1-1-2008

Harnessing the Wisdom of Crowds in Wikipedia: Quality through Coordination

Aniket Kittur
Carnegie Mellon University

Robert E. Kraut
Carnegie Mellon University

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 involving large numbers of contributors who improve the accuracy, completeness and clarity of articles while reducing bias. However, because of the high coordination needed to collaboratively write an article, increasing the number of 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 set direction by doing the majority of the work. 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, collaboration, coordination, social computing.

ACM Classification Keywords
H.5.3 [Information Interfaces]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work, Web-based interaction, H.3.5 [Information Storage and Retrieval]: Online Information Systems, K.4.3 [Computers and Society]: Organizational Impacts – Computer-supported collaborative work

INTRODUCTION

Wikipedia, the online encyclopedia that anyone can contribute to, is one of the most heralded success stories of peer collaboration. It consistently ranks in the top ten most visited sites on the Internet according to Alexa.com. Since its inception in 2001 it has grown exponentially, and now includes over two million pages contributed to by more than six million registered user accounts in the English Wikipedia alone. Studies have found content of comparable quality to traditional encyclopedias [7], and that vandalism and inaccuracies are often reverted within a matter of minutes [18][31]. Its success has spurred the application of an open approach to knowledge building to a variety of domains, ranging from science (www.scholarpedia.org) to the 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 many judgments has been studied since at least 1907, when Galton showed that the weight of an ox at a county fair could be well approximated by the combined independent judgments of many observers [6]. The internet makes aggregating judgments much easier, leading to systems of collective intelligence ranging from markets for accurately predicting the results of presidential elections [2], to thousands of volunteers classifying craters on Mars' surface resulting in work virtually indistinguishable from expert geologists [16]. Most models of collective intelligence are premised on aggregating the efforts of many users processing information independently, colloquially known as harnessing "the wisdom of crowds" [28].

However, many of the tasks involved in the collaborative editing of articles in Wikipedia violate assumptions of independent judgments and automatic aggregation. While some tasks, such proofreading an article or contributing diverse 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 are theoretically capable of [26]. Adding more heads often has diminishing returns due to increased coordination requirements or decreased member motivation, especially for complex tasks or those with strong interdependencies [10][25]. 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" [4].

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 what the article will include and what it won't; which points of view will be represented and to what degree; wording and prose; and its structure and organization. Coordination costs are a substantial and growing proportion of all work done in WP: nearly 40% of all edits in Wikipedia involve indirect work such as communication, consensus building, and development of policies and procedures [18]. Much of this work takes place on dedicated discussion pages where changes are often discussed before being implemented in the article itself [32]. Thus coordination factors may play an important role in Wikipedia's success, separate from the influence of the number of contributors.

Empirical evidence regarding the effects of increasing numbers of contributors on the quality of Wikipedia content is scarce. The most definitive study to date 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 [34]. 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 many editors producing higher quality articles). Furthermore, the coarse metric of quality used (featured vs. non-featured) meant that few high-quality articles were sampled (fewer than 2000 featured vs. more than 2,000,000 non-featured) and the ones that were sampled went through a highly stringent, peer review process not representative of most articles in Wikipedia. These factors potentially limit the generality of the results.

The present research uses longitudinal data to examine the degree to which the number of editors contributing to an article in Wikipedia affects its quality and more specifically, 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 lifecycle 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, they need to coordinate their efforts on interdependent aspects of the article, like its structure and stylistic coherence.

Jamie Zawinski, creator of the open-source Mozilla application, characterized the coordination problem in groups as: "great things are accomplished by small groups of people who are driven, who have unity of purpose. The more people involved, the slower and stupider their union is" [36]. However, there are many ways to effect a "unity of purpose" in group collaboration. Theories of group coordination suggest a basic distinction between *explicit coordination*, based on direct communication and verbal planning, and *implicit coordination*, based on unspoken expectations and shared mental models of the task to be accomplished [24][35]. Below we examine evidence of both types of coordination in Wikipedia.

