NBER WORKING PAPER SERIES

COLLECTIVE INTELLIGENCE AND NEUTRAL POINT OF VIEW: THE CASE OF WIKIPEDIA

Shane Greenstein
Feng Zhu

Working Paper 18167
http://www.nber.org/papers/w18167

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2012

Collective Intelligence and Neutral Point of View: The Case of Wikipedia
Shane Greenstein and Feng Zhu
NBER Working Paper No. 18167
June 2012
JEL No. L17,L3,L86

ABSTRACT

We examine whether collective intelligence helps achieve a neutral point of view using data from a decade of Wikipedia's articles on US politics. Our null hypothesis builds on Linus' Law, often expressed as "Given enough eyeballs, all bugs are shallow." Our findings are consistent with a narrow interpretation of Linus' Law, namely, a greater number of contributors to an article makes an article more neutral. No evidence supports a broad interpretation of Linus' Law. Moreover, several empirical facts suggest the law does not shape many articles. The majority of articles receive little attention, and most articles change only mildly from their initial slant.

Shane Greenstein
The Elinor and Wendell Hobbs Professor
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL 60208-2013
and NBER
greenstein@kellogg.northwestern.edu

Feng Zhu
University of Southern California
fzhu@marshall.usc.edu

An online appendix is available at:
http://www.nber.org/data-appendix/w18167

1. Introduction

The diffusion of the Internet and digital media has significantly increased the number of opportunities for individuals around the globe to collaborate with each other. As a result, an increasing number of organizations or communities today are harnessing the power of collective intelligence—a shared intelligence that emerges from the collaboration of many individuals—to tackle problems that are too big to be solved by themselves. For example, Wikipedia, the largest encyclopedia on the Web, relies entirely on free contributions. Threadless.com uses a community of over 500,000 people to design and select T-shirts (Malone et al. 2009). Many open-source software products are developed, distributed and supported on a voluntary basis by and for users (e.g., Lakhani and von Hippel 2003; Ma et al. 2008; Roberts et al. 2006; Singh et al. 2011).

Collective intelligence has generated tremendous interest both in practice and in academic literature for a number of reasons. Studies have documented that collective intelligence can lead to more accurate predictions. For example, Galton (1907) and Surowiecki (2004) show that the median estimate of a group when guessing the weight of an ox, stock prices, or winners in political elections can be more accurate than experts' estimates. Antweiler and Frank (2004) find that the information content from Internet stock message boards can predict the behavior of financial markets. Wu and Brynjolfsson (2009) demonstrate that search engines like Google can take advantage of queries submitted by their users to provide highly accurate predictions of future housing prices. Dewan and Ramaprasad (forthcoming) show that aggregate data from music blogs are correlated with consumers' sampling behavior of music tracks.

Several studies also find that collective intelligence can lead to high-quality output. von Hippel (2005) finds that in some industries such as kitesurfing, the collective product-design and testing work of user innovation communities are a better source of innovation than companies' own in-house development teams. Several studies find that open source software products have fewer bugs than competing closed-source software (e.g., Shankland 2003; Lemos 2004). In a similar vein, Giles (2005) finds that Wikipedia is about as good a source of accurate scientific information as Britannica, an encyclopedia authored by experts.

While these prior studies find that collective intelligence performs well in the context of uncontroversial and verifiable information, it is unclear whether a production model based on collective intelligence will perform equally well when information is controversial, subjective, and unverifiable. It is

also unclear whether collective intelligence can be harnessed to produce any desirable outcome in such settings.

In this study, we examine whether collective intelligence in Wikipedia helps achieve a neutral point of view in US politics. A "Neutral Point of View" (NPOV) is one of the tenets that all Wikipedia articles aspire to achieve, along with "verifiability" and "the absence of original research." If an article reflects NPOV, then conflicting opinions are presented next to one another, with all significant points of view represented. Achieving NPOV has been a goal for Wikipedia's contributors and editors since its founding.

One would expect that NPOV should not be difficult to achieve when articles cover uncontroversial topics loaded with objective information that can be verified against many sources. Such a setting characterizes the vast majority of Wikipedia articles about established scientific topics, for example. What about topics lacking these ideal features? What biases arise in topics where information is controversial, subjective, and unverifiable? In the context of Wikipedia, although most contributors try to diffuse issues with a fair representation, collective intelligence bias (CIB), the opposite of NPOV, may arise for a number of reasons. For example, some issues are simply too complex for contributors to resolve a dispute, such as in the case of interpreting the science behind global warming. Anyone can verify the same objective data, but generating a consensus about what it all means takes considerable effort and expertise. CIB can also survive because of the difficulties editing subjective information that is costly to verify.

The study sheds light on Wikipedia's revision processes. We use Linus' Law to build our null hypothesis about this process. "Linus' Law" often is expressed as the slogan, "Given enough eyeballs, all bugs are shallow" (Raymond 1998). Many editors and contributors of Wikipedia believe that it governs the emergence of NPOV during revision, and many participants in open-source communities regard the law as foundational. According to a narrow interpretation of Linus' Law, articles should come closer to NPOV as more contributors scrutinize them and make contributions. In a broad interpretation, a more widely dispersed set of contributors also should contribute to the production of NPOV.

The near-decade of experience at Wikipedia provides sufficient variance to test Linus' Law. With Wikipedia's size and scale, not all articles receive the same amount of attention and contributions. Some have accumulated many contributions over time while others have not. Articles also vary in the concentration of contributors they attract. In addition, Wikipedia retains prodigious records of its revisions, which allow for a detailed statistical analysis.

We apply the null hypothesis to a sample of 28,382 entries about US political topics in 2011. We select these articles for two reasons. First, it is an interesting place to look. Achieving NPOV faces challenges when articles cover controversial topics, and rely on subjective information that is costly to verify. Hence, we presume that Linus' Law has its highest probability of failing. Alternatively, if Linus' Law succeeds here, it is likely to succeed for other topics. Second, it is feasible to measure bias, building on an approach pioneered by Gentzkow and Shapiro (2010), and the broader literature examining content bias. Those gains come with one drawback. It is not available for all articles about US politics (over 70,000 in 2011), so we must consider whether zero slant and bias signals merely lack of information or sample selection issues.

Wikipedia's history also provides some interesting context for this study. Greenstein and Zhu (2012) show that in its earliest years, Wikipedia's political entries lean Democrat, on average, and tend to be biased. Both of these traits diminish over time. By the most recent observation, on average, Wikipedia's entries lack much slant and contain (moderately) less bias than observed earlier. What role did revision play in these trends? Oversimplifying somewhat, if Linus' Law holds, then older articles could lose their slant through more revision, diminishing bias and slant. Alternatively, if Linus' Law does not hold, then another mechanism, such as the entry of articles with opposite slant, is responsible for the aggregate decline in average slant over time.

This study shows that the evidence supports a narrow interpretation of Linus' Law. Only one feature of the revision process—namely, number of reviewers—shape the slant or bias of an article in the direction that the Law predicts. Moreover, several facets of the revision process do not shape revisions in the anticipated direction, and no evidence supports a broad interpretation. The evidence further points to persistence of bias in many articles, potentially consistent with the presence of CIB. This is partly a vintage effect and partly a result of the topic covered by the article. Some topics, such as entries on civil rights, tend to lean Democrat, and some, such as trade, lean Republican. Finally, we show that a fundamental resource issue constrains the implementation of Linus' Law, namely, due to not enough contributors, only a small percentage of articles ever get enough contributions to enable the law to have an opportunity to work.

Readers who are interested in open source will take an interest in this study. The vast majority of research on open source examines the production of programming code (e.g., MacCormack et al. 2006; Haefliger et al. 2008), not contributions to content development, and little examines Linus' Law directly. Also, little work considers content production from an aggregation of contributions from a large numbers of volunteers, as observed in Wikipedia. Wikipedia is a natural subject for the open source

community to examine because of the attention it receives. In most countries with developed Internet sectors Wikipedia ranks among the top-ten web sites visited by households.[1] In the US Wikipedia is one of the most popular web sites in which user-generated content plays a prominent role.

Our research also contributes to the broad literature examining content bias. Scholars have identified various sources of bias in media content, such as pressure from advertisers or the government (e.g., Price 2003; Besley and Prat 2006; Reuter and Zitzewitz 2006), the media's partisan bias (Larcinese et al. 2007), and readers' desire to reinforce their prior beliefs (e.g., Groseclose and Milyo 2005; Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006; Bernhardt et al. 2008; Gal-Or et al. 2010; Gentzkow and Shapiro 2010). Unlike these prior studies, we examine user-generated content and focus on factors that cause content bias to change over time.

