Paper to be presented at

DRUID15, Rome, June 15-17, 2015

(Coorganized with LUISS)

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

Greenstein Shane
Northwestern University
Kellogg School of Management
greenstein@kellogg.northwestern.edu

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

Abstract

Which source of information contains greater bias and slant?text written by an expert or that constructed via collective intelligence? Do the costs of acquiring, storing, displaying and revising information shape those differences? We evaluate this question empirically by examining slanted and biased phrases in content on US political issues from two sources?Encyclopædia Britannica and Wikipedia. Our overall slant measure is less (more) than zero when an article leans towards Democrat (Republican) viewpoints, while bias is the absolute value of the slant. Using a matched sample of pairs of articles from Britannica and Wikipedia, we show that, overall, Wikipedia articles are more slanted towards Democrat than Britannica articles, as well as more biased. Slanted Wikipedia articles tend to become less biased than Britannica articles on the same topic as they become substantially revised, and the bias on a per word basis hardly differs between the sources. These results have implications for the segregation of readers in online sources and the allocation of editorial resources in online sources using collective intelligence.

Jelcodes:L33,L00

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

Abstract

Which source of information contains greater bias and slant—text written by an expert or that constructed via collective intelligence? Do the costs of acquiring, storing, displaying and revising information shape those differences? We evaluate this question empirically by examining slanted and biased phrases in content on US political issues from two sources—Encyclopædia Britannica and Wikipedia. Our overall slant measure is less (more) than zero when an article leans towards Democrat (Republican) viewpoints, while bias is the absolute value of the slant. Using a matched sample of pairs of articles from Britannica and Wikipedia, we show that, overall, Wikipedia articles are more slanted towards Democrat than Britannica articles, as well as more biased. Slanted Wikipedia articles tend to become less biased than Britannica articles on the same topic as they become substantially revised, and the bias on a per word basis hardly differs between the sources. These results have implications for the segregation of readers in online sources and the allocation of editorial resources in online sources using collective intelligence.

1. Introduction

On March 13, 2012, Encyclopædia Britannica (hereafter, 'Britannica') announced that it would cease publishing in book format after 244 continuous years, and henceforth its services would only be available in digital format.[1] At its peak two decades earlier, Britannica served half the market share for household encyclopedias, and was widely regarded as authoritative (Greenstein and Devereaux 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 online encyclopedia Wikipedia (Bosman 2012).

Founded in 2001, Wikipedia had become the most widely-sourced reference site online. Non-profit and free to users, it had grown to become a top ten Internet site in over two dozen countries by 2012,[2] attracting over four hundred million hits a month. Wikipedia has also attracted notice as the world's largest experiment in "collective intelligence" and one of the largest human projects to ever bring information into one source.[3]

[1] The announcement is at http://www.britannica.com/blogs/2012/03/change/, accessed May 2014.
[2] Source: www.alexa.com, accessed April 2014.
[3] The English language Wikipedia has more than 4 million articles with over 2.5 billion words. Wikipedia in all languages has over 32 million articles. This is not yet the largest collection in human history. The British Library has over 150 million items, and the Library of Congress has over 155 million, with 12 million searchable.

As sources that aspire to provide comprehensive information, Britannica and Wikipedia face similar conflicts over the length, tone, and factual basis of controversial, unverifiable, and subjective content. These conflicts are pervasive in the production of encyclopedic knowledge about current events and political debates. Controversy is prevalent in topics such as the proper taxation level for government, the success of health care policies, or the biographical details of presidential candidates. Similar issues arise in the presentation of scientific knowledge that touches on persistent ideological divides, including topics such as the forecasts for climate change, the consequences of diffusing genetically-modified crops, or the funding of stem-cell research.

Our study begins with a simple observation: Each source resolves these conflicts differently in distinct production processes. Britannica sources its material from experts, and fosters a reputation for being an "august repository of serious information" (Melcher 1997). It produces its final content after consultation between editors and experts. In contrast, Wikipedia relies on the "crowd" for its content, receiving contributions from tens of millions of individual users. Conflicts are addressed in a highly decentralized process during countless arguments about whether a passage reflects a "neutral point of view." Do these distinct production processes differ in the presentation of controversial, unverifiable, and subjective information? One might expect so.

In this study, we examine the political slants and biases of pairs of Britannica and Wikipedia articles, and use the analysis to make inferences about the properties of the two production processes. We focus on three questions: Do different production processes shape the political leaning of the content, and if so, in which direction? Does collective intelligence result in greater or fewer biases? Do lower costs of acquiring, storing, displaying, and revising information shape those differences?

Addressing these questions contains a number of inherent challenges. It is challenging to construct a yardstick for judging a source's bias, slant, or neutrality. We also need to compare output from two distinct production processes and control for the bias and slant of some topics. To overcome these challenges, this study directly compares the appearance of biased and slanted phrases in pairs of articles that appear at the same point in time in both sources and cover (nearly) identical topics. We are able to find 3,918 pairs of articles about US politics that appeared in both outlets. US politics are particularly well suited for this comparison because such topics involve controversial, subjective, and unverifiable information, and, accordingly, display many biases and slants. Consistent with our reasoning, Yasseri et al. (2014) find that politics contains the highest percentage (25%) of the top 100 most controversial topics in Wikipedia.

We adapt the notions of 'bias' and 'slant' to online material, modifying a method developed by

---

http://en.wikipedia.org/wiki/Wikipedia:Size_comparisons, accessed August 29, 2014. The online version of Encyclopædia Britannica, which we use in this study, has 120,000 articles and 55 million words.

Gentzkow and Shapiro (2010). 'Slant' indicates which way a particular piece of knowledge 'leans' (and is thus positive or negative, accordingly), 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, and use that to measure the political leanings of news providers. 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.'

We find that the slanted and biased content sourced from collective intelligence differs significantly from expert-based production. In a general comparison of the two sources, we find that Wikipedia is more likely to mildly lean towards Democratic opinion than is Britannica. This general tendency varies across topics, with the number leaning Democrat outnumbering the number leaning Republican by a small amount. We also find a general tendency across most topics for Wikipedia articles to be more biased than those published in Britannica, with only a few exceptions. These findings motivate deeper analysis about why the content differs from the two sources.

