# Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination

**Aniket Kittur** **Robert E. Kraut**
Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213
{nkittur, robert.kraut}@cs.cmu.edu

**ABSTRACT**
Wikipedia's success is often attributed to the large numbers of contributors who improve the accuracy, completeness and clarity of articles while reducing bias. However, because of the coordination needed to write an article collaboratively, adding contributors is costly. We examined how the number of editors in Wikipedia and the coordination methods they use affect article quality. We distinguish between explicit coordination, in which editors plan the article through communication, and implicit coordination, in which a subset of editors structure the work by doing the majority of it. Adding more editors to an article improved article quality only when they used appropriate coordination techniques and was harmful when they did not. Implicit coordination through concentrating the work was more helpful when many editors contributed, but explicit coordination through communication was not. Both types of coordination improved quality more when an article was in a formative stage. These results demonstrate the critical importance of coordination in effectively harnessing the "wisdom of the crowd" in online production environments.

**Author Keywords**
Wikipedia, wiki, coordination, social computing, collective intelligence, distributed cognition, collaboration.

**ACM Classification Keywords**
H.5.3 [Information Interfaces]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work, Web-based interaction, K.4.3 [Computers and Society]: Organizational Impacts – Computer-supported collaborative work

**INTRODUCTION**
Wikipedia, the online encyclopedia to which readers can contribute, is one of the most heralded success stories of peer collaboration. It consistently ranks in the top ten visited sites on the Internet according to Alexa.com. Since its inception in 2001 it has grown exponentially, and now comprises almost 2.5 million pages contributed by more than six million registered user accounts in the English Wikipedia alone. Studies have found that its content is of comparable quality to traditional encyclopedias [8], and that vandalism and inaccuracies are often reverted within a matter of minutes [19][36][27]. Its success has spurred the application of an open approach to knowledge building in a variety of domains, ranging from science (www.scholarpedia.org) to enterprise (www.socialtext.com).

Despite Wikipedia's success, we know little about why it has been so effective. One possibility is that having many contributors results in higher quality and less biased articles. The benefits of aggregating judgments from many people have been observed since at least 1907, when Galton showed that averaging independent judgments of many observers estimated the weight of an ox at a county fair better than experts could [7]. The Internet makes aggregating judgments much easier, leading to systems of collective intelligence ranging from markets for accurately predicting presidential elections [3] to systems where volunteers classify craters on Mars' surface, resulting in work virtually indistinguishable from that of expert geologists [17]. Most models of collective intelligence are premised on aggregating the independent contributions of many people, colloquially known as harnessing "the wisdom of crowds" [32].

However, many of the tasks involved in the collaborative editing of articles in Wikipedia violate assumptions of independent contribution and automatic aggregation. While some tasks, such proofreading an article or adding facts, could benefit from having many independent contributors, other tasks, such as planning the structure of an article or developing a cohesive point of view, require significant coordination and interaction among contributors.

In coordination-intensive tasks, increasing the number of contributors incurs process losses; that is, the effectiveness of the group will be lower than what the members could ideally produce [30]. Adding more heads often yields diminishing returns because of increased coordination


*CSCW* 08, November 8–12, 2008, San Diego, California, USA.

requirements or decreased member motivation, especially for complex tasks or those with strong interdependencies [11][29]. In the extreme, adding more people may have negative consequences, as Brooks Law states in the domain of software projects: "Adding manpower to a late software project makes it later" [5].

There is significant evidence that collaboratively writing articles in Wikipedia requires a high degree of coordination between users. Contributors to a page work to achieve consensus on a variety of issues, such as its structure and organization; what the article will include and what it won't; which points of view will be represented and to what degree; and wording and style. Coordination costs are a substantial and growing proportion of all work done in Wikipedia: nearly 40% of all edits in Wikipedia involve indirect work such as communication, consensus building, and development of policies and procedures [19]. Much of this work takes place on dedicated "talk pages" where changes are often discussed before being implemented in the article itself [37]. Thus coordination may play an important role in Wikipedia's success, separate from the influence of the number of contributors.

Empirical evidence about the effects of increasing numbers of contributors on the quality of Wikipedia content is scarce. The most definitive study to date, by Wilkinson & Huberman, has demonstrated that high-quality articles in Wikipedia ("featured" articles) have substantially more editors involved than do run-of-the mill, non-featured articles, even after controlling for article popularity [39]. However, the correlational nature of the analyses used leaves open the possibility of reverse causation (i.e., higher-quality articles may attract additional editors instead of large numbers of editors producing higher-quality articles). Furthermore, the coarse metric of quality used (featured vs. non-featured) means that few high-quality articles were sampled (fewer than 2,000 featured vs. more than 2,000,000 non-featured), and the ones that were sampled went through a stringent peer review process not representative of most articles in Wikipedia. These factors may limit the generality of the results.

The present research uses longitudinal data to examine the conditions under which adding contributors to a Wikipedia article improves its quality. It shows that the effectiveness of adding contributors is critically dependent on the degree and type of coordination those contributors use, as well as the life cycle of the article and the interdependence of the tasks involved in editing it.

## COORDINATION AND QUALITY

Coordination is essential to the effective functioning of groups, and is especially important as those groups grow in size. Each new editor working on an article in Wikipedia has the potential to contribute new knowledge with which to flesh out an article, insight into how the article should be written and vigilance to discover errors in fact, grammar or judgment. To harness these contributions, however, editors need to coordinate their efforts on interdependent aspects of the article, like its structure and stylistic coherence.