Explicit coordination

Editors can explicitly coordinate with each other in a number of ways in Wikipedia. The most commonly used mechanism for communication is the discussion page, although other mechanisms such as user talk pages are growing in importance [18]. The discussion page is a dedicated page associated with each article that provides a forum for coordinating changes to the article, discussing policies and procedures, and eliciting assistance from other editors [32].

Figure 1 shows the topics from the "Glock" article, a typical discussion page. 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 (in "Misconceptions", whether certain myths and misconceptions should be included or excluded). There is often a complex interplay of coordination between the discussion page and the article page, with editors proposing changes to be made to the article page, as well as discussing changes already made.

Figure 1. Topics of the discussion page for the "Glock" article.

To gain insight into the overall impact of explicit coordination activity we conducted an analysis of the proportion of discussion page activity to article activity for all articles in Wikipedia. The "proportion of coordination activity" in Figure 2 refers the ratio of discussion to article edits for each week of an article's life. This analysis shows an increase in the importance of explicit coordination over the years as the number of contributors has grown. Furthermore, explicit coordination is especially important early in an article's lifecycle: more than half of all edits in the first week of an article are made to the discussion page rather than the content of the article.

Implicit coordination

Editors can also implicitly coordinate in Wikipedia, either by playing a leadership role in setting the direction and scope of an article, providing a framework to integrate the work of less involved editors, or by taking on tasks that would

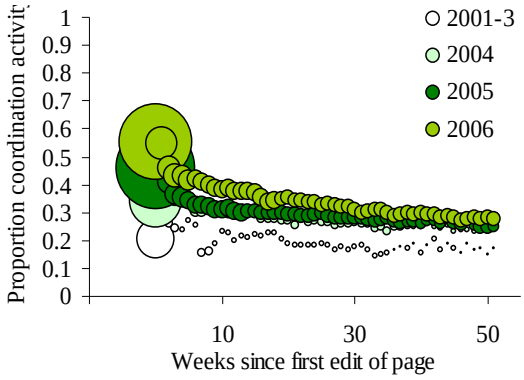

Figure 2. Proportion of coordination to production activity for the first 50 weeks of an article's life across all Wikipedia articles. Colors correspond to different years; the size of each circle is proportional to the log of the number of edits.

otherwise require high coordination between contributors. One marker of implicit coordination common to all cases is a few editors doing the lion's share of the work.

It is important to note that editing concentration is distinct from the number of editors: articles may have the same number of editors and even edits but very different distributions of work. 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.

Examples of implicit coordination can be found in the "Music of Italy" article. As shown in Figure 4, the user TUF-KAT does a large proportion of the work in the article during the first few months of the article's life. An examination of her first edits reveals the creation of the initial structure of the article (see Figure 3A) that is refined by herself and others to set the scope and direction of the content (Figure 3B). This structure remains substantively constant, with the addition of few sections of content, for the duration of the period marked as B in Figure 4. Note that during this period the proportion of work actually done by TUF-KAT is relatively low, indicating that others are continuing to follow the structure that she initiated.

Concentrating work in fewer individuals also decreases coordination requirements by reducing the need for explicit communication: highly interdependent tasks that would require significant coordination between editors -- such as making an article stylistically coherent -- may be efficiently executed by a single individual. Continuing to use the "Music of Italy" article as an example, Figure 4 shows that the user Jeffmatt does a large proportion of all of the editing within the period of January to July 2006 (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 Figure 3C to Figure 3D. These organizational changes were not made in a vacuum; after proposing and executing the changes Jeffmatt received explicit feedback from other highly involved members, notably TUF-KAT. However, this explicit coordination was largely limited to a few highly involved contributors rather than the general community. By concentrating work in a few individuals, the reorganization task could be accomplished while minimizing the overhead that would be required if a large group of editors was involved.

To quantify the concentration of work in an article we use the gini coefficient, a commonly used metric for measuring the skew of a distribution. Articles with edits are concentrated in few individuals will have gini coefficients close to one, while articles with edits evenly distributed across all editors will have coefficients close to zero.

Interactions with age and group size

Both explicit and implicit coordination are likely to be especially valuable when articles are at a formative stage, when the tasks involved are more likely to be interdependent. Editors in Wikipedia often use the discussion page to explicitly negotiate the scope and structure of an article and agree on standards to be used [32]. Additionally, early or heavy contributors may act as leaders by implicitly setting the direction, scope, and structure of an article, creating a framework for less involved editors to effectively fill in; indeed, Wikipedia directs users who create new articles to provide "enough information for [other] editors to expand upon" [10].

