Ideological Shift among an Online Crowd: Evidence from Wikipedians[*]

Shane Greenstein, Grace Gu, Feng Zhu

Preliminary, not for quotation without permission

Abstract

We analyze contributors to Wikipedia articles about U.S. politics. We find that on average, the slant in the content from contributors decline over time. Two factors cause the decline – an ideological shift, in which contributors contribute less slanted content, and a composition shift, in which extremely biased contributors contribute less content. We explain these patterns. We show that contributors tend to contribute to articles with slants that are the opposite of their own views. Interactions with opposite-slant contents or others' pushback lead the more extreme contributors to leave the community. They also become more neutral due to encounters more extreme content of the opposite slant, or when they receive more pushback from others. We also find some significant differences between Republicans and Democrats, and across topics. Our findings direct attention at the platform features that encourage participation and engagement from those who do not segregate their conversations.

[*] We thank Jana Gallus, Marco Iansiti, Gerald Kane, Karim Lakhani, Abhishek Nagaraj, Frank Nagle, Michael Norton, Brian Silverman, anonymous reviewers, and seminar participants at many universities, the INFORMS Annual Meeting 2015, the Conference on Open and User Innovation 2016, and the Academy of Management Annual Meeting 2018. We thank Justin Ng and John Sheridan of HBS Research Computing Services for their research assistance. We also thank Alicia Shems and Kate Adams for their editorial assistance. We gratefully acknowledge financial support from the Division of Research of the Harvard Business School.

1. Introduction

The growth of virtual communities that blur the boundaries between reader and writer has upended our understanding of processes for generating and consuming online content. Online communities bring together participants from disparate traditions, with different methods of expression, cultural and historical foundations for their opinions, and, potentially, with bases of different facts (e.g., Arazy et al. 2011). Over time a crowd renders its opinions. In spite of diversity of opinion, and sometimes due to it, the composition of participants evolves, and so too does the collective opinion, as participants experience alternative content and interact with points of view other than their own.

Why does a crowd's collective opinion vary over time in the presence of such diversity? In this study we identify and measure two factors that contribute to emergence of opinion in online crowds containing a diversity of points of view. Broadly, the first factor is the evolution of a participant's point of view, and the second factor is the composition of participation. A participant may hold to their point of view or change it. They may become more extreme in their existing view, more moderated, or changing their mind to the opposite view. In addition, each participant must decide whether to enter into a conversation or leave it, and over time participants with different opinions may stay or leave. The composition of a crowd may change if the participants with the most extreme views leave or stay, and, correspondingly, the collective opinion of the crowd may become more moderate or extreme. For either factor, a participant's experience may shape the evolution of their actions. Actions may or may not depend on their experience with other participants who hold similar or opposing views.

This study focuses on these factors in an important and understudied setting, one that contains *contested knowledge*—defined as topics involving subjective, unverifiable, or controversial information. In the presence of wide diversity of opinion, contested knowledge give rise to two distinct kinds of interactions among participants, which we will label as *segregated* or *unsegregated* conversation. Intuitively, a segregated conversation may become an "echo chamber" of like-minded views (or "EC" for short), while an unsegregated conversation ("non-EC") may mix views from a diverse group. More to the point, in a *segregated* conversation, like-minded participants self-select into supplying content for others with similar views and primarily read content from those with whom they already agree. Participants in a segregated conversation *evolve* into separate groups, and the crowd becomes an amalgam of independent silos of opinions. This behavior has gained attention because it polarizes information production and participants' interactions (e.g., Mullainathan and Shleifer 2005, Sunstein 2001), creating segregated "small villages" (e.g., Gentzkow and Shapiro 2003, Van Alstyne and Brynjolfsson 2005). The opposite behavior, an *unsegregated* conversation, involves contributors with diverse ideas and opposing views (Benkler 2006).

The crowd evolves to account for all positions. Participants may continue a conversation until they reach a consensus about how to create content that combines both views.

Participation in Wikipedia's articles about US politics provides the set of participants we study. The articles about politics provide a rich setting for studying participation in segregated conversations, where many debates in politics involve contested knowledge. Wikipedia states the ideals and norms to which it contributors should aspire, and considerable discretion sits with its participants with implementing its ideals. A key aspiration is Wikipedia's emphasis on aspiring for a neutral point of view in its content – succinctly summarized as "State facts about opinions instead of stating the opinions themselves." No algorithm determines the tendency towards segregation. It depends entirely on human behavior. Moreover, contributions to these debates has been well documented since 2001, because Wikipedia is one of the oldest and longest continuously operated communities producing online content from crowd-sourced contributions. That long life enables research into changes in the composition of the crowd, and into the evolution of segregated conversations, and it enables the observation of a variety of political topics.

This study operationalize a method for measuring whether (and when) an article achieves a neutral point of view or not, and gains insight into understanding whether (and when) contributors' actions and viewpoints moves content towards or away from mixing many points of view. Specifically, this study principally examines 66,389 articles about U.S. political topics, which receive contributions from 2,887,140 unique contributors. After cleaning, we analyze more than ten million edits by these contributors.

As with prior research (e.g., Greenstein and Zhu, 2016, forthcoming), this study characterizes all articles for bias and slant along a numerical yardstick by adapting the method developed by Gentzkow and Shapiro (2010) for rating newspaper editorials. In these ratings, *slant* denotes degree of opinion along a continuous yardstick. It can take on extreme degrees of red (e.g., Republican) and extreme degrees of blue (e.g., Democrat), and all the shades of purple in between. *Bias* is the absolute value of this yardstick from its zero point, and thus denotes the strength of the opinion.

The focus on the bias and slants of *participants* is the primary novelty of this study. This article operationalizes a method for measuring the slant and bias of contributors, and, by extension, enables measurement of the composition of the bias and slants of the crowd, contributor' participation in segregated and segregated conversations, and their evolution over time. As far as we know, this is the first study to ever develop an approach for following the evolution of the slant and bias of a population of contributors. Relatedly, it is the first to measure how interaction with other views leads participants to alter their views, and leads others to exit, changing the composition of the slants and opinions of participants.

Wikipedia's importance makes understanding its production interesting in its own right. Most reference information has moved online, and these online sources have displaced other sources of information in every developed country. Wikipedia is both a top-twenty site in several dozen developed countries, and, by far, the most popular and referenced online repository of comprehensive information in the developed world, with the English language version of Wikipedia receiving over eight billion page views per month, and over 500 million unique visitors per month.[1] Many firms also utilize Wikipedia as an input. Amazon (Alexa), YouTube, and Google (search), among others, use Wikipedia as a free source for neutral "facts" and as an unrestricted source for vocabulary in different languages.[2] Despite its importance, Wikipedia's emphasis on aspiring to a neutral point of view has not been tested by any social science or administrative science. Little is known about whether neutral content arises and, if so, why and how.

The study first shows that contributor behavior on Wikipedia tends to move the conversation towards less segregated conversation on most topics. We develop this broad finding in several steps. First, we find the presence of considerable heterogeneity in contributors and in their behavior. Contributors with every possible bias and slant contribute to articles containing every other possible bias and slant. Second, in spite of that variance, more contributors in Wikipedia exhibit a pattern of behavior consistent with what we define as "Non-EC" and not "EC." For example, a slanted contributor is on average 8% more likely to edit an article with the opposite slant than one with the same slant. Third, this tendency is pervasive. The most popular topics, which accounts for 19 out of the 26 topics and 66.7% of all the contributions, display non-EC outcomes while only 3 less popular topics, which accounts for only 6.1% of all the contributions, display EC. In other words, contributors with different political viewpoints tend to dialogue with each other during their editing of contestable knowledge, and that holds across the most popular political topics.

The study documents changes in the composition of the existing contributors and changes in the behavior of the existing contributors. It shows in a variety of analyses the exit of more biased contributors does shape the composition of participants. The largest declines in extreme views and the most frequent withdrawals from participation are found among contributors who edit or add content to articles that have more extreme biases. It also arises from those who experience more pushback from other contributors. The findings also show that editing articles reduces a contributor's slant, and editing more biased content leads contributors to offer less biased contributions later. These two mechanisms together reinforce the tendencies toward "the neutral point of view." Between the two, the exit is most significant driver of change. One simulation suggests it is responsible for approximately 90% of the change in participant's actions in the latter half of our sample.

[1] "Wikipedia vs. the small screen". http://www.nytimes.com/2014/02/10/technology/wikipedia-vs-the-small-screen.html?_r=1, assessed June 2016.

[2] See e.g., YouTube may add to the burdens of Humble Wikipedia, https://www.nytimes.com/2018/03/19/business/media/youtube-wikipedia.html.

To address concerns about endogeneity, we also examine a special circumstance, *mass edits*, where an article attracts an unusually high number of contributions in one day due to a sudden social event or breaking news about the topic. Such mass edits occurred to 5.1% of the articles in our sample. We interpret mass edits as exogenous events. While contributors involved in mass edits demonstrate significantly faster reduction in slant than others, we find that articles in mass edits also experience more flips in the slant in one day - from extremely blue/red to extremely red/blue. Contributors who edit during mass edits are 11.8% more likely to be exposed to contents of *both extremes* , this exposure explains why their slant moderates faster than those making normal edits.

This study's approach also enables the analysis of the speed of adjustment for different types of slants and biases in a contributor, and adjustment over time by type of contributor. For example, on average, our estimates suggest it takes extreme Republican content one year longer to a regular provider of neutral content than it does for extreme Democrat content. In the study we will trace this distinction to differences in the topics in which Democrats and Republican contributors participate, and to the topics that tend to have more/less EC behavior. Related, we find intriguing evidence that certain divisive topics contain EC behavior, such as taxes, border security, and healthcare, while others do not, such as abortion and gun control.

Because the study focuses on micro-behavior of contributors, it lends itself to tests of alternative explanations, and a battery of tests, which substantiate causal inference. We describe the overall contributor slant in Section 2 and 3, analyze its evolution in Section 4, discuss the properties and robustness in Section 5 and 6, and conclude in Section 7.

1.1. Relationship to Prior Work

These findings enhance the understanding of prior work. In contrast to Greenstein and Zhu (2012, 2016, forthcoming), which focus on measuring and characterizing the evolution of content, this study focuses on the evolution of participants, and the interactions among online contributors and content. Prior work finds that revisions in Wikipedia tends to lead to more neutrality in its content, but only very slowly. Past work did not measure the determinants of segregated conversations directly because it lacked an approach to measure the political slant and bias of contributors.

This study makes a contribution to the literature on segregated online conversations, often labelled as "online echo chambers." We share the same motivation found in prior research about such behavior (e.g., Sunstein 2001; Carr 2008; Lawrence et al. 2010; Gentzkow and Shapiro 2011; (Greenstein and Zhu 2012, 2016, Boxwell et al. 2017). Segregation can facilitate radicalization of some individuals and groups (Purdy 2015).[3] The persistence of many segregated conversations also can prevent bringing varying perspectives into a common view, and delay confrontation or a political discourse between contradictory facts and ideas. It also has been held responsible for discouraging interracial friendships, disconnecting different social segments, and stimulating social isolation. Concerns about the health and tenor of political conversation have also motivated prior work, and we share this motivation as well. In traditional media, ideological bias in news content affects political language (e.g., DellaVigna and Kaplan 2007; Stone 2009; Chiang and Knight 2011; Durante and Knight 2012). Closer to our study, Gentzkow and Shapiro (2011) focus on online conversations about political content and other topics, and Gentzkow and Shapiro (2010) starts from the premise that there are ideological tendencies that appear in the language of speakers. We borrow the latter insight for the measurement of a contributor's slant.

We differ in the analysis of causes of online segregated conversation. Prior work emphasizes the role and design of the social network. The direction and tenor of contributions can be shaped by, for example, the structure of online communities (e.g., Fan et al. 2005; Ahn et al. 2007), the laws for reusing knowledge (Nagaraj forthcoming), and the algorithms for search (e.g., Jeppesen and Frederiksen 2006; Chiu et al. 2006; Ma and Agarwal 2007; Xu and Zhang 2009, 2013; Slivko 2014; Slivko et al. 2016; Qiu et al. 2017). A key theme of prior study is that algorithms can reinforce segregated conversations on platforms. For example, Facebook or Reddit and other social networks encourage conversations among those with similar points of view. Algorithms produce unanticipated aggregate outcomes because algorithms encourage/discourage processes that users do not control. In contrast, algorithms play no role in this setting, and participants largely have discretion whether to contribute in directions that encourages or discourages segregated conversations. In contrast, this study emphasizes participant's choice. Most other work treats the sources of bias as isolated (e.g., Groseclose and Milyo 2005; Besley and Prat 2006; Reuter and Zitzewitz 2006; Bernhardt et al. 2008) and does not link them to contested knowledge, political discourse or aggregated knowledge, as this study does. In addition, no prior work measures whether participants change their slant over time, as this study does.

Prior work has examined partisanship in online media (e.g., Larcinese et al. 2007), and identified its importance for ideologically segregated conversations among those with different viewpoints (e.g., Carr 2008; Lawrence et al. 2010; Gentzkow and Shapiro 2011; Shore et al. 2016). Several studies examine settings where contested knowledge shapes the conversation, though these studies do not frame it as such. The setting that comes closest to this study are the paper about sharing links on twitter (Shore et al, 2016)

[3] See, for example, http://www.vice.com/read/we-asked-an-expert-how-social-media-can-help-radicalize-terrorists and http://www.rand.org/randeurope/research/projects/internet-and-radicalisation.html, accessed June 2017.

and a field experiment about following opinion leaders on twitter (Bail et al, 2018). The former examines whether participants share with others who are like-minded. Like our study, it analyzes the role of behavior, not algorithms, and focuses on dichotomous either/or behavior. Also like our study, both papers ask whether the behavior reinforces segregated conversation, but reach different conclusions. In contrast, we find that the most frequent contributors to Wikipedia display more neutral tendencies than the less frequent contributors. The data from Wikipedia also comes from a much longer period of time – a decade versus a few weeks or months – and a wider array of topics. This longer lens permits analysis of long run changes in contributors, and differences in prevalence across topics, which no study has examined. Finally, we also examine factors shaping the composition of participants, which has not receive attention in any other study.

This study also relates to the analysis of crowds and crowd behavior, which has grown into a large literature. Most empirical studies examine how online organizations aggregate contributions to solve collective problems (e.g., Kogut and Zander 1992; Lee and Cole 2003; Hargadon and Bechky 2006; Kuk 2006; Tucci and Villarroel 2007; Faraj et al. 2011; Ransbotham and Kane 2011; Afuah and Tucci 2012; Chen et al. 2012; Pierce 2012; Bassamboo et al. 2015). We share the broad agenda of the literature focused on user-generated content. We depart from emphasis on motivation for participation (e.g., Kogut and Metiu 2001; Rothaermel and Sugiyama 2001; Yang et al. 2009; Ransbotham et al. 2012; Kane et al. 2014; Gallus forthcoming; among many others), which examined its importance for a variety of tasks, such as software design, entrepreneurial finance, and engineering (e.g., Kogut and Metiu 2000; Von Krogh and Von Hippel 2006; Chesbrough 2006; Roberts et al. 2006; Ramasubbu and Balan 2012; Xu et al. 2015). Prior research left open questions about why users behave as they do in the presence of contested knowledge, where this study focuses attention. This study's focus on the evolution of the composition and opinions of the crowd is also novel.