Theories of group coordination suggest a basic distinction between *explicit coordination*, based on direct communication and verbal planning, and *implicit coordination*, based on workgroup structure, unspoken expectations and shared mental models of the task to be accomplished [28][40]. Below we examine evidence of both types of coordination in Wikipedia, focusing on direct communication and workgroup structure.

### Explicit coordination through communication

Editors can explicitly coordinate with each other in a number of ways in Wikipedia. The most commonly used mechanism for explicit coordination is communication on the article talk page. This is a dedicated page associated with each article that provides a forum for negotiating the scope and structure of an article, coordinating changes, discussing policies and procedures, and eliciting assistance from other editors [37].

Figure 1 shows the topics from the talk page associated with Glock, the handgun manufacturer. Discussion topics range from a relatively superficial (yet heated) debate about the capitalization of the name ("GLOCK vs. Glock debate") to more substantive coordination, building consensus around the scope of the article (e.g., in "Misconceptions," whether certain myths and misconceptions should be included or excluded). There is often a complex interplay between the talk page and the article page, with editors proposing changes to be made to the article page, as well as discussing changes already made.

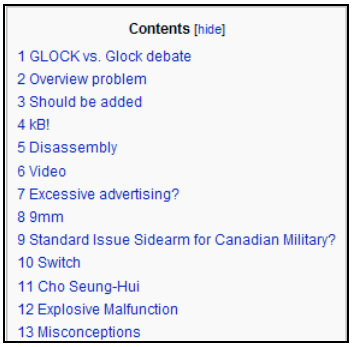
Contents [hide]
1 GLOCK vs. Glock debate
2 Overview problem
3 Should be added
4 kB!
5 Disassembly
6 Video
7 Excessive advertising?
8 9mm
9 Standard Issue Sidearm for Canadian Military?
10 Switch
11 Cho Seung-Hui
12 Explosive Malfunction
13 Misconceptions

**Figure 1. Topics on the talk page for the Glock article.**

To gain insight into the overall impact of coordination through communication we conducted an analysis of the proportion of talk page activity to article activity for all articles in Wikipedia. The "proportion of coordination activity" in Figure 2 refers to the ratio of talk-page to article edits for each week of an article's life[1]. This analysis shows

[1] These results are based on an analysis of all Wikipedia articles through 2006, including all editor types: anonymous and registered users, admins, and bots. Although it is possible that bot edits may be unrepresentative of coordination, bots have been shown to account for less than 5% of all editing activity [21].

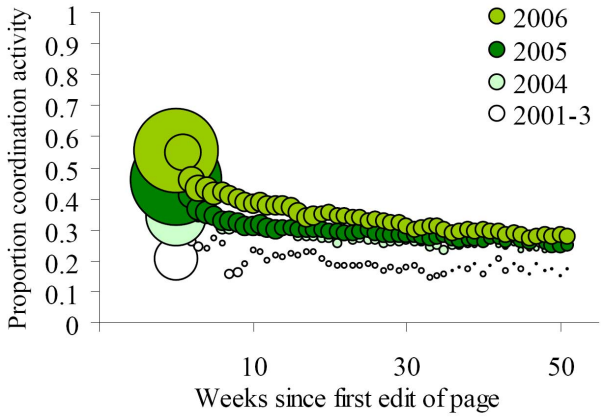


**Figure 2. Proportion of coordination to production activity for the first 50 weeks of an article's life across all Wikipedia articles. Colors correspond to different years; the size of each circle is proportional to the log of the number of edits.**

an increase in the importance of this type of coordination over the years as the number of contributors has grown. Furthermore, coordination through communication is especially important early in an article's life cycle: more than half of all edits in the first week of an article are made to the talk page rather than to the content of the article.

### Implicit coordination through structure

Editors can also coordinate implicitly by structuring their work in particular ways. In many conventional organizations, workgroup structure, such as the degree of managerial hierarchy, supervisors' span of control or division of labor, provides an important means of coordination, with powerful implication for its success (e.g., [23][34]). In the current research we examined the extent to which editing in an article is concentrated among a small subset of editors or spread equally among them.

While the concentration of contributions is a structural feature common to many online production environments, the degree of concentration varies with both the type of community and the type of production task. For example, the top 15 developers contributed 88% of new lines of code to the Apache server project, but only 66% of the bug fixes and 5% of the problem reports [26]. The present study examines how variations in concentration might influence the success of projects.

It is important to note that editing concentration is distinct from the number of editors involved in an article. To take an extreme example, an article with 100 editors and 1000 edits could be dominated by a single editor making 901 edits with the remaining 99 editors making a single edit each; alternatively, work could be evenly distributed, with each editor making 10 edits.

Having a core group of heavy contributors could improve coordination through several distinct mechanisms. First, needs for explicit coordination are reduced if most of the