We next compare pairs of articles directly against one another in which each covers the same topic. Direct comparison controls for otherwise unobservable slants and biases shared by the two articles, and which might be inherent to a topic. This exercise yields evidence that substantial revisions of Wikipedia articles reduce the differences in biases and slants to negligible statistical differences. In other words, the largest biases and slants arise on Wikipedia articles with fewer contributions. The rate of convergence due to revision is also comparatively slow, so there are many more Wikipedia articles with bias and slant than without. While articles receiving revision in the upper quartile (of our sample) get enough revision to achieve no difference in slant and bias, the median article (and lower quartile) does not receive enough revision.

Our evidence is consistent with the hypothesis that many of the differences arise because collective intelligence, as implemented by Wikipedia, faces comparatively lower acquisition, storage, and distribution costs. Those lower costs provide incentives for Wikipedia participants to add a new fact to an existing article, or to increase an article's length, which makes them longer than their Britannica counterparts. Longer articles make it more likely that an article includes more slanted phrases. Indeed, we find evidence that article length plays an important role in producing slant and bias. We also find that Wikipedia articles are slightly less biased on a per word basis than are Britannica articles, with little meaningful statistical difference.

1.1 Related Literature

Technological advances in the past few years have made it significantly easier for users to communicate and contribute to the same project. Scholars from many different fields have started to examine the "wisdom of crowds" in a variety of settings, such as the funding of startups (e.g., Zhang and Liu 2012; Agrawal et al. 2013; Kuppuswamy and Bayus 2013; Mollick 2014; Mollick and Kuppuswamy 2014), the development of new products (e.g., von Hippel 2005; Terwiesch and Ulrich 2009; Afuah and Tucci 2012), innovation tournaments (e.g., Lakhani et al. 2012), prediction markets (e.g., Wu and Brynjolfsson 2013) and scientific research (e.g., Lakhani 2009; Franzoni and Sauermann 2014).

Some of these studies have compared collective intelligence with expert-based models. A number of studies have shown that crowds could behave irrationally because they suffer from factors such as group thinking (e.g., Janis 1982), emotional contagion (Barsade 2002) and herding (e.g., Banerjee 1992; Bikhchandani et al. 1992). These findings support Mackay's viewpoint about the "madness of crowds" (Mackay1852). At the same time, several studies find that results from collective intelligence compare favorably to those from experts (e.g., Galton 1907; Shankland 2003; Antweiler and Frank 2004; Lemos 2004; Surowiecki 2004). These studies typically focus on one dimension of knowledge—its accuracy in a specific domain of knowledge. The evidence about Wikipedia's accuracy is mixed. While several studies find Wikipedia to be about as good a source of accurate scientific information as expert-based content (e.g., Giles 2005; Rajagopalan et al. 2011), several articles point to significant gaps in its coverage (Bragues 2007; Devgan et al. 2007; Brown 2011) and accuracy (e.g., Clauson et al. 2008; Hasty et al. 2014).

Recent research has begun to investigate other biases of "crowds," and comes closer to our motivation. The closest research in this vein investigates crowd-funding and compares it with traditional venture funding. The research finds that crowds can exhibit a taste or preference not otherwise present in traditional sources (Mollick and Nanda 2014). Our findings could be reinterpreted in a similar way: crowds display political preferences that experts do not display.

We also contribute to the growing literature examining bias and slant in news media. Our approach differs from the extant literature, which tends to be inspired by the "Fox News effect" (DellaVigna and Kaplan 2007). In this approach the fixed costs of maintaining a distinct production and distribution of information leads to distinct providers, each with divergent points of view. Consumers segregate between sources, and consume contents that confirm their prior views (e.g., see Mullainathan and Shleifer 2005). In light of the proliferation of web sites, some analysts have forecast an extreme form of self-selection among online news readers. For example, Sunstein (2001) frames the 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." Examining several sources of evidence, Gentzkow and Shapiro (2011) conclude that there is not strong evidence of more segregation online. The same concern applies to knowledge production on Wikipedia—if articles only attract contributors with similar ideologies, we expect Wikipedia to exhibit strong ideological biases.

This emphasis of the paper differs from prior research about the political biases and slants in Wikipedia (Greenstein and Zhu 2012, 2013), which emphasizes the history of biases and slants at Wikipedia. It shows that Wikipedia began its earliest years with a Democratic slant that diminished over time. It shows that revision plays a role in changing the slants and biases, but entry of new articles plays an equally important role. These studies lack an offline benchmark, so they fail to account for the intrinsic biases and slants of some topics, as the present study does.

The rest of the paper proceeds as follows: Section 2 provides a brief description of each production model. Section 3 describes our dataset. Section 4 makes general comparisons between the two sets of articles. Section 5 presents regression results and a number of robustness tests for the results. Section 6 concludes.

2. Background

2.1. Brief History of Britannica's Production

This study examines Britannica in its 244th year, when it ceased publication in book form. By this time (2012), the substantial bulk of its revenue came from CDs 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%.

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 a high-margin product, and this distribution cost comprised a major variable expense. A substantial fraction of the margin also went to covering the fixed costs of employing an editorial staff to produce and organize this content into book format. The organization was a privately held company, owned by the Benton Family, and after the death of William Benton, the organization was owned by a foundation that donated all its profits to the University of Chicago (Greenstein and Deveraux 2009).

A large fraction of content came to Britannica from experts at no expense. The publication has cultivated its reputation as the world's most comprehensive and authoritative encyclopedia (Evans and Wurster 2000). Experts jumped at the chance to write its entries, as it enhanced their own reputations. Accordingly, the firm's managers devoted considerable expense to maintaining its reputation as a comprehensive source of information. There were always concerns about becoming outdated. To prevent customers from perceiving Britannica as outdated, the publication 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, an initiative that continued into the new millennium. This required employing a large staff of researchers as well. Both programs fed into updating Britannica's entries, sometimes on a yearly basis, and sometimes on an irregular basis, particularly on its online entries.

