The Wisdom of Polarized Crowds

Feng Shi[1,4,+], Misha Teplitskiy[2,4,+,*], Eamon Duede[3,4] & James A. Evans[4,5,6,*]

[1]*Odum Institute for Research in Social Science, University of North Carolina at Chapel Hill, 208 Raleigh St., Chapel Hill, NC 27514*

[2]*Laboratory for Innovation Science, Harvard University, 175 N. Harvard Street, Suite 1350, Boston, MA 02134*

[3]*Committee on the Conceptual and Historical Studies of Science, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637*

[4]*Knowledge Lab, University of Chicago, 5735 S Ellis Ave, Chicago, IL 60637*

[5]*Department of Sociology, University of Chicago, 1126 E. 59th Street, Chicago IL 60637*

[6]*Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501*

[*]*Correspondence authors: mteplitskiy@fas.harvard.edu (MT); jevans@uchicago.edu (JAE)*

[+]*FS and MT contributed equally to this work*

As political polarization in the United States continues to rise [1–3], the question of whether polarized individuals can fruitfully cooperate becomes pressing. Although diverse perspectives typically lead to superior team performance on complex tasks [4,5], strong *political* perspectives have been associated with conflict, misinformation and a reluctance to engage with people and ideas beyond one's echo chamber [6–8]. Here, we explore the effect of ideological composition on team performance through analyzing millions of edits to Wikipedia's Political, Social Issues, and Science articles. We measure editors' online ideological preferences

by how much they contribute to conservative versus liberal articles. Editor surveys suggest that online contributions associate with offline (a) political party affiliation and (b) ideological self-identity. Our analysis reveals that polarized teams consisting of a balanced set of ideologically diverse editors produce articles of higher quality than homogeneous teams. The effect appears most strongly in Wikipedia's Political articles, but also in Social Issues and even Science articles. Analysis of article "talk pages" reveals that ideologically polarized teams engage in longer, more constructive, competitive, and substantively focused but linguistically diverse debates than ideological moderates. More intense use of Wikipedia policies by ideologically diverse teams suggests institutional design principles to help unleash the power of polarization.

Recent political events, including the 2016 presidential election, have underscored growing political divisions in American society. Political speech has become markedly more polarized in recent years [1], tracing a growing divergence between platforms of the major political parties [2] and leading to a state of political hyper-partisanship [3]. Yet the effects of political difference are not confined to the domain of politics alone. A growing literature documents how individual political alignments shape personal consumption of ostensibly non-political products, news, and cultural and scientific information [9–13]. This literature has converged on an alarming narrative: despite early promise of the world-wide-web to democratize access to diverse information [14], increased media choice and social networking platforms have led to the converse. Collaborative filtering allows individuals to passively enter "echo chambers" that limit the variety of information they observe and trust [15–17]. These can degrade the quality of individual decisions, including those that undergird basic democratic institutions [6–8]. Psychological mechanisms such as motivated reasoning

[18,19] and a tendency to discount identity-incongruent opinions [7,20] stimulate and reinforce polarizing information. Opposing social identities can foment conflict and even make communication counter-productive [21].

Nevertheless, a large literature documents the largely positive effect that social differences can exert on the collaborative production of information, goods, and services [4,5]. Research demonstrates that individuals from socially distinct groups embody diverse cognitive resources and perspectives that, when cooperatively combined in complex or creative tasks produce ideas, solutions, and designs that outperform those from homogeneous groups [22–25]. Collaborations between inventors from distinct social groups result in more creative patents [26], scientific teams representing distinct disciplines produce more highly cited papers [27], and gender diversity broadens the questions scientists ask [29].

The effect of political diversity on the collective production of knowledge, however, remains unclear. Insights from cognitive diversity research suggest that political diversity, like other forms of diversity, should positively impact the quality of group production because different perspectives, in the words of John Stuart Mill, "instead of being one true and the other false, share the truth between them" [30]. Literature on echo chambers, by contrast, suggests that political diversity may hamper productive cooperation as partisans perceive information held by opponents as not simply different, but wrong. In short, political diversity should increase access to fresh perspectives and information but may also undermine the quality of discourse and engagement required to enjoy the performance benefits typically obtained by diverse groups.

In order to assess the effect of political diversity on team performance, we studied the effect

of political polarization on the performance of approximately four hundred thousand online teams. Specifically, we focused on teams or communities of Wikipedia editors who worked on English-language articles in three large domains: Politics, Social Issues, and Science. We note that whether online editors collaborating on a Wikipedia article should be considered a *team* or *community* is an empirical question relating to the temporal and topical proximity of editor discussion and contribution. For clarity, we use the term *team* to refer to all such editing collectives, and *community* to characterize Wikipedia editors as a whole.

We observe a wide distribution of ideological alignments for Wikipedia editors (Figure 1 A). The peak at the center of the distribution comports with our observation that a large number of people only contributed minor edits to Wikipedia, such as correcting a typo. There are also two lower but significant peaks at the tails of the distribution, which identify editors who contribute substantial content to either liberal or conservative articles. The relative equality of these two divided populations suggests substantial polarization on Wikipedia. The variance of alignments across all editors of political articles is 0.04. This observed variance was significantly higher than those obtained in random simulations that assume that editors allocate their effort at random with ideological alignments of 0 (the bootstrap p-value for the observed variance 0.04 - the empirical probability of observing a variance larger than 0.04 in 100 random simulations - is less than .001; see Supplementary Methods 3 for details on simulations). We then measure the *polarization* of any given group of editors by the variance of their alignment scores.

As the number of editors for an article increases, their average political alignment decreases (Figure 2). This phenomenon is sometimes referred to as Linus' law – "with enough eyeballs, all

bugs are shallow"[28]. Articles attracting more attention tend to have more balanced engagement from editors along the conservative-liberal spectrum. This finding replicates those reported by Greenstein and Zhu [31,32] in their studies of bias in Wikipedia's US political coverage, showing that increased editor interaction reduced individual biases and yielded greater content neutrality. But this pattern raises a concern regarding the ideological alignment measure: if a *popular* political article, whether liberal or conservative, attracts more balanced engagement from liberal and conservative editors, then it should contribute less information to our inferred ideological alignment for those editors. We checked the robustness of our results after relaxing our dichotomous classification of Politics pages (see Supplementary Methods 2). Specifically, we adapted the Bayesian framework described above to update the alignment of liberal and conservative pages as a function of the balance of liberal and conservative editors who edit them. We recursively calculated the alignment of editors based on this updated alignment of pages and found that this measure produced results qualitatively the same as those shown below for our simple dichotomous measure (see Supplementary Table 3).