Our study's approach borrows from studies of "herding behavior," but differs in one important respect. Prior work examines whether contributors follow their predecessors in assigning a rating (Gao et al. 2015; Lee et al. 2015; Kwark et al. 2016; Wang et al. 2017), choosing products (e.g., Salganik et al. 2006), and whether one contributor's action triggers more future contributions (Aaltonen and Seiler 2015). Research has stressed the role of group thinking (e.g., Janis 1982), decreased communication cost (Rosenblat and Mobius 2004), emotional and social contagion (e.g., Barsade 2002; Sun et al. 2017), and, broadly, the occurrence of homophily in social networks (e.g., McPherson et al. 2001; Park et al. 2013; Gu et al. 2014). While some prior literature presumes an extrinsic reward and/or a range of motives for herding/departing from the consensus, the setting for our study rules out the possibility that extrinsic monetary motives play a major role. Building on evidence for recognition (Gallus forthcoming), and prosocial and reciprocal behavior among contributors (Algan et al. 2013), we presume intrinsic motives – i.e., an inherent desire to express their slanted opinions.[4] Our measurement strategy differs accordingly. We start with models that assume a "fixed" intrinsic viewpoint from each contributor, and then consider models that varies a contributor's viewpoint over time.

Relatedly, we depart from research about crowds that presumes a single "right" answer exists for a crowd to discover. These studies examine whether (and how) online crowds reach that single "right" answer (Page 2007), or presume the existence of a single "consensus forecast." We build on analysis that examines whether contributors herd around the consensus or deliberately choose "extreme" positions to influence the consensus (Laster et al. 1999; Zitzowitz 2001). In contrast, we cannot presume the existence of a single "right answer" in a setting with contested knowledge; At most, this setting gives rise to a focal point at a "neutral" position, which participants work out for themselves – usually either by sampling all slanted points of view or including no biased points of view. It is a subtle but key difference. The closest prior research asks: Does a participant's rating/assessment align with an aggregated report of prior ratings/assessments (e.g., Muchnick et al. 2013)? By comparison, we ask: Does a contributor add to content with a slant which matches (or differs from) his or her own, and how does that behavior change over time? The latter accounts for the potential for two possibilities, a segregated or unsegregated conversation. We also link the conversation to neutral or non-neutral outcome.

Finally, this study contributes to the literature on platform design. The setting for this study employs an architecture that gives participants considerable discretion in achieving platform-wide ideals and aspirations. While many participants inside Wikipedia believe its principles and processes help its online communities meet the ideals to which the site aspires, little quantitative evidence or controlled experimentation either confirms or refutes this belief. Like other online communities, Wikipedia has adopted explicit aspirations, rules, norms, policies (Forte et al. 2009; Jemielniak 2014; Schroeder et al. 2012), and quality assurance procedures (Stvilia et al. 2008), which appear to shape behavior. Many online communities have adopted schemes of access privileges that formally define roles in the organization (Arazy et al. 2015; Burke et al. 2008; Collier et al. 2008; Forte et al. 2012), and so has Wikipedia. These lead to a myriad of coordination mechanisms (Kittur et al. 2007a; Kittur and Kraut 2008; Kittur et al. 2007b; Schroeder and Wagner 2012), social interactions (e.g., Halfaker et al. 2011; Forte et al. 2012), and behaviors aimed at conflict resolution (Arazy et al. 2011). Our findings confirm that online platform can develop aspirations that actually do discourage segregated conversation in practice, even in the presence of highly decentralized contributions. While these findings suggest Wikipedia's mechanisms work as desired, our findings heighten questions about which specific mechanisms or norms are primarily responsible.

[4] Political bias is one form of bias. Scholars have examined other forms of online bias such as racial and gender bias (e.g., Edelman and Luca 2014; Cui et al. 2016; Carnahan and Greenwood 2017).

2. Defining terms and the setting

As in Gentzkow and Shapiro (2010) and Greenstein and Zhu (2016, forthcoming), we first define the *slant* of content. This indicates which way a particular piece of content "leans." It takes a numerical value, bounded on the interval [–D, R], D > 0, and R > 0. We normalize a neutral point of view to 0. *Bias* of content is the absolute value of slant. We define the slant and bias of a contributor in an analogous fashion.

2.1. Simple models of Slant in a crowd

One standard model of a crowd presumes a single objective answer, and a platform aggregates contributions from the crowd. In many models the results improve with a larger sample of contributions (Page 2007). In the appendix we modify this model for a setting in which two groups of contributors with intrinsic political views aspire to improve a controversial topic and do not agree on a single objective answer. That guides our empirical approach.

The intuition can be summarized as follow. First, neutrality cannot emerge from a random draw of opinions. The composition of the preferences of contributors plays a key role in shaping the overall results. Second, behavior of contributors is a key factor to consider. Third, reinforcing behavior leads to segregated conversations, i.e., contributions from those with similar slant will appear to be segregated. Third, unsegregated conversations will result from a process that does not reinforce existing slant and draws on opposite opinions. Fourth, segregated conversations are associated with more biased outcomes than unsegregated conversations, and the latter are associated with a comparatively moderate slant near the neutral point of opinion. Finally, slant only settles down in a single place after the number of suggestions reaches a large number.

It is an empirical question whether the composition of a crowd and reinforcing behavior will or will not be affiliated with segregated conversations. It is also an empirical question what a "large number of suggestions" means in practice. Finally, it is also a measurement challenge to characterize how segregated conversations differ from unsegregated conversations.

2.2. Empirical setting

Founded in 2001, Wikipedia positions itself as "the free encyclopedia that anyone can edit"—that is, as an online encyclopedia entirely written and edited via user contributions. Topics are divided into unique pages, and users can select any page to revise—expertise plays no explicit role in such revisions. It has become the world's largest "collective intelligence" experiment and one of the largest human projects ever to bring information into one source. The website receives enormous attention, with over eight billion page views per month in the English language, and over 500 million unique visitors per month.[5]

Contributions come from tens of millions of dedicated contributors who participate in an extensive set of formal and informal roles.[6] Some of these roles entail specific responsibilities in editing tasks; however, the Wikimedia Foundation employs a limited set of people and largely does not command its volunteers. Rather it helps develop a number of mechanisms to govern the co-production process by volunteers (Kane and Fichman 2009; Te'eni 2009; Zhang and Zhu 2011, Hill 2017). All these voluntary contributors are considered editors on Wikipedia. The organization relies on contributors to discover and fix passages that do not meet the site's content tenets. No central authority tells contributors how to allocate editorial effort.

The reliance on volunteers has many benefits and comes with many drawbacks. Among the latter, there is a long-standing concern that interested parties attempt to rewrite Wikipedia to serve their own parochial interests. Despite the persistence of such concerns, there is little systematic evidence pointing in one direction or another. Available evidence on conflicts suggests that contributors who frequently work together do not get into as many conflicts as those who do not, nor do their conflicts last as long (Piskorski and Gorbatâi 2017). While such behavior could lead to edits from contributors with different points of view, there is no direct evidence that it leads to more content that finds compromises between opposite viewpoints.

While the Wikipedia community tries to attract a large and diverse community of contributors, there is recognition that it invites many slanted and biased views, and the openness of Wikipedia's production model (e.g., allowing anonymous contributions) is subject to sophisticated manipulations of content by interested parties. So there is widespread acceptance of the need for constant vigilance and review.

A key aspiration for all Wikipedia articles is a "neutral point of view" or NPOV (e.g., Majchrzak 2009, Hill 2017). To achieve this goal, "conflicting opinions are presented next to one another, with all significant points of view represented" (Greenstein and Zhu 2012). When multiple contributors make inconsistent contributions, other contributors devote considerable time and effort debating whether the article's text portrays a topic from a NPOV. Because Wikipedia articles face few realistic limits to their number or size[7] (due to the absence of any significant storage costs or any binding material expense), conflicts can be

[5] "Wikipedia vs. the small screen". http://www.nytimes.com/2014/02/10/technology/wikipedia-vs-the-small-screen.html?_r=1, assessed June 2016.

[6] See https://en.wikipedia.org/wiki/Wikipedia:User_access_levels, accessed June 2017.

[7] Over time a de facto norm has developed that tends to keep most articles under six to eight thousand words. This arises as editorial teams debate and discuss the length of the article necessary to address the topic of the page. Of course, some articles grow to enormous lengths, and editor contributors tend to reduce their length by splitting them into sub-topics. Prior work (Greenstein and Zhu 2016) finds that the average Wikipedia article is shorter than this norm (just over 4,000 words), but the sample does include a few longer articles (the longest is over 20,000 words).

addressed by adding more points of view to articles instead of eliminating them (e.g., Stvila et al. 2008). In general, the vast majority of such disputes settle without elimination from the center.[8]

2.3. Measuring contributor slant and bias

A number of measurement challenges arise when analyzing a participant's slant. We adopt a "yardstick" approach, and specify its requirements. First, the yardstick characterizes a participant's slant and bias. Second, the yardstick must characterize how the slant and bias of articles changes as contributors revise articles.[9] Third, the yardstick must enable a measurement of whether a participant selects content with a slant similar to or different from their own slant. Lastly, the yardstick must enable measurement of a participant's change in slant as they gain experience editing articles with slants and biases similar to or different from their own.

Such an approach for Wikipedia's content was pioneered in Greenstein and Zhu (2016), and here we extend it to participant's point of view. This approach relies on a modification of an existing method, developed by Gentzkow and Shapiro (2010), for measuring slant and bias in newspapers' political editorials.[10] 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."[11] Similarly, we compute an index for the slant of each article from each source, tracking whether articles employ these words or phrases that appear to slant toward either Democrats or Republicans.

Gentzkow and Shapiro (2010) select such phrases based on the number of times they appear in the text of the 2005 *Congressional Record*, and apply statistical methods to identify those phrases that separate Democrat and Republican representatives. Their approach rests on the notion that each group uses a distinct "coded" language to speak to its respective constituents.[12] Each phrase is associated with a cardinal value that represents the degree to which each word or phrase is slanted. After offering considerable supporting evidence, Gentzkow and Shapiro estimate the relationship between the use of these phrases and the ideology of newspapers. As shorthand we refer to their words and phrases as "code phrases."

[8] Like all matters at Wikipedia, contributors have discretion to settle disputes on their own. The organization offers a set of norms for the dispute resolution processes, which today can be quite elaborate, including the three-revert edit war rule, as well as rules for the intervention of arbitration committees and mediation committees. Administrators can also decide to freeze an article under contention.

[9] This is a property that Greenstein and Zhu (2012) confirmed in their study of Wikipedia articles.

[10] Gentzkow and Shapiro (2010) characterize how newspapers also use such phrases to speak to constituents who lean toward one political approach over another.

[11] Several studies have applied their approach in analyzing political biases in online and offline content (e.g., Greenstein and Zhu 2012; Jelveh et. al. 2014; Shore et al. 2016). In addition, although Budak et al. (2014) use alternative approaches to measure ideological positions of news outlets, their results are consistent with Gentzkow and Shapiro (2010).

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

This approach has several key strengths in that it has passed many internal validity tests, and provides a general yardstick for measuring the bias of newspaper articles. The approach is also effective when analyzing information on other online platforms (e.g., Shore et al. 2016) or examining political bias in articles in economic journals (Jelveh et al. 2014), which we believe can be transferred to the context of Internet articles. We must assume that Wikipedia's contributors did not use this yardstick to target these words for editing, but, instead, chose to include or exclude them when endeavoring to represent or exclude a specific point of view. The method also leads to a quantifiable measure of "neutral," because the numbers are additive for finding the total slant of an article, and the range of slants can be normalized at the mean. An article is deemed unslanted or unbiased either when it includes no code phrases from many opposing points of view or when its use of Republican and Democrat code phrases equal the same cardinal value.[13]

The way of defining a contributor's slant builds on this approach: initially we assume that a contributor's slant is constant throughout the years. Then we define a contributor's slant as the average slant of all the contributions that the person made in our sample. A participant's bias is, again, the absolute value of this slant. When the contributor's slant is allowed to evolve over time, the measure is computed based on the contributions in each year instead of throughout the sample period.

In general, just as there is no definitive way to measure the "true bias" of a newspaper article in Gentzkow and Shapiro (2010), there is no definitive way to measure the true bias of a Wikipedia contributor or an online encyclopedia article. It is valid under specific assumption that the underlying weights for the code phases represent an invariant yardstick.

To construct our sample, we focus on broad and inclusive definitions of U.S. political topics, including all Wikipedia articles that include the keywords "Republican" or "Democrat." We start by gathering a list of 111,216 relevant entries from the online edition of Wikipedia on January 16, 2011. Eliminating the irrelevant articles and those concerning events in countries other than the United States[14] reduces our sample to 70,305. Our sample covers topics with many debates over contestable knowledge, ranging from the controversial topics of abortion, gun control, foreign policy, and taxation, to the less disputed ones relating to minor historical and political events and biographies of regional politicians. We next collect the revision history data from Wikipedia on January 16, 2011, which yields 2,891,877 unique contributors. For

[13] Greenstein and Zhu (2016) find no evidence that these two types of unslanted articles differ in their underlying traits. Hence, in this paper we treat them as identical.

[14] 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-U.S. countries.

much of the statistical work below we exclude articles with only one contribution, which is the contribution to create the article. We general work with 66,389 articles where we can observe an article over time.

Our key dependent variable is *Contributor Slant*. This measure is developed in two steps. First, we take the 70,305 articles in our sample, and every article on Wikipedia has a revision history that, for every edit, records a pre-edit and post-edit version. We compute the slant index for both the pre- and post-edit article versions, take the difference between the two, and use this difference as the *slant change* for an edit. We obtain the slant change of every edit. For sequential edits from the same contributor that happened consecutively and without anyone else editing between them, we treat the sequence of edits as one single edit.[15]

Next, we focus on individual contributors. We identify and measure the types of changes each contributor makes to Wikipedia articles. We assign each edit to each contributor, and assign a slant value for each edit. Under the assumption that every contributor has one fixed type of slant, we compute the *Contributor Slant* as the average value of the slant index of this contributor. A zero value of *Contributor Slant* means the user's edits either contain a balanced set of Republican/Democratic words (weighted by their cardinal values) or do not include any of the slanted phrases. A negative or positive value of *Contributor Slant* means the contributor is Democrat-leaning or Republican-leaning, respectively. Accordingly, the absolute value of a contributor's slant equals the *Contributor bias*. In our sample, 2,678,626 out of 2,891,877 unique contributors (92.6%) have a zero contributor slant, while over 225 thousand contributors make at least one slanted contribution. As it turns out, the vast majority of contributions to Wikipedia come from one of the contributors with a measurable slant or bias (57.5%). (What percentage?) (maybe we say "As it turns out, these contributors with a measurable slant or bias make 57.5% of all the contributions on Wikipedia." ?)