We also provide empirical evidence on whether the Internet is increasing ideological segregation (e.g., Sunstein 2001; Carr 2008; Lawrence, Sides and Farrell 2010; Gentzkow and Shapiro 2011). Our results support the view that prominent articles are not isolated. Our results are consistent with the view that contributors with different political viewpoints have dialogues with each other, and that diminishes the slant of articles. In addition, the general movement in Wikipedia's overall slant suggests entry of new opinions is not precluded. On the other hand, most Wikipedia articles only mildly change their slant, consistent with the view they might receive more attention from readers with similar viewpoints.

This study also adds to the empirical studies of Wikipedia, a topic of interest in its own right. Prior work emphasizes the social network behind editing (Zhang and Zhu 2011; Ransbotham and Kane 2011), the dynamics of contributions (Chi et al. 2007, Halfaker et al. 2009), the accuracy of articles (Giles 2005; Brown 2011), the social influences on the gamesmanship among editors (Piskorski and Gorbetai 2010), allocation of effort among topics (Gorbetai 2011) and the impact of Wikipedia on market information environment (Xu and Zhang forthcoming).[2] Our study is the first to develop statistical tests for whether Wikipedia articles achieve NPOV, and to translate widely discussed ideas about Linus' Law into testable propositions. It is also the first to raise questions about limitations of Linus' Law, and, related, this study frames several new open questions about revision, such as feedbacks between an article's bias and further contributions.

[1] See the rankings at Alexa.com. Wikipedia is the fifth or sixth most-visited web site in the United States, behind Google, Facebook, Yahoo, YouTube, and, perhaps, eBay (accessed May 2011).

[2] See a list of academic studies about Wikipedia at http://en.wikipedia.org/wiki/Wikipedia:Academic_studies_of_Wikipedia (accessed June 2012).

The rest of the paper is organized as follows. We first provide a background of Wikipedia in Section 2 and discuss the role of NPOV on Wikipedia in Section 3. We then develop our hypotheses in Section 4. We present our data and summary statistics in Section 5. Section 6 presents our regression results, and Section 7 concludes.

2. The Emergence of Wikipedia

The first wiki was developed in 1995 by Ward Cunningham, a software engineer from Portland, Oregon. Wikis were first developed and intended for documenting software development. Says Cunningham (Levine, 2006), "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."

Wikipedia was founded in 2001 when Wikipedia began to position itself as "the free encyclopedia that anyone can edit," that is, as an online encyclopedia that is entirely written and edited through user contributions. As of November 2011, it supports 3.7 million articles in English and well over 20 million articles in all languages. It hosts content that hundreds of millions of readers view each month.

Wikipedia's production defies simple characterization. Because it relies so heavily on user-generated content, it does not fit existing models of production, in which a fixed sequence of activities produces an output following a pre-specified design. Instead, Wikipedia uses a commons-based approach to aggregate and revise information from a widely dispersed set of contributors and it produces non-proprietary information.

This study examines Wikipedia just prior to its tenth birthday, which it is the largest Wiki on the planet. 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. Until 2006 the foundation operated with a minimal staff of two programmers, under the supervision of Jimbo Wales, but by 2010, the staff had grown to include a full-time professional manager and several dozen employees. Virtually all the content continues to come from volunteers.

Wikipedia operates under an open-source license. When Wikipedia first began, most images and other content were covered by the GNU free documentation license (GFDL), a variant on the more popular GPL designed for manuals, textbooks, and reference materials. With the latter, contributions remained the property of their creators, whereas the GFDL license ensured the content would remain freely distributable and reproducible. More recently, most content is dual-licensed under both the GFDL

and/or the Creative Commons Attribution-Sharealike 3.0 Unported License (CC-BY-SA).[3] Copies can be sold commercially, but if produced in larger quantities, then the original document or source code must be made available.

Wiki server technology allows the creation of hypertexts with nonlinear navigation structures. Each page contains a series of cross-links to other pages. The reader decides how to navigate through the site. Contributing to Wikipedia is easy and transparent. Contributors do not need specialized knowledge.

As there is no editorial control from the center, Wikipedia relies on users for fixing errors. Wikipedia started with almost no contribution restrictions, and as it grew, it developed a few restricted privileges to facilitate administration. It primarily relies on civility and transparency to govern contributors. Any entry can change if a contributor thinks that changes are warranted. As stated by a long-time editor who tested a number of articles: "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 (Schwarz, 2006)."

Wikipedia contains many articles that do not differ markedly from those in a printed encyclopedia, such as entries devoted to basic history or science. It also has many entries for general topics in geography and politics. Yet many Wikipedia entries do not neatly fit into a single category, many too obscure for attention in a traditional encyclopedia. It faces no limits on the number or size of articles, though a norm developed to keep articles under 6-8 thousand words.

3. The Role of NPOV

The site is organized in a way that presumes all errors will be corrected given enough review. This follows a shared assumption among all major participants: Wikipedia follows Linus' Law, "Given enough eyeballs, all bugs are shallow," which Eric Raymond stated in "The Cathedral and the Bazaar."[4]

Many participants in open-source communities consider Linus' Law to be a foundational principle. For example, ask an editor for Wikipedia about whether Linus' Law works well, and the answer is likely to emphasize the editing process; it comes back to believing in the power of an open-revision process that enables multiple users to edit any passage. Wikipedia's own page about contributing reads, "Many

[3] http://en.wikipedia.org/wiki/Wikipedia:About (accessed August 2011).

[4] See Raymond (1998), who was rephrasing Linus Torvald, founder of the open-source operating system, Linux. Torvald's rule No. 8 is: "Given a large enough beta-tester and co-developer base, almost every problem will be characterized quickly and the fix obvious to someone."

users of Wikipedia consult the page history[5] of an article in order to assess the number of people who have contributed to the article. An article can be considered more likely to be accurate when it has been edited by many different people."[6] Founder of Wikipedia, Jimbo Wales, reiterated the idea in his public comments: "I think the day will come in the future when people will look at an article in Britannica and say, 'This was written by one person and reviewed by two or three more? That's not sufficient. I need an article that's been reviewed by hundreds of people (National Public Radio, 2005).'"

Wikipedia has policies in place to nurture revisions. First, since founding Wikipedia has asked all contributors to aspire to write or edit with a NPOV, representing views fairly and without bias. Conflicting opinions are supposed to be presented alongside one another, not asserted in a way that is meant to be convincing. This sometimes was boiled down to the principle to "assert facts, including facts about opinions—but do not assert the opinions themselves."

The cost of representing additional viewpoints was low, so the judgment of the editors created the primary limit on multiple viewpoints. According to Wales: "If a viewpoint is held by an extremely small (or vastly limited) minority, it does not belong in Wikipedia regardless of whether it is true or not and regardless of whether you can prove it or not."[7]

Verifiability is the second aspiration for contributors. Any reader must be able to check an article's contents and verify against reliable sources. Editors have to be able to cite these sources in their articles and provide links if possible. Editors understand that verifiability is not equivalent to truth; the editor is not responsible for determining whether the information in a newspaper article he or she cites is true, as long as the newspaper is a reliable, peer-reviewed source.

Finally, contributors are asked not to include original research in their contributions. All material must have been previously published by a reputable source. Alternatively, a reasonable adult should understand the concept (i.e., a "vegetable" does not need to be published by a reliable source to be permitted an article in Wikipedia). This policy was put in place in order to avoid a "novel narrative or historical interpretation" of a subject.[8]

Enforcing these policies and aspirations created many challenges. Over time, the site has adopted a design that makes it simple for contributors to monitor each other. Editors and contributors can subscribe to follow (or "watchlist") articles to check if they have been changed. "Being very transparent encourages good behavior," Wales said (Hyatt, 2006). Furthermore: "Everything is very carefully

[5] Page histories allow a reader to trace the history of edits in reverse chronological order.
[6] http://en.wikipedia.org/wiki/Wikipedia:Who_writes_Wikipedia (accessed April 2012).
[7] http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view (accessed April 2012).
[8] Jimmy Wales, private correspondence, August 28, 2006.

monitored by a core community who is constantly watching the site, constantly discussing, reviewing changes that are coming in . . . If [a user] is something of an outsider to the community, his changes when they come in will be noticed as, oh, well, this is somebody we don't know and we'll check it over and if it seems fine, it'll stand. Otherwise, it can be removed very quickly (National Public Radio, 2005)."