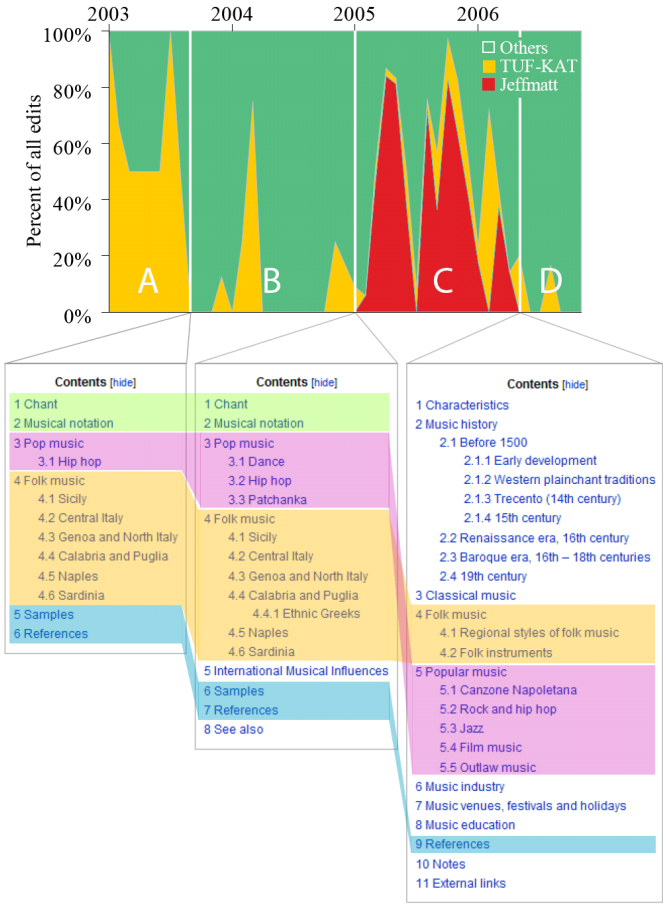


**Figure 3. User activity for the "Music of Italy" article. Callouts show the structure of the article's table of contents, with colors indicating the position of structural groups over time.**

interdependent work is done by a small group, since only members of this subgroup need to discuss issues. Second, this core group is likely to develop a shared mental model, whereby they can maintain a common view of the article and its structure without explicit communication. Finally, the core group could play a leadership role by setting the direction and scope of the article and providing a framework to which peripheral editors could effectively contribute

The "Music of Italy" article illustrates these different types of implicit coordination. . As shown in Figure 3, during the first few months of the article's life (period A), the user TUF-KAT does a large proportion of the work . Examing these contributions show that her first edits set the initial structure of the article and its scope and direction of the content, which she and others later refine . Her framework provides a scaffold around which others contribute. During period B others continue to add content to the article but the structure remains substantively constant, as shown in the similarity of the topic structure in the callouts before and after period B.

Figure 3 shows that the user Jeffmatt does a large proportion of all of the editing during period C. His contributions

involve a major reorganization and expansion of the article's structure and contents, which can be seen by the change in topic structure from the callouts before and after period C. These organizational changes were not made in a vacuum; after proposing and executing the changes Jeffmatt received explicit feedback from other members, including TUF-KAT. However, this explicit coordination was largely limited to a few highly involved contributors rather than the larger group editing the article. By concentrating work in a few individuals, the reorganization task could be accomplished while minimizing the overhead that would be required if a large number of editors were involved.

#### Interactions with age and group size

Coordination is likely to be especially valuable when articles are at a formative stage, because tasks involved are likely to be more interdependent then. Early contributors may act as leaders by implicitly setting the direction, scope, and structure of an article, creating a framework for subsequent editors to fill in effectively. Indeed, Wikipedia urges users who create new articles to provide "enough information for [other] editors to expand upon" [10]. Thus both explicit communication and concentration of editing may positively impact the article early on and when it is of low initial quality.

However, these forms of coordination may interact differentially with increases in group size. Explicit coordination is especially valuable in small groups engaged in complex tasks that require a high degree of synchronization between members [40]. As the size of the group grows, however, the amount of communication needed to effectively coordinate increases super-linearly, leading to process loss, inefficiency, and a reliance on pre-determined plans [30][35]. This suggests that increasing communication between editors may not be an efficient coordination mechanism as groups grow in size.

Conversely, coordination through editor concentration avoids the overhead of discussion. It can reduce coordination needs, lead to shared mental models or provide a framework into which contributions can be embedded, allowing a larger number of editors to aggregate their efforts towards a common goal more efficiently [28].

## QUALITY IN WIKIPEDIA

To assess article quality we took advantage of Wikipedia's article assessment project, which, in preparation for creating an off-line release of the encyclopedia, has organized the evaluation of over 900,000 articles into six gradations of quality, ranging from "stub" to "featured-article" status [15]. Concrete guidelines for assessing each class include factors such as how well-written, factually accurate, verifiable, comprehensive, neutral, and stable the article is [16]. For example, a stub is described as an article which is "very short … or a rough collection of information". A B-class article "usually [has] a *majority* of the material needed for a comprehensive article. Nonetheless, it has some gaps or missing elements or references, needs editing for language usage or clarity, balance of content, or contains other policy problems." An A-class article "provides a well-written, reasonably clear and complete description of the topic" with appropriate length, well-written introduction, good illustration, and information cited from reputable sources".

| Maximum quality | N of Articles |
|---|---|
| Not rated | 36,239 |
| Stub | 27,633 |
| Start | 51,983 |
| B-Class | 27,263 |
| GA-Class | 1,657 |
| A-Class | 1,132 |
| Featured | 1,453 |
| Total articles | 147,360 |

**Table 1. Distribution of articles by maximum quality.**

#### External validity of quality ratings

To check on the external validity of the community ratings we conducted a survey among non-Wikipedians. The guidelines on assessment grading in Wikipedia [13] provide an example page for each assessment quality level, including "Coffee table book," "Munich air disaster," "Real analysis," "International space station," "Durian," and "Agatha Christie: And then there were none." For each of these pages we used Amazon's Mechanical Turk market (mturk.com) to collect users' ratings of article quality. Articles were rated on a 7-point Likert scale ranging from "Very low quality" to "Very high quality." We employed the procedure detailed in [20] of asking verifiable questions (including the number of sections, images, and references) before asking for the subjective quality evaluation, which has been shown to promote valid responding.

