Coordination in Collective Intelligence:
The Role of Team Structure and Task Interdependence

Aniket Kittur   Bryant Lee   Robert E. Kraut

Carnegie Mellon University

5000 Forbes Ave, Pittsburgh, PA 15213 USA

{nkittur, bryantl, robert.kraut}@cs.cmu.edu

ABSTRACT

The success of Wikipedia has demonstrated the power of peer production in knowledge building. However, unlike many other examples of collective intelligence, tasks in Wikipedia can be deeply interdependent and may incur high coordination costs among editors. Increasing the number of editors increases the resources available to the system, but it also raises the costs of coordination. This suggests that the dependencies of tasks in Wikipedia may determine whether they benefit from increasing the number of editors involved. Specifically, we hypothesize that adding editors may benefit low-coordination tasks but have negative consequences for tasks requiring a high degree of coordination. Furthermore, concentrating the work to reduce coordination dependencies should enable more efficient work by many editors. Analyses of both article ratings and article review comments provide support for both hypotheses. These results suggest ways to better harness the efforts of many editors in social collaborative systems involving high coordination tasks.

Author Keywords

Wikipedia, wiki, social collaboration, collective intelligence, coordination, social computing.

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

The Internet in general and Web 2.0 applications in particular have enabled social collaboration on a massive scale, in which large numbers of people come together to produce services and artifacts that benefit the producers themselves and the larger public. These applications include

del.icio.us for collaboratively tagging websites, Digg for evaluating web content, Netflix for evaluating movies, YouTube for sharing videos, the Iowa Electronic Market for predicting future events, and open source production systems for creating economically valuable artifacts such as the Apache web server and Wikipedia. The novelty of these technology-supported collaboration systems and their widespread use make them worthy of scientific study in their own right, both to understand how they work and to improve them. In addition, these new systems provide a source of data to explore decades-old problems in group and organizational behavior concerning the ways that collections of people coordinate their work to achieve goals.

Many commentators, including Tim O'Reilly, who coined the term Web 2.0, treat the types of applications just mentioned with a broad brush and attribute their success in large part to an "architecture of participation" that encourages many people to contribute content that, when appropriately aggregated, benefits the group as a whole [25]. Thus, for example, people may use del.icio.us people to tag websites with keywords as a personal memory aid, but as a side effect they collaboratively produce a taxonomy of the web. These systems are generally characterized by low costs of participation, individuals' self-selecting tasks to work on, and easy or algorithmic ways of combining their contribution. The rationale for many of these sites is that having more contributors leads to more complete or better quality artifacts [5][32]. Efforts to combine the contributions of multiple people so that their judgments and efforts surpass the work of individual contributors are not new; scientists have tried to understand this issue for over 100 years. For example, in 1907 in an early demonstration of the wisdom of the crowd, Francis Galton showed that aggregating the independent judgments of many observers at a county fair led to estimates of the weight of an ox that were more accurate than experts' judgments [8].

Coordinating work

However, even casual inspection shows that the work that people engage in using these technology-enabled collaboration systems differs greatly in its coordination requirements, from tasks that are relatively independent and can be done with little coordination to highly interdependent ones, which require significant coordination amongst contributors [22]. For example, the collective

evaluations of movies generated by subscribers to netflix.com depend only on each subscriber honestly giving an evaluation of a movie, independently of other subscribers. In contrast, making a change to the Linux kernel is a task with high coordination requirements, since the change will affect many other software components. Collaboratively editing a Wikipedia article sits between these extremes of interdependence[1].

When tasks require high coordination because the work is highly interdependent, having more contributors can increase process losses, reducing the effectiveness of the group below what individual members could optimally accomplish [29]. Although having more contributors increases the resources available to accomplish a group's goals (e.g., time, energy and expertise), it may also lead to increased process losses due to the need to coordinate the work of more people. In the limit, coordination costs can overwhelm the benefits of added personnel, noted as Brooks's Law in the domain of software projects: "Adding manpower to a late software project makes it later" [6].

Psychologists have examined how factors such as the nature of the task, the size of the group, and the diversity of group composition influence how well groups can coordinate their work. Hill provides an early review of the literature [13], showing that groups perform many tasks better than any of their individual members but frequently perform less well than an optimum pooling of their resources should allow. For example, in problem-solving tasks, groups frequently ignore correct solutions if only a single individual member identifies it.