Explicit and implicit coordination may differentially interact with increasing group size. Explicit coordination is especially valuable in small groups engaged in complex tasks that require a high degree of synchronization between members [35]. 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 [26][30]. This suggests that increasing communication between editors may not be an efficient coordination mechanism as groups grow in size.

Conversely, implicit coordination involves the tacit development of a shared mental model among group members, allowing them to more efficiently aggregate their efforts towards a common goal [24]. Since implicit coordination avoids the overhead of discussion, it may scale with group size better than explicit coordination. Implicit coordination through fewer editors doing more of the work also reduces the total amount of interpersonal coordination needed, as interdependent tasks can be accomplished by single individuals.

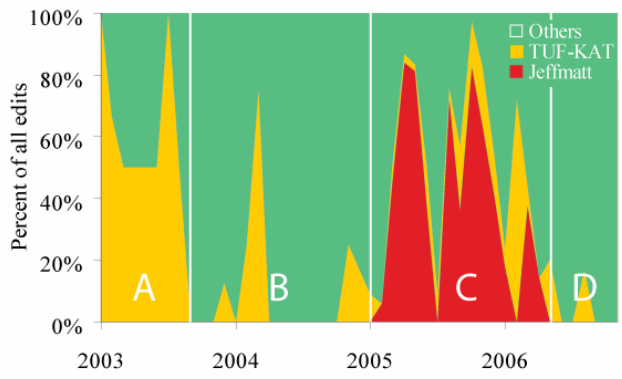

Figure 4. User activity for the "Music of Italy" article.

QUALITY IN WIKIPEDIA

To assess article quality we took advantage of Wikipedia's article assessment project, which in preparation for creating