Table 1 presents the distribution of types of contributors over ten years. When computing the number of Democratic, Republican, and Neutral contributors to Wikipedia each year, we count each user ID only once—even if the user contributes many times in a year. There are 2,891,877 unique contributors in our sample. We define a contributor as *core* if his or her total number of edits is distributed in the top 10% of all contributors' total number of edits, which in this case equals a total of no less than three contributions in our sample. Core contributors comprise 10% of contributors, but they make 74% of the contributions in the entire sample. In other words, most of the edits in the sample come from experienced contributors – these are the contributors who we expect to be savvy about reading the existing slant of the articles and

[15] These consecutive edits tend to be highly correlated, or could be several parts of a complete contribution, such as where the contributors saved their work several times. As a robustness check, we exclude deletions from a contributor's edits if the deletion does not bring an article's slant from left/right leaning to right/left leaning, or from less to more extreme. In this way, deleting biased content to make an article more neutral would not be seen as a biased edit. All our results still hold.

responding to that slant. Furthermore, while the number of neutral contributors who contribute each year is more than ten times that of contributors who have a slant, the proportion of core contributors in the neutral slant group (15.9% for the full sample) is much smaller compared to the proportion of core contributors in the other two groups (63.8% and 65.5% for the full sample). In summary, slanted contributors are more core than neutral contributors, and much of the slanted content comes from contributors making many edits.

For the sake of analyzing participant behaviors, we drop the first version of all articles in our sample, since we do not have a prior article slant and cannot observe EC or Non-EC effect for such contributions. This reduces the number of observations in the sample to 10,878,391, the number of articles to 66,389, and the number of unique contributors to 2,887,140. Unless pointed out otherwise, this paper uses this analysis sample throughout the paper.

2.4. Composition Effect or Ideological Shift?

A few simple graphs illustrate the evolution of contributors' evolution. To begin, we compute each contributor's slant over time. Figure 1 is a scatter plot of the average bias of any contributor who contributed at least once in that given year. Later figures are based on the same sample with 10,878,391 contributions on 66,389 articles if not explained otherwise. Also, in the graphs we do not plot the observations in 2001, because the first year contains many outliers. Average bias of contributors decline over the years.

Two types of changes could be contributing to the decline in bias, either change in the composition of contributors and/or ideological shift. That is, it is possible that new contributors with moderate slant show up each year to join Wikipedia, and, relatedly, the existing extreme contributors edit less over time or gradually stop participating. Alternatively, contributors could become less biased in their contributions, and our assumption of a fixed slant for each contributor requires modification.

First consider the possibility that the average contributor slant declines over time due to changes in the slants of people joining Wikipedia in different years. We compute the average slant of contributors entering in different years and plot the results in Figure 2. There is no obvious pattern across years. Contributors who entered earlier are not systematically more slanted compared to those who entered later.

Next consider an existing contributor on Wikipedia. Figure 3 displays the average number of contributions of the extreme contributors each year. A contributor is considered as "extreme" if the person's slant is more than 2 standard deviations away from the center. While there seems to be an increase in the number of edits from the contributors' first year to their second year on Wikipedia, it is followed by a declining pattern as these contributors stay past two years. As these extreme contributors stay longer they become less active over time. That finding is intriguing. Does that pattern hold more generally? We consider that further below.

For those extreme contributors who keep contributing over the years, we next ask whether we observe a decline in slant in their contributions. In Figure 4 we plot the average contributor bias each year for those who entered as “extreme.” If we redefine the slant and bias each year (based on their changes in that year), there is a constant declining pattern in the slant from these extreme contributors. The data suggest that extreme editors become less slanted over time. Again, we consider more below.

Overall, these graphs suggest a change in the composition of extreme contributors and their participation, mostly coming from exit of participants and an ideological shift favoring more neutral contributions. Do these patterns survive more careful statistical scrutiny? What factors drive the observed behavior patterns? We next define variables in preparation for investigating these questions.

Finally, we summarize the distribution of contributors’ total number of edits over the ten years in Figure 5. Our sample reflects the well-known skewness of contributions to Wikipedia. More than 75% of the contributors in our sample contributed only once in the entire ten-year period. 97.5% of the contributors contributed fewer than 10 times, averaging to less than one contribution per year. Only 1% of the contributors contributed more than 30 times in our sample. We also show the number of edits, the number of contributors and the average number of edits per contributor by the contributors’ years of experience in Figures 6-8, respectively. While contributors with 4 to 5 years of experience comprise the larger part of our sample compared to the rest both in terms of the number of contributors and the total number of edits, the average number of edits per contributor does not vary much with years of experience except for the 0.18% contributors who joined in January 2011.[16]

3. Additional Definitions

3.1. Variables

Contributor Slant assumes contributors have the same slant over their lifetime. We next define *Contributor Yearly Slant* which divide contributors’ edits by year and for each year use the same calculation as for *Contributor Slant*, that is, we compute the average slant change of all the edits a contributor has made within a given year. If a contributor’s numeric value for slant remains unchanged throughout the years, then his or her *Contributor Yearly Slant* equals *Contributor Slant*.[17] Relatedly, *Contributor Yearly Extreme*

[16] Dropping this group of contributors in our analysis does not change our results qualitatively.

[17] If a contributor does not make any contribution in a given year, his or her *Contributor yearly slant* has a missing value in that year.

Slant contains only edits that make an article more extreme, i.e., edits that move an article in a more biased direction. It is a more robust version that captures biases.

Prior Article Slant denotes an article's slant before a particular edit. This variable is essential for analyzing an article's (mis)match with a contributor's slant.

We count the contributor's number of edits to date that targets extreme opposite-slant articles. This is *Opposite-Slant Article Edits To Date.* Similarly, the number of edits to date that targets extreme same-slant articles is labeled as *Same-Slant Article Edits To Date*. This captures the amount of extreme content that he or she has interacted with.

A *revision war* is defined as a contributor's edit being reverted immediately by another contributor. If it then is immediately followed by the original contributor editing back the same contribution, it counts as a "fight back". We count the edits that a contributor makes during such pushbacks so far as *Revision War Edits To Date*.

Throughout the study, unobservable features of articles are a central concern. We add additional measures that may have attracted editors, and otherwise had spurious correlation with the slant or bias of an article. *Prior Article Length* and *Prior Refs* are two variables for this. We measure the length of the articles using the number of words in an article prior to a certain edit. This is *Prior Article Length.* We measure the number of the article's external references. This is *Prior Refs*. Articles that are longer may incorporate more viewpoints, which then, in turn, tends to attract more contributors. Also, Wikipedia requires citations from major third-party sources as references for its article content (often listed at the bottom of the page), so articles with more references are also more likely to incorporate more outside arguments or controversial views at the time.

In a similar spirit, additional controls measure unobservable features of editors. One such variable is *Number of Edits*, the total number of edits *to date* that the contributor has made on Wikipedia. Another variable is *Starting Number of Edits.* It equals the total number of edits the contributor made in his or her first two years after joining Wikipedia. *Contributor Years* is the number of years the contributor has been on Wikipedia before he or she made an edit. Finally, *Starting Contributor Slant* denotes the contributor's *Contributor Yearly Slant* in his or her first year after joining Wikipedia.

3.2. Descriptive Statistics

In Table 2, we provide summary statistics of all variables used in our analysis. The unit of analysis in this table is contributor-edits, and the total number of observations is 10,878,391. Edits from all contributors who have ever contributed to the articles in our sample are included in this table. Note the contrast with Table 1. Table 1 summarizes the level of *contributors*, while Table 2 focuses on all the *edits* made by the contributors within the entire time period.

In general, the average *Contributor Slant* in our sample is negatively close to zero, while the average *Contributor Category* is positively close to zero. The summary statistics indicate that (1) Democrat-leaning contributors are, on average, more slanted than Republican-leaning contributors, and (2) all article versions in our sample exhibit a Democrat-leaning slant, with similar absolute values of extreme slant on both ends. There is also substantial variation across article versions for each of the three control variable measures, and we use the logarithm of these three control variables in our models since they are highly skewed.

4. Analyzing Contributor Slant

Motivated by the observations of the slant-declining pattern in the raw data, this section will quantify how the contributors' contributions change with their edits. It analyzes whether (and how) their editing experience affects their slant decline.

4.1. Contributors' Participation Pattern on Wikipedia

To understand the changes in contributors' slant, we first investigate the type of content contributors interact with on Wikipedia. For every edit in our sample, we estimate the following regression model:

$$Contributor\ Slant_{ijt} = \alpha_0 + \alpha_1 Prior\ Article\ Slant_{it} + X_{it}B + \sigma_i + \eta_t + \varepsilon_{it}\ . \qquad (1)$$

In this baseline specification we set the contributor slant to one value, which does not change over time, even though we can observe the same contributor multiple times. The coefficient α_1 identifies whether the average contribution follows EC or Non-EC, as earlier noted. To address concerns about unobservable factors influencing the choice, we include X_{it}, a vector of the article's characteristics and control variables, and σ_i, an article fixed effect to control for any fixed differences among articles (despite potential changes over many years), and η_t, a year fixed effect to control for any common trend in media/macroeconomic shocks that may differentially affect articles of different years. In an alternative approach described in the text, we also use *Contributor Category* as the dependent variable, with *Prior Article Category* as the explanatory variable, estimating standard models for categorical choice. We note that the key exogenous variable is measured with considerable noise, and we worry that that will induce attenuation bias in the estimate.[18] Hence, we view the result as an underestimate, at best.

In Table 3, we report estimation results of Equation (1) using Ordinary Least Square (OLS) regressions. Models (1) through (3) use *Contributor Slant* as the dependent variable. Model (1) includes only *Prior Article Slant* as the explanatory variable. Model (2) adds in control variables *Log (Prior Article Length)* and *Log(Prior Refs).* Model (3) replicates Equation 1, with article- and year- fixed effects included. The coefficients on *Prior Article Slant* is negative and significant in all three models. This indicates that an increase in the article's slant is associated with a decrease in the slant of its next contributor; namely, when the article is more Republican-leaning, it tends to attract a more Democrat-leaning user as its next contributor. That is consistent with Non-EC behavior.

Models (4)-(6) repeat the analyses in Models (1)-(3) but replace *Contributor Slant* with *Contributor Category* as the dependent variable, and replace *Prior Article Slant* with *Prior Article Category* as the explanatory variable. Again, we find that the coefficients for the categorical explanatory variable *Prior Article Category* is negative and significant in all cases, suggesting that the slant category of the next contributor is significantly negatively correlated with the slant category of the prior article. Results are similar across models and in line with our findings from Models (1)-(3). To sum up, it appears that contributors target at opposite slant articles; that is, we observe much fewer Echo Chamber than Non-Echo Chamber on Wikipedia.

To further illustrate the Non-EC pattern, we estimate a simpler model. Define *Prior Article Category* by categorizing *Prior Article Slant* into -1, 0, and 1 for articles with slant two standard deviations below mean, in between, and above mean, respectively. The endogenous variable is defined in an analogous categorical fashion. This definition lends itself to a multinomial logistic regressions on the relationship between *Contributor Category* and *Prior Article Category*, with control variables and fixed effects similar to the specifications in Equation 1.

Model (1) of Table 4 includes only *Prior Article Category* as the explanatory variable. Model (2) adds in control variables *Log (Prior Article Length)* and *Log(Prior Refs).* Model (3) includes fixed effects. Unsurprisingly, the results continue to support our previous findings of a greater Non-EC effect than EC effect in contributors' online participation. The coefficients for *Prior Article Category* are all statistically significant and have opposite signs with the categorical dependent variable. Take the coefficients of *Prior Article Category* in Model (1) as an example. The coefficient for *Prior Article Slant* is 2.10 when the

[18] This is a standard concern with poorly measured exogenous variables. See, e.g., Draper, Smith, 1998, pg 19.

Contributor Category is -1, which leads to an 8% increase[19] in the probability of attracting a next contributor whose *Contributor Category* equals -1 when the article's prior slant changes from Democrat to Republican. Similarly, the coefficients in Models (2) and (3) suggest that the increase in the probability of attracting a subsequent contributor with an opposite slant is even higher than it was without control variables or year fixed effects.

4.2. Ideological Shift: How Does Editing Experience Change Contributions from Contributors?

Why do extreme contributors become moderate? Does interacting with extreme content have an effect on contributors? To address this question we measure the contributor's number of edits to date that targets extreme opposite-slant articles as *Opposite-Slant Article Edits To Date*, and similarly the number of edits to date that targets extreme same-slant articles as *Same-Slant Article Edits To Date*. Pushbacks also may shape contributors. An extreme example of such pushbacks is the *revision war*, in which a contributor "fights back" by making the same edit again. We include *Revision War Edits To Date*, to test how such pushbacks affect contributors' likelihood of exit.

We thus estimate the following equation:

$$Contributor\ Yearly\ Extreme\ Slant_{ijt} = \beta_0 + \beta_1 OppositeSlant\ ArticleEditsToDate_{ij}$$
$$+\beta_2 SameSlant\ ArticleEditsToDate_{ijt} + \beta_3 Revision\ War\ Edits\ ToDate_{ijt} + Z_t B + \mu_j + \epsilon_{it}. \qquad (2)$$

In this specification the unit of analysis is each contribution. The coefficient β_1 are our central interest. Z_t are the year dummies, to control for the effect due to different years when the contribution is made. μ_j is a contributor fixed effect. It is not possible to estimate μ_j, a contributor fixed effect, for contributors who make one contribution. The dependent variable is the absolute value of *Contributor Yearly Extreme Slant*. We take the absolute value to capture how far away the contributor yearly slant is from neutral, regardless of its sign. Also, we use *Contributor Yearly Extreme Slant* instead of *Contributor Yearly Slant* because we want to exclude contributions that make an article more neutral. For example, a contribution which moves an article's slant from -0.2 to -0.5 has a slant of -0.3, whereas another contribution which moves an article from 0.5 to 0.2 also has a slant of -0.3. While we consider the former contribution to be left-leaning, it is

[19] $\frac{e^{-5.11+2.07}}{1+e^{-5.11+2.07}+e^{-5.25-2.41}} - \frac{e^{-5.11}}{1+e^{-5.11+2.07}} = 0.0456 - 0.0058 = 0.0398$ when the prior article's category increases by 1. When the prior article's category changes from -1 to 1, the estimated probability increases by 8%.

unclear whether the latter contribution is actually left-leaning or neutral. Therefore, we avoid such ambiguity by excluding the latter type of contributions when measuring contributor yearly extreme slant.

Table 5 reports the regression results. We partition the contributors by their frequency of edits and examine whether core contributors behave similarly as the full sample. *Core* contributors, as defined earlier, are the top 10% of contributors in terms of each contributor's total number of edits. In our sample, this means making at least 3 edits in total. On average, for each article 84.2% edits are from core contributors, and 15.8% edits are from peripheral contributors.

Model (1) is based on the full sample of all contributions, whereas Model (2) includes only contributions from the core contributors. Both models control for contributor fixed effect and year dummies. In both models, the estimated coefficients of *Opposite-Slant Article Edits To Date* and *Revision War Edits To Date* are negative and statistically significant, whereas the coefficients of *Same-Slant Article Edits To Date* is positive. In summary, encountering extreme contents of the opposite slant (rather than the same slant), or receiving pushback from other contributors, significantly reduces the contributor's own bias.