The 6-category quality scale for Wikipedia articles ranges from "Featured article" (highest quality) to "Stub" (lowest quality). Figure 1 B plots the relationship between average team polarization (i.e., variance of alignments) and quality for Political, Social Issues, and Science articles.

In all three corpora – Political, Social Issues, and Science articles – higher polarization is associated with higher quality. To establish this relationship statistically, we estimated an ordinal logistic regression model at the article level with article quality as outcome and polarization as main independent variable. We controlled for the absolute value of average team alignment and article

and editor features that may plausibly confound the relationship between polarization and quality. These article features include length, number of edits, and number of editors for each article. We specifically controlled for editors' editing experience by adding the average number of previous edits for each team.

As expected, number of edits, length of article, and number of editors significantly predict article quality (see Table 1 for regression results). The coefficient for the $|alignment|$ term suggests that quality decreases when editors are biased, on average, in either direction. Most critical is that polarization, the variance of political alignments, is positively and substantially associated with quality: a 1-unit increase in polarization multiplies the odds of moving from lower- to higher-quality categories by a factor of 18.57 for Political articles, 2.06 for Social Issues articles and 1.90 for Science articles.

To explore mechanisms by which ideologically polarized teams outperform homogeneous teams, we examine Wikipedia 'talk pages'. Each Wikipedia article has an associated talk page where 'backstage' knowledge assemblage occurs. Here, editors debate proposed additions and deletions, identify shortcomings, and attempt to persuade their fellow editors regarding content for the public facing, 'frontstage', Wikipedia article [33]. Using text from these talk pages, we examine relationships between political polarization and the following aspects of debate: (1) debate intensity, (2) information diversity, and (3) use of Wikipedia institutions—policies and guidelines—to discipline discussion. We investigate pairwise correlations between polarization and these debate mechanisms, then we estimate regression models to test the effect of polarization on these mechanisms separately, and finally, assemble them into a structural equation model that allows us to

identify their relative influence on article quality. All statistical analyses yield consistent results regarding mechanisms of collaboration, as discussed below and detailed in the Supplementary Methods 6 and 7.

Studies of team diversity and performance argue that the diversity of information individuals hold is the key driver of superior performance [5]. Nevertheless, information diversity is almost never measured directly, particularly in non-laboratory settings. The concept of information diversity confounds the substance and form of information, which are independently critical in ideological conflicts where interlocutors holding different beliefs frame the same issues in distinct language with distinct meanings. To address this challenge, we decompose information diversity into two measurable dimensions: *lexical* and *semantic* diversity. *Semantic diversity* captures distinct issues discussed on a talk page, while *lexical diversity* captures the number of ways in which editors discuss those issues. Because political polarization is predictably related to political content, we expect that political polarization among Wikipedia editors will focus debate on the subset of politically relevant topics within their article, but frame those issues in diverse and conflicting ways. If we are correct, greater political polarization among editors will yield lower semantic diversity and higher lexical diversity on their talk pages. We measure the lexical diversity of each talk page as a function of its distinct and distinguishing words. We measure the semantic diversity of a page as a function of the dispersion of words on that page in a latent semantic space defined by all Wikipedia articles, such that higher semantic diversity indicates more Wikipedia topics were debated. (See Supplementary Methods 4 for details on the two diversity measures.) We find that high political polarization narrows debate by reducing talk page semantic diversity but generates alternative framings traced by greater lexical diversity, as illustrated in Figure 3.

Diverse information should be more difficult to integrate, particularly if contested. We measure two core aspects of debate intensity including *volume* and *temperature*. Following previous research that found talk page length associated with article quality [34,35], we measure debate volume as a function of talk page length and distinct edits. Polarized teams may attempt to integrate more diverse information, requiring more talk, which yields greater article quality. Integrating diverse perspectives on contested and value-laden topics could be acrimonious, but a balance of liberals and conservatives could lower the *temperature* of potentially volatile collaborations, following research that links competitive imbalance to emotional aggression and violence [36]. We measure debate temperature using the `Detox` tool, developed by Wikimedia to identify harassment in the Wiki community. `Detox` detects toxic comments using a sophisticated machine learning classifier [37], which we apply to all talk page edits. We find that polarized teams generate a larger volume of debate (polarization's coefficient $\beta = 0.37, CI = [0.33, 0.40], z(86) = 20.20, p < .001$, 2-sided z-test, in the structural equation model discussed below and in Supplementary Methods 7) and their balance of political perspectives reduces flare-ups in debate temperature ($|alignment|$'s coefficient $\beta = 0.08, CI = [0.04, 0.13], t(205737) = 3.67, p < .001$, 2-sided t-test, in the linear regression model that predicts debate temperature using $|alignment|$, polarization, numbers of editor and edits, and page lengths).

Finally, we explore the self-governance of contested knowledge through use of Wikipedia policies and guidelines. Policies and guidelines are invoked so frequently that they have a standard nomenclature (https://en.wikipedia.org/wiki/Wikipedia:Shortcut_directory). For example, an editor who believes that part of an article is biased may invoke "NPOV" (the "Neutral Point of View" policy) in the article's talk page. The "NPOV" policy requires claims made on Wikipedia

to be free from editorial bias and inclusive of all significant views reported in reliable sources. Wikipedia also relies on a collection of less binding guidelines that refer to desired qualities of Wikipedia pages and the editorial process. These include that articles should cite sources ("CITE") and avoid and/or disclose any conflicts of interest ("COI"). We expect editors within polarized teams to encounter differences not easily resolved and, when debate fails, to discipline or challenge collaborators by invoking Wikipedia's policies and guidelines. Indeed, the numbers of policy and guideline mentions are found to increase with polarization. When disaggregated, we find that "NPOV" (Neutral point of view) and "OR" or "NOR" (No original research) are the most frequently cited policies, and each significantly correlates with polarization (for "NPOV", Pearson $\rho = 0.10, t(4671) = 6.87, p < .001, CI = [0.072, 0.128]$, and for "NOR," Pearson $\rho = 0.08, t(6498) = 6.47, p < .001, CI = [0.056, 0.104]$, both for 2-sided *t*-tests).