We collected 20 ratings for each of the 6 community-assessed example articles. Responses were checked for validity of verifiable questions and sufficient time on task (>30 seconds). Only one response was excluded due to failing to satisfy the above requirements. Thirty-four users participated in the study, with all but one reporting that they were not expert users of Wikipedia. A Spearman rank correlation of article ratings from external raters with Wikipedia community assessments was positive and highly significant ($r(5) = .54$, $p < .001$), indicating substantial agreement between external and Wikipedia ratings.

## DATA

The two datasets used in the analyses were provided by the MediaWiki foundation. For editing information we used a dataset from October 2007 that included only the metadata of the revisions (e.g., who edited which page, which category a page is in, etc.) but not the full text. As the large size of the Wikipedia dataset has made full text dumps

difficult for the MediaWiki foundation to create, archive, and publish, a slightly older full text dump was used that included the full text of all Wikipedia revisions up to April 2007 (approximately 160 million).

We sampled articles in each quality level based on their status in the latest (October 2007) dataset. Articles that were simply lists of other articles were excluded, as these were unrepresentative of the typical Wikipedia article. For each of these articles we used data from the full-text dump for the final analysis, which enabled us to determine the quality level of an article over time, as described below. We split the data for each article into two six-month intervals, from April 2006 to April 2007. Table 1 shows the distribution of the maximum quality for articles in the sample.[2]

For each interval and for each article we computed the article's age in months since its first edit, the number of article editors who made at least one edit during the interval, explicit coordination, editing concentration, and the article's assessed quality at the beginning and end of the interval. Because age, the number of article editors, and the number of discussion edits all have highly skewed distributions, we transformed these measures by taking their log to the base 2 to make the distributions more normal.

Explicit coordination was operationalized as the number of edits made to the talk page of an article. To quantify implicit coordination through structure we used the gini coefficient to quantify the concentration editing across editors. This measure is commonly used in studies examining the inequality of distribution, such as the distribution of wealth in a country [2]. Articles with edits concentrated in a few individuals have gini coefficients closer to one, while articles with edits evenly distributed across all editors have coefficients closer to zero.

We computed the quality of an article at both the beginning and end of each interval. The primary way the Wikipedia community assesses article quality is by placing a template on the talk page of an article. This template also automatically places an article into the relevant quality class. To gather longitudinal information on changes in article quality over time, the date on which a given assessment template was added to an article is needed. Unfortunately, the Wikipedia database does not directly store the date on which a specific template was added to a specific article. To acquire this information the Hadoop distributed file system [9] was used on a grid computing cluster to analyze the full text of every revision for all articles in Wikipedia in parallel (approximately 1.6 TiB). The highest-level quality assessment template prior to the time period was used to determine the starting quality of the article, and the highest-level assessment template within the time period was used to determine the end quality. Table 1 shows the highest quality rating achieved by the 147,360 articles in the sample. Of these, only 23,619 had both beginning and ending quality assessments for any time period; these are the articles used for the primary analyses reported below.

[2] These data included all articles from the top three quality levels, and 75,000 from the lower-quality levels. Due to the difference in time between the two datasets used, some articles identified with a given quality level in the October dataset had not been rated as such by the end of the April dataset. Thus, although 75,000 articles were sampled for each quality level, some levels had fewer articles that reached that quality level by April 2007.

## APPROACH

We initially conducted a similar cross-sectional analysis as in [39] but using multiple levels of the quality assessment scale. Consistent with [39], the total number of editors who ever worked on an article was moderately correlated with the quality of the article ($r=.40$, $p < .001$), and this association remained when one controlled for the age of the article and the number of page views the article received (partial correlation = .22, $p < .001$)[3].

However, these cross-sectional correlations provide little insight into the influence of number of editors on article quality because of problems with reverse causation (high quality articles attracting more editors, rather than participation by more editors improving article quality) and uncontrolled confounding factors like article importance.

Using a longitudinal approach enables stronger causal claims than would be possible with cross-sectional correlations and avoids problems with reverse causation. By examining changes in the quality of the same articles over time, longitudinal analyses also control for time-invariant unobserved qualities of the articles, like their topic, its importance, and the availability of secondary sources.

Our goal is to examine how the number of contributors to an article and the differing coordination techniques they use result in *changes* in the quality of the article during a given time period. To do so, we use the lagged, multiple regression approach for modeling change recommended by Cohen et al. [6]. We predict changes in article quality between two time periods ($QualityT_n$-$QualityT_{n-1}$) from the number of editors involved and the coordination techniques they used during this interval, holding constant the article quality at the first time period ($QualityT_{n-1}$). This procedure removes the potential influence of the initial quality on the relationship between the predictors and changes in quality, controls for regression towards the mean, and controls for unobserved, stable article characteristics.

