results show that, indeed, crowds can, but do not necessarily, perform better than experts in every dimension. Our results also suggest that the allocation of editorial time and user contributions is key to minimizing differences in bias and slant between production models. If editorial time and attention tend 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. There is no reason to assume that such a highly decentralized organization would result in an optimal allocation of time and attention. Hence, our findings motivate questions about how organizations that depend on collective intelligence should prioritize allocation of editorial time and user contributions.

Concerns about contested knowledge arise in many fields other than politics, from art to astronomy. As Wikipedia increasingly becomes many online readers' primary source of comprehensive information, there may be a strong incentive for those with strong opinions to manipulate its content to promote their own points of view. With the shift from reliance on expert-based sources to collectively produced intelligence, managerial trust in widely used information sources could be problematic. Their slants and biases are not widely understood, nor are the properties of the organizational forms by which their output is produced.

While this study focuses on a setting in which we can implement a viable empirical strategy, the same concern arises in many other online communities. For example, the largest forprofit wiki, Wikia,[4] 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 Wikipedia's (Greenstein et al. 2009). Knowledge communities have also been based on 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: (a) representing multiple sides of an issue typically takes many contributions and revisions and (b) length by itself is not usually sufficient to guarantee a balanced view without considerable revision.

Aggregation efforts within private firms (Majchrzak et al. 2009; Surowiecki 2004; Wagner 2005; Wagner and Majchrzak 2006) face similar concerns. Many private firms use wikis or other knowledge management technologies to organize their internal knowledge management (e.g., 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. An organization is likely to have at least a few employees with strong views; these can dominate a text if only a few employees participate regularly in the crowd and if there are few revisions. Our findings suggest that managers must do more than offer guidelines—they must make sure the community generates effective principles for participation in contested areas of knowledge and states effective principles for resolving disputes without intervention. Management also must make allowance for intervention to resolve disputes in the

[4] Source: http://www.wikia.com/, accessed March 2017.

cases when those principles break down, and this intervention is essential.

Limitations and Future Research

Our study has several limitations. First, we focus on a large online community for knowledge production; it is not clear that our results are generalizable to small online communities. In such communities, contributors may know each other or share similar social contexts, so it may be easier to develop a mutual understandings of neutral content and to enforce norms. Applying our approach to studying small communities would be an interesting area of future research.

Second, we examine two specific knowledge repositories: Wikipedia and Encyclopædia Britannica. These two organizations were chosen for their prominence and importance, and the comparison illustrates general issues. Nonetheless, the observed slant and bias are shaped by each organization's norms and policies for a neutral point of view. Our study does not conclude which particular elements of governance mechanisms result in the difference in the two production models. Future research could look for policy changes that shift some key governance mechanism in one production model or collect data from a large number of organizations with different governance mechanisms to establish this linkage.

Finally, we focus on ideological bias, but bias can take other forms, such as ethnic, racial, and gender bias (Hinnosaar 2015; Reagle and Rhue 2011), which can be consequential to the culture of an organization and of society. Other forms of bias can coexist in online communities; for example, Wikipedia does not present knowledge traditionally associated with women with the same depth and attention as it does knowledge traditionally 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

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

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. "Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict," *Journal of Management Information Systems* (21:4), pp. 71–98.

Arazy, O., Ortega, F., Nov, O., Yeo, L., and Balila, A. 2015. "Functional Roles and Career Paths in Wikipedia," *Computer Supported Cooperative Work (CSCW)*, pp. 1092–1105.

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

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

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

Barwise, P. 1995. "Good Empirical Generalizations," *Marketing Science* (14:3), pp. 29–35.

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

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

Burke, M., and Kraut, R. 2008. "Mopping Up: Modeling Wikipedia Promotion Decisions," in *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW)*, pp. 27–36.

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

Chesney, T. 2006. "An Empirical Examination of Wikipedia's Credibility," *First Monday*, (11:11), ISSN 13960466. Available at: *<http://journals.uic.edu/ojs/index.php/fm/article/view/1413>*.

Collier, B., Burke, M., Kittur, N., and Kraut, R. 2008. "Retrospective versus Prospective Evidence for Promotion: The Case of Wikipedia," in *2008 Annual Meeting of the Academy of Management.*

Eppler, M J., and Wittig, D. 2000. "Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years," in *Proceedings of the 5th International Conference on Information Quality*, pp. 83–96.

Faraj, S., Jarvenpaa, S. L., and Majchrzak, A. 2011. "Knowledge Collaboration in Online Communities," *Organization Science* (22:5), pp. 1224–1239.

Forte, A., Kittur, N., Larco, V., Zhu, H., Bruckman, A., and Kraut, R. E. 2012. "Coordination and Beyond: Social Functions of Groups in Open Content Production," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, pp. 417–426.

Forte, A., Larco, V., and Bruckman, A. 2009. "Decentralization in Wikipedia Governance," *Journal of Management Information Systems* (26:1), pp. 49–72.

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.

Gallus, J. Forthcoming. "Fostering voluntary contributions to a public good: A large-scale natural field experiment at Wikipedia," *Management Science*.