Enforcing NPOV has become the focal point for discussion by those constructing entries in Wikipedia. Many of the back-channel conversations on Wikipedia-dedicated Internet Relay Chat (IRC) channels concern whether particular passages reflect this principle. In general, the vast majority of entries settle on approaches that the wide community of contributors agrees to, either because such agreements reflect a consensus or because those with minority opinions got the passage they wanted in additional test or a dissident gave up.

Could a NPOV ever exist on any of the most controversial topics? Wikipedia's editors point to the triumph of civility on even the most controversial topics, arguing that the results display a more neutral view than any printed entry. They argue that the process takes multiple views into account, achieving something printed encyclopedias do not do as well by relying on a single author.

4. Hypotheses

This study develops a statistical approach for measuring NPOV in the context of Wikipedia articles. That research goal requires translating the collective activities of many contributors, as well as the beliefs of Wikipedia's editors, into testable propositions. This study uses classical statistical approaches, employing a narrow or broad interpretation of Linus' Law, which will constitute the null. We then test predictions consistent with that null.

We presume an article is the unit of observation, both at any point in time and over time. Although there are mild exceptions to the constant identity of an article—because some articles are merged or eliminated, etc.—this is a good working assumption for the vast majority of articles. Wikipedia facilitates this approach by assigning numerical identities to articles and maintaining prodigious histories of edits, which helps identify when contributors create new articles and alter (even minor) aspects of existing articles. This also helps make it possible to measure the variance in the ages of articles and their condition over time.

As with other studies of media bias, this study posits that there exists a uni-dimensional yardstick for measuring neutrality bias. Call this aspect of an article, Y, where Y is a real number that measures its political slant. As normalization, let zero be neutral, and loosely speaking, negative is Democrat while positive is Republican. Cardinal numbers have meaning along this yardstick, with larger numbers

denoting more extreme values. Such a yardstick provides two related but somewhat different definitions for neutral/not neutral. One notion is "slant," namely, comparing Y = 0 to negative or positive numbers. Another notion compares "NPOV" to "bias," namely, comparing Y = 0 to Abs(Y), the absolute value of the slant. The first definition leads to a "slant index" and measures the size of bias *and* its direction. The second definition leads to "bias size" and measures *only* the size of bias.

This study characterizes the statistical relationship between contributions and Y or Abs(Y). One set of predictions arises intuitively from Linus' Law and its role with NPOV – "Given enough eyeballs, all bugs are shallow." It is possible to proceed under the hypothesis that this law captures a feature of the revision process, namely, that revision attenuates bias. Empirically that means thinly edited pages will have a higher likelihood of bias than thickly edited pages. Stated narrowly:

> **Hypothesis 1:** All other things equal, an article that has attracted more contributions and contributors over its lifetime will be less extreme than one that has attracted fewer contributions and contributors. Less extreme articles will have a level of Y or Abs(Y) closer to zero.

Note that Linus' Law allows for a narrow and broad interpretation. The narrow version focuses solely on the amount of contributions. The broad interpretation focuses on related aspects, such as the range of contributions or its dispersion. Based on the broad interpretation, we can have the following hypothesis:

> **Hypothesis 2:** All other things equal, an article that has attracted a more diverse set of contributors over its lifetime will be less extreme than one that has attracted a similar set of contributors. Less extreme articles will have a level of Y or Abs(Y) closer to zero.

These hypotheses are free of historical context, and many variables will try to control for factors related to vintage and year, such as changes to Wikipedia's size. Such controls are necessary because the Wikimedia Foundation has altered the site over time to enable participation from an increasingly larger group of participants. In addition, many contributors have access to improved broadband technologies, which facilitate online activities, so the composition of online readers has dramatically changed over the decade. There is no reason to expect Wikipedia's contributors to favor one or another

political persuasion, on average, so our approach is agnostic with respect to party. The test for Linus' Law will allow for more bias or less bias over time, as the number of contributors increases.

Though not labeled as such, prior work offers insight into the mechanisms that might generate NPOV or CIB. For example, the majority of statistical studies to date stress the importance of the social networks behind the editing process.[9] Frequent editors and contributors develop social ties, and these generate informal norms for when it is appropriate to edit an article. These social ties also may generate formal and informal norms about what constitutes NPOV in an article. That social understanding does not necessarily have to settle at a place that another set of observers would regard as unbiased.

A related mechanism posits a two-stage model of production with a feedback loop.[10] At the first stage, some topics attract readership, and these readers provide small edits. At the second stage, the articles that attract more interest from readers then attract more interest from editors, who provide a large number of contributions and serve as arbiters in disputes over NPOV. In turn, well-edited articles attract more readers, and so on. In this model of production, NPOV will be achieved primarily at the second stage, when it attracts editors with interests on all sides of a topic. If an article attracts strong interest from those with one view, it is possible for an article not to settle at a NPOV aligned with the views of contributors at the first stage, but at a place that the second set of editors agree upon. In that case, the articles might display CIB.

5. Data and Summary Statistics

This study's data come from Wikipedia on January 16, 2011. We develop methods to produce a data set that meets these three criteria: (1) it is possible to measure the NPOV; (2) it is possible to measure the editing process; and (3) within a set of articles, each individual article differs from the others in the amount of attention received.

5.1. Assembling a sample

This study employs a process to maximize the likelihood that at least a few of the articles contain some controversial material, or lack objective data that can be easily verified against outside sources. The initial sample of articles focuses on a broad and inclusive definition of US political topics. It examines the latest version of each article in January 2011, selecting all articles with keywords "Republican" or

[9] See Zhang and Zhu (2011), Ransbotham and Kane (2011), and Piskorski and Gorbetai (2010).

[10] Gorbetai (2011) discusses why some topics attract more interest, but does not apply this model to NPOV.

"Democrat," resulting in a list of 111,216 articles. Many of these cover countries other than the United States, necessitating further culling.[11] From this set, we obtain a list of 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. It also includes many articles that lack anything controversial, such as undisputed historical accounts of minor historical political events and biographies of comparatively obscure regional politicians.

We compute a slant index for each article. This index applies the methods and estimates developed by Gentzkow and Shapiro (2010), hereafter G&S, who developed a method for measuring the biases of US newspapers. Related to G&S, we ask whether a given Wikipedia article uses phrases favored by more Republican members or more Democratic member of Congress. G&S select 1,000 phrases based on the number of times these phrases appear in the text of the 2005 *Congressional Record,* applying statistical methods to identify words and phrases that separate Democratic representatives from Republican representatives, under the model that each group speaks to its respective constituents with a distinctly coded language. For example, G&S find that Democratic representatives are more likely to use words such as "war in Iraq," "civil rights," and "trade deficit," while Republican representatives are more likely use words such as "economic growth," "illegal immigration," and "border security."[12] After offering considerable supporting evidence, G&S estimate the relationship between the use of each phrase and the ideology of newspapers, using those 1,000 phrases to identify whether newspapers tend to use phrases more aligned with Democrats or Republicans. We label the 1,000 words from the G&S lexicon as "code words."

This approach has several strengths. It has been tested on newspapers and has passed many internal validity tests. In addition, as with newspapers, this provides a general yardstick for measuring the bias of articles, and it removes many subjective elements from that yardstick. Moreover, Wikipedia's contributors are unlikely to have targeted these 1,000 words for editing with this yardstick as a goal, though they might have included or excluded these phrases to try to represent their own views or edit another's views.

[11] 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.

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

This benefit comes with one potential limitation (as the study will show). Although newspapers contain hundreds or thousands of code words over time, the measure is quite noisy in a setting with few code words, as occurs on many Wikipedia pages. In one interpretation of G&S, a lack of code words directly indicates that an article lacks bias. In another interpretation, it simply means an article's slant cannot be measured, and it signals little except that the slant index is uninformative. The latter interpretation requires correction for selection.

The first step of this study is to follow the methods outlined by G&S for measuring the slant of a newspaper.[13] The procedure is identical to that in G&S with a few slight modifications to accommodate some features of this setting. First, in G&S, articles with no code words have a slant index of 0.49, and articles with slant indices below (above) 0.49 are Democrat-leaning (Republican-leaning). For convenience, we center the slant index for articles with no codes at zero by subtracting 0.49 from all slant indices. We can thus compute the bias size of an article directly as the absolute value of its slant index. Second, the method applies some trimming to account for outliers. The 1,000 phrases exhibit a few words (e.g., "civil rights" and "illegal immigration") with unusual values for their slant, and in light of the many articles with only a few code words, these outliers could have an inordinate influence on all results. To mitigate their effect, we reset the parameter values for each extreme phrase, namely, the nine most Democrat-leaning phrases and nine most Republican-leaning phrases. We make the value for these phrases equal to the tenth-most left-leaning and tenth-most right-leaning phrase, respectively.