There was no set rule for the length of Britannica articles, but concerns about the length of the volumes played a significant role in the length of its articles. The sales department regarded additional length as a negative attribute, arguing that customers resisted buying books that took up too much shelf space, so there was strong resistance to adding extra length to articles. Management fixed the total physical length of the encyclopedia for decades, so any new addition generated a subtraction somewhere else. Editors were hired on the basis of their ability to 'edit to fit,' i.e., to make an entry fit a prescribed length (Greenstein and Deveraux 2009).

Britannica had begun experimenting with different formats long before Wikipedia was launched. Some years after it had turned down a Microsoft initiative to produce a CD-ROM in 1985, Britannica executives started building a multimedia encyclopedia and eventually launched its first multimedia CD-ROM encyclopedia under the Compton Brand. DOS and Mac versions became available in 1991 in an experiment that preceded Microsoft's launch of Encarta. Britannica eventually offered CD-ROMs of its name-brand encyclopedia, but these CDs never became a major commercial success in comparison to its book sales.

Financial and strategic challenges interfered with Britannica management's ability to translate these early technical successes into commercial products. Book sales began to decline in the mid-1990s under competition from Encarta, eroding Britannica's ability to invest in new initiatives. The encyclopedia changed ownership in 1996, and, by the time of Wikipedia's founding in 2001, the size of its sales force had been greatly reduced. In the midst of that decline in the number of employees, there was never a move to turn away from relying on experts for content. The 2012 content used in our study is still an excellent representative of "expert-based content."

2.2. 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 that is entirely written and edited via user contributions. Users selected which content they would revise and how much. Expertise played no role in revising content. By November 2011, it was the world's largest Wiki, supporting 3.7 million articles in English and well over 20 million articles in all languages. Some articles received large numbers of revisions, and some did not.

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 Web site and related efforts. At no point has the web site ever accepted advertising, nor does it have plans to do so. It is entirely supported by donations.

The expenses for Wikipedia's production have changed over the years. In its earliest days, servers, storage, and data-transmission comprised the largest expense. In more recent times, the programming costs and administrative costs have grown to an important percentage of operational expense. Until 2006 the foundation operated with a minimal staff of two programmers, under the supervision of Jimmy Wales. By 2012, it had grown to include a full-time professional manager and several more employees, many of whom are devoted to raising donations and serving the community of volunteer contributors.

Wikipedia pays no licensing fees for software. Wikipedia is the largest user of wiki technology, which was developed in 1995 by Ward Cunningham, a software engineer from Portland, Oregon. Wikis were first developed and intended for documenting open software development. Cunningham claims, "It's a medium that allows people to collaborate more easily than they could in systems that are modeled after the pre-computer world, like e-mail" (Levine 2006). Cunningham designed Wikis so that contributors required no special training. Wiki server technology allows the creation of hypertexts with nonlinear navigation structures: each page contains a series of cross-links to other pages, so readers can decide how to navigate through the site for themselves. Today, Wikipedia uses a modified version of Cunningham's invention, with most of the alterations designed to handle the increased scale of contributions from a world-wide dispersed set of contributors.

At no time ever in its history has Wikipedia paid for content. Contributions come from tens of millions of sources, and tens of thousands of dedicated volunteers make the bulk of changes to existing entries. While any entry can be changed if a contributor thinks that changes are warranted, the distribution of effort tends to be skewed. "An outsider makes one edit to add a chunk of information, then insiders make several edits tweaking and reformatting it. In addition, insiders rack up thousands of edits doing things like changing the name of a category across the entire site—the kind of thing only insiders deeply care about. As a result, insiders account for the vast majority of the edits. But it's the outsiders who provide nearly all of the content" (Swartz 2006).

There are tens of thousands of editors in the Wikipedia community, and they are not under any central editorial control from the Wikimedia Foundation. The organization relies on users to discover and fix passages that do not meet the site's content tenets, and no central authority compels editors or users how to allocate editorial time and attention. At most there are gentle reminders or "stubs" which editors may attach to articles to suggest changes. Wikipedia started with almost no restrictions on contributions, and as it grew, developed a few restricted privileges to facilitate administration. Registered users tend to

provide most of the editorial contributions and their contributions tend to survive editorial changes at a higher rate. Available evidence about conflicts suggests that editors who frequently work together do not get into as many conflicts, such as "revision wars," 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 articles is a 'Neutral Point of View' (NPOV). 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, editors can devote considerable time and energy debating whether an article's text correctly portrays a topic from a NPOV. As Wikipedia articles face no limits to their number or size, due to the absence of any significant storage costs, conflicts are often addressed through adding more points of view instead of eliminating points of view.

The Wikimedia Foundation makes no claims that the end product is ever finished. It invites further revisions, and the website proudly declares itself as the 'encyclopedia that anyone can edit.'[4] It advises users to check the recent history of revision, and not to treat any passage as definitive. In practice, hundreds of millions of its readers do treat it as comprehensive and definitive. It is often the first organic listing in a Google search, and for many searches Google will display the Wikipedia answer in a formatted box. Other question-answer sites also source from Wikipedia.

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

In summary, the production process for Wikipedia does not fit the standard neoclassical model of production. Output price plays no role, and there is minimal organizational hierarchy. The final good does not follow a fixed design, and quality control is allocated to voluntary labor. The entire system, its processes and popularity, have been replicated in dozens of languages where the Internet is widely used.

3. Data

We gather relevant entries from the online edition of Britannica and from Wikipedia on June 8, 2012.[5] The initial sample of articles from Wikipedia focuses on a broad and inclusive definition of US political

[4] Source: Wikipedia.org, http://en.wikipedia.org/wiki/Main_Page, accessed April 2014.
[5] We checked the online edition of Britannica to ensure that, just like Wikipedia, it is constantly updated to incorporate the latest information.

topics, and includes all articles in Wikipedia that included the keywords 'Republican' or 'Democrat,' resulting in a list of 111,216 articles. Many of these articles concern events in countries other than the United States, which necessitates further assessment for relevance,[6] which reduced the list to 70,668 articles about US politics. This sample covers an enormous array of topics, including many controversial ones—such as entries 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 the list of all (120,000+) articles in the Britannica's online edition, and are able to identify 3,918 pairs of matching articles. In 73% of the pairs the titles are identical. In the remainder the titles are nearly identical and we manually check that the pairs cover similar topics. As we show below, these 3,918 articles are drawn from a representative sample of topics within 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—negative (positive) numbers represent a slant towards Democratic (Republican) 'view.' The degree of bias is the absolute value of the slant, so larger numbers indicate more bias than smaller numbers.