Because only a subset of Wikipedia articles have been evaluated for quality, and the same factors, such as the number of editors working on an article, can influence both whether it will be evaluated and its quality, it is necessary to control for selection biases when modeling changes in quality. To do so, we used the Heckman 2-step selection

[3] Pageview data was taken from the partial feed of the Wikipedia server logs described in [27].

| | | Descriptive Statistics | | | Correlations | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Median | Std | Quality change | Initial Quality | Article Age | # Article Editors | Editor Concen-tration | # Talk Editors |
| 1 | Quality change | .09 | .00 | .55 | | | | | | |
| 2 | Initial Quality | 2.36 | 2.00 | 1.14 | -.20 | | | | | |
| 3 | Article Age | 25.90 | 21.73 | 17.41 | .00 | .29 | | | | |
| 4 | # Article Editors | 48.31 | 11.00 | 108.73 | .08 | .43 | .51 | | | |
| 5 | Editor Concentration | .26 | .25 | .18 | .20 | .27 | .21 | .61 | | |
| 6 | # Talk Editors | 6.00 | 2.00 | 16.28 | .14 | .47 | .41 | .78 | .52 | |

**Table 2. Descriptive statistics before log transformation and correlations after log transformation. Quality ranges between 1 (Stub) and 6 (Featured Article). Editor concentration is measured by the gini coefficient, which ranges from 0 (equally distributed) to 1 (highly concentrated).**

model [10], which uses logistic regression to generate a derived variable predicting whether the article will receive at least one quality evaluation during the study interval. This likelihood of being evaluated is then used as a control variable in the second stage of the analysis, predicting changes in article quality.

When modeling an article's likelihood of being evaluated for quality, we included in the first stage of the Heckman analysis the cumulative number of edits the article had before the start of the observation period, along with the article's age, the number of editors working on the article during the period, the editor concentration, and the number of talk edits it received. A robustness check, in which we added the number of page views an article received, led to similar results.

Of the 23,619 articles in the sample, we could compute change scores across two time periods for 890 articles and a change scores across one time period for 22,729 articles. To deal with the non-independence in the data, with some articles having change scores in more than one time interval, we used a robust technique that adjusts the standard errors for intra-article correlation.

## RESULTS

Table 2 presents the means, medians and standard deviations of the variables used in the analysis (before the log transformation) and correlations among variables (after the log transformations) for the 23,619 articles that had both a beginning and end quality assessment in at least one interval. One interesting observation from this table is that during each 6-month period, many more contributors actually edit the article (median=11) than participate in discussions on the article's talk page (median=2). This observation suggests that even explicit coordination via discussion has an implicit structural component: planning is done by a small subset of contributors.

Table 3 presents results from the second stage of the Heckman regression, predicting changes in article quality from initial quality, article age, the number of editors working on an article and the coordination techniques they used. To reduce the multi-colinearity between main effect and interactions in regression analyses, we centered all independent variables [1]. Thus, the coefficients presented in Table 3 should be interpreted as the effect of a unit increase in an independent variable (i.e., increasing initial quality by one class, increasing concentration from completely even to completely concentrated, and doubling an article's age in

| | Model 1 # Editors | | | Model 2 Editor Concentration | | | Model 3 Talk Edits | | | Model 4 All | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | SE | P | Coef. | SE | P | Coef. | SE | P | Coef. | SE | P |
| Intercept | .442 | .025 | *** | .329 | .031 | *** | .362 | .028 | *** | .304 | .033 | *** |
| Initial Quality | -.149 | .005 | *** | -.138 | .005 | *** | -.151 | .005 | *** | -.146 | .005 | *** |
| Article Age | -.024 | .004 | *** | -.005 | .004 | | -.017 | .004 | *** | -.006 | .004 | |
| # Editors | .053 | .002 | *** | .003 | .003 | | -.005 | .003 | | -.020 | .003 | *** |
| Editor Concentration | | | | .791 | .041 | *** | | | | .600 | .038 | *** |
| Editors X Concentration | | | | .248 | .017 | *** | | | | .216 | .020 | *** |
| Quality X Concentration | | | | -.236 | .032 | *** | | | | -.222 | .035 | *** |
| Age X Concentration | | | | -.066 | .027 | * | | | | -.041 | .028 | |
| # Talk Edits | | | | | | | .113 | .005 | *** | .087 | .004 | *** |
| Editors X Talk | | | | | | | -.003 | .001 | * | -.010 | .002 | *** |
| Quality X Talk | | | | | | | -.012 | .003 | *** | -.001 | .003 | |
| Age X Talk | | | | | | | -.009 | .003 | ** | -.003 | .003 | |

**Table 3. Nested lagged regression analysis of the number of editors, coordination metrics (editor concentration and talk edits), and article life cycle on change in article quality.**

**Note: *** p<.001, ** p<.01, * p<.05**

months, the number of editors involved, and the number of edits on its talk page) when other variables are at their mean level.

Model 1 in Table 3 is the base model, predicting change in article quality over a 6-month period from the article's initial quality, its age, and the number of editors who worked on it during that period. On average, articles improved .44 quality classes over a six-month period when all other variables were at their mean level. An article's initial quality was inversely associated with changes in its quality (i.e., initially poor articles got better and high quality articles got worse). This is a result of both floor and ceiling effects (stubs can't get worse and featured articles can't get better), higher standards in Wikipedia over time, and regression towards the mean. In addition, a project's age was associated with a small decline in its quality, again reflecting more stringent standards in Wikipedia over time.

More interestingly, the number of editors working on an article during a 6-month period was positively associated with increases in article quality during that period, even holding constant article age and initial quality. Each doubling in the number of editors was associated with a increase in .05 quality classes.

Model 2 adds implicit coordination (i.e., editor concentration) and its interactions with the number of editors, article age and initial article quality to the model. When concentration is added to the model, the influence of the number of editors no longer predicts increases in quality, while concentration does. Articles in which the work is highly concentrated in a small subset of editors improve in quality much more than do articles where the work is evenly divided.