It is difficult to interpret the above coefficients, so we use a Markov matrix to illustrate how the slant composition of contributors evolves. This matrix, reported in Figure 9, is constructed as follows: First, we divide in half every contributor's time that he or she has been on Wikipedia. Then, we divide the direction of this contributor's edits by attaching values (-1, 0, 1) to negative slant, zero slant, and positive slant edits. Based on the sum of these values for the first half and the second half of this contributor's activity, we can categorize the contributor as Democrat, Neutral, or Republican: If the sum of all edits in one half is negative (positive), the contributor is a Democrat (Republican), respectively. And, if the sum of all edits in this half is zero, the contributor is Neutral. We do this for each half of every contributor's activity on Wikipedia and accumulate them to get the overall transition probabilities in the entire community. We find that, for both democratic-leaning and republican-leaning contributors in the first half, there is more than a 70% chance that they will move to Neutral in the second half of their activities.

While the community of participants has a general tendency of moving towards neutral, Figure 9 does not provide any sense of whether this is faster or slower than the effects coming from changes in composition. We next characterize those composition effects, and then compare them with the above.

4.3. Composition Shift: Why Do Existing Extreme Contributors Leave Over Time?

To examine the determinants of composition, we define a dummy variable *Stay Dummy* at the contributor level. *Stay Dummy* equals 1 if the contributor made at least 1 edit in the last year period covered in our sample, i.e., year 2010 and Jan 2011. The dummy variable equals 0 otherwise.

Because the stay/exit analysis is at the contributor level, the explanatory variables change to accommodate the unit of observation. *Opposite Slant Article Edits Fraction, Same Slant Article Edits Fraction,* and *Revision War Edits Fraction* are computed the same as *Opposite-Slant Article Edits To Date, Same-Slant Article Edits To Date,* and *Revision War Edits To Date* but, instead, use their corresponding values for the contributor's entire time in sample, divided by the contributor's total number of edits.

We employ the following linear probability regression to examine each contributor's exit decision:

$$\begin{aligned} Stay\ Dummy = {} & \theta_0 + \theta_1 Starting\ Contributor\ Slant + \theta_2 Starting\ Number\ of\ Edits \\ & + \theta_3 OppositeSlant\ Article\ Edits\ Fraction + \theta_4 SameSlant\ Article\ Edits\ Fraction \\ & + \theta_5 Revision\ War\ Edits\ Fraction + \varphi + \omega, \end{aligned} \tag{3}$$

where the unit of analysis is each contributor. Here a linear probability model with standard robust errors follows previous literatures (e.g., Horrace and Oaxaca 2006, Angrist and Pischke 2008). We use this model because it allows us to interpret the coefficients easily while still obtaining unbiased estimates of the probabilities.[20]

We focus on the coefficients for *Opposite-Slant Article Fraction, Same-Slant Article Fraction*, and *Revision War Fraction*, which are similar to *Opposite-Slant Article Edits To Date, Same-Slant Article Edits To Date,* and *Revision War Edits To Date* in the previous analysis, except that these are the total number of such edits throughout the contributor's experience divided by the contributor's total number of edits in sample. We also control for φ, the contributor's year of entry, to captures the vintage effect.

Table 6 presents estimates. Model (1) includes all contributors, while Model (2) includes only the core contributors. *Starting Contributor Slant* has a significantly negative relationship with *Stay Dummy* in both models, which means compared to a neutral contributor, the probability for an extreme contributor to stay on Wikipedia before encountering any opposite-slant content or pushbacks from others is 1.3% smaller in general. This is consistent with our observation from the raw data: the existing extreme contributors tend to edit less or exit over time, which leads to the composition effect on overall slant.

The coefficients for *Opposite-Slant Article Fraction* and *Revision War Fraction* are negative and statistically significant, which means interacting with opposite slant extreme content or fighting in revision wars constantly reduces a contributor's likelihood of staying on Wikipedia by 1% and 20.5% in the long term, compared to a contributor who encounter no opposite slant content or revision wars at all. The effect of interacting with same slant contents, shown by the coefficient of *Same-Slant Article Fraction*, goes in the other direction, which increases the likelihood of a contributor staying. Putting the two sets of results

[20] We also repeat the analysis using logit regression specifications. The results do not qualitatively change.

together, we see that encountering extreme opposite-slant content or pushback from others lead contributors to either become moderate, or decide to leave in the long term.

4.4. *Which Effect is Contributing More to Changes in Contributor Slant?*

Which effect contributes more to overall trends, the ideological shift and the composition shift? We use the second half of our data, the 5-year period between 2005 and 2010 for a comparison of the size of each effect. This second half of the data is better than the first half because the second half has more participants than the first half, and it examines Wikipedia in its most "developed" state, where participants are familiar with all the norms and rules.

The goal is to calculate what proportion of the change in bias is due to ideological shift, and what portion is due to the composition shift. To begin, we compute the average bias - the absolute value of slant – of any contributor who contributed at least once in the first and the last year. We call the former crowd of contributors the *starting crowd*, and the latter crowd of contributors the *ending crowd*. The average bias of the starting crowd in 2005 is 0.00608, whereas the average bias of the ending crowd in 2010 is 0.00315. That is a decline of 0.00293.

Next we identify the composition of each crowd. In the starting crowd, 6,448 (2.67%) contributors are the staying contributors - those who still edit after five years, with an average bias of 0.00489; and 234,472 (97.33%) are the leaving contributors who do not edit after 5 years, whose average bias in the first year is 0.00611. The average bias of the starting crowd is thus 0.00608 = 0.00489*2.67%+0.00611*97.33%. In the ending crowd, these 6,448 staying contributors have an average bias of 0.00242. The rest 454,891 (98.60%) are new contributors, who joined during the 5-year period, whose average bias in 2010 is 0.00316. Next, we compute the proportion of the bias decline due to each effect.

To consider the importance of the composition effect, we simulate what would happen if only the composition effect shaped the outcome. If some contributors leave after five years but the remaining have the same slant as in the beginning, then the ending crowd's bias would equal the average of the staying contributors' beginning bias (0.00489) and the new contributors' bias (0.00316), yielding 0.00489*1.40%+0.00316*98.60%=0.00318. The five-year bias decline from the composition effect would be 0.00608 - 0.00318 = 0.00290.

For comparison, we simulate what would happen if only the ideological shift shaped the outcome. We simulate all contributors remaining active but their slant changes over time. In this case, the ending crowd's average bias would equal the average of the staying contributor's decreased bias (0.00242) and the new contributors' bias (0.00316), yielding 0.00242*34.6% + 0.00316*65.4% = 0.00290.[21] The five-year bias decline from the ideological shift only would be 0.00608 - 0.00290 = 0.00318.

The ideological effect is not common enough to have a big effect on the overall results. It involves too few people as shown in the above calculation. With an actual crowd bias decline of 0.00293 in the raw data, a simple equation solving for the proportion of each effect yields the estimate that roughly 89.3% of the change is due to the composition effect, and the remaining 10.7% is due to the ideological shift effect. While this is a simple back-of-the-envelope calculation, it provides an indicator that the decline in contributor slant is largely due to the composition effect.

5. Discussion

5.1. Mass Edits: Is the Change in Slant Robust Under Exogenous Conditions?

An ideal design to establish causality would employ some exogenous shocks, and observe how contributors' slants change before and after these shocks, comparing these actions with those that did not receive shocks. We operationalize this idea with the special circumstances of *mass edits*, where an article attracts an unusually high number of contributions in one day due to a sudden social event or due to breaking news about the topic. In these occasions, the article usually receives a large volume of searches online. Social events or breaking news is unpredictable, and, so are these mass edits.

We define mass edits using online search volumes from the Google Trends website.[22] Specifically, we use the article's title as the search keyword(s) in Google Trends, collect the global daily Google Search Index (GSI) for a two-week window around the potential "mass edit" date, and compare whether the average GSI of the 3-day window around the search date (day -1, day 0, day 1) is greater than the 3-day average GSI before the mass edit date (day -4, day -3, day -2) and the 3-day average GSI after the mass edit day (day 2, day 3, day 4). Mass edits are defined as the contributions to an article during a day when the article: 1) receives more than 10 contributions, and 2) its title has an abnormal search peak in GSI during the 3-day window around that day.

Mass Edits Dummy is aggregated to the article-date level, which equals 1 if the article: 1) receives more than 10 contributions on that date, and 2) its title has an abnormal search peak in GSI during the 3-day window. *Mass Edits Dummy* equals 0 otherwise.

---

[21] 34.6% = (6,448+234,472) / (6,448+234,472+454,891).

[22] See here for descriptions of the Google Trends website and Google Search Index: https://support.google.com/trends/?hl=en#topic=6248052, accessed May 2018.

We define a "flip" of an article's slant as the changes from extremely left/right leaning, i.e., more than 2 standard deviations away left/right from neutral, to extremely right/left leaning. Aggregated to the article-date level, *Flip Dummy* equals 1 if an article has at least one flip on a given day, and equals 0 if the article has no slant flip on that day.

We estimate the following equation to understand how contributor slant changes during mass edits:

$$Contributor\ Slant\ by\ Year_{ijt} = \beta_0 + \beta_1 Contributor\ Years_{ijt} + Z_{it}B + \mu_j + \epsilon_{it}\ . \qquad (4)$$

The contributor slant changes over time. The coefficient β_1 can identify whether and how the contributor's slant changes over time. Here Z_{it} controls for time-varying differences among contributors, such as *Number of Edits.* μ_j is a contributor fixed effect. As with prior estimates, because it is not possible to estimate μ_j, a contributor fixed effect, for contributors who make one contribution, the number of observations that enter the regression with contributor fixed effect are actually smaller by 2,180,327. We try estimates with and without this effect.

In Table 7, we present the analysis results based on Equation (4) using both the full sample of contributions and the subsample of mass edits. The OLS regression results in Column (2) show that, compared to the same analysis in Column (1) for all edits, during mass edits contributors become neutral more than twice *faster* than during normal edits, as shown by the estimated coefficients for *Contributor Years*. The results still hold after dropping the edits made during the first 3 months of each article.

In addition, we also see in Table 8 Column (1) that, compared to the same analysis in Table 3 Column (2) for the full sample, the estimated coefficient for *Prior Article Slant* has an absolute value twice as large, indicating that contributors who get attracted during mass edits are "more opposite" than those who get attracted to the article during normal edits. The findings still hold if we focus on mass edits for only more mature articles, after we drop the contributions during the first 3 months after each article is created.

Why do contributors' slants change faster during mass edits? We do a simple mean comparison using T-test of the average article slant, comparing mass edits and normal edits. This comparison shows that the article versions after each mass edit are significantly *more biased* than article versions after each normal edits ($p < 0.0000$). This is consistent with our findings on the moderating effects of biased contents on contributor slant. Next we examine a "flip." Again, we find a significantly higher frequency of article slant flips on mass edits days in comparison to normal days ($p < 0.0000$).

We estimate the likelihood an article has at least one slant flip on mass edit days, and compare it to normal edit days. The logit regressions are in Table 9. Each observation is an article-date. After converting the estimated coefficients, during mass edits an article is 11.8% more likely to experience slant slips than during normal edits. We conclude that contributors who contribute during mass edits are more likely to be exposed to contents of *both extremes*, and that explains why their slant changes faster than normal periods. The interpretation could be that during mass edits, contributors change faster because their arguments are settled faster.

5.2. Rate of Slant Change: How Long Will It Take for Contributors to Become Neutral?

We next estimate how long it takes for a contributor's slant to gradually become neutral if this tendency continues. Given the potential for attenuation bias, this is, at best, an underestimate of the speed it takes.

We use a Markov Chain Process to simulate the evolution. Although a contributor's slant exhibits long-term trend over the years, it fluctuates frequently, and this should be accounted for. We divide slant into different bins and investigate how a contributor's slant changes from one bin to another. *Contributor yearly slant* is divided into seven bins, divided by the ± 0.5, ± 1.5, and ± 2.5 standard deviations intervals. The middle bin represents a neutral slant; the first and last bins represent extreme slants. We then compute a transition matrix for contributor slant based on our empirical data: For each year, we compute the proportions of contributors whose yearly slant moves from one slant bin to another, and fill the probabilities in the transition matrix for this year. Averaging the transition matrices among all years gives us the final transition matrix we use in our simulation, reported in Figure 10.

In this transition matrix, the rows denote the starting bins and the columns denote the ending slant. Bin 4 represents a neutral slant, defined as a slant index ranging from -0.5 to 0.5 standard deviations away from the mean. We find that: (1) the probabilities on the diagonal are large. As expected, contributors tend to have a higher chance of staying near their original slant; and (2) the farther the end bins are from the start bins, the smaller the probabilities. This indicates that contributor slant change is a gradual and a cumulative process, and it is not likely that the contributor's slant would suddenly jump from one extreme to another.

Next, we use the transition matrix to simulate the contributor slant change process over time (see Table 13). We compute the time it takes for a contributor to have a greater than 50% probability of moving to neutral. As expected, the length of time depends on the contributor's original slant: Extremely slanted contributors spend a longer time moving to neutral than slightly slanted contributors. More surprisingly, we find that on average, it takes one more year for the Republicans to become neutral than for Democrats.

We test for possible reasons why Republican contributors tend to a neutral slant slower than Democratic contributors. First, do Republican contributors display more EC behavior than Democratic contributors? Regression results of Equation (1) using the two groups respectively do not support this explanation. Republican contributors show more Non-EC behavior than Democratic contributors.

Second, Republican contributors might choose to edit less extreme articles compared to Democratic contributors, so that they are less influenced during their interaction with online content. However, we find no statistically significant difference between the level of content extremeness for the articles edited by Republicans or Democrats. The distributions contain similar bias and variance.

A third possible reason might stem from the contributors' number of edits. Republican contributors make fewer edits than Democrats, so their experience has less of an effect on the overall tendency, and may differ in some way. Summary statistics provide evidence for this explanation. In our sample, the total number of edits from Democratic contributors is about 1.5 times that of Republican contributors.

This motivates examining whether the two types of contributors examine different topics, and whether each of these topics display different EC/Non-EC behavior. We characterize the heterogeneity of Non-EC/EC among different topics, using Wikipedia's classification for articles. This exercise is also of independent interest for the potential to derive insight from which topics generate EC for which types.

We create dummy variables for each topic categories and modify Equation (1), adding these dummies and their interactions with $Prior\ Article\ Slant$. We then compute the EC effect for each topic category using the regression results. There are 24 categories of topics, and these are not mutually exclusive. Articles can speak to one or more topics, and these rarely change over the lifetime of an article. We estimate this modification to Equation (1) for the entire sample, and for two sub-samples, one consisting of Republican contributors and one for Democrat Contributors. We report the results in Table 11.

Consistent with our overall findings, the majority of topics display Non-EC for contributors from both parties. For example, the four topics with the most edits – Foreign Policy, Government, War and Peace, and Biographies – display an overall pattern of Non-EC. Overall, the general finding about Non-EC is reflected in most subgroups.

The absence of EC is the most striking finding, so its strong presence in any topic is noteworthy. Three topics—Homeland Security, Energy, and Tax—display evidence of a segregated conversation, where both parties engage in EC. These are not in the top ten in terms of the number of edits, so they do not shape the overall patterns very much. In these three topics, however, the EC effect of Republican contributors is much stronger than that of Democrats, indicating that Republicans' edits are the relatively stronger force that contributes to these segregated conversations. This topic also is noteworthy, given later Republican action on taxation. It is harder to interpret on the other two topics.