As in G&S, we ask whether Wikipedia or Britannica articles use 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 language to speak to its respective constituents.[7] Each phrase is associated with a cardinal value that represents 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.'

This approach has several key strengths. G&S has been tested on newspaper editorials and has passed many internal validity tests. It provides a general yardstick for measuring the bias of newspaper articles, which we can employ in the context of Internet articles, and avoids many subjective elements. Moreover, 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.

[6] 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.
[7] See Table I in Gentzkow and Shapiro (2010) for more examples.

In this method, an article is considered to be 'unslanted' or 'unbiased' when it includes no code words. It also contains no slant when it uses an equal number of Republican/Democrat code words with the same cardinal values.

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 article is paired with its own match, so we can net out the 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, and in which direction. In addition, by comparing articles on the same topics from the two sources at the same point in time, we are able to control for topic-specific and time-specific unobservables that affect bias and slant of these articles.

This study focuses on the features of Wikipedia in 2012, a decade after Wikipedia's founding. This is useful for this study because it comes at a time after enormous variance in revision. Some articles have received extensive revision, while many have not.

4. Comparing Slants and Biases

We first consider the representativeness of our sample by comparing it against the much larger initial set of Wikipedia articles. Table 1 lists the topics covered, using category information defined in Wikipedia. The first column lists the 3,918 articles that match across the two knowledge sources, which we label the 'analysis sample': since a Wikipedia article may have more than one topic affiliated with it, the categories are not mutually exclusive.

We observe no troubling features in our analysis sample. Overall, the analysis sample and the original sample are roughly in proportion with one another, with a few obvious exceptions. The most common topic is 'Government,' followed by 'War and Peace,' 'Foreign Policy,' and 'American Politicians.' The analysis sample has a large representation of American political biographies, which we think is a by-product of how easy it is to exactly match biographies of individuals across the two samples. The analysis sample also comparatively over-represents entries about Government, although that seems innocuous since the label is used so frequently in Wikipedia. Areas where there is some underrepresentation—such as Education, Foreign Policy, War and Peace, and Infrastructure and Technology—again do not reveal any obvious selection issues.

Tables 2 and 3 show the slants and biases of articles in the analysis sample, computing the mean and standard deviations for the average slant and bias for all articles in each category. Both Britannica and Wikipedia's articles display considerable variance in the levels of slant and bias across topics. The two sources also track one another: in 19 of the 23 categories, the difference in slant between the two sources is insignificant—but they 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 democratically slanted than those published in Britannica.

The findings for bias show that the articles from Wikipedia are often more biased than those from Britannica. In only five categories of topics are these differences insignificant—in many topics they are considerable, with Wikipedia articles displaying more bias in every instance. Overall, we conclude that Wikipedia articles are generally much more biased.

At first sight, this finding about bias reflects the different frequencies of appearance of code words across the two sources. A much higher percentage of Wikipedia articles (73%) have at least one more code word than those published in Britannica (34%). This initial evidence suggests that the difference may be associated with the length of the articles. Wikipedia articles are longer than their Britannica matches, as can be expected given Wikipedia's cheaper storage costs and different editorial processes. Although Britannica has the longest single article in our dataset, on average Wikipedia articles in the analysis sample contain 4,113 words, while Britannica articles are only 43% as long (1,778 words). As a result, code words are more likely to appear in Wikipedia articles because of their greater length.

Table 4 provides descriptive statistics for the entire matched sample dataset, and shows again that Wikipedia articles are more slanted towards Democratic viewpoints than are Britannica articles. Again, the Wikipedia articles are also more biased. The table also shows that there is wide variance in the number of contributors per article, with the average Wikipedia article in this sample having 839 contributors (s.d. = 1,077) and 1,924 revisions. Prior studies of revisions show that biases and slants in Wikipedia articles do not change over time without hundreds, if not tens of thousands of revisions (Greenstein and Zhu 2013). Because revisions are skewed, this number suggests the upper quartiles may get sufficient revision to yield big changes in the slant and bias, but the majority of articles near or below the median may not.

5. Regression Results

We next examine the differences in slant and bias in 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 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 the article's 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 the articles' unobserved underlying slant or bias.

Models (1) and (2) in Table 5 use *slant* as the dependent variable. We find that, overall, Wikipedia articles are more Democratic-slanted than are Britannica articles. Controlling for article length matters for the result, suggesting Wikipedia's longer articles play an important role. Once we control for length in model (2) we also find that longer articles from both sources are more Democratic. The estimated coefficient on length is of moderate size: a doubling in length (i.e., adding approximately 4,000 words on average to each article), generates a change in the Democratic direction of approximately -0.01 in Wikipedia. Even with this control, the Wikipedia articles are still more Democratic overall (-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 are more biased than Britannica articles. 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 first columns tried to account for the skewed distribution of article length. Alternatively, we normalize our slant and bias measures by the length of the article to capture slant and bias per word, and use them as our dependent variables in Models (5) and (6). In Model (5), we find that the sign of the *Wikipedia* variable reverses—so Wikipedia articles are now more right-leaning at the per word level. But, since both Wikipedia and Britannica articles exhibit overall Democratic slants at the per word level (Table 4), this result suggests that Wikipedia articles are closer to neutral than their Britannica counterpart at the per word level. Similarly, results from Model (6) confirm that Wikipedia articles are less biased than Britannica articles at the per word level.

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 numbers 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 (which will be tested in the robustness check section). We include two explanatory variables related to the revision process at Wikipedia. The first is the logarithm of the total number of revisions the article has already received, *Log(Revisions)*, and as 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 categories of the articles.

Table 6 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 which have received more revisions tend to be more neutral. While the article length continues to contribute to the slant difference between Wikipedia and Britannica articles, the number of revisions influences this difference. Revisions can affect the level of article bias, but not as much as article length: doubling the number of revisions reduces bias by -0.01, while doubling article length increases bias by 0.03.Yet 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, the typical article receives 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 in all years in Model (4), which suggests that the greatest differences between Britannica and Wikipedia appear in the oldest articles that have not yet been revised. In summary, the biases from the two sources converge when articles receive large numbers of revisions, even when they come from vintages with large biases.