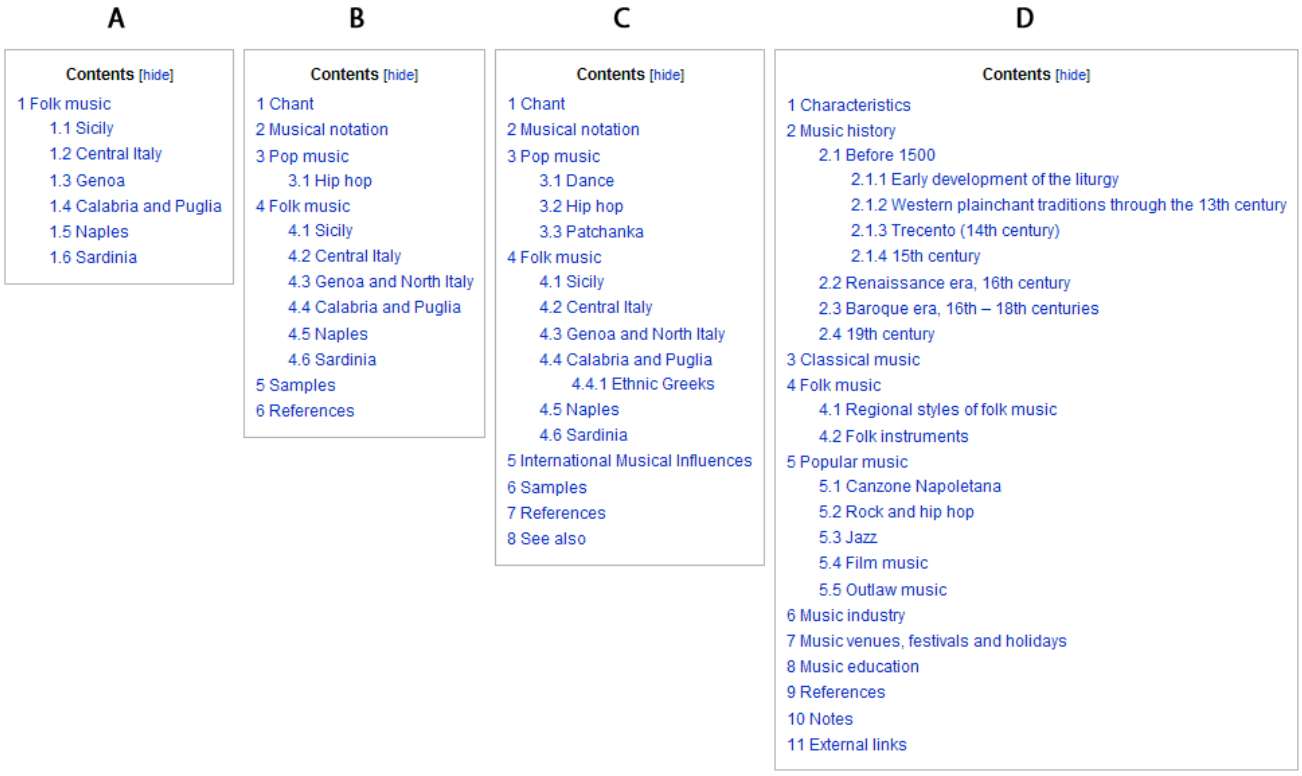

Figure 3. Evolution of topic structure for the "Music of Italy" article. Letters above the topic structures correspond to the beginning of the periods in Figure 4: creation of the article (A), the end of TUF-KAT's initial editing (B), the beginning of Jeffmatt's editing (C), and the end of Jeffmatt's editing (D).

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 [14]. Concrete guidelines for assessing each class include factors such as how well written, factually accurate, verifiable, comprehensive, neutral, and stable the article is [15]. 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.

To check on the external validity of the community ratings we conducted a survey among non-Wikipedians. The guidelines on assessment grading in Wikipedia [12] provides 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 user's 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 [19] 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 120 ratings of the 6 example articles, each representing a different community-assessed quality level. 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 up to 75,000 articles in each quality level (which included all articles for the top three quality levels) from the October 2007 dataset. Articles that were simply lists of other articles were excluded, as these were unrepresentative of the typical Wikipedia article. 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.[1]

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 and implicit coordination metrics, and the article's assessed quality at the beginning and end of the interval. To quantify explicit coordination we used the number of edits made to the discussion page. To quantify implicit coordination we used the gini coefficient, which provides a measure of the concentration of edits within the distribution of editors. 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.

We computed the quality at the beginning and end of each interval. The primary way that the Wikipedia community assesses article quality is by placing a template on the discussion page of an article. The template used to indicate quality also automatically places an article into the relevant quality class.

To gather longitudinal information on changes in article quality over time we need to know the date on which a given assessment template was added to an article. Unfortunately, the Wikipedia database does not directly store the date on which a specific template was added to specific article. To acquire this information the Hadoop distributed file system [8] was used to analyze the full text of every revision for all articles in Wikipedia (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 a beginning and ending quality assessment for any time period; these are the

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.

	Descriptive Statistics			Correlations					
	Mean	Median	Std	Quality change	Initial Quality	Article Age	# Article Editors	Editor Concen-tration	# Talk Editors
Quality change	.09	.00	.55						
Initial Quality	2.36	2.00	1.14	-.20					
Article Age	25.90	21.73	17.41	.00	.29				
# Article Editors	48.31	11.00	108.73	.08	.43	.51			
Editor Concentration	.26	.25	.18	.20	.27	.21	.61		
# Talk Editors	6.00	2.00	16.28	.14	.47	.41	.78	.52	
# Talk Edits	15.84	2.00	88.21	.17	.45	.38	.76	.57	.97

Table 2. Descriptive statistics before log transformations and correlations after log transformation.

articles used for the primary analyses reported below.

APPROACH

We initially conducted a similar cross-sectional analysis as in [34] but using multiple levels of the quality assessment scale. Consistent with [34], 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 controls for the age of the article and the number of page views the article receives (partial correlation = .22, p <.001).

However, these cross-sectional correlations provide little insight into the influence of the 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, importance, and availability of secondary sources.