Since the departures from overall non-EC are rare, even the weak presence of EC is striking. Among the ten topics receiving the most edits, three topics – Budget and Economy, Civil Rights, and Crime – display an interesting pattern: Non-EC overall, with either Democrats displaying EC and Republicans displaying Non-EC, or no significant pattern. This arises because the Democratic contributors resist changing content when Republicans try to insert their points of view. (Yet, it is unclear why these three topics are the focus of Democratic conversation, except for Civil Rights, which is a staple of the Democratic coalition.) A similar but opposite pattern, with Democrats displaying Non-EC and Republicans displaying EC, occurs on only one topic with much fewer edits—Healthcare. (This is noteworthy as a sign of the later passionate views coming from Republicans about US healthcare.)

Overall, Table 11 suggests Republican and Democratic contributors do occasionally have different experiences, selecting among different groups of articles to edit. The weight of experience results in Non-EC overall, with Republican editors experiencing (somewhat) segregated conversations less frequently. To say it another way, Republicans evolve more slowly to neutral because of the proportion of time they find themselves on content of their same slant compared to Democrats. The findings again supports our primary conclusions that (1) online experiences change contributors' slant and (2) there is a tendency for Wikipedia contributors' slants to moderate in the majority of articles.

We note one intriguing feature of Table 11. Some of the historically most divisive topics of US politics, such as Abortion and Gun Control, do not display EC behavior. The areas with the most EC behavior – namely, where participants from the two political parties do not achieve a neutral point of view – are topics that did play a prominent role in the Trump campaign's approach to the 2016 election, which occurs at least five years after our sample. These topics include budget & economy and taxes (after the election the new administration implemented as a tax cut), healthcare (implemented as attempts to repeal Obamacare), and homeland security (implemented in immigration bans and attempts to build a southern border wall). It is a little less clear how to link the other two topics which contain EC behavior, namely, civil rights and crime, to the administration's behavior. We label this finding "intriguing" because considerable effort would be required to link this behavior to the preferences of voters, the substance of campaign, and actual policy. That is beyond the scope of this study.

6. Robustness

We further corroborate our findings by performing the following robustness tests.

6.1. Is the Measure of Contributor Slant Representative of Ideologies?

First, one might be concerned about whether the measure of slant in Wikipedia is representative of contributors' real-world political ideologies. Also, a neutral article in our sample can either be interpreted as having no slanted words at all or as having equal numbers of very slanted words. These concerns might lead to questioning the external validity of the slant measure.

To address this concern we use an alternative measure of slant and bias of contributors. We match the voting data from the 2004 Presidential Election to locations affiliated with IP addresses of contributors.[23] Because Wikipedia only reveals IP addresses for contributors without user IDs, we restrict our sample to contributors who are not logged in when editing the articles and also drop contributors whose IP addresses indicate that they are located outside the United States. Using OLS regressions, we then test the relationship between the voting record and *Prior Article Slant*. Note that this analyzes the behavior of a different population of contributors than the contributors we have examined thus far.[24] This regression is valid under the assumption that a contributor has – on average – the political tastes of the regions in which they live.

Table 12 presents the results. *RepPerc* denotes the percentage of Republican votes in the contributor's county. As we use positive values in the slant index to indicate Republican-leaning ideologies for Wikipedia users and articles, the negative and statistically significant coefficient of *Prior Article Slant* suggests that a contributor from a county with higher percentage of Republican votes tends to target a Democratic-leaning article when he or she contributes on Wikipedia. The results show a Non-EC pattern in the contributing process and are qualitatively similar to the prior estimates. This also provides support that the measure of contributors' slant reflects contributors' real political ideologies.

We also collected talk pages for articles. These are used by contributors to discuss edits and to achieve consensus. We find that the total size of an article's talk pages has a correlation of 0.22 with the average bias of the article over time, suggesting that our bias measure does capture how contested an article is.

6.2. What Else Could Be Driving the Non-EC behavior?

The effect of Non-EC in contributors' voluntary editing behavior indicates that contributors are more likely to edit articles with the opposite slant. However, apart from the interpretation of contributors being attracted by the article slant, this could also be due to a "correcting" behavior between contributors, which might have little to do with the article's slant. On Wikipedia, we sometimes see edits that are reverted and added back within a short time, which are called "edit wars." Could these edit wars be driving the Non-EC effect? We address this question by including only the initial edits of every contributor when they revise an article for the first time. Doing so rules out edit wars or any possible correcting behavior later in the edits.

We observe from Table 13 that the signs and statistical significance of the estimated coefficients do not change, and the magnitude of the coefficients becomes even larger, indicating an even stronger Non-EC effect than when investigating all edits. The results further strengthen the robustness of the Non-EC effect.

---

[23] The data on geolocation of IP comes from MaxMind. We match on county records.

[24] The identities of contributors are known after they register, and when they edit after logging on. An anonymous edit comes from either an unregistered contributor or from an editor who chose not to logon before editing. Hence, it is possible for the samples to include some of the same contributors, but it is not possible to know what fraction.

We also conduct several additional robustness checks to make sure the Non-EC effect is not driven by alternative explanations. First, our slant index is measured on the basis of frequently used phrases, or code phrases, favored by party representatives. It may be the case that longer articles tend to contain more code phrases and are therefore more measurable. In this case, long articles could drive our results. To rule out this explanation, we eliminate outlying long articles from our full sample, that is, articles that are more than two standard deviations above the mean article length. We obtain similar results.

Second, the articles whose titles contain code phrases might tend to show greater biases in our sample simply because these code phrases are more likely to be used repetitively in the article content. To check our findings against this concern, we exclude from our sample all articles whose title contains code phrases, which is 1.77% of all articles. Again, we find a significant Non-EC effect from the results.

Third, it is possible that certain code phrases are chosen simply because these words do not have other commonly-used synonyms that are neutral or of the opposite slant. In this case, as our measure captures the contributor's choice of words describing the same concept for a given topic, one's contribution may be slanted merely because he or she could not find neutral substitutes of the code phrases to choose from. We rely on the experiences of a legal and copyediting professional to identify these instances in our dictionary and leave only code phrases with natural substitutes. After re-measuring the slant index for articles and contributors, we repeat our analyses and find no significant change in our results. Therefore, the Non-EC effect is not driven by instances where contributors do not have a choice for substitute phrases.

Fourth, because contributors' edits to popular articles tend to have greater impact than those to less popular ones, their political slants measured from these popular articles could carry more weight. Therefore, we use articles' page views as weights when computing the average contribution slant and repeat our analysis using the weighted contributor slant. We continue to find significant Non-EC patterns.

We are also concerned that contributors blocked by Wikipedia administrators may affect our results.[25] These contributors may create extremely biased content initially and drop out of the dataset after being blocked. As a result, contributors overall may become more neutral over time. This problem is mitigated by our approach of assigning missing values to *Contributor yearly slant* when a contributor makes no edits in a year. As a robustness check, we repeat our analysis after dropping all 56,329 contributors who have ever been blocked (temporarily or permanently) and the associated 480,960 edits from our sample. Again, the results remain unchanged.

Finally, we test if the Non-EC effect is driven only by extremely slanted articles. We eliminate from our full sample articles with slant index two standard deviation points away from the mean. Changing this

[25] Blocks are used to prevent damage or disruption to Wikipedia. Contributors may be blocked for reasons such as vandalism and edit warring. See https://en.wikipedia.org/wiki/Wikipedia:Blocking_policy for the detailed policy, accessed August 2017.

threshold to articles without slant in the top and bottom 10% does not differ qualitatively in results. The estimated coefficients with subsamples have the same signs but larger absolute values.

7. Conclusion

This research shows that Wikipedia has a record of bringing opposing opinions into the same conversation. Our findings point toward patterns that lead contributors to offer content to those with different points of view, avoiding micro-behavior that contributes to EC. We measure both an ideological shift, in which contributors contribute less slanted content over time, and a composition shift, in which extremely biased contributors contribute less content over time. Investigating the mechanism shows evidence of extreme contributors either becoming more moderate, or tending to leave after interacting with opposite-slant contents or encountering pushbacks from others frequently. A contributor's slant also becomes neutral faster as he or she is exposed to biased contents of both extremes. These effects reinforce the prevalence of unsegregated conversations at Wikipedia over time. Contributors interact with those of opposite viewpoints much more frequently than they silo themselves and participate in echo chambers.

Our study offers a two-step method for identifying the mechanisms contributing to polarization that distinguishes selection from evolution. Nothing in these methods presumes the results; the method can flexibly measure contributions to (un)segregated conversations in a variety of settings.

These findings have implications for when online communities could be hampered by a crowd's enthusiasm or frenzy. Collective intelligence should be more trustworthy when mechanisms encourage confrontation between distinct viewpoints. It also should adopt processes, as Wikipedia contributors have, which retains contributors who learn to moderate their contributions from their experience.

It is not as if Wikipedia avoids its share of disagreements and confrontations, so the findings also raise a subtle question: How does Wikipedia transform controversial topics into arguments that include many points of view and sustain the community over time? We believe that this success arises from the institutions that help overcome the challenges affiliated with aggregating contested knowledge. For one, the aspiration of achieving NPOV directs attention to specific areas. No side can claim exclusive rights to determine the answer, which allows every contributor to add another paragraph if it diffuses an issue by giving voice to dissent. In addition, miniscule storage and transmission costs reduce the cost of listing another view on a web page. Our results also suggest that the conflict resolution mechanisms and the mix of informal and formal norms at Wikipedia play an essential role in encouraging a community that works towards a neutral point of view. This finding is consistent with theories that articles go through a lifecycle, settle into a consensus, which contributors subsequently "defend" (see e.g., Kane et al, 2014).

These findings also raise questions for the market design literature about other online social media – such as Facebook, Twitter, and Reddit. We speculate that some simple design differences may have profound consequences for (un)segregating conversations. For example, Wikipedia contributors can both add material and remove material or refine the content in myriad ways, whereas contributors on Facebook/Twitter only add additional content on top of what is already there. Allowing for removing or editing anyone's contributions can change how the reader and writer choose to direct the conversations, resulting in contributions from different points of view. Some platforms also aggregate contributions in ways that shape the prevalence of segregation. For example, on Yelp (e.g., rating restaurants) or Rotten Tomatoes (e.g., rating movies) additional material can be added without limit, the platform provides a numerical summary that can direct conversations between readers and reviewers. Our results frame questions about whether a numerical summary motivates others with views that differ from the summary or attracts more reviews from those who agree with it.

These findings also highlight the importance of platform design of algorithms. For example, on Facebook, an algorithm selects content for users, and its design increases the chance that participants read and write contents only in a community of like-minded people. In contrast, Wikipedia contributors have the option to be exposed to different opinions and can make the choice of reading and writing any content on the platform. Future work can focus on the heterogeneous effect of online participation on different contributor subgroups—for example, with interest in different political topics, or participation in different types of online platforms, such as resource-sharing platforms versus communities of innovation. In addition, existing literature on open communities investigates the content production more frequently than the contributors themselves. Given the huge number of volunteers on Wikipedia, as well as the enormous attention this community gets from around the globe, we hope to see more research on Wikipedia's online participation and interactions, as well as on the mechanisms behind changes to its content.

References

Aaltonen, A., and Seiler, S. 2015. Cumulative growth in user-generated content production: Evidence from Wikipedia. *Management Science* 62(7): 2054–2069.

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

Ahn, Y-Y., Han, S., Kwak, H., Moon S., and Jeong H. 2007. Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th international conference on World Wide Web* (WWW'07). ACM, New York, NY, USA, 835–844.

Algan, Y., Benkler, Y., Morell, M.F. and Hergueux, J. 2013. Cooperation in a peer production economy: Experimental evidence from Wikipedia. *Workshop on Information Systems and Economics*.

Angrist, Joshua, J. Pischke. 2008. "Mostly Harmless Econometrics." *Princeton, NJ: Princeton University*.

Arazy, O., Nov, O., Patterson, R., and Yeo, L. 2011. Information quality in Wikipedia: The effects of group composition and task conflict. *Journal of Management Information Systems* 21(4): 71–98.

Arazy, O., Ortega, F., Nov, O., Yeo, L., and Balila, A. 2015. Functional roles and career paths in Wikipedia. *Computer Supported Cooperative Work (CSCW):* 1092–1105.

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.F., Lee, J., Mann, M., Merhout, F. and Volfovsky, A., 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115(37): 9216–9221.

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

Bassamboo, A., Cui, R., and Moreno, A. 2015. The wisdom of crowds in operations: Forecasting using prediction markets. Working paper. SSRN: https://ssrn.com/abstract=2679663.

Benkler, Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom.* Yale University Press.

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

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

Boxell, L., Gentzkow, M., and Shapiro, J.M. 2017. Is the internet causing political polarization? Evidence from demographics. *National Bureau of Economic Research* Working paper No. 23258.

Budak, C., Goel, S., and Rao, J. M. 2014. Fair and balanced? Quantifying media bias through crowdsourced content analysis. Working paper. SSRN: http://ssrn.com/abstract=2526461.

Burke, M., and Kraut, R. 2008. Mopping up: Modeling Wikipedia promotion decisions. *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW)*: 27–36.

Carnahan, S., and Greenwood, B.N. 2017. Managers' political beliefs and gender inequality among subordinates: Does his ideology matter more than hers?. *Administrative Science Quarterly*.

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

Chen, S.S., Chuang, Y.W., and Chen, P.Y. 2012. Behavioral intention formation in knowledge sharing: Examining the roles of KMS quality, KMS self-efficacy, and organizational climate. *Knowledge-Based Systems* 31: 106–118.

Chesbrough, H.W. 2006. *Open Innovation: the New Imperative for Creating and Profiting from Technology*. Harvard Business Press.

Chiang, C.F. and Knight, B. 2011. Media bias and influence: Evidence from newspaper endorsements. *Review of Economic Studies* 78(3): 795–820.

Chiu, C-M., Hsu, M-H., and Wang, E. 2006. Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision support systems* 42(3): 1872–1888.

Collier, B., Burke, M., Kittur, N., and Kraut, R. 2008. Retrospective versus prospective evidence for promotion: The case of Wikipedia. *2008 Meeting of the Academy of Management.*

Cui, R., Li, J., and Zhang, D.J. 2016. Discrimination with incomplete information in the sharing economy: Evidence from field experiments on Airbnb. Working paper. SSRN: https://ssrn.com/abstract=2882982

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

Draper, N.R.; Smith, H. 1998. *Applied Regression Analysis (3rd ed.).* John Wiley.

Durante, R. and Knight, B. 2012. Partisan control, media bias, and viewer responses: Evidence from Berlusconi's Italy. *Journal of the European Economic Association* 10(3): 451–481.

Edelman, B.G. and Luca, M. 2014. Digital discrimination: The case of Airbnb.com. *Harvard Business School NOM Unit* Working Paper No. 14-054.

Fan, M., Tan, Y., and Whinston, A.B. 2005. Evaluation and design of online cooperative feedback mechanisms for reputation management. *IEEE Transactions on Knowledge and Data Engineering* 17(2): 244–254.

Faraj, S., Jarvenpaa S.L., and Majchrzak, A. 2011. Knowledge collaboration in online communities. *Organization Science* 22(5): 1224–1239.

Forte, A., Kittur, N., Larco, V., Zhu, H., Bruckman, A., and Kraut, R. E. 2012. Coordination and beyond: Social functions of groups in open content production. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work:* 417–426.