Past research on contingency models of coordination has shown that there is no single best way to coordinate work [e.g., 34]. The best coordination techniques depend both on the nature of the environment in which the work is done (e.g., its uncertainty) and the nature of the task that the contributors are trying to accomplish (e.g., task interdependence). Much of the focus among organizational scholars has been on the role of uncertainty in both the environment and the task [2] [12][33]. However, in a recent meta-analytic review, Stewart [30] describes how "team design features" — the ways that teams are constituted and organized — interact with the tasks they perform to influence their performance. According to this review, for example, larger teams generally perform better than smaller ones, but the benefits of team size are greater for production teams (engaged in low-coordination work) than for project teams (engaged in tasks requiring a high degree of coordination).

In the current paper we examine how the number of people contributing to a task and their coordination techniques influence performance for tasks varying in their coordination requirements. We examine Wikipedia, one of the most successful examples of large-scale, technology-enabled collaboration. Writing Wikipedia articles involves tasks with a variety of coordination requirements. Our results suggest that the benefits of having more editors[2] critically depend on the type of task: more editors benefit independent tasks but can negatively effect tasks with high interdependencies. We also find that reducing coordination requirements by concentrating the work in fewer editors more efficiently harnesses the efforts of the crowds for highly interdependent tasks but not for independent tasks.

Coordination in Wikipedia

Wikipedia is an online encyclopedia built through the efforts of volunteer editors. The English Wikipedia alone includes over 2.5 million pages and 1 billion words, created by more than six million registered user accounts, as well as anonymous edits from over 11 million distinct I.P. addresses[3]. These editors work together to create content, at least some of which has been found to approach the quality of traditional encyclopedias [11]. Even though anyone —even unregistered, anonymous users — can make changes to almost any article, vandalism and inaccuracies are repaired quickly, often within minutes [17][26]. The popularity of Wikipedia has led to the adoption of the wiki approach in a variety of domains ranging from science (scholarpedia.org) to government (Intellipedia) to the enterprise (socialtext.com).

Recent empirical evidence about the effects of increasing numbers of editors on the quality of Wikipedia content highlights the importance of coordination. Wilkinson and Huberman found 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 [35]. However, Kittur and Kraut [20] demonstrated that the benefits of having additional editors depended on the ways in which the editors coordinated their work, by directly communicating with each other or by concentrating most of the editing among a subset of the editors.

Kittur and Kraut demonstrated that use of appropriate coordination techniques is essential for harnessing the wisdom of crowds so that a project can benefit from the resources additional editors provide. For example, coordinating by direct communication among editors improved articles on articles with relatively few editors but actually harmed quality when many editors were involved. However, Kittur and Kraut could not test hypotheses about

---

[1] Note that in this discussion, we have treated coordination requirements as being determined by the type of product being produced. However, as attempts in software engineering to modularize software components indicate, coordination requirements themselves can be altered, even within a common task.

[2] In Wikipedia the contributors are called editors, and we will use this term for them throughout this paper.

[3] These numbers were calculated from the March 2008 Wikipedia database dump.

interdependence, because they did not differentiate different types of editing nor the quality of material that these edits produced. Instead, they simply counted the number of edits made to an article, and their outcome measure was the overall quality of the entire article.

Coordination requirements
Studies have differentiated dimensions of the quality of written documents, such as how comprehensive, accurate, unbiased, well-structured and well-written they are [21][31][15]. The tasks needed to achieve these different types of quality differ in their interdependence. A fact-checker or copy-editor can work relatively independently of an article's authors to improve its accuracy, grammar or spelling. However, efforts to improve the article's structure or bias require more interdependent work by the article's authors and editors to determine a single flow for the article or to create a unified point of view.

Writing Wikipedia articles involves a variety of tasks which vary in their interdependence. Fixing spelling or grammatical errors are examples of independent tasks: they have few dependencies on other information in the article. On the other hand, restructuring an article by consolidating two sections is an example of a highly interdependent task involving significant dependencies in both grammatical and conceptual context, as the editor must attend to much of the article's content in order to maintain a unified voice and consistent structure.