Our goal is to examine how the number of contributors to an article and the differing coordination techniques they use results 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 [5]. 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 model [9], 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 propensity to be 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 propensity to be evaluated for quality, we included in the first stage of the Heckman analysis the cumulative number of edits the article had prior to the start of the observation period along with the article 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 discussion page (median=2). This observation suggests that even explicit coordination via discussion has an implicit component to it: 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 variables (i.e., increasing initial quality by one class, increasing concentration from completely even to completely concentrated, and doubling an article's age in 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 5, increasing the number of editors improved quality when work was highly concentrated (i.e., above the median), but detrimental when work was more evenly distributed (i.e., below the median).

In addition, Model 2 also shows negative interactions between concentration and the initial quality of an article (see Figure 7) and it age. Both interactions indicate 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 must outline 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 lifecycle.

In contrast to the case for implicit coordination, the interaction between explicit coordination and the number of editors is significantly negative (see Figure 6). This indicates that although more communication is helpful when there are few editors, the benefits of communication decline when there as 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

	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 lifecycle on change in article quality.

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

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 discussion page edits is .57, making it hard to distinguish the independent influence of each of these coordination techniques.

DISCUSSION

Although some tasks in Wikipedia may be driven by large numbers of contributors working independently, many other tasks involve significant coordination between contributors. The idea that Wikipedia works through the "wisdom of crowds" is only partially supported in this research. As [34] showed and we have replicated, articles that have had many editors working on them 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 with articles and time periods with many contributors while explicit coordination was especially valuable in 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 lifecycle, when the article is young and has little content. It is at this period when tasks are most interdependent and coordination needs are highest, and the articles 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 lifecycle, 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.

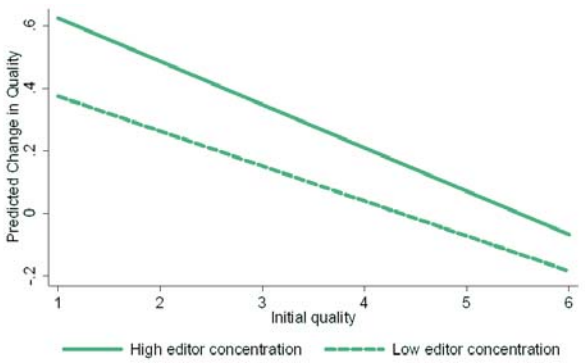

Figure 7. Joint influence of initial quality and concentration of editors on changes in quality.

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 [29]. Decades of research in organizations shows that communication as the basis for coordination is especially important in tasks that are highly uncertain, unconstrained, and subject to many changes (e.g., [17]). These conclusions are consistent with the observations in the current research that explicit communication through coordination was most beneficial in an article's formative stages, when its structure is highly unconstrained. In the beginning, no one knows either the

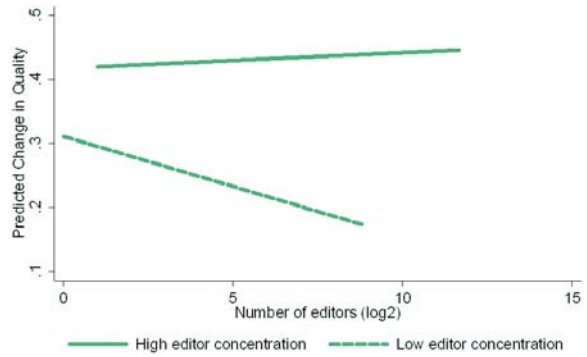

Figure 5. Joint influence of number and concentration of editors on changes in quality.

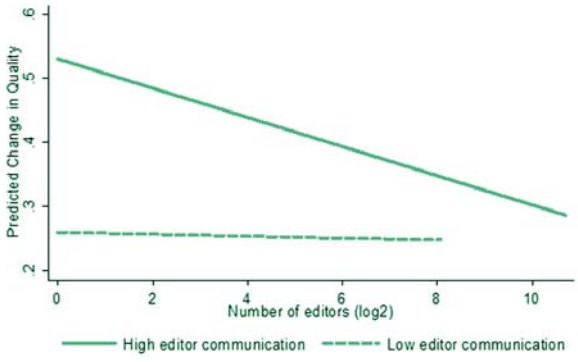

Figure 6. Joint influence of number of editors and communication on changes in quality.

content or structure of an article, but later in its lifecycle 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. 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 [23]. 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 ideology of Wikipedia is 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 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 formally recognize their roles. 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 such 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.