Just as there is no definitive way to measure the "true bias" of a newspaper in G&S, there is no definitive way to measure the "true bias" of a Wikipedia article. Rather, this study uses the distinct words of Republicans/Democrats to measure biases and looks for a series of internal consistency and internal validity tests. In this sense, "unbiased" and "unslanted" means an equal number of Republican/Democrat words with the same cardinal values.

Of the 70,668 articles observed in January 2011, it is possible to measure the bias for 28,382 articles (40.2%). As it turns out, 3.68% have more than 10 code words by this final date. This variance is not surprising, given an oversampling on a wide array of political articles. It is also evidence of skewness in attention at Wikipedia and should not come as a surprise to a frequent Wikipedia participant. Wikipedia includes many articles about obscure political events and individuals that engender little or no attention (e.g., the biography of a mayor of almost any major US city). It also contains another group of political articles about controversial topics (e.g., George W. Bush, Barack Obama, the Iraq War, health-care

[13] We provide the details in the appendix.

legislation) that might attract considerable attention. By this measure of bias, that group of articles attracting the majority of the attention numbers around several thousand, give or take.

5.2. Descriptive statistics

Table 1 presents the descriptive statistics of the resulting slant index for these 28,382 articles in January 2011, the last period in which we observe them. The table also shows these statistics for different categories of topics in that same year. These categories are not mutually exclusive. Articles can have more than one category attached to them. These categories are assigned by editors and contributors, typically early in an article's life, changing very little over time.[14] The table shows the most commonly used categories.

On average, these 28,382 articles have a Democrat bias (-0.09). Most categories have a bias that differs significantly from zero. For example, articles about civil rights tend to have a Democrat bias (-0.16), while trade tends to have a Republican bias (0.06), and articles about energy tend not to be biased, on average (-0.02). At the same time, seemingly controversial topics, such as abortion and drugs, are centered at zero. Moreover, in addition to considerable variance across topics, the standard deviation is large within most categories.

The 70,668 articles have a total of 17,270,274 revisions. As it is computationally infeasible to examine all these revisions, we take each article and divide its revisions into ten revisions of equal length. For articles with less than ten revisions, we keep all that are available, even if it is low (many of these are short and contain no code words). We retain all revisions, even when one of the 28,382 articles lacks any code word in a prior version, and also when the last version contains nothing but an earlier does. This effort results in 647,352 article observations. Of those, 409,363 of these contain no code words. At least one code word appears in 237,989 observations (36.8%). There is enormous variance in the last year, with 1,086 articles having 19 or more code words, but 11,524, or 40.6%, articles have only one. Although some articles tend to have more code words over time as a result of revision, most retain the same number of code words.

Alternatively, we can measure the change in the slant indices between the earliest and latest observations for each article. For the 68,253 articles for which we have more than one observation, we find that 46,187 articles (68%) have no change in slant.[15] Only 1,193 articles (9.2%) among the 12,902

[14] Table 1 does not show changes in averages over time. These tend not to change within any given category.

[15] For convenience, if an article has no code words in both its first and last observations, we assume that it has no change in slant index between the two observations.

articles that have observed slant indices in both the first and last observations change the sign of their slant indices between the two observations, and only 4 articles have a change of more than 1.0 in slant index. Generally, articles retain their general direction of bias, and if they transition from one state to another, it is a moderate transition.

Tables 2 and 3 show how the aggregate statistics vary over time.[16] This procedure produces noisiness (particularly in the first and last year).[17] It does not support definitive conclusions. Panel A of Table 2 shows there has been movement toward NPOV over time: Wikipedia's articles become less slanted, moving from a mean value of -0.53 in 2002 to a mean value of -0.18 in 2003, and moving gradually downward thereafter to -0.07 in 2010. That would be equivalent to the newspapers in G&S with a strong Democrat slant, such as the San Francisco Chronicle, Atlanta Constitution, or the Baltimore Sun.[18]

The standard deviation of this slant index remains large, however, with evidence of only a gradual decline, starting in 2002 (0.22), rising in 2003 (0.33), and gradually declining by 2010 (0.27). The absolute value of the slant, our bias size, has a similar characteristic, starting at 0.55 (in 2002) and 0.30 (in 2003), and eventually declining to 0.21 (in 2010). Once again, the standard deviation of bias size remains large throughout, showing evidence of only a gradual decline.

Panel B shows patterns for articles with more or less attention. Panel B shows a weighted average across the articles, where the weights come from the number of revisions an article receives in a given year.[19] The number of revisions serves as a proxy for the attention an article receives, and it is the best variable we could assemble that is available for all years and all articles.[20] The average slant is much lower in the weighted averages. The largest slant is -0.15 (in 2002), and it settles to around -0.05 in most of the later years. Consistent with those results, the largest bias is 0.17 (in 2002), settling to around 0.14 in the later years. In both columns, the weighted average is lower than the unweighted average, more

[16] Different versions of the same article can appear in the same year, so there is no reason to observe 27,000 articles each year. Moreover, the last revision of an article may not have been in January 2011, so there will not be a version of every article in 2011.

[17] Only 1,292 articles have ages between ages 9 and 10 years, i.e., a birth in 2001, because this was the first year of Wikipedia. There were not many political articles written in that year.

[18] See Gentzkow and Shapiro (2011), page 46-47. Note that the data in this paper have been demeaned, while Genzkow and Shapiro include the original estimates.

[19] The weight is the number of revisions plus one. Because the number of revisions per article is very skewed, this procedure differs little from the alternative, weighting these articles by zero.

[20] A more ideal weight, an article's number of views in a year, is available after 2007. It is highly correlated with revisions (> 0.5) across articles when both are available.

so in the slant than the bias. Panel B, therefore, suggests that some of the slant and bias in Panel A arises because articles receiving less attention tend to be more slanted and biased.

Panel A of Table 3 provides an overview of how slant and bias change with the age of articles. We have 70,636 observations for articles that are less than one year old. We obtain such a large number because some (very young) articles, all less than one year old, have multiple revisions with a measured bias. In that case, all revisions are included. We observe fewer at each successive year of age. The trend toward less bias and slant must partly result from the features of older/younger articles. Most of the older articles lean more Democrat. Every article over five years old except the oldest year (with the smallest sample) leans Democrat (-.016, -0.17, -0.22, -0.24, and -0.04, respectively), while every article under five years old leans Democrat but less strongly (-0.06, -0.05, -0.08, -0.11, -0.14, respectively). The bias size has a similar characteristic, with the older articles being more extreme than the younger ones, with the exception of the oldest year (i.e., 0.27, 0.27, 0.30, 0.31, and 0.16 for the older five versus 0.21, 0.20, 0.22, 0.23, and 0.25 for the younger five, respectively). In both cases, the standard deviation shows only a mild decline as articles become younger.

Panels B and C look at different vintages of articles at distinct ages. Both panels suggest that vintages play an important role and that this role is more important than age. The slant and bias are most pronounced for articles born in 2002 and 2003, with lower slants and bias in all subsequent years. These slants decline mildly with age, with the biggest decline resulting from small samples in the last year (an artifact of the data-collection method). The differences between vintages of articles released in 2002 and 2003 and other vintages also persist.[21]

To summarize, the average old political article in Wikipedia leans Democratic. Wikipedia's articles gradually have lost that disproportionate use of Democratic code words, moving to nearly equivalent use of words from both parties, akin to a NPOV, on average. Moreover, the words used are mildly less extreme over time. The number of recent articles far outweighs the number of older articles, so by the last measurement, Wikipedia's articles appear to be centered close to a middle point, on average. Overall, therefore, Tables 2 and 3 give rise to a question: Why did Wikipedia become less biased over time? What factors in the revision process shape the bias, and what factors determine the appearance of the code words themselves?

Table 4 provides an understanding about the skewness in the revision process. The table is organized around the total share of all revisions, which the first column shows the fraction of total

[21] Weighted averages indicate similar differences between 2002 and 2003 and other vintages, albeit at lower cardinal values. For the sake of brevity, these are not shown.

revisions, starting with 0.1 and ending with 0.99. The second column shows the smallest number of articles necessary to obtain the fraction of total revisions in the row. For example, it takes seven articles to account for one percent of the revisions. The next column expresses this same number as the share of articles, dividing by the total number of articles, 70,636. The second to last column provides the average age of articles meeting this criteria, denominated in years. The last column provides the average revisions per year for these articles.