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

Gallus, J. Forthcoming. Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia. *Management Science.*

Gao, G.G., Greenwood, B.N., Agarwal R., and McCullough, J.S. 2015. Vocal minority and silent majority: How do online ratings reflect population perceptions of quality? *MIS Quarterly* 39(3): 565–589.

Gentzkow, M., and Shapiro, J.M. 2003. Media, education, and anti-Americanism in the Muslim world. *Journal of Economic Perspectives* 18(3): 117–133.

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

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

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

Greenstein, S., and Zhu, F. 2016. Open content, Linus' Law, and neutral point of view. *Information Systems Research* 27(3): 618–635.

Greenstein, S., and Zhu, F. Forthcoming. Do experts or collective intelligence write with more bias? Evidence from Encyclopædia Britannica and Wikipedia. *MIS Quarterly.*

Groseclose, T., and Milyo, J. 2005. A measure of media bias. *Quarterly Journal of Economics* 120(4): 1191–1237.

Gu, B., Konana, P., Raghunathan, R., and Chen, HM. 2014. Research note—The allure of homophily in social media: Evidence from investor responses on virtual communities. *Information Systems Research* 25(3): 604–617.

Halfaker, A., Kittur, A., and Riedl, J. 2011. Don't bite the newbies: How reverts affect the quantity and quality of Wikipedia work. *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym'11)*: 163–172.

Hargadon, A.B., and Bechky, B.A. 2006. When collections of creatives become creative collectives: A field study of problem solving at work. *Organization Science* 17(4): 484–500.

Hill, M. 2017. Almost Wikipedia: Eight early Encyclopedia projects and the mechanisms of collective action. Working paper. https://mako.cc/academic/hill-almost_wikipedia-DRAFT.pdf, accessed January 2017.

Horrace, William C., Ronald L. Oaxaca. 2006. Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters* 90(3): 321–327.

Janis, I. L. 1982. Groupthink: Psychological studies of policy decisions and fiascoes. *Houghton Mifflin*.

Jelveh, Z., Kogut, B., and Naidu, S. 2014. Political language in economics. *Columbia Business School Research*, Paper No. 14-57.

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

Jeppesen, L.B., and Frederiksen, L. 2006. Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization Science* 17(1): 45–63.

Kane, G.C., and Fichman, R. G. 2009. The shoemaker's children: Using Wikis for information systems teaching, research, and publication. *MIS Quarterly* 33(1): 1–17.

Kane, G.C., Johnson, J., and Majchrzak, A. 2014. Emergent life cycle: The tension between knowledge change and knowledge retention in open online coproduction communities. *Management Science* 60(12): 3026–3048.

Kittur, A., Chi, E., Pendleton, B. A., Suh, B., and Mytkowicz, T. 2007. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World Wide Web* 1(2): 19.

Kittur, A., and Kraut, R. 2008. Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. *ACM Conference on Computer Supported Cooperative Work:* 37–46.

Kittur, A., Suh, B., Pendleton, B. A., and Chi, E. H., 2007. He says, she says: Conflict and coordination in Wikipedia. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems:* 453–462.

Kogut, B., and Metiu, A. 2000. The emergence of E-Innovation: Insights from open source software development. *Reginald H. Jones Center Working Paper, Philadelphia, PA*.

Kogut, B., and Metiu, A. 2001. Open-source software development and distributed innovation. *Oxford Review of Economic Policy* 17(2): 248–264.

Kogut, B., and Zander, U. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science* 3(3): 383–397.

Kuk, G. 2006. Strategic interaction and knowledge sharing in the KDE developer mailing list. *Management Science* 52(7): 1031–1042.

Kwark, Y., Lee, G.M., Pavlou, P., and Qiu, L. 2016. The spillover effects of user-generated online product reviews on purchases: Evidence from clickstream data. Working paper.

Larcinese, V., Puglisi, R., and Snyder, J.M. 2007. Partisan bias in economic news: Evidence on the agenda-setting behavior of U.S. newspapers. *Journal of Public Economics* 95(9): 1178–1189.

Laster, D., Bennett, P., and Geoum, I.S. 1999. Rational bias in macroeconomic forecasts. *Quarterly Journal of Economics* 114 (1): 293–318.

Lawrence, E., Sides, J., and Farrell, H. 2010. Self-segregation or deliberation?: Blog readership, participation, and polarization in American politics. *Perspectives on Politics* 8(1): 141–157.

Lee, G.K., Cole, R.E. 2003. From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. *Organization Science* 14(6): 633–649.

Lee, Y-J., Hosanagar, K., and Tan, Y. 2015. Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Science* 61(9): 2241–2258.

Ma, M., Agarwal, R. 2007. Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research* 18(1): 42–67.

Majchrzak, A. 2009. Where is the theory in Wikis? *MIS Quarterly* 33(1): 18–20.

McPherson, M., Smith-Lovin, L. and Cook, J.M. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*: 415–444.

Muchnik, L., Aral, S., and Taylor, S.J. 2013. Social influence bias: A randomized experiment. *Science* 341(6146): 647–651.

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

Nagaraj, A. Forthcoming. Does copyright affect reuse? Evidence from Google Books and Wikipedia. *Management Science*.

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

Park, J., Konana, P., Gu, B., Kumar, A., and Raghunathan, R. 2013. Information valuation and confirmation bias in virtual communities: Evidence from stock message boards. *Information Systems Research* 24(4): 1050–1067.

Pierce, L. 2012. Organizational structure and the limits of knowledge sharing: Incentive conflict and agency in car leasing. *Management Science* 58(6): 1106–1121.

Piskorski, M.J., and Gorbatâi A. 2017. Testing Coleman's social-norm enforcement mechanism: Evidence from Wikipedia. *American Journal of Sociology* 122(4): 1183–1222.

Purdy, W. 2015. Radicalization: Social media and the rise of terrorism. *Testimony presented before the House Committee on Oversight and Government Reform's Subcommittee on National Security*, https://oversight.house.gov/wp-content/uploads/2015/10/10-28-2015-Natl-Security-Subcommittee-Hearing-on-Radicalization-Purdy-TRC-Testimony.pdf , accessed January 2017.

Qiu, L., and Kumar, S. 2017. Understanding voluntary knowledge provision and content contribution through a social-media-based prediction market: A field experiment. *Information Systems Research*.

Ramasubbu, N., Balan, R.K. 2012. Overcoming the challenges in cost estimation for distributed software projects. In *Proceedings of the 34th International Conference on Software Engineering*: 91–101.
Ransbotham, S., and Kane, G.C. 2011. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quarterly* 35(3): 613–627.
Ransbotham, S., Kane, G.C., and Lurie, N.H. 2012. Network characteristics and the value of collaborative user-generated content. *Marketing Science* 31(3): 387–405.
Reuter, J., and Zitzewitz, E. 2006. Do ads influence editors? Advertising and bias in the financial media. *Quarterly Journal of Economics* 121(1): 197–227.
Roberts J.A., Hann, I.H., and Slaughter, S.A. 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.
Rosenblat, T.S., and Mobius, M.M. 2004. Getting closer or drifting apart? *Quarterly Journal of Economics* 119(3): 971–1009.
Rothaermel, F.T., and Sugiyama, S. 2001. Virtual Internet communities and commercial success: Individual and community-level theory grounded in the atypical case of TimeZone.com. *Journal of Management* 27(3): 297-312.
Salganik, M.J., Dodds, P.S., and Watts, D.J. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762): 854–856.
Schroeder, A., and Wagner, C. 2012. Governance of open content creation: a conceptualization and analysis of control and guiding mechanisms in the open content domain. *Journal of the American Society for Information Science and Technology* 63(10): 1947–1959.
Shore, J., Baek, J., and Dellarocas, C. 2016. Network structure and patterns of information diversity on Twitter. Working paper.
Slivko, O. 2014. Peer effects in collaborative content generation: The evidence from German Wikipedia. ZEW - Centre for European Economic Research Discussion Paper No. 14-128.
Slivko, O., Kummer, M., and Saam, M. 2016. Identification of causal effects in the context of mass collaboration. In *Mass Collaboration and Education*: 391–411. Springer International Publishing.
Stone, D.F. 2011. Ideological media bias. *Journal of Economic Behavior & Organization* 78(3): 256–271.
Stvilia, B., Twidale, M. B., Smith, L. C., and Gasser, L. 2008. Information quality work organization in Wikipedia. *Journal of the American Society for Information Science and Technology* 59(6): 983–1001.
Sun, T., Viswanathan, S., and Zheleva, E. 2017. Creating social contagion through firm mediated message design: Evidence from a randomized field experiment. *Robert H. Smith School Research Paper No. RHS 2543864*.
Sunstein, C.R. 2001. *Echo Chambers: Bush v. Gore, Impeachment, and Beyond.* Princeton University Press.
Te'eni, D. 2009. Comment: The Wiki way in a hurry—The ICIS anecdote. *MIS Quarterly* 33(1): 20–22.
Tucci, C.L., and Villarroel, J.A. 2007. Open Source Corporate Strategy (OSCS): Reaching outside the firm's boundaries to tap upon global talent. In *Academy of Management Annual Meeting* (No. CSI-CONF-2007-001).
Van Alstyne, M., and Brynjolfsson, E. 2005. Global village or cyber-balkans? Modeling and measuring the integration of electronic communities. *Management Science* 51(6): 851–868.
Von Krogh, G., and Von Hippel, E. 2006. The promise of research on open source software. *Management Science* 52(7): 975–983.
Wang, Y., Goes, P., Wei, Z., and Zeng, D.D. 2017. Production of online word-of-mouth: Peer effects and the moderation of user characteristics. Working paper. SSRN: https://ssrn.com/abstract=2899452.
Xu, L., Niculescu, M.F., and Wu, D.J. 2015. When to use the open business model for software products under network effects? Working paper.
Xu, S.X., and Zhang, X.M. 2009. How do social media shape the information environment in the financial market? *ICIS 2009 Proceedings*: 56.
Xu, S.X., and Zhang, X.M. 2013. Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction. *MIS Quarterly* 37(4): 1043–1068.

Yang, Y., Chen, P.Y., and Pavlou, P. 2009. Open innovation: An empirical study of online contests. *ICIS 2009 Proceedings*: 13.
Zhang, X.M., and Zhu, F. 2011. Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review* 101(4): 1601–1615.
Zitzewitz, E. 2001. Measuring herding and exaggeration by equity analysts and other opinion sellers. Working Paper, Stanford Graduate School of Business. http://www.dartmouth.edu/~ericz/chapter1.pdf.

Table 1: Distribution of Different Types of Contributors over Years

Year	Democrat Contributors	Core Democrat Contributors	Republican Contributors	Core Republican Contributors	Neutral Contributors	Core Neutral Contributors	Total # of Contributors Contributed in the Year
2001	26.4%	18.1%	20.0%	12.5%	53.6%	9.9%	800
2002	9.9%	7.5%	9.6%	7.4%	80.4%	17.6%	4,364
2003	8.5%	6.5%	8.8%	6.9%	82.6%	18.3%	14,951
2004	7.8%	5.7%	7.7%	5.9%	84.5%	17.3%	66,867
2005	7.0%	4.7%	6.7%	4.6%	86.3%	15.6%	242,121
2006	5.7%	3.6%	5.7%	3.6%	88.6%	14.7%	584,438
2007	5.3%	3.2%	5.2%	3.3%	89.5%	13.8%	706,195
2008	5.2%	3.1%	5.3%	3.2%	89.5%	13.9%	640,871
2009	4.7%	3.1%	4.7%	3.2%	90.5%	14.1%	526,255
2010	4.2%	2.8%	4.2%	2.9%	91.6%	13.2%	461,663
2011	9.5%	8.5%	10.8%	9.9%	79.6%	19.4%	26,886

Notes: "Democrat/Republican/Neutral contributors" shows the percentage of contributors with negative/zero/positive *Contributor Slant* among all contributors who contribute in that year to the articles in our sample. "Core Democrat/Republican/Neutral contributors" shows the percentage of that year's "Democrat/Republican/Neutral contributors" whose total number of edits is distributed in the top 10% of all contributors' total number of edits. Final year, 2011, is sampled in January, which accounts for the low numbers in that year.

Table 2: Summary Statistics of Variables Used in the Main Analyses

Variable	Mean	Std. dev.	Min	Max
Contributor Slant	-0.0001	0.023	-1.229	0.998
Contributor Category	0.001	0.112	-1	1
Prior Article Slant	-0.057	0.208	-0.605	0.624
Prior Article Category	-0.058	0.265	-1	1
Prior Article Length	4,075.930	3,850.209	0	1,963,441
Prior Refs	34.189	61.040	0	1,636
Contributor Yearly Extreme Slant	-0.00086	0.021	-1.229	0.998
Opposite Slant Article Edits To Date	106.031	960.329	0	13143
Same Slant Article Edits To Date	37.468	231.295	0	6018
Revision War Edits To Date	36.774	145.788	0	2737
Number of Edits	1,175.720	7,567.790	1	122,264
Contributor Years	1.040	1.366	0.003	9.797
Stay Dummy	0.242	0.428	0	1
Opposite Slant Article Edits Fraction	0.0030	0.0497	0	1
Same Slant Article Edits Fraction	0.0758	0.2543	0	1
Revision War Edits Fraction	0.0077	0.0573	0	0.944
Starting Contributor Slant	0.0008	0.031	-0.768	0.784
Starting Number of Edits	13.774	114.361	1	48,192
Flip Dummy	0.0008	0.028	0	1
Mass Edits Dummy	0.004	0.065	0	1
RepPerc	0.457	0.142	0.093	0.920

Notes: Number of observations in this table is 10,878,391 except for: *Stay Dummy*, *Opposite Slant Article Edits Fraction*, *Same Slant Article Edits Fraction*, *Revision War Edits Fraction*, *Starting Contributor Slant*, and *Starting Number of Edits*, which are constructed at the contributor level with 2,887,140 observations; *Flip Dummy* and *Mass Edits Dummy*, which are constructed at the article-date level with 5,804,714 observations; *RepPerc*, which has 2,438,628 observations.

Table 3: OLS Regressions on the Relationship between Contributor Slant and Prior Article Slant

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Contributor Slant	Contributor Slant	Contributor Slant	Contributor Category	Contributor Category	Contributor Category
Prior Article Slant	-0.0075*** [0.0001]	-0.0074*** [0.0001]	-0.0167*** [0.0004]			
Prior Article Category				-0.0123*** [0.0002]	-0.0124*** [0.0002]	-0.0197*** [0.0009]
Log(Prior Article Length)		0.0005*** [0.0000]	0.0009*** [0.0001]		0.0014*** [0.0000]	0.0017*** [0.0003]
Log(Prior Refs)		-0.0003*** [0.0000]	-0.0009*** [0.0001]		-0.0008*** [0.0000]	-0.0024*** [0.0004]
Observations	10,878,391	10,878,391	10,878,391	10,878,391	10,878,391	10,878,391
Adjusted R-squared	0.005	0.006	0.006	0.001	0.001	0.001
Year FE	No	No	Yes	No	No	Yes
Article FE	No	No	Yes	No	No	Yes
Number of Articles	66,389	66,389	66,389	66,389	66,389	66,389

Notes: Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%. Observations in this panel are all the edits of the Wikipedia articles in our sample from 2001 to 2011. *Contributor Slant* is defined as the average slant change of all edits a contributor has made on Wikipedia. *Prior Article Slant* is the slant of the article before a particular edit. *Log(Prior Article Length)* is the logarithm of the article's total number of words. *Log(Prior Refs)* is the logarithm of the number of external references in the article plus one.