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 the 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. Berg, J. E., Forsythe, R., and Rietz, T.A. The Iowa Electronic Market. In *Blackwell Encyclopedic Dictionary of Finance*, Blackwell, Oxford UK (1997).
3. Benkler, Y. Coase's penguin,or Linux and the nature of the firm. *Yale Law Journal 112* (2002), 367-445.
4. Brooks, F.P. *The Mythical Man Month.* Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA (1975).
5. 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).
6. Galton, F. Vox populi. *Nature, 75* (1907), 7.
7. Giles, G. Internet encyclopaedias go head to head. *Nature, 438* (2005), 900-901.
8. Hadoop Project. http://lucene.apache.org/hadoop/.
9. Heckman, J. Sample selection bias as a specification error. *Econometirica, 47,* 1 (1979), 153-162.
10. Hill, G. W. Group versus individual performance. Are n + 1 heads better than one? *Psychological Bulletin, 91* (1982), 517-539.

11. Hinsz, V. B., Tindale, R. S., and Vollrath, D. A. The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*, 1 (1997), 43-64.
12. http://en.wikipedia.org/wiki/Category:Wikipedia_1.0_assessments
13. http://en.wikipedia.org/wiki/Wikipedia:Stub
14. http://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team/Assessment
15. http://en.wikipedia.org/wiki/Wikipedia:WIAFA
16. Kanefsky, B., Barlow, N.G., and Gulick, V.C. Can Distributed Volunteers Accomplish Massive Data Analysis Tasks? In *Proc. Lunar and Planetary Science* (2001).
17. 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.
18. Kittur, A., Suh, B., Chi, E., and Pendleton, B. A. He says, she says: Conflict and coordination in Wikipedia. In *Proc. CHI 2007*, ACM Press (2007), 453-462.
19. Kittur, A., Chi, E., Suh, B. Crowdsourcing for Usability: Using Micro-Task Markets for Rapid, Remote, and Low-Cost User Measurements. In *Proc. CHI 2008*, ACM Press (2008).
20. 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.
21. Malone, T. W. Modeling coordination in organizations and markets. *Management Science, 33*, 10 (1987), 1317-1332.
22. March, J. G., & Simon, H. A. *Organizations*. New York: Wiley (1958).
23. 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.
24. 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.
25. Shepperd, J. A. Productivity loss in performance groups: A motivation analysis. *Psychological bulletin, 113*, 1 (1993), 67-81.
26. Steiner, I. D. *Group process and productivity*. New York: Academic Press (1972).
27. Stvilia, B., Twidale, M. B., Smith, L. C., and Gasser, L. Assessing information quality of a community-based encyclopedia. In *Proc. ICIQ* (2005), 442-454.
28. Surowiecki, J. *The wisdom of crowds*. New York: Doubleday (2005).
29. Thompson, J. *Organizations in action*. New York: McGraw-Hill (1967).
30. Van de Ven, A. H., Delbecq, A. L., & Koenig, R., Jr. Determinants of coordination modes within organizations. *American Sociological Review, 41* (1967), 322-338.
31. Viégas, F. B., Wattenberg M., and Dave K. Studying cooperation and conflict between authors with history flow visualizations. In *Proc. CHI 2004*, ACM Press (2004), 575-582.
32. Viégas, F. B., Wattenberg, M., Kriss, J., and van Ham, F. Talk Before You Type: Coordination in Wikipedia. In *Proceedings of HICSS* (2007).
33. Viégas, F. B., Wattenberg M., and McKeon, M. M. The hidden order of Wikipedia. In *Proc. HCII* (2007).
34. Wilkinson, D.M., and Huberman, B.A. Assessing the value of cooperation in Wikipedia. *First Monday, 12*, 4 (2007).
35. Wittenbaum, G.W., Vaughan, S.I., and Stasser, G. Coordination in task-performing groups. In *Theory and Research on Small Groups*, Plenum Press, NY (1998).
36. Zawinsky, J. Resignation and postmortem. http://www.jwz.org/gruntle/nomo.html (1999).