Tasks with greater dependencies incur greater coordination costs when they involve more contributors [7]. This suggests that the most effective strategies for coordination should depend on the task being performed. Compared to interdependent tasks, tasks that can be performed independently and have low coordination requirements should benefit from having more contributors involved. In the examples above, catching and fixing spelling errors—independent tasks with low coordination requirements—should be efficiently performed by large numbers of people. A similar intuition underlies Linus' law in the domain of open source software development: "With enough eyeballs all bugs are shallow" [27]. Conversely, interdependent tasks, which have high coordination requirements, should not benefit as much from having additional editors. As coordination costs in these interdependent tasks scale super-linearly with the number of participants [6], the costs of coordination may even outweigh the benefits of having additional editors.

Hypothesis 1: Compared to interdependent tasks, independent tasks will benefit more from increasing numbers of editors. Growing coordination costs for interdependent tasks may even result in *negative* effects from increased numbers of editors.

Strategies of coordination
Baecker et al. [4] provide a taxonomy of control strategies in collaborative writing: centralized document control, with one person having final responsibility for the text; relay control, where the document passes from one author to another; independence, where different authors maintain control of different segments of the document; and shared control, where anyone can change the document. These coordination patterns are not necessarily mutually exclusive: for example, in many open source communities different developers work independently on different software modules, but their changes need to be approved by a select group of 'committers,' who have responsibility for reviewing and approving changes to the software release. In the extreme case, centralized coordination rests on a single individual who has final say (e.g., Linus Torvalds in the early days of Linux).

In both interview data and laboratory studies Baecker et al. found that the independent method was used most frequently. Similarly, the most common strategy in document creation was to have separate writers. At first glance, this looks very different from the shared control in wikis, which allows anyone to edit anything. However, while wikis technically enable editing by anyone, in reality, the system of control that emerges from the norms of community behavior may be very different. For example, ad hoc centralized control is sometimes seen in Wikipedia when some editors make suggestions on an article's talk page that are implemented by a single editor, a strategy often seen in the peer review or featured article review process. A small core of editors in Wikipedia accounts for a large proportion of both the edits and the viewed content [19][26], again suggesting the possibility of more centralized coordination control strategies.

This is consistent with the intuition that the simplest way to maximize the work of many editors is to divide the document up and have each work on separate sections that are later combined. However, there still exists the problem of how to combine the separately created parts. Thus the commonly used "divide and conquer" methodology may be efficient for tasks such as generating content, but may be problematic for having a cohesive, unified document.

Instead, it may be more efficient when completing high-coordination tasks to reduce coordination requirements by concentrating the work in a core of editors. This strategy reduces the overhead of communication between many editors and enables other editors to work more efficiently by providing an outline for the article which they can flesh out and by promoting shared mental models [28]. Mockus, Fielding, & Herbsleb [23] find support for this coordination strategy in the domain of open-source software, where a large set of peripheral developers work on low coordination tasks such as finding bugs, while a smaller set of core developers builds new features. In Wikipedia, concentrating the work among a subset of all the editors working on an article may be an effective coordination strategy that scales up to large numbers of editors. In contrast, communication as a coordination strategy does not scale as well [20].

Hypothesis 2: Compared to the case of independent tasks, in interdependent tasks concentrating work among a small number of editors should reduce coordination costs. This should be revealed as a statistical interaction between work concentration and the interdependence of the task. The effect of work concentration could happen through two mechanisms. H2a: Concentrating the work may directly improve an article's quality, revealed statistically by a main effect of editor concentration on quality. H2b: Work concentration may indirectly affect quality by enabling a larger group of editors to contribute more effectively. This would be revealed by the statistical interaction of the number of editors and concentration.

We test these hypotheses in two studies in which we examine the effects of the number of editors and their concentration on the coverage of Wikipedia articles, a quality of articles requiring relatively independent work among editors, and on article readability, a quality that requires interdependent work among the contributors.