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. 2016. "The Reference Wars: Encyclopædia Britannica's Decline and Encarta's Emergence," *Strategic Management Journal* (38:5), pp. 995–1017.

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.

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. 2016. "Open Content, Linus' Law, and Neutral Point of View," *Information Systems Research* (27:3), pp. 618–635.

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.

Halfaker, A., Kittur, A., and Riedl, J.. 2011. "Don't Bite the Newbies: How Reverts Affect the Quantity and Quality of Wikipedia Work," in *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*, pp. 163–172.

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.

Janis, I. L. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*, Houghton Mifflin.

Jelveh, Z., Kogut, B., and Naidu, S. 2014. "Political Language in Economics," Columbia Business School Research Paper No. 14-57.

Jemielniak, D. 2014. *Common Knowledge?: An Ethnography of Wikipedia*. Palo Alto, CA: Stanford University Press.

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), pp. 1–17.

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.

Kitchin, R. 1998. *Cyberspace: The World in the Wires*. Chichester, UK: Wiley.

Kittur, A., Chi, E., Pendleton, B. A., Suh, B., and Mytkowicz, T. 2007. "Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie," *World Wide Web* (1:2) pp. 19.

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

Kittur, A., and Kraut, R. 2008. "Harnessing the Wisdom of Crowds in Wikipedia: Quality through Coordination," in *ACM Conference on Computer Supported Cooperative Work*, pp. 37–46.

Kittur, A., Suh, B., Pendleton, B. A., and Chi, E. H. 2007. "He Says, She Says: Conflict and Coordination in Wikipedia," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 453–462.

Klein, B. D. 2001. "User Perceptions of Data Quality: Internet and Traditional Text Sources," *Journal of Computer Information Systems* (41:4), pp. 9–18.

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

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

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).

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

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

Madnick, S. E., and Wang, R. Y. 1992. "Introduction to the TDQM Research Program," Total Data Quality Management Research Program Working Paper #92-01.

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.

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.

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

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

Miller, H. 1996. "The Multiple Dimensions of Information Quality," *Information Systems Management* (13:2), pp. 79–82.

Mollick, E. R., and 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.

Naumann, F., and Rolker, C. 2000. "Assessment Methods for Information Quality Criteria," in

Proceedings of 5th International Conference on Information Quality, pp.148–162.

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.

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.

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.

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

Raymond, E. 1998. "The Cathedral and the Bazaar," *First Monday*, http://tinyurl.com/bqfy3s, accessed May 2016.

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

Rector, L. H. 2008. "Comparison of Wikipedia and Other Encyclopedias for Accuracy, Breadth, and Depth in Historical Articles," *Reference Services Review* (36:1), pp.7–22.

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

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

Schroeder, A., and Wagner, C. 2012. "Governance of Open Content Creation: A Conceptualization and Analysis of Control and Guiding Mechanisms in the Open Content Domain," *Journal of the American Society for Information Science and Technology* (63:10), pp. 1947-1959.

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).

Stvilia, B., Twidale, M. B., Smith, L. C., and Gasser, L. 2008. "Information Quality Work Organization in Wikipedia," *Journal of the American Society for Information Science and Technology* (59:6), pp. 983–1001.

Sun, M., Zhang, X., and Zhu, F. 2017. "Nonconformity in Online Social Networks," Working paper.

Sun, M., and Zhu, F. 2013. "Ad Revenue and Content Commercialization: Evidence from Blogs," *Management Science* (59:10), pp. 2314–2331.

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

Swartz, A. 2006. "Who Writes Wikipedia?" September 4, http://www.aaronsw.com/weblog/whowriteswikipedia (accessed July 2016).

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.

Wang, R., and Strong, D. 1996. “Beyond Accuracy: What Data Quality Means to Data Consumers,” *Journal of Management Information Systems*, (12:4), pp. 5–34.

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.

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.

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.

Table 1: Comparing Slants in Wikipedia and Britannica Articles

Topic categories	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***
Employment	256	-0.03	0.19	-0.01	0.15	-0.01
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
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
Taxation	21	-0.15	0.22	-0.21	0.27	0.06
Trade	104	0.03	0.17	0.04	0.13	-0.01
Value	165	-0.05	0.22	-0.03	0.16	-0.03
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: Our slant measure is less (more) than zero when an article leans towards Democratic (Republican) viewpoints. We report both means and standard deviations for Wikipedia and Britannica articles; the last column shows the difference in means (i.e., slant of Wikipedia articles – slant of Britannica articles). We also conduct two-tailed t-tests to examine 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

Topic categories	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***
Employment	256	0.13	0.14	0.07	0.14	0.06***
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***
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***
Taxation	21	0.20	0.17	0.23	0.26	-0.02
Trade	104	0.13	0.11	0.09	0.10	0.04***
Value	165	0.17	0.16	0.08	0.15	0.09***
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: Our bias measure is the absolute value of the slant. We report both means and standard deviations for Wikipedia and Britannica articles; the last column shows the difference in means (i.e., bias of Wikipedia articles – bias of Britannica articles). We also conduct two-tailed t-tests to examine whether the difference is significantly different from zero. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: Summary Statistics