Correlations between all modeled variables are presented in Supplementary Figure 3 and are consistent with the regressions and structural equation model described below (also see Supplementary Methods 6 and 7). We also note interesting associations *between* talk page measures, suggesting micro-mechanisms of conflict and coordination, such as the negative correlation between debate temperature and talk page length (Pearson ρ=-0.08, $n = 205744$, $p < .001$ in 2-sided *t*-tests). This is relevant to the growing literature about online "trolling" behavior [38,39], suggesting that interactional toxicity is associated with foreshortened debate and a decreased collective capacity to construct quality Wikipedia pages.

We present results from a structural equation model in Figure 4, which allowed us to evaluate the combined impact of political polarization on article quality through mechanisms of col-

laboration. We summarize the model estimates here. (See Supplementary Methods 7 for additional details.) The model is fitted to 205,744 pages. Compared with ideologically homogeneous or skewed teams, polarized teams (1) debate *fewer* topics ($\beta = -0.65, z(86) = -37.28, p < .001, CI = [-0.68, -0.61]$, 2-sided z-test), (2) with *more* competing terminology and framings ($\beta = 0.20, z(86) =, p < .001, CI = [0.17, 0.23]$, 2-sided z-test). They engage in (3) *more* debate ($\beta = 0.37, z(86) = 20.20, p < .001, CI = [0.33, 0.40]$, 2-sided z-test), which is (4) *less* acrimonious ($\beta = -0.60, z(86) = -18.28, p < .001, CI = [-0.656, -0.53]$, 2-sided z-test). They also (5) more frequently appeal to Wikipedia policies and guidelines to govern these interactions ($\beta = 0.20, z(86) = 8.20, p < .001, CI = [0.14, 0.24]$, 2-sided z-test).

Lastly, to assess whether editors' ideological preferences measured on Wikipedia correspond to their ideological preferences offline, we fielded two surveys, focusing on editors' political party affiliation and ideological self-identity (see Supplementary Discussion 1 for details). In both surveys, approximately 45% of the contacted editors reported living in the U.S., and we report results from their answers here. Our computational ideological alignment measure correlates moderately well, and significantly predicts self-reported party affiliation on the democrat - republican spectrum (Pearson $\rho = 0.35, p = 0.035, n = 28$; one-tailed permutation test, performed because the correlation is predicted to be positive; AUC=0.71). Associations between computational and self-reported ideology (on the liberal - conservative scale) are slightly weaker, with a Pearson correlation of $\rho = 0.25$ ($p = 0.06, n = 41$ in a one-tailed permutation test) and a prediction AUC of 0.65. The lower associations with self-identified ideology are not surprising, considering that self-identity may be constructed in reference to one's peers rather than being anchored along a global spectrum. Moreover, the ideology survey was fielded after the inauguration of Donald Trump as President,

which may have influenced individuals' political self-identities or willingness to report them in surveys. Finally, responses to this second survey were highly imbalanced, with 7 individuals self-identifying as "extremely liberal" and 0 self-identifying as "extremely conservative."

Taken together, the survey results provide some evidence that editing preferences measured online are correlated with, and predictive of, party affiliations and ideological preferences offline. Further research, including lab and field experiments, is necessary to establish these suggestive relationships with causal certainty.

Mechanisms of polarized collaboration are echoed by editors in their survey responses. One third of respondents indicated awareness of politically motivated conflicts, and two thirds of those described them in detail. Conflicts typically entailed the encounter of biased content (e.g. "The page read like anti-Russian propaganda"), or having one's own content revised by editors perceived as biased (e.g. "My neutral edits regarding a particular political group were moved lower in the article to show negative opinions of this group first"). Many such conflicts were resolved through debate. One respondent recalled a conflict over the meaning of the word "refugee", which was resolved "by legal arguments that would convince an impartial observer." Another related an intense conflict on a page about homosexuality, but admitted that as a result "the article is in a better state." Other conflicts were resolved through administrator intervention. One respondent reported editing a page about a far-right politician that other editors would repeatedly vandalize; administrators intervened and protected the page from further edits. Unbalanced political competition, however, where lone editors sought to de-bias articles maintained by politically like-minded communities (e.g., with a perceived "right wing slant" or "anti-Russian bias") led to more acrimo-

nious conflict that often resulted in editor bans. Editing contested topics required toughness and endurance, which was ameliorated by balanced conflict. It is precisely these engagements that are missing from segregated "echo chamber" platforms, and which channel Wikipedia editors' diverse perspectives into articles of superior quality.

This study provides empirical, real-world evidence that ideological polarization can lead to productive, high-quality collaboration. Wikipedia teams comprised by a balance of ideologically polarized individuals perform better than groups comprised of political partisans and even moderates. Positive effects from polarization are observed in *Political*, *Social Issues*, and even *Science* articles. The intensified effect of ideological polarization on pages with greater political content suggests that diversity is not universally beneficial, but assists when directly or indirectly relevant to the topics considered. We demonstrate how frequent, intense disagreement within ideologically polarized teams foments focused debate [40] and, as consequence, higher quality edits that are more robust and comprehensive.

While our study revealed a statistically significant relationship between polarization and collaborative outcomes, we note several limitations of our analysis. First, our passive, computational assessment of editors' political alignments, on which our focal polarization measure relies, is fundamentally indirect. It captures political *interest* through engaged editorial participation, which modestly predicts and correlates with both political party affiliation and self-described ideology. A direct measure of political ideology or one containing greater signal of ideology and/or affiliation, however, would tighten our inferences linking polarization to productive collaboration.

Second, although we have controlled statistically for many factors known to influence article

quality in our models, confounding factors that we did not consider may remain. For example, most contentious articles have been protected at some point and can only be edited by administrators or senior editors during protection. It is possible that the effect of polarization we observed is driven by this protection mechanism. We ruled out this possibility by controlling for protection status in our models (see Supplementary Discussion 2 for details), but it is beyond our capacity to exhaust all possible factors. We anticipate this to be an important area for future research.