The table demonstrates the skewness of the allocation of revisions. Just over 1.1% of the articles account for 30% of the revisions, and just over 3.3% of the articles account for half of the revisions. Less than 14.5% (27.8%) account for 80% (90%) of the revisions. In short, a small percentage of articles get most of the revisions.

Some of this skewness result reflects the composition of the age of articles, for example, whether the group includes old or newer articles. Older articles are more likely to have received a larger number of revisions merely by virtue of being around long enough to collect them. Yet, that cannot be the primary explanation for the skewness. The second to last column suggests that age does not matter much at the most interesting tale of the distribution, where articles get the most revisions. The top 5% are 8.6 years old and the top 50% are 8.1 years old. A half a year hardly accounts for the average difference between these in revisions.

The last column provides a better explanation. As the next column shows, revisions per year differs significantly in each group. For example, the difference of the top 5% and top 50% is 1746 and 536. In other words, some articles simply get more revisions per year than others, which is a symptom that some articles simply attract more attention from contributors.

The large change in skewness in revisions per year continues at all levels. For example, in the row for those articles receiving 80% of the revisions per year is 216, contrasting with the 536 received by the top 50%. That implies that the additional 7904 articles (e.g., 10237-2333) had to have an average at 122 to bring the average down to 216. [22] That is more than one quarter the rate of revisions as the top 50%.

To summarize, Table 4 shows one simple reason why Linus Law will not work with all the articles, namely, a large fraction of articles do not receive much attention or revision. Table 4 also suggests why an econometric approach is required. There is considerable variance in the number of revisions an article receives.

[22] (216.64*10237-536.10*2333)/7904 = 122.34

5.3. Explanatory variables

We classify the key explanatory variables into three groups. The first group examines a narrow interpretation of Linus' Law, which we label "attention and editing." The second group examines the broad interpretation of Linus' Law, focusing on the dispersion of contributions. The third group measures features of articles that act as controls. Table 5 presents descriptive statistics for all variables.

In the first group we have three variables. We expect that more attention and editing lead to more NPOV. We also use *Unique identifiers* to measure the number of unique users who edited this article in the past. Users are identified by their user ids and Internet protocol (IP) addresses. Different (same) IP addresses are counted as different (same) users.[23] We use *Total revisions to date* to measure the total number of revisions an article had to date. Finally, we use *Pageviews* to measure the number of page views in that month for this article. Unfortunately, we have data for this variable only after February 2007, when it first began to be collected. Hence, *Pageviews* limits available data, with 415,836 revisions of articles have a non-missing *Pageviews*. Of these, 259,417 have no measure of bias, leaving 156,419 (37.6%) observations for which we can observe the bias of an article with a measure of *Pageviews*.

Both *Pageviews* and *Unique identifiers* come closer to measuring "eyeballs" than *Total revisions to date.* The latter reflects the controversy of an article and potentially can be inflated by reversion wars or editors who artificially inflate their revisions with many small changes.

The second group contains two variables. We use *Revisions per contributor*, defined as *Total revisions to date/Unique identifiers*, as one measure of the dispersion of contributions. We also use *Herschman-Herfindahl-Index (HHI)*, based off the concentration of *Unique identifiers.* If all revisions are edited by one user in the past, *HHI* will be 1 (just as a monopoly in the industry). A small *HHI* index indicates less concentration, or more dispersion.

Many variables are included in the control group. *Total frequency* measures the number of code words contained in a version of the article. *Words* measures the number of words in the observed version of an article. *Total frequency* and *Words* are highly correlated, especially for the sample of data in which *Words* > 0, so only one can be used in a regression. Articles are longer mostly because they attract more attention and more editing. Linus' Law would predict that greater *Total frequency* or *Words* leads to more NPOV. However, because slant arises from the sum of codes words, whether two or twenty or in between, *Total frequency* or *Words* also measures whether more code words tends to slant an index as a statistical artifact.

[23] This is correlated with the total number of IP addresses, and total number of minor revisions.

References measures the number of references in this version. *References per word*, defined as *References/Words*, measures the extent of verification per length of article. A larger number of references should lead to more NPOV under the null. However, references are also easily manipulated and inflated, so the coefficients estimates need to be interpreted cautiously.

Dummies indicate the year in which the article was created. Under the null, the older vintages have had more opportunity for more attention and more editing, so articles with older vintages should have more NPOV. However, the changing composition of participants on Wikipedia and change knowledge of those participations about NPOV norms could lead to different interpretations. For related reasons we also add year-specific effects as a further control.

Lastly, we create dummies for the categories listed in Table 1 and year of observation to control for category. These controls are for the statistical tendency of some categories to slant in certain directions, or contain a large bias.

Under a narrow interpretation of Linus' Law, only the amount of attention and editing matters. Under a broader interpretation, the dispersion of contributions also matters.

6. Regression Analysis

We use the slant index or the bias size of each article as our dependent variable. These variables distinguish between extreme and mild slant, but that gain requires controlling for selection as many articles do not contain any code words.

Our selection model for the slant index assumes a production function for slant, $Y^* = f(X)$, where X is the list of explanatory variables as discussed in Section 5.3, and Y is measured with error, so $Y = Y^* + u$. The model assumes a function for observing Y, using the same exogenous variables, as in a standard "type-2" Tobit (Amemiya 1985). Hence, in our first stage, we regress the probability of observing any code words in an article on X. Similarly, the selection model for bias size assumes $\ln[Abs(Y^*)] = h(X)$. Once again, this model becomes a standard "type-2" Tobit.

Tables 6 and 7 present the key results.[24] Table 6 presents the estimates for the first stage of the selection model. One estimate includes *Pageviews* and the other does not, with a corresponding change in the sample size. Having more revisions lowers the probability of having code words, while having more unique identifiers (e.g., more contributors) works in the opposite direction, increasing the probability of having code words. More attention, as measured by *Pageviews*, lowers the probability of

[24] We report coefficents of category dummies for all regression tables in the appendix and omit them in the paper for the sake of brevity.

having code words. In each case, the value of the coefficient suggests a small overall effect, but one that matters when the variable reaches a value nears its maximum.

The results from dispersion also do not point in one direction about selection. These results indicate that a wider community of contributors produces a variety of effects. More revisions per contributor lead to a higher likelihood of having code words, and the effect is small. A one-standard-deviation increase in revisions per contributor leads to little change. Only a value near the maximum leads to a large change. Less concentration among the number of contributors (lower *HHI*) reduces the likelihood of code words, but the effect is unimportant except at the maximum values. Overall, these results do not suggest that the breadth of contributors has much impact on the likelihood of having code words.

Features of articles do predict the appearance of code words. Longer articles (more words) have more code words. A one-standard-deviation increase in length leads to a 0.7 increase in the likelihood of having code words (in the lowest estimate). That is a big effect, very big at maximum values. Articles from the oldest two vintages also are likely to have code words. Having more references increases the likelihood of having code words near its maximum value.

The aforementioned ambiguity contrasts with the results for many of the controls. For example, some topics, such as abortion, civil rights, energy, trade, and gun control, are more likely to result in the appearance of code words. Articles on these topics are more likely to have code words than every other article, all other things equal, with coefficients all above 0.5. Fixing other variables at their means, being related to civil rights increases the probability of having code words by 20.47%. Some topics also come close to this, such as foreign policy, social security, and tax reform, with coefficients above 0.3. Some topics are less likely to have any code words, such as drugs, family & children, and infrastructure & technology. The topic, infrastructure & technology, lowers the probability of having code words by 3.80%.

Overall, these results suggest that selection cannot be dismissed as a concern for a few controversial topics and for long articles. In most articles, however, it is not likely to be important.

The first two columns of Table 7 present the results for the second stage of the model of the determinants of the slant index. More attention and editing make an article more Democrat. Having more revisions produces more Democrat bias, as does having fewer unique identifiers (e.g., fewer contributors). More attention, as measured by *Pageviews*, has the same effect. All are quantitatively small in importance, however. They do not matter except at values near the maximum.

The results from dispersion do not point in one direction. Having more revisions per contributor leads to more Democrat words, but less concentration among the number of contributors (lower *HHI*)

leads to more Republican words. The effects are small, however, except at maximum values. Overall, these results do not suggest that the breadth of contributors affects the slant very much.