5.1. Robustness Checks

We conduct several checks to ensure our results are robust to alternative explanations. Our first concern is that article lengths exhibit significant variations (as Table 4 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 (Tables 7a and 7b). We conclude that outlying article length does 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 are 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 (Tables 8a and 8b). We conclude that the presence of code words in articles' titles do not drive our 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 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). By design, our study focuses much more on the former than the latter, i.e., the choice of phrasing conditional on the topic, with the result that some phrases in G&S (e.g., 'Saddam Hussein,' 'World Trade Organization,' and 'Endangered Species Act') do not have natural variants that have 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 other code words). Our results (Tables 9a and 9b) continue to hold, so we conclude that choice of phrasing does not drive our results.

The last of our robustness checks tests 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 and it is possible that biases in Britannica articles arise 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 was founded) which must be exogenous, by design. Of the 3,918 Britannica articles in our dataset, 2,860 exist in the 2001 edition. When we repeat the analysis using only these 2,860 articles and their matched Wikipedia versions, we obtain similar results (Tables 10a and 10b). This supports our assumption that the biases in Britannica articles are exogenous to those processes that create bias in Wikipedia articles.

6. 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 is 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 it would face its greatest challenge—where the topics are inherently controversial, subjective, and unverifiable.

We focus on the factors in production that generate bias and slant in text. That research objective leads us to ask a novel question about the effects of segregation on the characteristics of content. We ask whether the evidence suggests contributors with different ideologies engage in fruitful online conversations with each other about how the same source of information displays their respective points of view. Related, we are also the first to recognize that the costs of producing, storing, and distributing

knowledge shape different biases and slants in collective intelligence and the expert-based model.

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 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 also 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 and that is still not enough for eliminating bias. So, Wikipedia falls short of its ideal, and it does so because it takes a lot of "eyeballs" 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 go around.

These results suggest that the allocation of editorial time and user contributions is central to the minimization of differences in bias and slant between production 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 with biases and slants from 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 and 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 production process as Wikipedia. There is some evidence that allocation of editorial attention is only weakly correlated with reader interest in Wikipedia in general for numerous reasons (Gorbetai 2014). Hence, our finding frames an open question about how to conceptualize the optimal allocation of editorial time and user contribution in production models based on collective intelligence.

We also conclude that many of the differences between Wikipedia and Britannica arise because Wikipedia faces insignificant storage, production, and distribution costs. This leads to longer articles with greater coverage of more points of view. In the best cases, it reduces slant and bias to a negligible difference with an expert-based model.

We see no reason to be sanguine. Concerns about bias and slant arise in discourse in a wide array of current events and scientific topics. As Wikipedia becomes the primary source of comprehensive information for many online readers, it might be suggested that there is a strong incentive for those with strong views to seek to manipulate the site's content to foster specific points of view. As the world moves from reliance on expert-based production of knowledge 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 the production model fully understood.

While this study focuses on a setting where we can implement a viable empirical strategy, the dilemma arises in many communities outside of Wikipedia and Britannica. 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 many Wikipedia alumni to help communities interested in topics that Wikipedia's communities considers inappropriate, such as cooking, celebrity gossip, popular music, movies, gaming, and hobbies, to name a few such topics. It uses principles and norms similar to those at Wikipedia to guide its communities of contributors and attract readers (Greenstein et al. 2009). Our results imply two normative pieces of advice for such a community site: representing both sides of an issue typically takes a lot of contributions and considerable revision; and length by itself is usually insufficient to get a balanced view unless it also involved considerable revision.

A similar piece of managerial advice goes for aggregation efforts inside of closed communities within private firms (Surowiecki 2004; Wagner 2005). Many private firms today use Wikis to organize internal efforts at knowledge management. These tools are viewed as well suited for aggregating information from many unverifiable sources, and our results imply that this strength is also a potential weakness in the absence of close managerial oversight. Among practitioners there is considerable debate about how closely to moderate the activities because strong views can take over text if only a few employees regularly participate, and without many revisions, the knowledge can easily become biased and slanted. Our findings favor the view that managers must do more than just offer guidelines. Community managers can work towards a balanced view if intervention alleviates disputes and generates the right kind of participation.

Appendix

Procedure for Computing Slant Index

In G&S, for each Congressperson c, they observe their ideology y_c and phrase frequency f_{pc}, the number of times phrase p appears in Congressperson c's speech, for each phrase p. For each phrase p, G&S regress the relative frequency $\overline{f_{pc}}$, where $\overline{f_{pc}} = f_{pc} / \sum_{p \in P} f_{pc}$, on y_c, and obtain the intercept and slope parameters a_p and b_p, for each phrase p. The parameter values, together with the 1,000 phrases identified by G&S, are available at http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/26242.

[8] Source: http://www.wikia.com/, accessed September 2014.

The 1,000 phrases exhibit heterogeneous slants. To mitigate the effect of outlier phrases (e.g., "African American" and "illegal immigration"), we set the parameter values for the 9 most left-leaning phrases and 9 most right-leaning phrases to be the same as the 10th most left- and most right-leaning phrases, respectively.

For each Wikipedia article n, we regress $\overline{f_{pn}} - a_p$, where $\overline{f_{pn}}$ is the relative frequency of phrase p in the article, on b_p for the 1,000 phrases, and obtain the slope estimate $\overline{Y_n} = \frac{\sum_{p \in P} b_p (\overline{f_{pn}} - a_p)}{\sum_{p \in P} b_p^2}$. When an article has none of the 1,000 phrases, $\overline{Y_n}$ is 0.4975, so we set $Y_n = \overline{Y_n} - 0.4975$, and use Y_n as our slant index for article n.

References

Afuah, A., C. Tucci. 2012. Crowdsourcing as a solution to distant search. *Academy of Management Review* **37**(3) 355–375.

Agrawal, A., C. Catalini, A. Goldfarb. 2013. The simple economics of crowdfunding. *Innovation Policy and the Economy* **14**(1) 1–46.

Algan, Y., Y. Benkler, M. F. Morell, J. Hergueux. 2013. Cooperation in a peer production economy: Experimental evidence from Wikipedia. Working paper, NBER Workshop on the Economics of IT and Digitization.