Third, the observational nature of this study places constraints on interpreting the relationship between ideological polarization and quality as a causal one. We observed only the behavior of those editors who voluntarily cooperated with others of contrary politics to produce articles of higher quality, or those who avoided such collaborations and produced lower quality articles. It is possible that these are different kinds of people, and so we cannot rule out the possibility that *randomly* assigned ideologically polarized teams may not outperform more homogeneous ones. Causal identification of this relationship between Wikipedia's design and productive collaboration will demand laboratory and field experiments that enable both randomization and control. Nevertheless, concerns of extreme self-selection on Wikipedia are allayed by its "encyclopedic monopoly". As the fifth most visited website in the world with more than 5 million articles on a wide range of topics, Wikipedia represents an effective monopoly of reference attention. Efforts have been made to produce politically skewed alternatives[41], but no viable substitutes exist. More importantly, Wikipedia contains only a single version of an article for a given topic. Consequently, if someone wishes to influence public knowledge on topics such as "Climate change" or "Free market" through Wikipedia, they must collaborate with existing editors who hold differing views but equal motivation. This is particularly salient for articles on contested topics, and frames a dramatic contrast with

segregated "echo chambers" in the blogosphere. Previous research on Wikipedia also suggests that cross-party collaboration is the norm rather than the exception [42].

Ideologically diverse collaborations are not without costs. One major obstacle to creating well functioning, diverse teams is that such teams produce outputs that may appear worse to the team members themselves [43]. Membership in homogeneous teams also *feels* better as participation affirms prior beliefs [44] and shelters contributors from aggressive interaction. Respondents to our survey echoed this sentiment by reporting pervasive displeasure in having to convince obstinate, competing partisans of points that they took to be self-evident. Balanced competition softened the emotional edge of ideological conflict, however, by allowing members to police tone and content with the omnipresent policies and norms of Wikipedia [45]. Use of these named/acronymed norms and policies was woven into the practice of Wikipedia conversation, signaled membership in the overarching Wikipedia project, and likely helped compensate for a reduced vocabulary of norm enforcement available in online, as opposed to face-to-face, interaction [46]. Unlike many online settings, when norms and policies break down, powerful moderators may step in and revert edits, lock pages and execute bans, but it is the success and not the failure of Wikipedia norms that predicts quality. Excluding the crowd by restricting pages to senior editor contributions had a strong negative association with quality content (see Supplementary Discussion 2 for details).

Previous research suggests that very high levels of diversity in teams may deteriorate the quality of teamwork. To explore whether political diversity has an upper bound beyond which polarization *hampers* performance, we re-estimated the regression models of quality with a quadratic polarization term. A negative coefficient on the quadratic polarization term would suggest that

very high polarization degrades article quality. We find that the coefficient is indeed negative ($\beta = -12.66, t(223151) = -11.64, p < .001$, in a 2-sided *t*-test). Nevertheless, the polarization level at which the association between polarization and quality becomes negative is very high, and not realized by 95% of the teams in this study. Furthermore, a regression of quality on polarization estimated on only the 5% most polarized teams shows no statistically significant pattern between polarization and quality ($\beta = 0.12, t(11152) = 0.23, p = .81$, in a 2-sided *t*-test). In sum, we do not find evidence that very high levels of political polarization hampers Wikipedia performance.

This study raises the possibility that in crowd-sourcing *contested* knowledge, the most motivated contributors are those with a biased perspective—an idiosyncratic take or angle on the disagreement at hand. Conducting debates on platforms like Wikipedia can demand high levels of motivation and patience. For example, the top editor of Hillary Clinton's Wikipedia page estimated spending 15 hours per week on protecting it from vandals (https://newrepublic.com/article/63288/wiki-woman. Last accessed 2018-12-18). Neutral users lacking partisan motivation may choose to allocate their time elsewhere. It is plausible that for voluntary crowd-sourcing platforms there exists an *optimal, non-zero* amount of user bias. Platforms that discourage all user bias may be inefficient or unsustainable.

Insofar as ideological diversity can improve the quality of politically relevant crowd-sourced knowledge, it is important to consider whether platforms should intervene to promote or even impose such diversity where missing. Our work suggests that for contested knowledge, platforms should seek not only high *numbers* of experts, but those with balanced, diverse perspectives to construct an environment through which motivated conflicts can be disciplined by enforceable policies

and guidelines. Just as institutional designs to promote gender diversity have proven valuable for fairness and performance in a variety of domains [47–49], designing for political diversity may become an increasingly important priority. Our study suggests that designing for political diversity may allow the digital age to grapple with John Stuart Mill's admonition that "not the violent conflict between parts of the truth, but the quiet suppression of half of it, is the formidable evil; there is always hope when people are forced to listen to both sides..."[30].

Methods

Overview. Using edit histories, we measured the online political preferences - *ideological alignment* - of 605,359 Wikipedia editors by the relative quantity of content they contributed to conservative versus liberal political articles. In addition, two surveys were conducted with n=500 and n=327 random samples of Wikipedia editors for whom we had calculated the index. We then used a machine learning algorithm developed by Wikimedia's internal researchers to measure the quality of Wikipedia articles [50] and related article quality to the political diversity of teams. Finally, we sought insight regarding mechanisms of collaboration among polarized teams by computationally exploring characteristics of article "talk pages" where the work of editing, debate and persuasion occurs.

Data collection. We extracted data from the complete English Wikipedia database dump on 12/01/2016. Data includes all edits made to all English Wikipedia articles since its start until 12/01/2016. Within this dump, we focused on three sets of articles: politics (20,947 articles), social issues (162,085 articles) and science (49,530 articles), which represent approximately 5%

of all English Wikipedia articles. Summary statistics of the three corpora may be found in the Supplementary Table 1. Users' total numbers of edits ever made to Wikipedia were collected through Wikipedia's online API (http://en.wikipedia.org/w/api.php).

The corpus of Political articles consists of two sub-corpora, Liberal and Conservative articles. The Liberal sub-corpus consists of all articles categorized under the "Liberalism in the United States" category (https://en.wikipedia.org/wiki/Category:Liberalism_in_the_United_States. Last accessed 2018-12-18.) and its subcategories. For instance, the article "New Deal coalition" is directly under the root category, while "The New Republic" is located under the sub-category "Liberalism in the United States $>$ Modern liberal American magazines". The Conservative sub-corpus was collected in a similar fashion starting with the " Conservatism in the United States" category (https://en.wikipedia.org/wiki/Category:Conservatism_in_the_United_States. Last accessed 2018-12-18). For instance, "Conservatism in the United States" links to "Economic liberalism," which links to "Market economy," and all three articles are in the "Conservative" sub-corpus. 406 articles (approximately 4%) appear in both corpora and were removed.