Many of the controls work as expected, but their effects are small. The features of articles do predict their slant, however. Older vintages lean Democrat in comparison to later vintages, with a difference over 0.3, which is a big effect, equal to one standard deviation in the endogenous variable. That Democrat leaning is especially pronounced in Wikipedia's second year, 2002.

Many of the category controls also take on interesting directions. Abortion articles lean Republican (0.1), civil rights articles lean Democrat (-0.14), and foreign policy leans Republican (0.1), while the typical biography leans Republican (between 0.05 and 0.7). These effects are big in light of the size of the standard deviation of the endogenous variable, which is 0.3.

Overall, the biggest predictors of slant come from the features of articles, such as their topic and vintage. The best predictor of leaning Democrat is the vintage. Consistent with Table 2, older articles lean Democrat, as do articles observed at earlier moments, and the latter tendency diminishes in recent articles. Hence, the best predictor of leaning Republican is a recent observation of an article with a late vintage on a topic, such as abortion. The editing process, such the number of contributors or their spread, does not have much effect.

The last two columns of Table 7 present the results for bias size. There is some evidence for the narrow interpretation of Linus' Law. *Pageviews* does not matter, but the other measure of eyeballs, *unique identifiers*, does matter. While having more revisions increases the movement away from zero, having more participants decreases it. The former matters, but latter is more important. A one-standard-deviation increase in *unique identifiers* results in an 18% decline in the bias. Holding all other variables at their mean values, one-standard-deviation increase in *Total revisions to date* results in a 12% increase in the bias. This result for *Total revisions to date* is consistent with the presence of revision wars, interpreting large numbers as a measure of an article's controversy.

The results for the broad interpretation of Linus' law are weak. *HHI* is negative, which means more dispersion leads to more bias. A one-standard-deviation increase in dispersion leads to a 6% increase in bias. *Revisions per contributor* is too small to matter, except at maximum values.

The biggest predictors of bias come from the features of articles, such as their topic, (some) vintage, and year of observation. The first three vintages of articles display the largest bias, as do the first three years in which those articles are observed. The effects are large. There is a 90% decline in bias between the 2002 vintage and 2010 vintage and a 20% decline between 2003 and 2010. The differences among

vintages between 2004 and 2010 are comparatively small. The 2001 and 2011 vintages also display extreme differences, but this is identified by small samples in each case.

Some topics, such as civil rights, trade, and government, are also more biased, while topics such as abortion, corporations, gun control, social security, and biographies tend to be unbiased. This is the one place where the slant and bias estimates do not line up. The cardinal value for bias yields distinct insight.

The results in Tables 6 and 7 suggest that a few features of the revision process do shape slant and bias, such as the number of unique contributors. This is consistent with a narrow interpretation of Linus' Law. None of the evidence is consistent with a broad interpretation.

In unreported regressions, we conduct robustness checks by estimating probit models that examine changes to slant and bias size, controlling for the selection effects. The inferences are similar.[25] We also found no qualitative difference estimating the basic equations on a sample that removed the youngest articles and those with the least active revision rates. The results supporting a narrow interpretation of Linus' Law also are robust to including or excluding the tests for the broad interpretation, as well as other features of the articles, such as *words, references and references per word*.[26]

7. Discussion and Conclusion

In this study, we empirically examine whether Linus' Law, a principle that many participants in open-source communities consider foundational, shapes content production. We conduct this test on US political articles in Wikipedia, where Linus' Law would face challenges due to the presence of controversial topics and lack of verified and/or lack of objective information.

Our first set of findings pertains to the general characteristics of Wikipedia's slant and bias over time. In broad terms, in its earliest years, Wikipedia's political entries lean Democrat and tend to be biased. Over time, both traits diminish, on average. By the most recent observations, Wikipedia's entries lack much slant and contain less bias than observed earlier.

Our second set of finding points toward persistent bias, inconsistent with Linus' Law. This arises partly from a vintage effect, partly from the skewed attention of contributors, and partly as a result of the topic. Overall this second set of findings is consistent with only a narrow interpretation of Linus' Law (i.e., Hypothesis 1), one that emphasizes the number of contributors or reviewers. No evidence supports a broad interpretation of Linus' Law (i.e., Hypothesis 2).

[25] See the appendix for results.

[26] Available upon request.

Our third set of findings points towards a number of empirical patterns inconsistent with Linus' Law shaping most articles. The majority of articles receive little attention, and most articles change only mildly from their initial slant. Altogether, this is a potential indicator of collective intelligence bias.

These findings can be reconciled with the historical facts. The general tendency toward more neutrality in Wikipedia's political articles appears to arise largely not from revision, but from the entry of later vintages of articles with an opposite point of view from earlier articles. Wikipedia achieves something akin to a NPOV across articles, but not necessarily within them.

The study demonstrates a broad approach for estimating the relationship between features of articles and the revision process, based on testing both narrow and broad interpretations of Linus' Law. We consistently find that evidence for a narrow interpretation, at best, and we find no support for a broad interpretation. This result has implications for measuring production involving large number of contributors, suggesting it is possible to find regularities (around the number of contributors) but that this effect should not be regarded as the sole determinant of neutrality.

Broadly, our study provides empirical evidence on the limit of collective intelligence. Many managers believe that they could improve products by taking advantage of the wisdom of crowds. We show that in the case of Wikipedia, there are aspects such as NPOV that collective intelligence does not help achieve successfully. A natural question for future research is whether an alternative production model could perform better in this dimension. For example, would a production model centered on experts, such as the one employed by Encyclopedia Britannica, do a better job than the one used by Wikipedia at achieving NPOV?

Some readers may not conclude that Linus' Law fails to hold, but, rather, that we did not measure the revision process with a proper set of statistics. As with any econometric research, we do not consider our research to be definitive. In that light, these results motivate a number of potential questions. For example, how frequently do articles with distinct biases link to one another? What factors shape the entry of new articles, particularly articles with bias? Does Linus' Law become weaker in practice due to feedback between an article's existing biases and the biases of the revisions it attracts? Hence, we hope that our attempt to measure Linus' Law and NPOV, and to formulate an approach using classical statistical methods, motivates further work on the operation of key principles and their operation within open-source production.

References

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Harvard University Press: Cambridge, MA.

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

Bernhardt, Dan, Stefan Krasa, Mattias Polborn. 2008. Political polarization and the electoral effects of media bias. *Journal of Public Economics* 92(5-6) 1092–1104.

Besley, Timothy, Andrea Prat. 2006. Handcuffs for the grabbing hand? Media capture and government accountability. *American Economic Review* 96(3) 720–736.

Blake, Aaron. 2006. Wikipedia site attempts to make politics healthier. *The Hill*, http://tinyurl.com/6qycehm (accessed January 2012)

Brown, Adam R. 2011. Wikipedia as a data source for political scientists: accuracy and completeness of coverage. *Political Science & Politics*, 44, 339–343

Carr, Nicholas. 2008. *The Big Switch: Rewiring the World, from Edison to Google*. New York: Norton.

Chi, Ed, Aniket Kittur, Bryan A. Pendleton, Bongwon Suh, Todd Mytkowicz. 2007. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *Computer/Human Interaction 2007 Conference*.

Dewan, Sanjeev, Jui Ramaprasad. Forthcoming. Consumer blogging and music sampling: long tail effects. *Information Systems Research*.

Galton, Francis. 1907. Vox populi. *Nature* 75:7.

Gal–Or, Esther, Tansev Grelani, Tuba Pinar Yildirim. 2010. The impact of advertising on media bias. *Working paper*.

Gentzkow, Matthew, Jesse M. Shapiro. 2006. Media bias and reputation. *Journal of Political Economy* 114(2) 280–316.

Gentzkow, Matthew, Jesse M. Shapiro. 2010. What drives media slant? Evidence from U.S. daily newspapers**.** *Econometrica* 78(1) 35–71.

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

Giles, Jim. 2005. Internet encyclopaedias go head to head. *Nature* 438 900–901.

Gorbetai, Andreea. 2011. Aligning collective production with social needs: evidence from Wikipedia. *Working paper, Harvard Business School*.

Greenstein, Shane, Michelle Devereux. 2009. Wikipedia in the spotlight. *Kellogg Case 5-306-507*, *Kellogg School of Management, Northwestern University*.

Greenstein, Shane, Feng Zhu. 2012. Is Wikipedia biased? *American Economic Review, Papers and Proceedings,* 102(3) 343–348.