Antweiler, W., M. Z. Frank. 2004. Is all that talk just noise? The information content of Internet stock message boards. *The Journal of Finance* **59**(3) 1259–1294.

Banerjee, A. V. 1992. A simple model of herd behavior. *Quarterly Journal of Economics* **107**(3) 797–817.

Barsade, S. G. 2002. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly* **47**(4) 644–675.

Bikhchandani, S., D. Hirshleifer, I. Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* **100**(5) 992–1026.

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

Bragues, G. 2007. Wiki-philosophizing in a marketplace of ideas: Evaluating Wikipedia's entries on seven great minds. *MediaTropes eJournal* **2**(1) 117-158.

Brown, A. R. 2011. Wikipedia as a data source for political scientists: Accuracy and completeness of coverage. *Political Science and Politics* **44**(2) 339-343.

Clauson, K. A., H. H. Polen, M. N. K. Boulos, Dzenowagis. 2008. Scope, completeness, and accuracy of drug information in Wikipedia. *The Annals of Pharmacotherapy* **42**(12) 1814–21.

DellaVigna, S., E. Kaplan. 2007. The Fox news effect: Media bias and voting. *Quarterly Journal of Economics* **122**(3) 1187-1234

Devgan, L., N. Power, B. Blakey, M. Makary. 2007. Wiki-surgery? Internal validity of Wikipedia as a medical and surgical reference. *Journal of the American College of Surgeons* **205**(3) S76-S77.

Evans, P., T. S. Wurster. 2000. *Blown to Bits: How the New Economics of Information Transforms Strategy.* Harvard Business School Press, Boston.

Franzoni, C., H. Sauermann. 2014. Crowd science: The organization of scientific research in open collaborative projects. *Research Policy* **43**(1) 1-20.

Galton, F. 1907. Vox populi. *Nature* **75**(1949) 450-451.

Gentzkow, M., J. M. Shapiro. 2010. What drives media bias? Evidence from newspapers. *Econometrica* **78**(1) 35-71.

Gentzkow, M., J. M. Shapiro. 2011. Ideological segregation online and offline. *Quarterly Journal of Economics* **126**(4) 1799–1839.

Giles, J. 2005. Internet encyclopaedias go head to head. *Nature* **438**(7070) 900–901.

Gorbatai, A. 2014. The paradox of novice contributions to collective production: Evidence from Wikipedia. Working Paper, University of California, Berkeley.

Greenstein, S., M. Devereux. 2009. Wikipedia in the spotlight. Kellogg School of Management Case No. 5-306-507, Northwestern University, Chicago.

Greenstein, S., R. Frazzano, E. Meagher. 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., F. Zhu. 2012. Is Wikipedia biased? *American Economic Review: Papers and Proceedings* **102**(3) 343–348.

Greenstein, S., F. Zhu. 2013. Neutral point of view and collective intelligence bias: The case of Wikipedia. NBER Working Paper 18167.

Hasty, R. T., R. C. Garbalosa, V. A. Barbato, P. J. Valdes, D. W. Powers, E. Hernandez, J. S. John, G. Suciu, F. Qureshi, M. Popa-Radu, S. San Jose, N. Drexler, R. Patankar, J. R. Paz, C. W. King, H. N. Gerber, M. G. Valladares, A. A. Somji. 2014. Wikipedia vs. peer-reviewed medical literature for information about the 10 most costly medical conditions. *Journal of the American Osteopathic Association* **114**(5) 368-373.

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

Kuppuswamy, V., B. L. Bayus. 2013. Crowdfunding creative ideas: The dynamics of project backers in Kickstarter. Working Paper, UNC Kenan-Flagler Research Paper No. 2013-15.

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

Lakhani, K. R., D. A. Garvin, E. Lonstein. 2012. TopCoder (A): Developing Software through Crowdsourcing. Harvard Business School Case 610-032, Harvard University, Cambridge.

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

Levine, R. 2006. New web sites seeking profit in wiki model. *The New York Times*, September 4, 2006, available at http://www.nytimes.com/2006/09/04/technology/04wiki.html.

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

Melcher, R. A. 1997. Dusting off the Britannica. *Business Week,* October 20, 1997.

Mollick, E. 2014. The dynamics of crowdfunding: Determinants of success and failure. *Journal of Business Venturing* **29**(1) 1–16.

Mollick, E. R., R. Nanda. 2014. Wisdom or madness? Comparing crowds with expert evaluation in funding the arts. Working Paper, Harvard Business School, Boston.

Mollick, E., V. Kuppuswamy. 2014. Crowdfunding: evidence on the democratization of startup funding. K. Lakhani, D. Harhoff, eds. *Revolutionizing Innovation: Users, Communities and Openness*. MIT Press, Cambridge.

Mullainathan, S., A. Shleifer. 2005. The market for news. *American Economic Review* **95** (4) 1031–1053.

Piskorski, M. J., A. Gorbatai. 2013. Testing Coleman's social-norm enforcement mechanism: Evidence from Wikipedia. Working Paper, Harvard Business School, available at http://ssrn.com/abstract=1926284.

Rajagopalan, M. S., V.K. Khanna, Y. Leiter, M. Stott, T.N. Showalter, A.P. Dicker, Y. R. Lawrence. 2011. Patient-oriented cancer information on the Internet: A comparison of Wikipedia and a professionally maintained database. *Journal of Oncology Practice* **7**(5) 319–23.

Shankland, S. 2003. Study lauds open-source code quality. *CNET News*, February 19, 2003, available at http://news.cnet.com/Study-lauds-open-source-code-quality/2100-1001_3-985221.html.

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

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

Swartz, A. 2006. Who writes Wikipedia?. September 4, 2006, available at http://tinyurl.com/7n84bsh.

Terwiesch, C., K. Ulrich. 2009. *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. Harvard University Press, Boston, 256.

Von Hippel, E. 2005. *Democratizing Innovation*. MIT Press, Cambridge, MA.

Wagner, C. 2005. Supporting knowledge management in organizations with conversational technologies: Discussion forums, weblogs, and wikis. *Journal of Database Management* **16**(2) i-viii.