Titles of Social Issues articles were collected starting from the page "Category:Social issues" (https://en.wikipedia.org/wiki/Category:Social_issues. Last accessed 2018-12-18). We collected all articles and subcategories linked from the page; repeating this process in every subcategory of Social Issues, stopping 4 levels down from the root. Social Issues include articles relating to human welfare and justice, including "Homelessness," "Teenage pregnancy," and "Social services." These articles tend to be relatively controversial and politically salient. Titles of science articles were collected similarly, following the category structure of scientific disciplines in Wikipedia, starting from

the page "Category:Scientific disciplines" (https://en.wikipedia.org/wiki/Category:Scientific_disciplines. Last accessed 2018-12-18) and following the iterative procedure pursued for Social Issues articles.

Measurement of article quality. We measured the quality of Wikipedia articles algorithmically using a prominent approach that draws on features derived from article content alone and not information about editors or their collaboration patterns [51]. Wikipedia editors have scored hundreds of articles on quality, but human-generated ratings for most of Wikipedia's millions of articles do not exist and necessitate an algorithmic approach. In particular, we used the `wikiclass` algorithm, developed by Wikimedia research staff [50] and trained on Wikipedia articles scored by active editors for quality using a six-class scale, which ranges from "Featured Article" (highest quality) to "Stub" (lowest quality). The `wikiclass` algorithm predicts the correct quality class in 62.9% of cases and is off by at most one quality class in 90.7% of cases [50].

Measurement of ideological preferences. We measure editors' online ideological preferences – *alignments* – by the fraction of bytes they choose to contribute to "Conservative" (red) versus "Liberal" (blue) articles on the English-language Wikipedia, with a Bayesian framework to account for limited or random edits. The corpus of conservative articles consists of all articles categorized under "Conservatism in the United States," and similarly for "Liberalism in the United States." (See Data collection above.) This procedure scores editors as ideologically neutral (≈ 0) if they contribute equally to both sets of articles or little to either set, and closer to -1 or +1 the more exclusively they contribute to liberal or conservative articles, respectively.

Specifically, we model the total bytes an editor contributed to red articles (X) as a random variable satisfying a binomial distribution $X \sim Binomial(K, p)$, where K is the total number of

bytes contributed to political articles (red or blue) and p is the probability of contributing to red articles. This probability p represents our measure of *ideological alignment* for the editor, after rescaling it to the range -1 (most liberal) to +1 (most conservative). The parameter p is an unknown quantity to be estimated from observations X and K. We estimated it through a conservative, Bayesian framework described in Supplementary Methods 1.

The quantity of primary interest is the variance of alignments among a group of editors, which quantifies the spread of editors across the liberal-conservative spectrum. We used the variance in ideological alignments as a measure of polarization for any group of editors editing the same Wikipedia page. Previous research has found that this measure most directly captures the cognitive diversity of a group along a particular demographic dimension [52].

Surveys of editors. To assess the extent to which our measure of online ideological preferences corresponds to individuals' offline preferences and self-identities, we conducted two surveys on n=500 and n=327 random samples from the set of editors for whom we had estimated alignment scores. The first survey focused on political party affiliation, while the second focused on ideological self-identity. Both surveys allow us to explore the mechanisms of collaborations within polarized groups by asking editors about relevant editing experiences. We worked directly with the Wikipedia community and Wikimedia staff to carry out the surveys, including the development of a research page on the Wikimedia "Meta-Wiki" site and direct engagement with those expressing concerns therein (https://meta.wikimedia.org/wiki/Research:Wikipedia_%2B_Politics. Last accessed 2018-12-18). The arrived-upon process required a single member of our team (E.D.) to personally post the survey link on each of the randomly selected editors' pages along with an ex-

planation. The number of solicitations we could make per day (and the total number) was capped. In the end, we were able to post 500 solicitations for the first survey and received 118 responses. For the second survey, we posted 327 solicitations and received 100 responses. We did not record any demographic data on our respondents beyond whether their residence was inside or outside of the U.S. Participants were shown a consent script prior to any questions. All questions on the survey were optional. The surveys' methods were approved by the University of Chicago's Institutional Review Board (IRB17-0679). More information may be found in Supplementary Discussion 1.

Data availability. Data used in the study are available at https://github.com/KnowledgeLab/wisdom-of-polarized-crowds.

Code availability. Code used to gather, process and analyze the data is available at https://github.com/Knowl
of-polarized-crowds.

Acknowledgements

We acknowledge funding from the National Science Foundation SBE-1829366, John Templeton Foundation to the Metaknowledge Network, Air Force Office of Scientific Research FA9550-15-1-0162, and computation support from Cloud Kotta. We also acknowledge support from the Data@Carolina initiative, and thank undergraduate RA Han Guo for the help with some computations. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author Information

Feng Shi and Misha Teplitskiy contributed equally to this work.

Competing Interests

The authors declare no competing interests.

Contributions

All authors designed the research, interpreted the results and drafted the paper. FS, MT, ED gathered the data. FS designed the code and FE and MT analyzed the data.

References

1. Gentzkow, M., Shapiro, J. & Taddy, M. Measuring polarization in High-Dimensional data: Method and application to congressional speech. Tech. Rep., National Bureau of Economic Research, Cambridge, MA (2016).

2. Fiorina, M. P. & Abrams, S. J. Political polarization in the american public. *Annual Review of Political Science* **11**, 563–588 (2008).

3. Campbell, J. E. *Polarized: Making Sense of a Divided America* (Princeton University Press, 2016).

4. Joshi, A. & Roh, H. The role of context in work team diversity research: A Meta-Analytic review. *Acad. Manage. J.* **52**, 599–627 (2009).

5. Page, S. E. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton University Press, 2008).

6. Sunstein, C. *Republic : divided democracy in the age of social media* (Princeton University Press, Princeton, 2017).

7. Mutz, D. C. *Hearing the Other Side: Deliberative Versus Participatory Democracy* (Cambridge University Press, 2006).

8. Bishop, B. *The Big Sort: Why the Clustering of Like-minded America is Tearing Us Apart* (Houghton Mifflin Harcourt, 2009).

9. DellaPosta, D., Shi, Y. & Macy, M. Why do liberals drink lattes? *AJS* **120**, 1473–1511 (2015).

10. Gauchat, G. Politicization of science in the public sphere: A study of public trust in the united states, 1974 to 2010. *Am. Sociol. Rev.* **77**, 167–187 (2012).

11. Sarewitz, D. How science makes environmental controversies worse. *Environ. Sci. Policy* **7**, 385–403 (2004).

12. Zhao, X., Leiserowitz, A. A., Maibach, E. W. & Roser-Renouf, C. Attention to Science/Environment news positively predicts and attention to political news negatively predicts global warming risk perceptions and policy support. *J. Commun.* **61**, 713–731 (2011).