Strikingly, there was a significant positive interaction between the degree of concentration and the number of editors. As shown in Figure 4, increasing the number of editors improved quality when work was highly concentrated (i.e., above the median), but was detrimental when work was more evenly distributed (i.e., below the median).

In addition, Model 2 also shows negative interactions between concentration and the age of an article (Figure 5) and its initial quality (Figure 6). Both interactions indicate

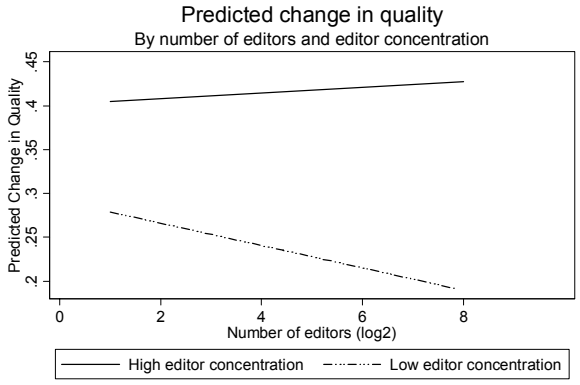


**Figure 4. Joint influence of number and concentration of editors on changes in quality.**

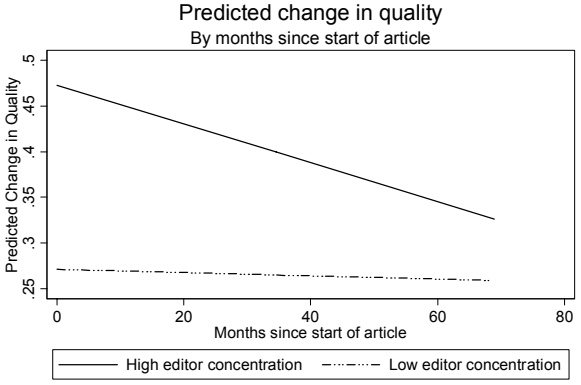


**Figure 5. Joint influence of article age and concentration of editors on changes in quality.**

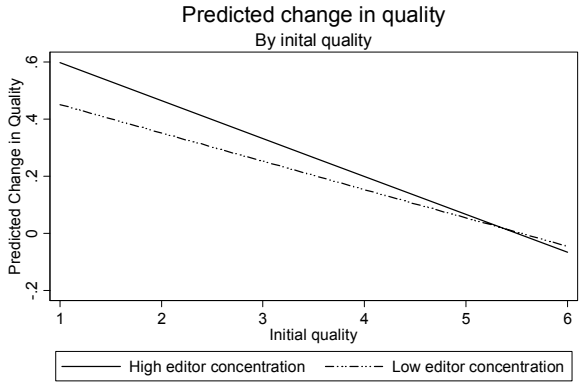


**Figure 6. Joint influence of initial quality and concentration of editors on changes in quality.**

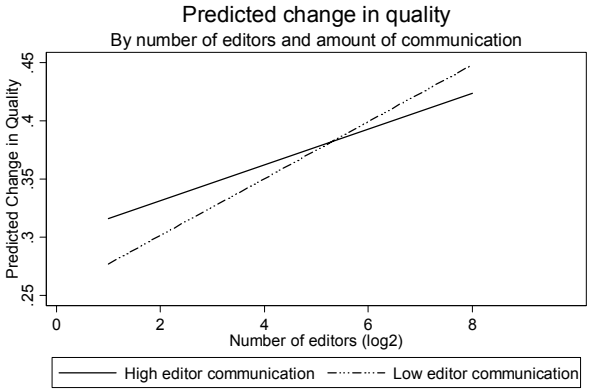


**Figure 7. Joint influence of number of editors and communication on changes in quality.**

that the benefits from implicit coordination are greater early in the article's life history, when it is still a 'stub', the least complete article. During this phase, when editors do interdependent tasks such as outlining the article's structure, having a subset of them do the direction-setting leads to greater increases in quality.

Model 3 adds explicit coordination (i.e., talk edits) and its interactions with the number of editors, article age, and initial article quality to the base model. Articles with more talk page communication improve in quality more than articles with little communication. As in Model 2, when talk edits are added to the base model, the number of editors no longer predicts increases in quality, while communication does. The significant negative interactions between talk page communication and both article age and initial quality indicate that, like implicit coordination, explicit coordination is most beneficial early in an article's life cycle.

In contrast to the case for implicit coordination, the interaction between explicit coordination and the number of editors is significantly negative (see Figure 7). Although high communication between editors (i.e., above the median) is helpful when there are few editors, the benefits of communication decline when the number of editors involved grows.

Finally, Model 4 combines the effects and interactions of the number of editors, implicit coordination, and explicit coordination. The results for this combined model are largely consistent with the models looking at the effects of implicit and explicit communication separately. However, articles and periods in which a few editors do most of the work are also ones in which these editors talk to each other on the article talk page. As seen in Table 2, the correlation between article concentration and talk page edits is .57, making it hard to distinguish the independent influence of each of these coordination techniques.

## DISCUSSION

Wikipedia differs from traditional collective intelligence systems in that while some tasks are driven by large numbers of contributors working independently, many other tasks involve significant coordination between contributors. This suggests that simply adding more contributors may incur coordination costs and process losses, and that coordination and communication may play a critical role in harnessing the "wisdom of crowds."