Wu, L., E. Brynjolfsson. 2013. The future of prediction: How Google searches foreshadow housing prices and sales. Working Paper.

Yasseri, T., A. Spoerri, M. Graham, J. Kertesz. 2014. The most controversial topics in Wikipedia: A multilingual and geographical analysis. P. Fichman, N. Hara, eds. *Global Wikipedia: International and Cross-Cultural Issues in Online Collaboration.* Rowman & Littlefield, Lanham, MD, 25-48.

Zhang, J., P. Liu. 2012. Rational herding in microloan markets. *Management Science* **58(**5) 892–912.

Table 1: Comparing Analysis Sample to Initial Wikipedia Sample

Category	Analysis 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%

Table 2: Comparing Slants in Wikipedia and Britannica Articles in Analysis Sample

	No. of Obs.	Wikipedia		Britannica		Difference
Abortion	13	-0.14	0.23	-0.06	0.18	-0.07
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
American Politicians	438	-0.05	0.20	-0.05	0.19	0.00

*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 3: Comparing Biases in Wikipedia and Britannica Articles in Analysis Sample

	No. of Obs.	Wikipedia		Britannica		Difference
Abortion	13	0.19	0.19	0.08	0.18	0.11*
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***
American Politicians	438	0.14	0.15	0.10	0.17	0.04***

*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 4: Summary Statistics

Wikipedia Articles					
Variable	Obs.	Mean	Std. Dev.	Min	Max
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
Contributors	3,918	839.50	1,077.40	1.00	14,160
Revisions	3,918	1,924.23	2,826.28	1	44,880
Slant/Length	3,918	-0.00003	0.00019	-0.0042	0.0013
Bias/Length	3,918	0.00007	0.00018	0	0.0042

Britannica Articles					
Variable	Obs.	Mean	Std. Dev.	Min	Max
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 5: 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]	[0.000]	[0.000]
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	3918	3,918

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

Table 6: 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

*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%.*

Table 7a: Fixed Effects Regressions to Compare 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	-0.037***	-0.011**	0.076***	0.020***	0.000***	-0.000***
	[0.004]	[0.005]	[0.003]	[0.005]	[0.000]	[0.000]
Log(Length)		-0.015***		0.031***		
		[0.002]		[0.002]		
Observations	7,626	7,626	7,626	7,626	7,626	7,626
Adjusted R-squared	0.027	0.035	0.132	0.169	0.002	0.026
Number of Articles	3,813	3,813	3,813	3,813	3,813	3,813

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

Robustness: We exclude all matched articles if either its Wikipedia or Britannica version has a length longer than two standard deviations above the mean. In total, out of 3,918 pairs of the articles, 105 pairs are excluded. We then replicate the analysis in Table 5.

Table 7b: OLS Regressions to Examine the Impact of Revisions on Biases of Wikipedia Articles

Model	(1)	(2)	(3)	(4)
Dependent Variable	Wikipedia Bias	Wikipedia Bias	Wikipedia Bias	Wikipedia Bias
Britannica Bias	0.238***	0.245***	0.262***	0.223***
	[0.022]	[0.022]	[0.022]	[0.022]
Log(Length)	0.022***	0.033***	0.031***	0.025***
	[0.002]	[0.004]	[0.004]	[0.003]
Log(Revisions)		-0.010***	-0.017***	-0.014***
		[0.003]	[0.004]	[0.004]
Average Revisions Per Contributor		-0.001	0.008*	0.006
		[0.004]	[0.004]	[0.004]
Year Created = 2002			0.065***	0.050***
			[0.007]	[0.007]
Year Created = 2003			0.005	0.006
			[0.008]	[0.008]
Year Created = 2004			0.009	0.007
			[0.010]	[0.010]
Year Created = 2005			-0.023*	-0.020
			[0.013]	[0.013]
Year Created = 2006			-0.033*	-0.026
			[0.017]	[0.017]
Year Created = 2007			-0.034*	-0.024
			[0.020]	[0.020]
Year Created = 2008			-0.062***	-0.046**
			[0.020]	[0.020]
Year Created = 2009			-0.094***	-0.083***
			[0.015]	[0.017]
Year Created = 2010			-0.109***	-0.085***
			[0.017]	[0.018]
Year Created = 2011			-0.191***	-0.165***
			[0.016]	[0.020]
Dummies for Categories	No	No	No	Yes
Observations	3,813	3,813	3,813	3,813
Adjusted R-squared	0.071	0.073	0.111	0.186

*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%.*

Robustness: We exclude all matched articles if either its Wikipedia or Britannica version has a length longer than two standard deviations above the mean. In total, out of 3,918 pairs of the articles, 105 pairs are excluded. We then replicate the analysis in Table 6.

Table 8a: Fixed Effects Regressions to Compare 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	-0.037***	-0.013***	0.075***	0.024***	0.000***	-0.000***
	[0.004]	[0.005]	[0.003]	[0.004]	[0.000]	[0.000]
Log(Length)		-0.014***		0.030***		
		[0.002]		[0.002]		
Observations	7,736	7,736	7,736	7,736	7,736	7,736
Adjusted R-squared	0.027	0.034	0.130	0.168	0.002	0.024
Number of Articles	3,868	3,868	3,868	3,868	3,868	3,868

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

Robustness: We identify all articles whose titles contain code words, and exclude these articles from the analysis. Fifty pairs (1.3%) are excluded. We then replicate the analysis in Table 5.