13. Shi, F., Shi, Y., Dokshin, F. A., Evans, J. A. & Macy, M. W. Millions of online book co-purchases reveal partisan differences in the consumption of science. *Nature Human Behaviour* **1**, 0079 (2017).

14. Benkler, Y. *The wealth of networks: How social production transforms markets and freedom* (Yale University Press, 2006).

15. Bakshy, E., Messing, S. & Adamic, L. A. Political science. exposure to ideologically diverse news and opinion on facebook. *Science* **348**, 1130–1132 (2015).

16. Pariser, E. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think* (Penguin, 2011).

17. Del Vicario, M. *et al.* The spreading of misinformation online. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 554–559 (2016).

18. Taber, C. S. & Lodge, M. Motivated skepticism in the evaluation of political beliefs. *Am. J. Pol. Sci.* **50**, 755–769 (2006).

19. Miller, J. M., Saunders, K. L. & Farhart, C. E. Conspiracy endorsement as motivated reasoning: The moderating roles of political knowledge and trust. *American Journal of Political Science* **60**, 824–844 (2016).

20. Kahan, D. M., Peters, E., Dawson, E. C. & Slovic, P. Motivated numeracy and enlightened self-government. *Behavioural Public Policy* **1**, 54–86 (2017).

21. Hart, P. S. & Nisbet, E. C. Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communic. Res.* **39**, 701–723 (2012).

22. Mannix, E. & Neale, M. A. What differences make a difference? the promise and reality of diverse teams in organizations. *Psychol. Sci. Public Interest* **6**, 31–55 (2005).

23. Hong, L. & Page, S. E. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16385–16389 (2004).

24. Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N. & Malone, T. W. Evidence for a collective intelligence factor in the performance of human groups. *Science* **330**, 686–688 (2010).

25. Nielsen, M. *Reinventing Discovery: The New Era of Networked Science*. Reinventing Discovery: The New Era of Networked Science (Princeton University Press, 2012). URL `https://books.google.com/books?id=afqfFW8WV9cC`.

26. Fleming, L., Mingo, S. & Chen, D. Collaborative brokerage, generative creativity, and creative success. *Adm. Sci. Q.* **52**, 443–475 (2007).

27. Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).

28. Raymond, E. The cathedral and the bazaar. *Knowledge, Technology & Policy* **12(3)**, (23–49 (1999).

29. Nielsen, M. W. *et al.* Opinion: Gender diversity leads to better science. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 1740–1742 (2017).

30. Mill, J. S. & Mehta, U. *On liberty* (Henry Regnery Company, 1955).

31. Greenstein, S. & Zhu, F. Is wikipedia biased? *Am. Econ. Rev.* **102**, 343–348 (2012).

32. Greenstein, S. & Zhu, F. Open content, linus' law, and neutral point of view. *Information Systems Research* **27**, 618–635 (2016).

33. Viegas, F., Wattenberg, M., Kriss, J. & Ham, F. Talk before you type: Coordination in wikipedia. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)* (2007).

34. Wilkinson, D. M. & Huberman, B. A. Cooperation and quality in wikipedia. In *Proceedings of the 2007 international symposium on Wikis - WikiSym '07* (2007).

35. Kittur, A. & Kraut, R. E. Harnessing the wisdom of crowds in wikipedia. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08* (2008).

36. Collins, R. *Violence: A micro-sociological theory* (Greenwood Publishing Group, 2009).

37. Wulczyn, E., Thain, N. & Dixon, L. Ex machina: Personal attacks seen at scale. In *Proceedings of the 26th International Conference on World Wide Web*, 1391–1399 (International World Wide Web Conferences Steering Committee, 2017).

38. Cheng, J., Danescu-Niculescu-Mizil, C. & Leskovec, J. How community feedback shapes user behavior. *CoRR* **abs/1405.1429** (2014). URL `http://arxiv.org/abs/1405.1429`. `1405.1429`.

39. Cheng, J., Danescu-Niculescu-Mizil, C. & Leskovec, J. Antisocial behavior in online discussion communities. *CoRR* **abs/1504.00680** (2015). URL `http://arxiv.org/abs/1504.00680`. `1504.00680`.

40. Mercier, H. The argumentative theory: Predictions and empirical evidence. *Trends Cogn. Sci.* **20**, 689–700 (2016).

41. Sobel Fitts, A. Welcome to the wikipedia of the alt-right. *Wired* (2017). URL https://www.wired.com/story/welcome-to-the-wikipedia-of-the-alt-right/.

42. Greenstein, S., Gu, Y. & Zhu, F. Ideological segregation among online collaborators: Evidence from wikipedians. Tech. Rep. (2016).

43. Phillips, K. W., Liljenquist, K. A. & Neale, M. A. Is the pain worth the gain? the advantages and liabilities of agreeing with socially distinct newcomers. *Pers. Soc. Psychol. Bull.* **35**, 336–350 (2009).

44. Hannah Nam, H., Jost, J. T. & Van Bavel, J. J. "not for all the tea in china!" political ideology and the avoidance of Dissonance-Arousing situations. *PLoS One* **8**, e59837 (2013).

45. Jemielniak, D. *Common knowledge?: An ethnography of Wikipedia* (Stanford University Press, 2014).

46. Sproull, L. & Kiesler, S. Computers, networks and work: Electronic interactions differ significantly from face-to-face exchanges. as a result, computer networks will profoundly affect the structure of organizations and the conduct of work. In *Readings in Human–Computer Interaction*, 755–761 (Elsevier, 1995).

47. Bohnet, I. *What works: Gender equality by design* (Harvard University Press, 2016).

48. Dahlerup, D. *The impact of gender quotas* (Oxford University Press, 2012).

49. Besley, T., Folke, O., Persson, T. & Rickne, J. Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. *American Economic Review* .

50. Halfaker, A. Interpolating quality dynamics in wikipedia and demonstrating the keilana effect. In *Proceedings of the 13th International Symposium on Open Collaboration - OpenSym '17* (2017).

51. Warncke-Wang, M., Cosley, D. & Riedl, J. Tell me more: an actionable quality model for wikipedia. In *Proceedings of the 9th International Symposium on Open Collaboration*, 8 (ACM, 2013).

52. Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L. & Briggs, A. L. Getting specific about demographic diversity variable and team performance relationships: A Meta-Analysis. *J. Manage.* **37**, 709–743 (2010).