The idea that Wikipedia improves through the aggregation of many contributors' efforts is only partially supported in this research. As Wilkinson and Huberman showed[39] and we have replicated, articles with many editors are generally better than those with fewer editors. However, adding more editors to a page seems to improve its quality only when appropriate coordination techniques are used. Having more editors work on an article was associated with increases in article quality only if the editors used implicit coordination, so that most of the work was done by a small subset of them, with others playing a supporting role. Having more editors was not associated with improvements in article quality when the work was distributed evenly among editors or when they used explicit communication on the article talk page to coordinate. Phrased another way, both implicit coordination, through editor concentration, and explicit coordination, through communication, were valuable and were generally associated with improvement in article quality during the periods when they were used. However, implicit coordination was especially valuable for articles and time periods with many contributors, while explicit coordination was especially valuable for articles and time periods with few contributors.

In addition, this research has demonstrated that both implicit and explicit coordination have stronger associations with increases in article quality early in an article's life cycle, when the article is young and has little content. This is the period when tasks are most interdependent and coordination needs are highest, and this is when the article's creator(s) need to provide a structure for the article to which others can contribute. For example, our data suggest that it is important to have a small number of contributors setting the direction, structure, and scope of the article at the beginning of its life cycle, either implicitly by actually doing the writing or by explicitly communicating and coordinating. As the article matures and coordination requirements ease, tasks may be more effectively distributed to a larger group of contributors.

It is no surprise that articles in which the authors communicate with each other on the article's talk page improve more than articles in which editors work independently, each making a contribution without discussing changes or getting advice from collaborators. Interpersonal communication is perhaps the most general coordination technique available and is the paradigmatic case of what March and Simon (1958) describe as "coordination through mutual adjustment," in which members reveal their current states and intentions and adjust their behavior to others' goals and actions [33]. Decades of research in organizations show that communication as the basis for coordination is especially important in tasks that are highly uncertain, unconstrained, and subject to many changes (e.g., [18]). These conclusions are consistent with the observations in the current research that explicit communication through coordination is most beneficial in an article's formative stages, when its structure is highly unconstrained. In the beginning, no one knows either the content or the structure of an article, but later in its life cycle the existing material constrains what new editors can write.

The route by which the concentration of work leads to better coordination and improvements in quality is less clear. One possibility is that concentrating editing in fewer editors may enable a small core of committed contributors to focus on complex, highly interdependent tasks while allowing infrequent contributors to add value on simple, stand-alone tasks or those that benefit from diverse knowledge. For example, it may be effective for a small core group of users

to work on tasks such as organizing an article and making it cohesive, whereas tasks with low coordination requirements—such as fixing grammar, combating vandalism, or creating links—may be more effectively distributed to peripheral contributors. This strategy may be an efficient way to take advantage of a large number of contributors while at the same time minimizing process losses.

A similar participation pattern is found in other successful peer collaboration projects; for example, in most open source software (OSS) development projects a relatively small core of participants do the bulk of the work. However, this distribution depends upon the nature of the work. As cited earlier, in the Apache server project a core group of 15 developers contributed 88% of the new lines of code, but only 66% of the bug fixes, a less interdependent task [26]. To make an addendum to Linus' law about distributed work in OSS, that "with enough eyeballs, all bugs are shallow," our results suggest that while low-coordination tasks such as finding bugs may benefit from many eyeballs, high-coordination tasks such as planning new features may best be done by a small group of core users.

*Management implications.* Both the public perception and the ideology of Wikipedia are of a free and open environment, where everyone is equally eligible to contribute. However, even a free and open environment needs coordination, and peer-to-peer communication is ineffective if too many people are involved. Our results show that it is highly beneficial to have people who can set direction and provide a structure to which others can contribute. In Wikipedia this leadership group emerges naturally and informally based on editors' pre-existing interest in the topic, expertise, availability, and commitment. However, in other peer production environments it may be valuable to appoint leaders to positions of authority and control and to recognize their roles formally. This may be especially true for tasks where it is critical to get things right. For example, in the Linux OSS project, any change to the code of the central kernel has the potential to break the entire project, and a structure has evolved so that a single individual (Linus Torvalds) has the power to disapprove any change deemed inappropriate. Similarly, authoritative management techniques are found in critical knowledge bases such as the Gene Ontology (www.geneontology.org), in which researchers may propose new changes but expert curators are responsible for approving them and maintaining a cohesive structure.

*Limitations.* The use of longitudinal analyses enables us to make stronger causal claims than cross-sectional analyses would allow. However, both types of analyses utilize correlational data and involve assumptions that can be challenged. Our analyses take into account individual differences among articles and quality at the initial time period. However, the use of longitudinal analyses to assess causation rests upon an assumption that all the relevant variables have been measured. It is still possible that some unmeasured variable that co-varies with quality *change* and the use of coordination techniques may account for what appears to be a direct relationship between use of these techniques and improvement in quality.

Similarly, the assumptions underlying the measures used here can be challenged. Using quantifiable markers of coordination such as the number of discussion edits and the concentration of editing enables large-scale rigorous testing of the influence of coordination methods; however, coordination is a complex phenomenon which may not be fully captured by these metrics. Although we have tried to demonstrate how these metrics can relate to actual coordination behaviors (e.g., in the "Music of Italy" article), further research is necessary to better understand coordination behavior at the micro level of article editing.

*Summary.* Wikipedia is both an existence proof and a model for how complex cognitive tasks with high coordination requirements can be effectively achieved through distributed methods. Such methods are already beginning to become more widely used in both science and enterprise. Our results demonstrate the critical importance of coordination in effectively harnessing the "wisdom of the crowd" in such environments.