Groseclose, Tim, Jeffrey Milyo. 2005. A measure of media bias*. Quarterly Journal of Economics* 120(4) 1191–1237.

Haefliger, Stefan, Georg von Krogh, Sebastian Spaeth. 2008. Code reuse in open source software. *Management Science* 54(1) 180–193.

Halfaker, Aaron, Aniket Kittur, Robert Kraut, John Riedl. 2009. A jury of your peers: Quality, experience and ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration.*

Hann, Il-Horn, Jeff Roberts, Sandra Slaughter. 2004. Why developers participate in open source software projects: An empirical investigation. *ICIS 2004 Proceedings.*

Hyatt, Josh. 2006. The wonder of Wikipedia: How to motivate—and control—an army of 30,000 volunteer workers. *Fortune*, June 12.

Lawrence, Eric, Jon Sides, Henry Farrell. 2010. Self-segregation or deliberation? Blog readership, participation and polarization in American politics. *Perspectives on Politics* 8(2010) 141–157.

Larcinese, Valentino, Riccardo Puglisi, James M. Snyder. 2007. Partisan bias in economic news: Evidence on the agenda-setting behavior of U.S. newspapers. *NBER working papers*, *National Bureau of Economic Research, Inc*.

Lakhani, Karim R., Lars Bo Jeppesen. 2007. Getting unusual suspects to solve R&D puzzles. Forethought. *Harvard Business Review* 85(5).

Lemos, Robert. 2004. Security research suggests Linux has fewer flaws. *CNET News*. December 13.

Levine, Robert. 2006. New Web sites seeking profit in wiki model. *New York Times*, September 4.

Ma, Liye, Pei-yu Chen, Tridas Mukhopadhyay. 2008. Voluntary open-source - the effect appropriability, externality, and uncertainty. *ICIS 2008 Proceedings.*

MacCormack, Alan, John Rusnak, Carliss Y. Baldwin. 2006. Exploring the structure of complex software designs: An empirical study of open source and proprietary code. *Management Science* 52(7) 1015–1030.

Malone, Thomas W., Robert Laubacher, Chrysanthos Dellarocas. 2009. Harnessing crowds: Mapping the genome of collective intelligence. MIT Sloan Research Paper No. 4732–09.

Margolis, Jonathan. 2009. Wicked-Pedia! Millions trust its every word. *Daily Mail*, February 15.

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

National Public Radio. 2005. Talk of the Nation, "Analysis: Wikipedia, open source, and the future of the Web," Transcript, November 2.

Piskorski, Mikolaj Jan, Andreea Gorbatai. 2010. Testing Coleman's social norm enforcement mechanism: Evidence from Wikipedia. *Working Paper*.

Price, Cindy J. 2003. Interfering owners or meddling advertisers: How network television news correspondents feel about ownership and advertiser influence on news stories. *Journal of Media Economics* 16(3) 175–188.

Ransbotham, Sam, Gerald C. Kane. 2011. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quarterly* 35(3) 613–627.

Raymond, Eric. 1998. The cathedral and the bazaar. *First Monday*, http://tinyurl.com/bqfy3s (accessed January 2012).

Reuter, Jonathan, Eric Zitzewitz. 2006. Do ads influence editors? Advertising and bias in the financial media. *Quarterly Journal of Economics* 121(1) 197–227.

Roberts, Jeffrey A, Il-Horn Hann, Sandra A Slaughter. 2006. Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the apache projects. *Management Science* 52(7) 984–999.

Singh, Param Vir, Yong Tan, Vijay S. Mookerjee. 2011. Network Effects: The influence of structural capital on open source project success. *MIS Quarterly* 35(4) 813–829

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

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

Swartz, Aaron (2006). Who Writes Wikipedia? http://tinyurl.com/7n84bsh (accessed January 2012).

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

von Hippel, Eric. 2005. *Democratizing Innovation.* The MIT Press: Cambridge, MA.

Wu, Lynn, Erik Brynjolfsson. 2009. The future of prediction: How Google searches foreshadow housing prices and sales. *Working paper*.

Xu, Sean, Xiaoquan Zhang. Forthcoming. Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction. *MIS Quarterly*.

Zhang, Xiaoquan, Feng Zhu. 2011. Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review* 101(4) 1601–1615.

Table 1: Summary Statistics for Slant Index by Category

	No. Obs	Mean	Std. Err.	One-tailed t-test
All Categories	28,382	-0.09	0.00	***
Abortion	71	0.02	0.03	n.s.
Bios	4,748	-0.05	0.00	***
Budget & Economy	1,109	-0.02	0.01	***
Civil rights	1,183	-0.16	0.01	***
Corporations	121	-0.06	0.02	***
Crime	1,257	-0.05	0.01	***
Drugs	105	-0.02	0.02	n.s.
Education	1,362	-0.05	0.01	***
Energy & Oil	270	-0.02	0.01	**
Families & Children	405	-0.06	0.01	***
Foreign Policy	2,094	0.02	0.00	***
Trade	399	0.06	0.01	***
Government	11,383	-0.14	0.00	***
Gun Control	56	-0.03	0.02	*
Health Care	556	-0.05	0.01	***
Homeland Security	490	-0.05	0.01	***
Immigration	372	-0.02	0.01	**
Infrastructure & Technology	1,143	-0.04	0.01	***
Jobs	693	-0.05	0.01	***
Principles & Values	614	-0.05	0.01	***
Social Security	5	-0.10	0.05	*
Tax Reform	95	-0.06	0.02	***
War & Peace	2,292	-0.02	0.00	***
Welfare & Poverty	323	-0.04	0.01	***

n.s. not significant, *** p<0.01, ** p<0.05, * p<0.1

Table 2: Summary Statistics for Slant Index and Bias Size by Year

Panel A: Unweighted Slant Index and Bias Size Over Time

Year	Slant Index		Bias Size		No. Obs.
	Mean	Std. Dev.	Mean	Std. Dev.	
2001	0.03	0.24	0.19	0.15	290
2002	-0.53	0.22	0.55	0.15	3,276
2003	-0.18	0.33	0.30	0.23	960
2004	-0.23	0.34	0.33	0.25	4,571
2005	-0.10	0.30	0.24	0.21	9,733
2006	-0.11	0.30	0.24	0.21	28,521
2007	-0.12	0.30	0.24	0.21	37,465
2008	-0.10	0.29	0.23	0.20	42,552
2009	-0.08	0.28	0.22	0.20	46,139
2010	-0.07	0.27	0.21	0.19	51,210
2011	-0.10	0.27	0.22	0.19	13,272

Panel B: Slant Index and Bias Size Over Time Weighted by Revisions in that Year

Year	Slant Index (Weighted Mean)	Bias Size (Weighted Mean)	No. Obs.
2001	0.00	0.05	290
2002	-0.15	0.17	3,276
2003	-0.03	0.07	960
2004	-0.04	0.11	4,571
2005	-0.03	0.13	9,733
2006	-0.05	0.15	28,521
2007	-0.05	0.15	37,465
2008	-0.06	0.15	42,552
2009	-0.05	0.14	46,139
2010	-0.05	0.14	51,210
2011	-0.05	0.14	13,272

Table 3: Summary Statistics to Examine Articles' Vintage Effects

Panel A: Slant Index and Bias Size of Wikipedia's Political Articles for Different Article Ages

Age (Year)	Slant Index		Bias Size		No. Obs.
	Mean	Std. Dev.	Mean	Std. Dev.	
[0, 1)	-0.06	0.28	0.21	0.19	70,636
[1, 2)	-0.05	0.27	0.20	0.18	28,946
[2, 3)	-0.08	0.29	0.22	0.20	28,412
[3, 4)	-0.11	0.30	0.23	0.21	28,614
[4, 5)	-0.14	0.30	0.25	0.22	27,461
[5, 6)	-0.16	0.31	0.27	0.22	21,348
[6, 7)	-0.17	0.30	0.27	0.22	15,398
[7, 8)	-0.22	0.31	0.30	0.23	10,043
[8, 9)	-0.24	0.31	0.31	0.23	5,839
[9, 10)	-0.04	0.21	0.16	0.14	1,292

Panel B: Slant Index of Wikipedia's Political Articles for Different Article Ages and Years

Age (Year)	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
[0, 1)	0.03	-0.53	-0.17	-0.03	-0.03	-0.05	-0.04	-0.04	-0.04	-0.04
[1, 2)	-0.11	-0.51	-0.10	-0.04	-0.05	-0.05	-0.05	-0.02	-0.03	-0.08
[2, 3)	0.02	-0.46	-0.09	-0.05	-0.05	-0.04	-0.04	-0.03	-0.05	.
[3, 4)	-0.01	-0.39	-0.09	-0.05	-0.05	-0.04	-0.04	-0.06	.	.
[4, 5)	-0.02	-0.37	-0.11	-0.06	-0.04	-0.03	0.02	.	.	.
[5, 6)	-0.02	-0.36	-0.10	-0.05	-0.04	-0.09	.	.	.	.
[6, 7)	-0.03	-0.33	-0.09	-0.05	-0.08	.	.	.	.	.
[7, 8)	-0.04	-0.33	-0.09	0.02	.	.	.	.	.	.
[8, 9)	-0.02	-0.29	-0.05	.	.	.	.	.	.	.
[9, 10)	-0.04	-0.06	.	.	.	.	.	.	.	.