Table 1: Odds ratios from ordinal logistic regression models predicting article quality. Statistical significance levels (p-values) are derived from 2-sided Wald tests. The columns present odds ratios estimated on Political, Social issues and Science articles, separately.

	Dependent variable: article quality					
Independent variable	*Politics*	p	*Social issues*	p	*Science*	p
polarization	18.88	$< .001$	2.06	$< .001$	1.79	$.006$
| alignment |	0.30	$< .001$	0.49	$< .001$	0.65	$.002$
editing experience	1.05	0.02	1.06	$< .001$	1.01	0.30
number of editors	0.41	$< .001$	0.51	$< .001$	0.56	$< .001$
article length	33.55	$< .001$	47.83	$< .001$	56.54	$< .001$
number of edits	3.26	$< .001$	1.71	$< .001$	1.69	$< .001$
N	12,570		161,070		49,995	

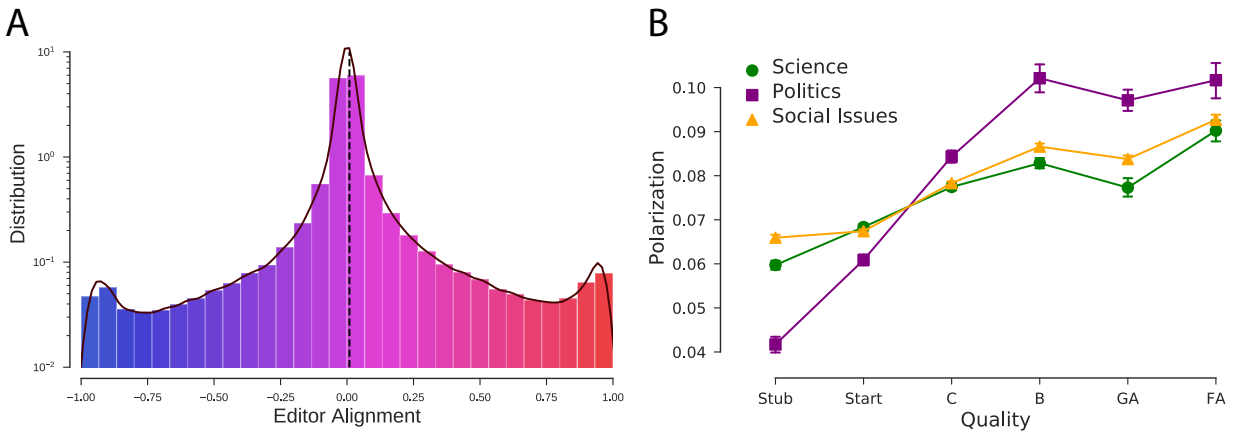

Figure 1: **A**. Distribution of editors' computationally measured ideological alignments, ranging from -1 (most liberal, where editors make intensive and exclusive contributions to liberal articles) to 1 (most conservative, where editors make intensive and exclusive contributions to conservative articles). **B**. Average polarization score of teams in each article quality level (Stub=lowest, FA=highest) for Politics (purple), Social Issues (orange) and Science (green) articles. Bands around each mean denote its 95% confidence interval. The actual numbers for the means and their CI's are included in Supplementary Table 3.

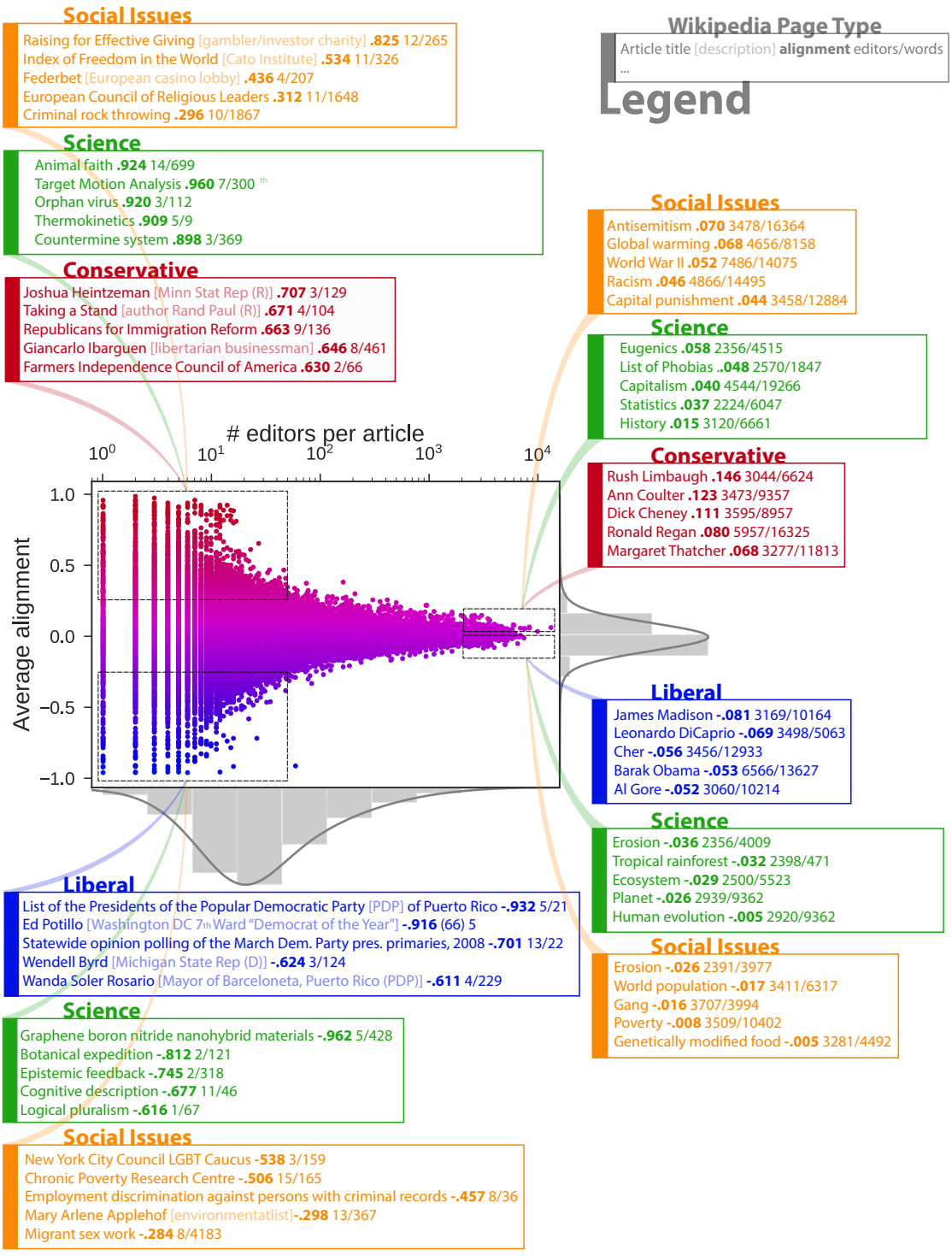

Figure 2: Scatter plot with each article's average editor alignment by number of editors. Average political alignment shrinks as number of editors increases, demonstrating the Linus Effect. Histograms on x and y axes reveal the density of articles at each level of editorial attention and average political alignment, respectively. Call-out boxes list five of the most "liberal" and "conservative" pages for articles receiving the most and least editorial attention, featuring article titles followed