## ACKNOWLEDGMENTS

We deeply thank Ed Chi, Bongwon Suh, and Bryan Pendleton for providing quality assessment data, and John Riedl and Aaron Halfaker for sharing page view data. Early directions in this research began while the first author was at Palo Alto Research Center in the Augmented Social Cognition group. This work was supported by NSF grants IIS0729286 and IIS0325049.

## REFERENCES

1. Aiken, L. and West, S. *Multiple regression: Testing and interpreting interactions.* Sage, CA (1991).
2. Atkinson, A.B. On the measurement of inequality. *Journal of Economic Theory, 2* (1970), 244-263.
3. Berg, J. E., Forsythe, R., & Rietz, T.A. The Iowa electronic market. In *Blackwell Encyclopedic Dictionary of Finance*, Blackwell, Oxford UK (1997).
4. Benkler, Y. Coase's penguin, or Linux and the nature of the firm. *Yale Law Journal, 112* (2002), 367-445.
5. Brooks, F.P. *The mythical man month.* Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA (1975).
6. Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. *Applied multiple regression/correlation analysis for the behavioral sciences.* Lawrence Erlbaum Associates, Mahwah, New Jersey (2003).
7. Galton, F. Vox populi. *Nature*, *75* (1907), 7.
8. Giles, G. Internet encyclopaedias go head to head. *Nature*, *438* (2005), 900-901.
9. Hadoop Project. http://lucene.apache.org/hadoop/.

10. Heckman, J. Sample selection bias as a specification error. *Econometrica, 47*, 1 (1979), 153-162.
11. Hill, G. W. Group versus individual performance. Are n + 1 heads better than one? *Psychological Bulletin, 91* (1982), 517-539.
12. Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*, 1 (1997), 43-64.
13. http://en.wikipedia.org/wiki/Category:Wikipedia_1.0_assessments
14. http://en.wikipedia.org/wiki/Wikipedia:Stub
15. http://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team/Assessment
16. http://en.wikipedia.org/wiki/Wikipedia:WIAFA
17. Kanefsky, B., Barlow, N.G., & Gulick, V.C. Can Distributed Volunteers Accomplish Massive Data Analysis Tasks? *Lunar and Planetary Science* (2001).
18. Katz, R., & Tushman, M. Communication patterns, project performance, and task characteristics: An empirical evaluation and integration in an R&D setting. *Organizational Behavior and Human Decision Processes, 23*, (1979) 139-162.
19. Kittur, A., Suh, B., Chi, E., & Pendleton, B. A. He says, she says: Conflict and coordination in Wikipedia. *CHI 2007,* ACM Press (2007), 453-462.
20. Kittur, A., Chi, E., & Suh, B. Crowdsourcing user studies with Mechanical Turk. *CHI 2008*, ACM Press (2008).
21. Kittur, A., Chi, E., Pendleton, B. A., Suh, B., & Mytkowicz, T. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *Alt.CHI, 2007*, San Jose, CA (2007).
22. Laughlin, P. R., VanderStoep, S. W., & Hollingshead, A. B. Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. *Journal of Personality and Social Psychology, 61*, 1 (1991), 50-67.
23. Leavitt, H. Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology, 46* (1951), 38-50.
24. Malone, T. W. Modeling coordination in organizations and markets. *Management Science, 33,* 10 (1987), 1317-1332.
25. March, J. G., & Simon, H. A. *Organizations*. New York: Wiley (1958).
26. Mockus, A., Fielding, R. T., & Herbsleb, J. D. Two case studies of open source software development: Apache and mozilla. *ACM Transactions on Software Engineering and Methodology, 11,* 3 (2002), 309-346.
27. Priedhorsky, R., Chen, J., Lam, S., Panciera, K., Terveen, L., & Riedl, J. Creating, destroying, and restoring value in Wikipedia. *Proc GROUP 2007*, ACM Press (2007), 259-268.
28. Rouse, W.B., Cannon-Bowers, J. A., & Salas, E. The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics, 22* (1992) 1296-1308.
29. Shepperd, J. A. Productivity loss in performance groups: A motivation analysis. *Psychological Bulletin, 113*, 1 (1993), 67-81.
30. Steiner, I. D. *Group process and productivity*. New York: Academic Press (1972).
31. Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. Assessing information quality of a community-based encyclopedia. In *Proc. ICIQ* (2005), 442-454.
32. Surowiecki, J. *The wisdom of crowds*. New York: Doubleday (2005).
33. Thompson, J. *Organizations in action*. New York: McGraw-Hill (1967).
34. Tushman, M. Work characteristics and subunit communication structure: A contingency analysis. *Administrative Science Quarterly, 24*, (1979), 82-98.
35. Van de Ven, A. H., Delbecq, A. L., & Koenig, R., Jr. Determinants of coordination modes within organizations. *American Sociological Review, 41* (1967), 322-338.
36. Viégas, F. B., Wattenberg M., & Dave, K. Studying cooperation and conflict between authors with history flow visualizations. *CHI 2004*, ACM Press (2004), 575-582.
37. Viégas, F. B., Wattenberg, M., Kriss, J., & van Ham, F. Talk before you type: Coordination in Wikipedia. *HICSS* (2007).
38. Viégas, F. B., Wattenberg, M., & McKeon, M. M. The hidden order of Wikipedia. In *HCII* (2007).
39. Wilkinson, D.M., & Huberman, B.A. Assessing the value of cooperation in Wikipedia. *First Monday, 12*, 4 (2007).
40. Wittenbaum, G.W., Vaughan, S.I., & Stasser, G. Coordination in task-performing groups. In *Theory and research on small groups*, Plenum Press, NY (1998).