Wikipedia articles					
Variable	Obs.	Mean	Std. dev.	Min	Max
Num of Code Phrases	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
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 Phrases	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
Slant/Length	3,918	-0.00006	0.00050	-0.0085	0.0063
Bias/Length	3,918	0.00015	0.00048	0	0.0085

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

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Slant	Slant	Bias	Bias	Slant/Length	Bias/Length
Wikipedia	-0.036***	-0.013***	0.074***	0.023***	0.00002***	-0.00008***
	[0.004]	[0.005]	[0.003]	[0.004]	[0.00001]	[0.00001]
Log(Length)		-0.013***		0.030***		
		[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	0.002	0.026
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	OLS	OLS	OLS	OLS	OLS	OLS

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

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

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Wikipedia bias	Wikipedia bias	Wikipedia bias	Wikipedia bias	Wikipedia bias	Wikipedia bias
Britannica bias	0.245***	0.262***	0.224***	0.247***	0.265***	0.226***
	[0.022]	[0.022]	[0.021]	[0.022]	[0.022]	[0.021]
Log(Length)	0.033***	0.030***	0.025***	0.033***	0.030***	0.025***
	[0.003]	[0.004]	[0.003]	[0.003]	[0.003]	[0.003]
Log(Revisions)	-0.011***	-0.018***	-0.014***			
	[0.003]	[0.004]	[0.004]			
Log(Num of Contributors)				-0.012***	-0.020***	-0.016***
				[0.003]	[0.004]	[0.004]
Average Revisions per Contributor	-0.002	0.007	0.004	-0.006*	0.001	-0.000
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Year Created = 2002		0.065***	0.049***		0.064***	0.048***
		[0.007]	[0.007]		[0.007]	[0.007]
Year Created = 2003		0.006	0.006		0.005	0.005
		[0.008]	[0.008]		[0.008]	[0.008]
Year Created = 2004		0.009	0.005		0.007	0.004
		[0.010]	[0.010]		[0.010]	[0.010]
Year Created = 2005		-0.023*	-0.021*		-0.025**	-0.022*
		[0.013]	[0.013]		[0.013]	[0.013]
Year Created = 2006		-0.033*	-0.027		-0.036**	-0.030*
		[0.017]	[0.017]		[0.017]	[0.017]
Year Created = 2007		-0.038*	-0.028		-0.041**	-0.031
		[0.019]	[0.019]		[0.019]	[0.019]
Year Created = 2008		-0.063***	-0.047**		-0.066***	-0.050**
		[0.020]	[0.020]		[0.020]	[0.020]
Year Created = 2009		-0.096***	-0.085***		-0.100***	-0.089***
		[0.015]	[0.017]		[0.015]	[0.017]
Year Created = 2010		-0.110***	-0.086***		-0.113***	-0.089***
		[0.017]	[0.017]		[0.017]	[0.017]
Year Created = 2011		-0.191***	-0.166***		-0.200***	-0.174***
		[0.016]	[0.019]		[0.016]	[0.020]
Dummies for categories	No	No	Yes	No	No	Yes
Observations	3,918	3,918	3,918	3,918	3,918	3,918
Adjusted R-squared	0.071	0.109	0.185	0.072	0.111	0.186

Note: Heteroskedasticity-adjusted standard errors in brackets. In Models (2), (3), (5), and (6), *Year Created* = 2001 is used as the benchmark group. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table 1: Comparing Coverage of Matched Sample to Initial Wikipedia Sample

Topic category	Matched sample		Full sample		Difference
Abortion	13	0.33%	71	0.25%	0.08%
American Politicians	438	11.18%	4,748	16.73%	-5.55%
Budget	249	6.36%	1,109	3.91%	2.45%
Civil Rights	263	6.71%	1,183	4.17%	2.54%
Corporations	28	0.71%	121	0.43%	0.28%
Drugs	39	1.00%	105	0.37%	0.63%
Education	311	7.94%	1,362	4.80%	3.14%
Employment	256	6.53%	693	2.44%	4.09%
Energy	52	1.33%	270	0.95%	0.38%
Family	126	3.22%	405	1.43%	1.79%
Foreign Policy	524	13.37%	2,094	7.38%	5.99%
Government	1183	30.19%	11,383	40.11%	-9.92%
Gun	9	0.23%	56	0.20%	0.03%
Health Care	120	3.06%	556	1.96%	1.10%
Homeland Security	132	3.37%	490	1.73%	1.64%
Immigration	99	2.53%	372	1.31%	1.22%
Infrastructure & Technology	277	7.07%	1,143	4.03%	3.04%
Social Security	0	0.00%	5	0.02%	-0.02%
Taxation	21	0.54%	95	0.33%	0.21%
Trade	104	2.65%	399	1.41%	1.24%
Value	165	4.21%	614	2.16%	2.05%
War & Peace	578	14.75%	2,292	8.08%	6.67%
Welfare & Poverty	109	2.78%	323	1.14%	1.64%