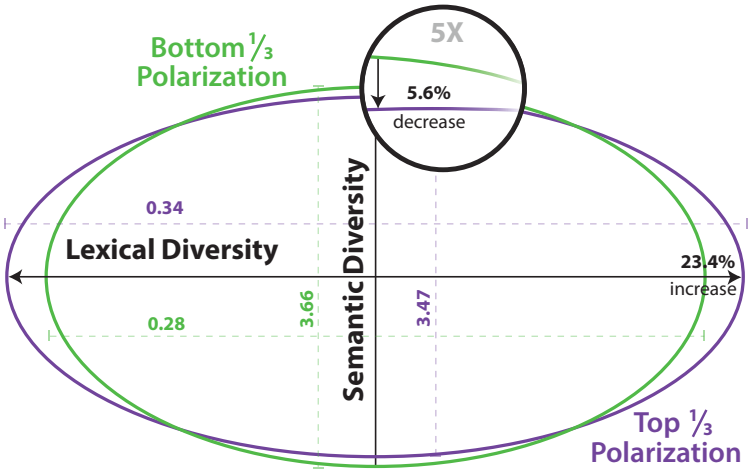

Figure 3: Illustration of the shift in "talk page" debate activity between teams in the bottom and top thirds of the political polarization distribution. Compared with the least polarized third of teams (green), the most polarized third (purple) semantically contract by 5.6% and lexically expand by 23.4%: they talk more about less, focusing on core politically-contested subjects, but framing them in distinctive ways.

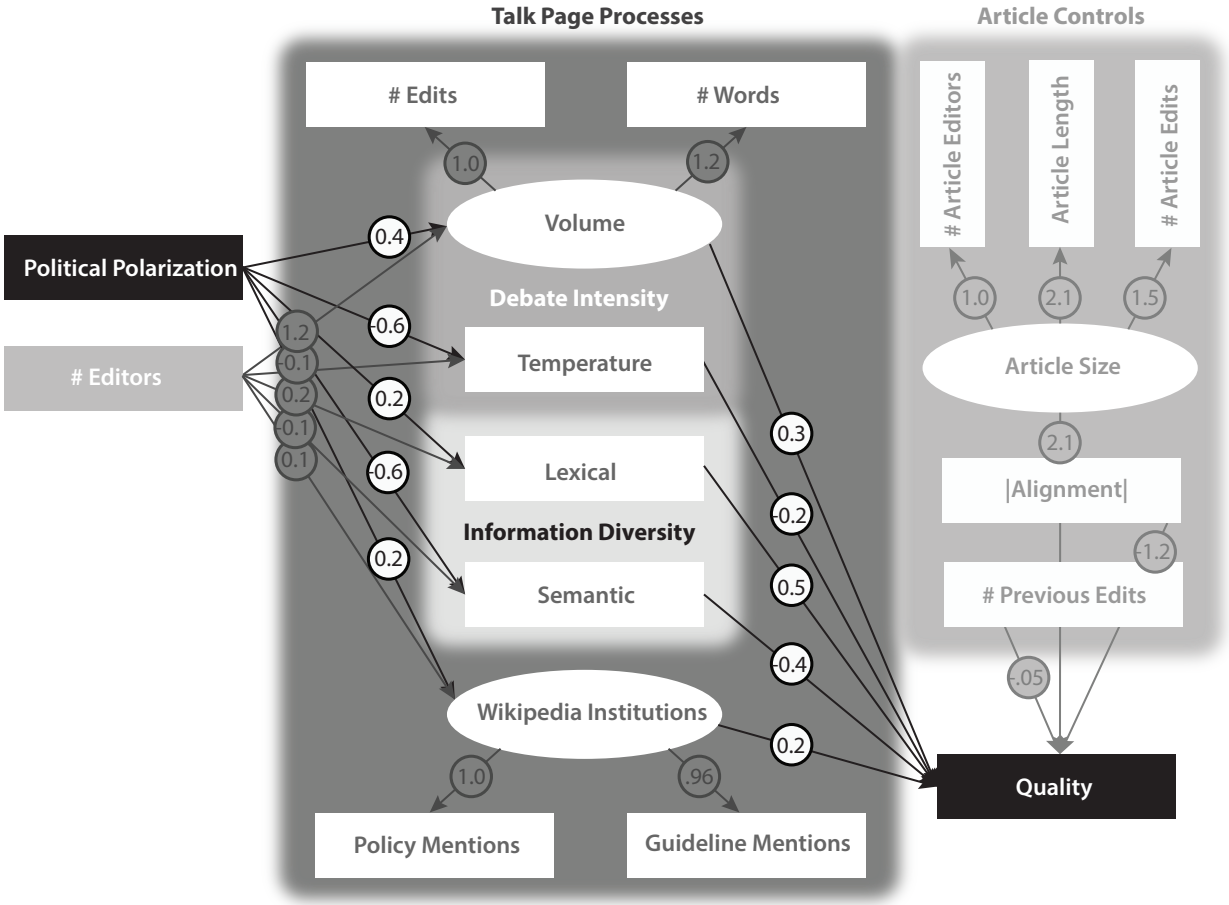

Figure 4: Estimated structural equation model linking *political polarization* (top left) with *article quality* (bottom right) through talk page debate intensity, information diversity and use of Wikipedia policies and guidelines. The model includes control variables associated with features of the articles themselves. Rectangles represent measured variables and ovals indicate latent variables. All coefficients are significant at the $p < .0001$ level, agreeing with individual models and bivariate correlations. See Supplementary Methods 7 for more details about the model and results.