Table 4: Logit Regressions on the Relationship between Contributor Category and Prior Article Category

Model	(1)		(2)		(3)	
Dependent Variable	Contributor Category=-1	Contributor Category=1	Contributor Category=-1	Contributor Category=1	Contributor Category=-1	Contributor Category=1
Prior Article Slant	2.0743*** [0.0266]	-2.4063*** [0.0135]	2.0819*** [0.0269]	-2.3404*** [0.0133]	2.1042*** [0.0270]	-2.2918*** [0.0132]
Log(Prior Article Length)			-0.0344 [0.0045]	0.1486*** [0.0052]	-0.0115 [0.0051]	0.1859*** [0.0058]
Log(Prior Refs)			-0.2232*** [0.0032]	-0.3128*** [0.0030]	-0.2851*** [0.0042]	-0.4079*** [0.0040]
Year FE	No		No		Yes	
Article FE	No		No		Yes	
Observations	10,878,391		10,878,391		10,878,391	
Pseudo R-squared	0.021		0.038		0.043	

Notes: Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: OLS Regressions on Contributor Slant and Contents the Contributor Interact With

Model	(1)	(2)
Sample	All Contributions	Core Contributors' Contributions Only
Dependent Variable	Abs(Contributor Yearly Extreme Slant)	Abs(Contributor Yearly Extreme Slant)
Log(Opposite-Slant Article Edits To Date)	-0.000084* [0.000049]	-0.000081* [0.000049]
Log(Same-Slant Article Edits To Date)	0.000246*** [0.000063]	0.000235*** [0.000063]
Log(Revision War Edits To Date)	-0.000113*** [0.000041]	-0.000108*** [0.000041]
Observations	10,878,391	8,019,333
R-squared	0.001	0.001
Contributor FE	Yes	Yes
Year Dummies	Yes	Yes

Notes: Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%. The unit of analysis is each contribution. *Abs(Contributor Yearly Extreme Slant)* is calculated the same as *Abs(Contributor Yearly Slant)*, but contains only edits that make an article more extreme, i.e., edits that move an article to left-/right-leaning instead of neutral. A *Revision war* is defined as a contributor's edit being reverted immediately by another contributor, and then is immediately followed by the original contributor editing back the same contribution, as a "fight back". We compute the number of edits that the contributor made so far during such pushback situations, denoted as *Revision War Edits To Date*. This variable is highly skewed, so we use the logarithm of its value plus 1 in the regressions.

Table 6: Linear Probability Regressions on Contributors' Likelihood of Staying on Wikipedia

Model	(1)	(2)
Sample	All Contributors	Core Contributors Only
Dependent Variable	Stay Dummy	Stay Dummy
Abs(Starting Contributor Slant)	-0.0134*** [0.0023]	-0.1212*** [0.0201]
Log(Starting Number of Edits)	0.0887*** [0.0004]	0.0686*** [0.0008]
Opposite-Slant Article Edits Fraction	-0.0099*** [0.0017]	-0.0423*** [0.0085]
Same-Slant Article Edits Fraction	0.0005** [0.0002]	0.0046*** [0.0028]
Revision War Edits Fraction	-0.2045*** [0.0011]	-0.3864*** [0.0028]
Observations	2,887,140	366,319
R-squared	0.845	0.329
Vintage Dummies	Yes	Yes

Notes: Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%. The unit of analysis is each contributor. Model (2) includes only core contributors, i.e., contributors who made at least 3 edits in the sample. The dependent variable *Stay Dummy* equals 1 if the contributor made at least 1 contribution in 2010 or 2011. *Starting Number of Edits* is the total number of contributions that a contributor made in the first two years after he or she joined Wikipedia. *Starting Contributor Slant* is the yearly contributor slant in the first year when the contributor joined Wikipedia. *Vintage* is the year in which the contributor joins Wikipedia, or the vintage that the contributor belongs to. The "fractions" are computed using the absolute value of the corresponding variable in the previous table, divided by the contributor's total number of edits.

Table 7: OLS Regressions on How Contributor Slant Changes during Mass Edits

Model	(1)	(2)	(3)
Sample	Full Sample	Mass Edits Sample	Mass Edits Sample, with First 3 Months Dropped
Dependent Variable	Abs(Contributor yearly slant)	Abs(Contributor yearly slant)	Abs(Contributor yearly slant)
Contributor Years	-0.0002*** [0.0000]	-0.0005*** [0.0000]	-0.0005*** [0.0000]
Log(Number of Edits)	-0.0005*** [0.0000]	-0.0007*** [0.0001]	-0.0007*** [0.0000]
Observations	10,878,391	488,631	458,064
Adjusted R-squared	0.004	0.008	0.008

Notes: Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%. Observations in Column (2) includes only contributions that happened on the mass edits days; observations in Column (3) excludes contributions made within the first 3 months after each article is created.

Table 8: OLS Regressions on the Relationship between Contributor Slant and Prior Article Slant, During Mass Edits

Model	(1)	(2)
Sample	Mass Edits Sample	Mass Edits Sample, with First 3 Months Dropped
Dependent Variable	Contributor Slant	Contributor Slant
Prior Article Slant	-0.0146*** [0.0004]	-0.0150*** [0.0004]
Log(Prior Article Length)	0.0007*** [0.0001]	0.0007*** [0.0000]
Log(Prior Refs)	-0.0005*** [0.0000]	-0.0005*** [0.0000]
Observations	488,631	458,064
Adjusted R-squared	0.014	0.014

Notes: Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%. Observations in Column (1) includes only contributions that happened on the mass edits days; observations in Column (2) excludes contributions made within the first 3 months after each article is created.

Table 9: Logit Regressions between Article Slant Flips and Mass Edit Events

Model	(1)	(2)
Dependent Variable	Flip Dummy	Flip Dummy
Mass Edits Dummy	1.3893*** [0.1151]	2.0174*** [0.1161]
Prior Article Slant		-0.5597*** [0.0657]
Log(Prior Refs)		-0.0833*** [0.0116]
Log(Prior Article Length)		-0.4635*** [0.0076]
Observations	5,804,714	5,804,714
Pseudo R-squared	0.001	0.042

Notes: Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%. Observations in this panel are at the article-date level. The dependent variable *Flip Dummy* equals 1 if the article experiences at least 1 slant flip during that day. *Mass Edits Dummy* represents whether the article is receiving mass edits on the given day.

Table 10: Time Needed for a Contributor to Have > 50% Probability of Moving to Neutral Slant

Starting Contributor Slant	**Number of Years**
Extremely Democratic	10
Democratic	6
Slightly Democratic	3
Neutral	0
Slightly Republican	4
Republican	7
Extremely Republican	11

Notes: Number of years calculated based on the Markov Chain Process. *Neutral* state includes contributor slant 0.5 standard deviation away from 0. *Slightly Democratic (Republican)* state includes contributor slant between 0.5 and 1.5 standard deviations below (above) 0. *Democratic (Republican)* state includes contributor slant between 1.5 and 2.5 standard deviations below (above) 0. *Extremely Democratic (Republican)* state includes contributor slant more than 2.5 standard deviations below (above) 0. On average, after about 30 years, the probabilities in all articles' end state reach stationary distribution, with the probability of contributor slant moving to *Neutral* being 87.4%.

Table 11: Heterogeneity of EC and Non-EC across Different Article Topics

		All sample		Republican contributors		Democratic contributors	
Article Topics	No. of Edits	Estimate	Pattern	Estimate	Pattern	Estimate	Pattern
Abortion	30,400	-0.0039*** [0.0012]	Non-EC	-0.0161*** [0.0044]	Non-EC	0.0003 [0.0012]	*n.s.*
Budget & Economy	765,729	-0.0019*** [0.0003]	Non-EC	-0.0125*** [0.0011]	Non-EC	0.0036*** [0.0003]	EC
Civil Rights	902,531	-0.0038*** [0.0002]	Non-EC	-0.0183*** [0.0008]	Non-EC	0.0009*** [0.0002]	EC
Corporations	54,709	-0.0009 [0.0008]	*n.s.*	0.0035 [0.0031]	*n.s.*	-0.0046*** [0.0007]	Non-EC
Crime	957,613	-0.0016*** [0.0002]	Non-EC	-0.0089*** [0.0009]	Non-EC	0.0015*** [0.0003]	EC
Drugs	164,330	-0.0029*** [0.0007]	Non-EC	-0.0163*** [0.0025]	Non-EC	0.0001 [0.0012]	*n.s.*
Education	864,373	-0.0064*** [0.0003]	Non-EC	-0.0270*** [0.0011]	Non-EC	-0.0028*** [0.0003]	Non-EC
Energy	183,598	0.0021*** [0.0004]	EC	0.0103*** [0.0015]	EC	0.0012* [0.0007]	EC
Family	434,980	-0.0013*** [0.0003]	Non-EC	-0.0112*** [0.0014]	Non-EC	0.0020*** [0.0004]	EC
Foreign Policy	1,883,375	-0.0038*** [0.0002]	Non-EC	-0.0079*** [0.0007]	Non-EC	-0.0048*** [0.0004]	Non-EC
Trade	442,561	-0.0038*** [0.0004]	Non-EC	-0.0028*** [0.0010]	Non-EC	-0.0125*** [0.0009]	Non-EC
Government	3,376,993	-0.0039*** [0.0000]	Non-EC	-0.0174*** [0.0004]	Non-EC	-0.0026*** [0.0001]	Non-EC
Gun	62,668	-0.0037*** [0.0009]	Non-EC	-0.0207*** [0.0033]	Non-EC	-0.0003 [0.0012]	*n.s.*
Healthcare	385,659	-0.0004 [0.0004]	*n.s.*	0.0027** [0.0014]	EC	-0.0028*** [0.0006]	Non-EC
Homeland Security	478,796	0.0021*** [0.0004]	EC	0.0045*** [0.0014]	EC	0.0025*** [0.0004]	EC
Immigration	255,461	-0.0035*** [0.0005]	Non-EC	-0.0031* [0.0019]	Non-EC	-0.0047*** [0.0007]	Non-EC
Infrastructure & Tech	920,016	-0.0017*** [0.0003]	Non-EC	-0.0009 [0.0009]	*n.s.*	-0.0034*** [0.0004]	Non-EC
Jobs	693,295	-0.0023*** [0.0003]	Non-EC	-0.0074*** [0.0011]	Non-EC	-0.0031*** [0.0004]	Non-EC

Principles & Values	562,908	-0.0027*** [0.0003]	Non-EC	-0.0017 [0.0012]	*n.s.*	-0.0071*** [0.0004]	Non-EC
Social Security	2,501	-0.0111** [0.0048]	Non-EC	-0.0365* [0.0190]	Non-EC	-0.0138*** [0.0029]	Non-EC
Tax	46,048	0.0058*** [0.0007]	EC	0.0177*** [0.0033]	EC	0.0039*** [0.0007]	EC
War & Peace	1,837,644	-0.0018*** [0.0002]	Non-EC	-0.0030*** [0.0007]	Non-EC	-0.0022*** [0.0003]	Non-EC
Welfare & Poverty	439,851	-0.0031*** [0.0004]	Non-EC	-0.0109*** [0.0014]	Non-EC	-0.0010** [0.0004]	Non-EC
Biographies	1,311,337	-0.0024*** [0.0002]	Non-EC	-0.0014* [0.0008]	Non-EC	-0.0027*** [0.0003]	Non-EC

Notes: Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%; n.s.: not significant.

Table 12: Regressions on the Relationship between Percentage of Republican in the Area and Prior Article Slant

Model	(1)	(2)
Dependent Variable	RepPerc	RepPerc
Prior Article Slant	-0.0009** [0.0004]	-0.0010** [0.0004]
Log(Prior Article Length)		0.0037*** [0.0001]
Log(Prior Refs)		0.0005*** [0.0001]
Observations	2,438,628	2,438,628
Adjusted R-squared	0.000	0.001

Notes: Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%.

Table 13: Relationship between Contributor Slant and Prior Article Slant, First Edits Only

Models	(1)	(2)
Dependent Variables	Contributor Slant	Contributor Slant
Prior Article Slant	-0.0092***	-0.0218***
	[0.0001]	[0.0004]
Log(Prior Article Length)	0.0007***	0.0011***
	[0.0000]	[0.0001]
Log(Prior Refs)	-0.0004***	-0.0011***
	[0.0000]	[0.0001]
Observations	7,113,130	7,113,130
R-squared	0.007	0.007
Year FE	No	Yes
Article FE	No	Yes
Number of Articles	66,389	66,389

Notes: Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%. Observations in this panel only include every contributor's first edit of an article.

Figure 1: Average Absolute Value of Contributor Slant over the Years

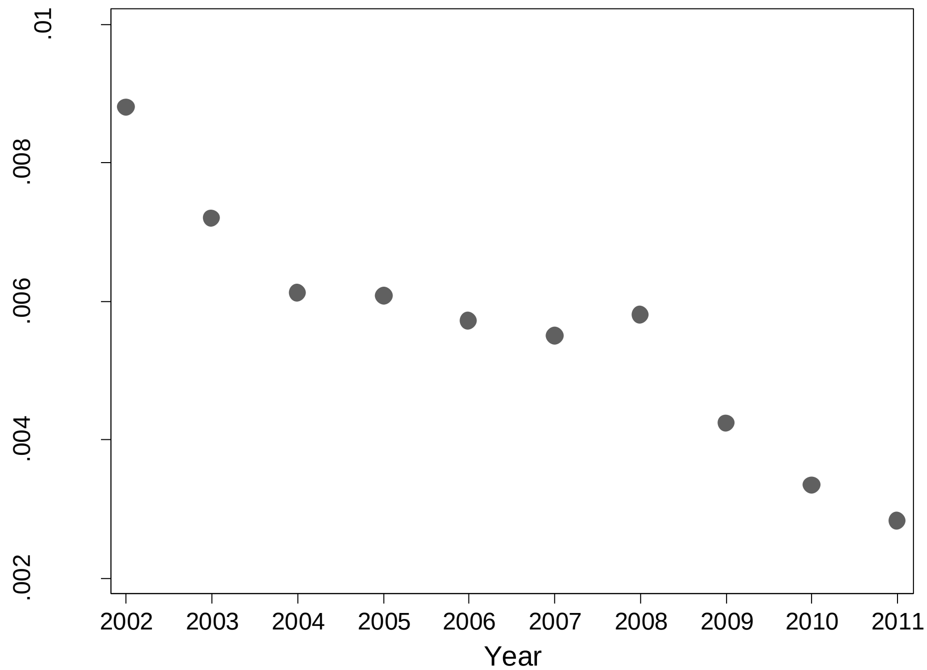

Figure 2: Vintage Analysis for Contributors Entering in Different Years

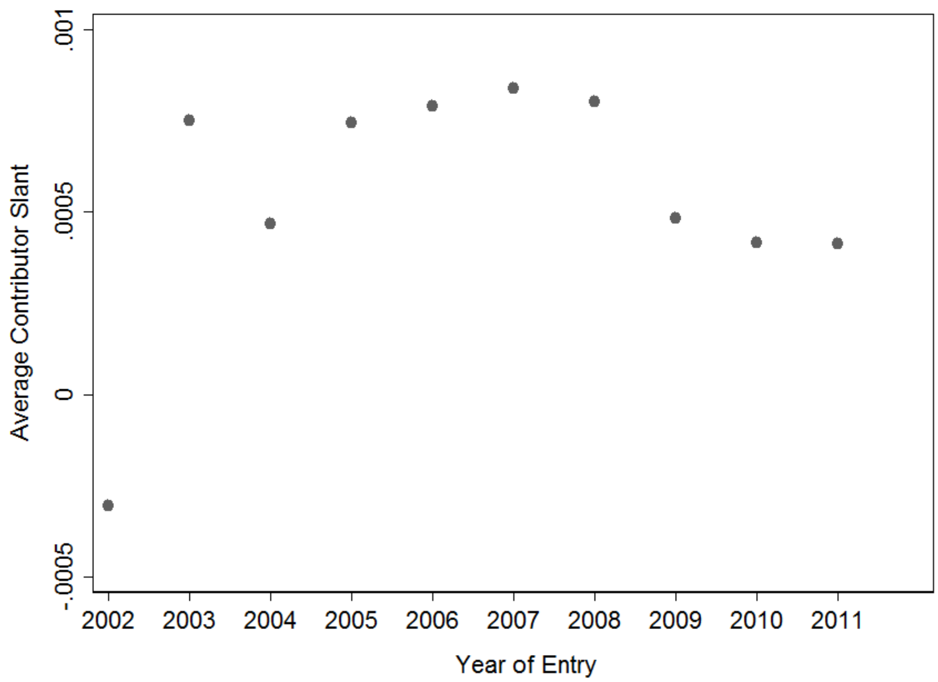

Figure 3: Average Number of Edits in the Year over Contributor's Years on Wikipedia, Extreme Contributors Only