Table 8b: OLS Regressions to Examine the Impact of Revisions on Biases of Wikipedia Articles

Model	(1)	(2)	(3)	(4)
Dependent Variable	Wikipedia Bias	Wikipedia Bias	Wikipedia Bias	Wikipedia Bias
Britannica Bias	0.232***	0.239***	0.257***	0.221***
	[0.022]	[0.022]	[0.022]	[0.022]
Log(Length)	0.020***	0.033***	0.030***	0.025***
	[0.002]	[0.003]	[0.004]	[0.003]
Log(Revisions)		-0.011***	-0.017***	-0.014***
		[0.003]	[0.004]	[0.004]
Average Revisions Per Contributor		-0.002	0.006	0.004
		[0.004]	[0.004]	[0.004]
Year Created = 2002			0.067***	0.051***
			[0.007]	[0.007]
Year Created = 2003			0.008	0.008
			[0.008]	[0.008]
Year Created = 2004			0.009	0.007
			[0.010]	[0.010]
Year Created = 2005			-0.022*	-0.019
			[0.013]	[0.013]
Year Created = 2006			-0.031*	-0.024
			[0.017]	[0.017]
Year Created = 2007			-0.036*	-0.025
			[0.019]	[0.019]
Year Created = 2008			-0.060***	-0.044**
			[0.020]	[0.020]
Year Created = 2009			-0.093***	-0.082***
			[0.015]	[0.017]
Year Created = 2010			-0.107***	-0.082***
			[0.017]	[0.017]
Year Created = 2011			-0.187***	-0.161***
			[0.016]	[0.019]
Dummies for Categories	No	No	No	Yes
Observations	3,868	3,868	3,868	3,868
Adjusted R-squared	0.065	0.068	0.107	0.186

*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%.*

Robustness: We identify all articles whose titles contain code words, and exclude these articles from the analysis. Fifty pairs (1.3%) are excluded. We then replicate the analysis in Table 6.

Table 9a: Fixed Effects Regressions to Compare 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	-0.071***	-0.024***	0.127***	0.042***	0.000	-0.000***
	[0.005]	[0.007]	[0.005]	[0.006]	[0.000]	[0.000]
Log(Length)		-0.027***		0.049***		
		[0.003]		[0.003]		
Observations	7,836	7,836	7,836	7,836	7,836	7,836
Adjusted R-squared	0.042	0.053	0.145	0.184	0.001	0.007
Number of Articles	3,918	3,918	3,918	3,918	3,918	3,918

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

Robustness: We use 638 words out of the 1,000 code words that are identified to have substitutes as slanted code words and repeat the analysis of Table 5.

Table 9b: OLS Regressions to Examine the Impact of Revisions on Biases of Wikipedia Articles

Model	(1)	(2)	(3)	(4)
Dependent Variable	Wikipedia Bias	Wikipedia Bias	Wikipedia Bias	Wikipedia Bias
Britannica Bias	0.257***	0.260***	0.286***	0.229***
	[0.026]	[0.026]	[0.027]	[0.027]
Log(Length)	0.041***	0.047***	0.043***	0.034***
	[0.003]	[0.005]	[0.005]	[0.005]
Log(Revisions)		-0.005	-0.019***	-0.010*
		[0.004]	[0.006]	[0.006]
Average Revisions Per Contributor		-0.002	0.012	0.007
		[0.007]	[0.007]	[0.007]
Year Created = 2002			0.096***	0.063***
			[0.013]	[0.013]
Year Created = 2003			-0.006	-0.007
			[0.014]	[0.014]
Year Created = 2004			-0.023	-0.029*
			[0.016]	[0.017]
Year Created = 2005			-0.042**	-0.038*
			[0.020]	[0.021]
Year Created = 2006			-0.053**	-0.041
			[0.026]	[0.027]
Year Created = 2007			-0.107***	-0.086***
			[0.026]	[0.027]
Year Created = 2008			-0.135***	-0.106***
			[0.024]	[0.024]
Year Created = 2009			-0.125***	-0.100***
			[0.025]	[0.028]
Year Created = 2010			-0.147***	-0.098***
			[0.032]	[0.031]
Year Created = 2011			-0.269***	-0.228***
			[0.027]	[0.033]
Dummies for Categories	No	No	No	Yes
Observations	3,918	3,918	3,918	3,918
Adjusted R-squared	0.070	0.071	0.105	0.188

*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%.*

Robustness: We use 638 words out of the 1,000 code words that are identified to have substitutes as slanted code words and repeat the analysis of Table 6.

Table 10a: Fixed Effects Regressions to Compare 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	-0.050***	-0.020***	0.087***	0.031***	0.000	-0.000***
	[0.004]	[0.006]	[0.004]	[0.005]	[0.000]	[0.000]
Log(Length)		-0.017***		0.032***		
		[0.003]		[0.002]		
Observations	5,720	5,720	5,720	5,720	5,720	5,720
Adjusted R-squared	0.049	0.060	0.162	0.205	0.001	0.014
Number of Articles	2,860	2,860	2,860	2,860	2,860	2,860

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

Robustness: We identify articles that exist in the 2001 Britannica edition and repeat the analysis using these 2,860 articles with their matched Wikipedia versions, as found in Table 5.

Table 10b: OLS Regressions to Examine the Impact of Revisions on Biases of Wikipedia Articles

Model	(1)	(2)	(3)	(4)
Dependent Variable	Wikipedia Bias	Wikipedia Bias	Wikipedia Bias	Wikipedia Bias
Britannica Bias	0.240***	0.254***	0.275***	0.232***
	[0.025]	[0.026]	[0.026]	[0.026]
Log(Length)	0.018***	0.033***	0.028***	0.022***
	[0.002]	[0.004]	[0.004]	[0.004]
Log(Revisions)		-0.016***	-0.021***	-0.014***
		[0.004]	[0.005]	[0.005]
Average Revisions Per Contributor		0.006	0.011**	0.006
		[0.005]	[0.006]	[0.005]
Year Created = 2002			0.075***	0.055***
			[0.009]	[0.008]
Year Created = 2003			0.006	0.007
			[0.011]	[0.010]
Year Created = 2004			-0.003	-0.003
			[0.013]	[0.012]
Year Created = 2005			-0.043***	-0.036**
			[0.016]	[0.016]
Year Created = 2006			-0.064***	-0.053**
			[0.022]	[0.021]
Year Created = 2007			-0.054	-0.028
			[0.035]	[0.037]
Dummies for Categories	No	No	No	Yes
Observations	2,860	2,860	2,860	2,860
Adjusted R-squared	0.055	0.061	0.109	0.211

*Heteroskedasticity-adjusted standard errors in brackets. All articles in the dataset are created on or before 2007. In Models (3) and (4), Year Created = 2001 is used as the benchmark group. * significant at 10%; ** significant at 5%; *** significant at 1%.*

Robustness: We identify articles that exist in the 2001 Britannica edition and repeat the analysis using these 2,860 articles with their matched Wikipedia versions, as found in Table 6.