Panel C: Bias Index of Wikipedia's Political Articles for Different Article Ages and Years

Age (Year)	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
[0, 1)	0.19	0.55	0.29	0.19	0.19	0.19	0.20	0.19	0.19	0.19
[1, 2)	0.22	0.54	0.23	0.20	0.19	0.19	0.20	0.18	0.18	0.16
[2, 3)	0.17	0.50	0.22	0.20	0.19	0.19	0.19	0.18	0.14	.
[3, 4)	0.17	0.44	0.22	0.19	0.19	0.18	0.19	0.17	.	.
[4, 5)	0.17	0.41	0.23	0.19	0.19	0.17	0.19	.	.	.
[5, 6)	0.16	0.41	0.22	0.19	0.19	0.21	.	.	.	.
[6, 7)	0.16	0.38	0.21	0.19	0.15	.	.	.	.	.
[7, 8)	0.17	0.38	0.21	0.16	.	.	.	.	.	.
[8, 9)	0.16	0.35	0.17	.	.	.	.	.	.	.
[9, 10)	0.16	0.16	.	.	.	.	.	.	.	.

Table 4: The Distribution of Revisions across Articles in January, 2011

Total share of all revisions	(n) Number of articles	(n/N) Share of articles	Average Age of articles (Year)	Average Revisions Per Year
0.01	7	0.0001	8.77	2815.52
0.05	59	0.0008	8.66	1746.02
0.10	150	0.0021	8.77	1343.76
0.20	413	0.0058	8.70	1005.58
0.30	821	0.0116	8.52	780.64
0.40	1433	0.0203	8.25	697.02
0.50	2333	0.0330	8.07	536.10
0.60	3747	0.0530	7.74	412.55
0.70	6072	0.0859	7.38	310.40
0.80	10237	0.1449	6.96	216.64
0.90	19604	0.2774	6.32	135.03
0.95	30627	0.4334	5.80	96.01
0.99	51365	0.7268	5.05	66.79

Definitions
Total share of all revisions = fraction of total revisions.
Number of articles = smallest number of articles to achieve fraction of total revisions.
Share of articles = smaller share of articles to achieve fraction of total revisions.
N = total number of articles, 70,636.
Average age of articles = average age of articles, in years.
Average revisions per year = average revisions per year for articles.

Table 5: Summary Statistics for Explanatory Variables[27]

Variable	Obs.	Mean	Std. Dev.	Min	Max
Attention and Editing					
Unique identifiers (10,000 IDs)	237,989	0.0121	0.0329	0.0001	1.3983
Total revisions to date (100,000 revisions)	237,989	0.00282	0.00816	0.000001	0.44193
Pageviews (100,000,000 of views)	156,419	0.000092	0.000683	0	0.06169
Dispersion					
Revisions per contributor (100,000 revisions)	237,989	0.0000284	0.0000489	0	0.00597
HHI (0.01 is perfectly concentrated)	237,989	0.0024	0.003	0	0.01
Article Features					
Total frequency (1000 of code words)	237,989	0.0039	0.00737	0.001	0.62
Words (10,000 article words)	237,989	0.193	0.244	0.0003	19.7806
References (100,000 references)	237,989	0.00016	0.00039	0	0.01022
References per word (100 references/word)	237,989	0.000074	0.0001	0	0.004516

[27] We re-scale some variables for the ease of displaying their coefficients in regression tables.

Table 6: Regression Results for the First Stage Selection Equation[28]

	(1)	(2)
Dependent Variables	Has code words	Has code words
Attention and editing		
Unique identifiers	5.297***	1.732***
	[0.551]	[0.480]
Total revisions to date	-35.582***	-20.211***
	[2.231]	[1.960]
Pageviews	-13.659*	
	[7.245]	
Dispersion		
Revisions per contributor	1,278.439***	1,665.096***
	[57.775]	[50.039]
HHI	-52.372***	-49.307***
	[1.073]	[0.752]
Article features		
Words	2.965***	3.328***
	[0.018]	[0.015]
Reference	711.030***	538.473***
	[15.477]	[13.628]
References per word	-160.131***	-53.585***
	[14.952]	[11.456]
Year created = 2002	0.650***	0.593***
	[0.025]	[0.017]
Year created = 2003	0.002	-0.099***
	[0.026]	[0.017]
Year created = 2004	-0.217***	-0.327***
	[0.025]	[0.017]
Year created = 2005	-0.334***	-0.415***
	[0.025]	[0.017]
Year created = 2006	-0.442***	-0.512***
	[0.025]	[0.017]
Year created = 2007	-0.484***	-0.521***
	[0.026]	[0.017]
Year created = 2008	-0.320***	-0.402***
	[0.026]	[0.018]
Year created = 2009	-0.206***	-0.292***
	[0.026]	[0.019]
Year created = 2010	-0.165***	-0.264***
	[0.027]	[0.019]
Year created = 2011	-0.570***	-0.665***
	[0.055]	[0.051]
Category Dummies	Included	Included
Year Dummies	Included	Included
Observations	415,836	647,352

Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

[28] Note that to make the coefficients easier to read, we have re-scaled the variables as shown in Table 5.

Table 7: Regression Results for Slant Index and Bias Size

	(1)	(2)	(1)	(2)
Dependent Variables	Slant index	Slant index	Bias size	Bias size
Attention and editing				
Unique identifiers	1.460***	1.351***	-6.571***	-6.151***
	[0.081]	[0.070]	[0.370]	[0.316]
Total revisions to date	-3.120***	-2.791***	16.167***	13.826***
	[0.324]	[0.279]	[1.483]	[1.258]
Pageviews	-2.072**		-2.979	
	[1.050]		[4.809]	
Dispersion				
Revisions per contributor	-13.874	-24.412**	301.312***	355.838***
	[13.625]	[12.065]	[62.404]	[54.432]
HHI	2.367***	1.840***	-18.135***	-14.774***
	[0.364]	[0.270]	[1.667]	[1.216]
Article features				
Total frequency	-1.462***	-1.081***	-0.007	-1.073***
	[0.098]	[0.079]	[0.447]	[0.358]
Reference	15.830***	4.280**	-70.922***	-20.222**
	[2.308]	[2.078]	[10.572]	[9.376]
References per word	4.224	18.391***	203.924***	184.997***
	[6.319]	[6.001]	[28.911]	[26.999]
Year created = 2002	-0.201***	-0.235***	0.697***	0.697***
	[0.004]	[0.003]	[0.014]	[0.014]
Year created = 2003	0.030***	0.008**	-0.077***	-0.077***
	[0.005]	[0.003]	[0.015]	[0.015]
Year created = 2004	0.089***	0.079***		
	[0.005]	[0.003]		
Year created = 2005-2011, [min, max] shown.	[0.103, 0.115]***	[0.085,0.103]***		
Year created = 2004-2010, [min,max] shown.			[-0.300, -0.381]***	[-0.289, -0.434]***
Category Dummies	Included	Included	Included	Included
Select Year dummies	2007 omitted	2001 omitted	2007 omitted	2001 omitted
	0.020*** in 2008	-0.251*** in 2002	-0.062*** in 2008	0.566*** in 2002
	0.021***in 2011	-0.138*** in 2003	-0.084*** in 2009	0.270*** in 2003
		-0.103*** in 2010		0.025 in 2010
		-0.109*** in 2011	-0.096*** in 2011	0.049 in 2011
Full set of Year Dummies	Included	Included	Included	Included
Inverse Mills ratio	-0.026***	-0.041***	0.157***	0.239***
	[0.003]	[0.002]	[0.013]	[0.010]
Observations	156,419	237,989	156,419	237,989

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1