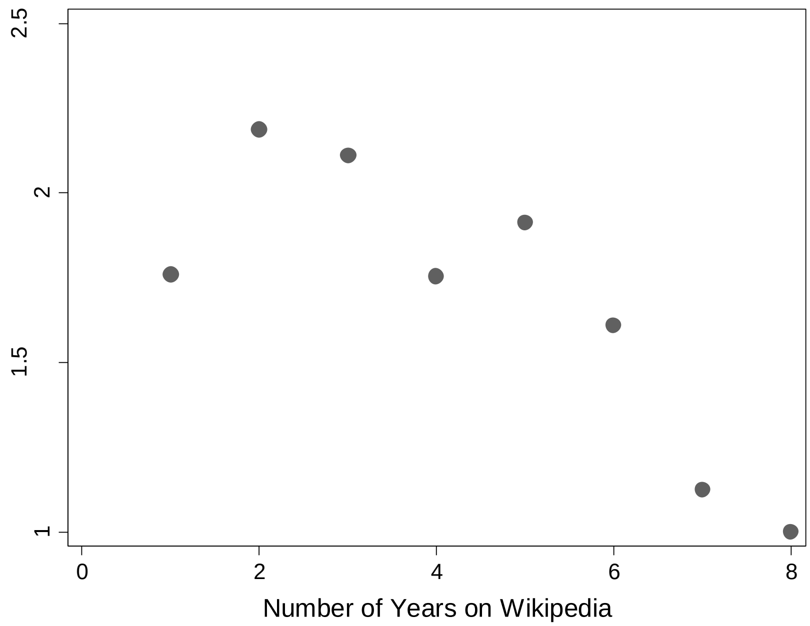

Figure 4: Average Contributor Slant Each Year over Contributor's Years on Wikipedia, Extreme Contributors Only

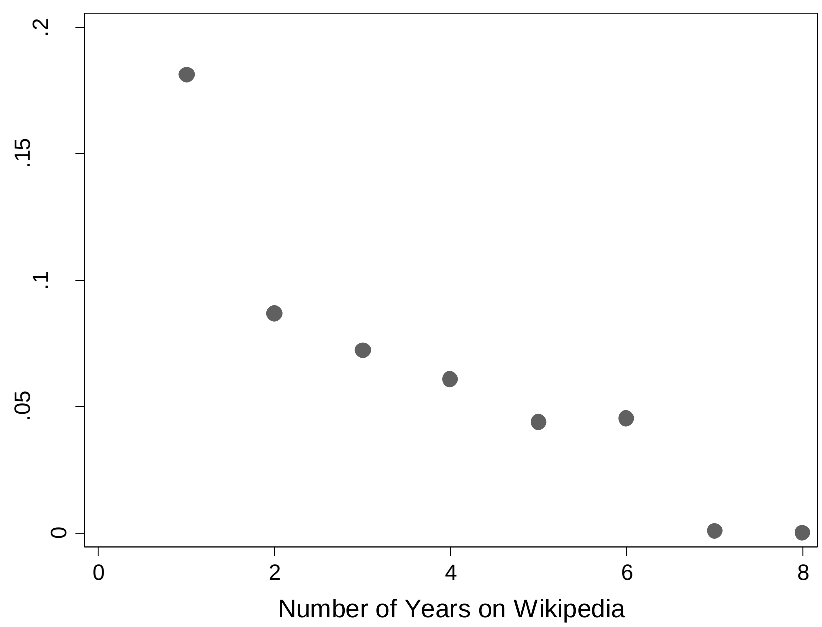

Figure 5: Distribution of Each Contributor's Total Number of Edits

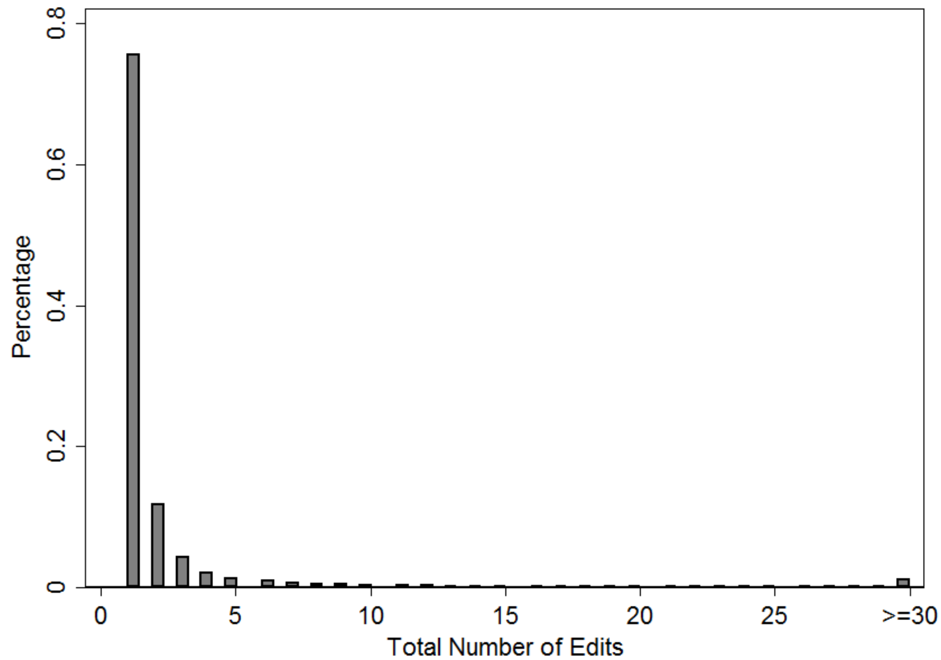

Figure 6: Distribution of All Edits in the Sample by Contributors' Years of Experience

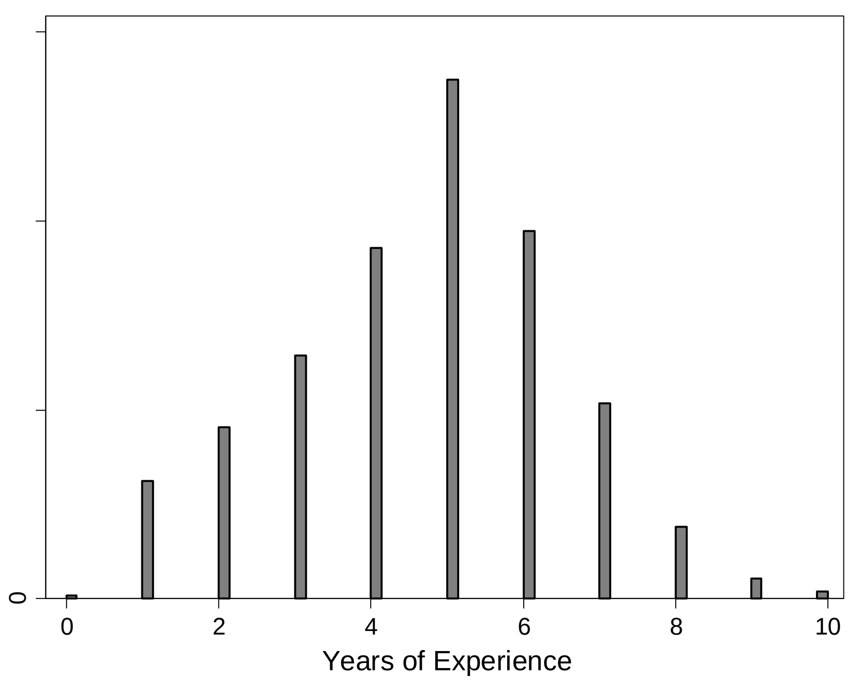

Figure 7: Number of Contributors by Years of Experience

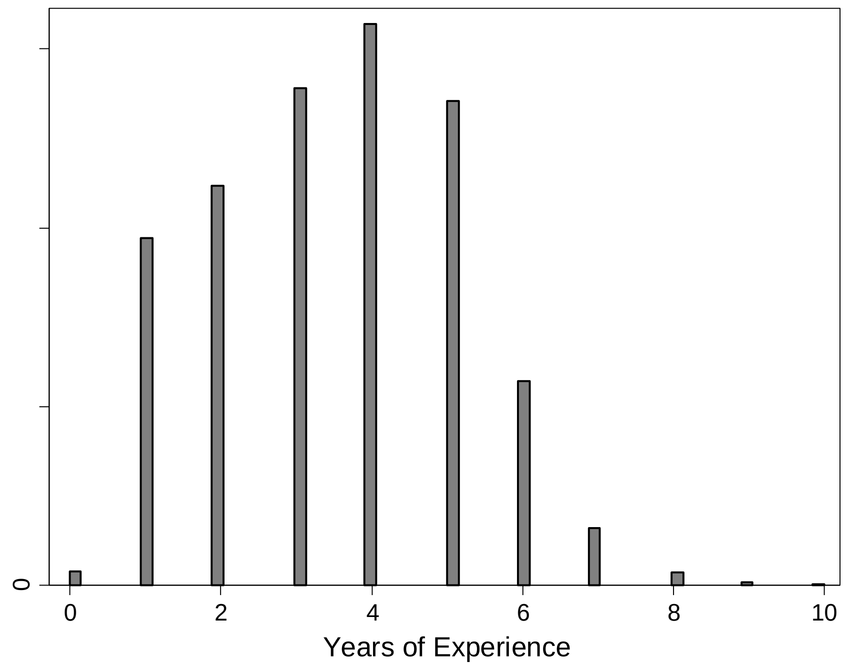

Figure 8: Distribution of Average Number of Edits per Contributor by Years of Experience

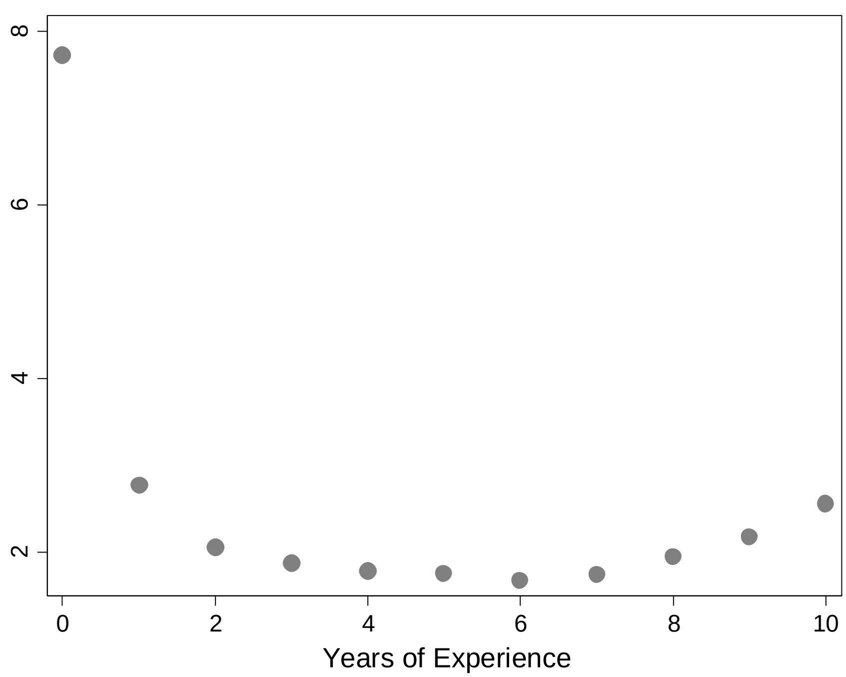

Figure 9: Transition Matrix of Contributor Slant Change in Wikipedia

		First half of activity		
		Democratic Type	Neutral	Republican Type
Second half of activity	Democratic Type	0.1407	0.0328	0.1145
	Neutral	0.7451	0.9333	0.7416
	Republican Type	0.1142	0.0339	0.1439

Notes: The sample is constructed by dividing every contributor's time in half. Then divide the direction of his or her edits, i.e. attach values (-1, 0, 1) to negative, 0, positive slant edits. Sum up the edits' values for the first half and the second half of his or her activity. If the sum of all edits in this half is negative, the contributor is a Democrat Type in this half. If the sum of all edits in this half is zero, the contributor is Neutral in this half. If the sum of all edits in this half is positive, the contributor is Republican Type in this half.

Figure 10: Transition Matrix of Contributor Slant Change over Time

			Start Slant						
			bin1	bin2	bin3	bin4	bin5	bin6	bin7
			[-1.229, -0.059)	[-0.059, -0.035)	[-0.035, -0.012)	[-0.012, 0.012)	[0.012, 0.035)	[0.035, 0.059)	[0.059, 1.000)
End Slant	bin1	[-1.229, -0.059)	0.8298	0.0139	0.0024	0.0011	0.0013	0.0008	0.0015
	bin2	[-0.059, -0.035)	0.0717	0.7242	0.0044	0.0020	0.0103	0.0019	0.0007
	bin3	[-0.035, -0.012)	0.0591	0.1745	0.7438	0.0055	0.0040	0.0149	0.0029
	bin4	[-0.012, 0.012)	0.0323	0.0713	0.2286	0.9795	0.2089	0.0531	0.0277
	bin5	[0.012, 0.035)	0.0036	0.0128	0.0177	0.0060	0.7545	0.1867	0.0624
	bin6	[0.035, 0.059)	0.0008	0.0014	0.0015	0.0033	0.0052	0.7222	0.0757
	bin7	[0.059, 1.000)	0.0028	0.0019	0.0018	0.0025	0.0158	0.0203	0.8291

Note: *Contributor yearly slant* is split by the ± 0.5, ± 1.5, and ± 2.5 standard deviations intervals. The middle bin represents neutral slant; the first/last bin represents extreme slant.