STUDY 1: COVERAGE & READABILITY OF ARTICLES

Study 1 examined the effects of the number of editors and their concentration of work on two different measures of quality: coverage and readability. Although there have been a number of metrics used to measure the quality of information [21][31], coverage and readability provide the most direct tests of our hypotheses. Coverage relies on the amount of information in an article rather than how integrated it is and therefore requires relatively little coordination. According to Hypothesis 1, it should benefit from the participation of many editors.

In contrast, producing a readable article is interdependent work requiring integration of both information and style. Hypothesis 1 predicts that the participation of many editors would have fewer benefits for such a task. Additionally, based on Hypothesis 2 we predict that readability but not coverage should benefit from concentrating the work among fewer editors, either because the core editors can easily coordinate or because their work provides a scaffold that enables the larger numbers of peripheral editors to work effectively.

Sampling Articles

To test these hypotheses we examined the association of number of editors, editor concentration and their interaction on the coverage and readability of 20 articles in the domain of "American rock singers." We sampled from a single category in order to minimize topic-dependent differences between articles. We selected this category because it had enough articles to sample from (over 600 in July 2007) and because the content allowed for a variety of writing styles, in contrast to a category such as "Cities," which was highly template-driven, covering specific topics (e.g., demographics, geography) in a specific order.

We used data from the July 2007 dump of Wikipedia provided by the MediaWiki foundation [14]. Articles with fewer than 20 edits or 300 words (140 total) were excluded due to concerns with interpreting the effects of the independent variables on short, rarely edited articles. Stratified sampling based on article Gini coefficient was used to select 20 target articles.

Variables and Coding

We calculated the following variable for each of the two articles:

Number of editors. The number of editors who made at least one edit from the beginning of the article to July, 2007. Because the number of editors had a highly skewed distribution, we used a log to the base 2 transformation to make the distribution more normal. The number of editors correlates .97 with the number of article edits, suggesting that both variables measure the same construct, i.e., the amount of work performed on the article. Although we cannot include both total editors and total edits in the same model because of multi-collinearity problems, by including total editors in the model we control for total work when examining the effects of other variables.

Editor concentration. Measured by the Gini coefficient [3], commonly used to measure the inequality of a distribution. To calculate this we compared the cumulative percentage of editors working on an article with the work they did. If all editors are contributing equally then, for example, 50% of the editors will have contributed 50% of the edits, but if the work is highly concentrated, then a small percentage of the editors will have contributed a large percentage of the work. The Gini coefficient takes values between 0 (completely equal) and 1 (completely unequal).

Coverage. Coverage was measured using both human and machine-assessed methods. For the human-coded coverage measure, an article was rated from 0-10, with one point given for each piece of information from a predetermined set of ten topics, including dates of birth and death; name of patron/music label that supported the artist; and name of musician whose work influenced the artist. A machine-assessed coverage score was also computed by summing the number of words, links, categories, references, and images in each article. As shown in Table 1, the two coverage measures correlated with each other .51.

Readability. Readability was coded by two coders who independently rated each article's readability on a 5-point scale, looking at paragraph- and sentence-level criteria such as cohesion, fluency, and word choice. Any disagreements in ratings were resolved by discussion between coders.

Analysis and Results

Table 1 shows the descriptive statistics and correlations for the variables in Study 1. Correlational analyses are consistent with Hypothesis 1: The correlations between both human-coded and computer-coded coverage, measures of relatively independent work, are higher with the number of editors than with editor concentration (both *p*s <.07, 1-tailed).

Variable	Mean	Std.	Correlations			
			1	2	3	4
1 Coverage (human-coded)	5.45	1.47				
2 Coverage (computer-coded)	186.54	99.79	0.51			
3 Readability	9.25	1.07	0.13	-0.10		
4 # Editors (log2)	6.31	1.53	0.35	0.50	-0.20	
5 Concentration (Gini)	0.37	0.14	-0.21	0.05	-0.41	-0.37

Table 1: Descriptive statistics and correlations for variables in Study 1. The descriptive statistics for # Editors were calculated before the log transformation, while the correlations were calculated after it. N = 20 articles. *P*<.05 if |*r*|>.43.

Multiple regression analyses predicting coverage and readability from the number of editors, editor concentration and their interaction lead to similar conclusions (see Table 2). To reduce the multi-colinearity between main effect and the interaction, we centered all independent variables [1]. Thus, coefficients should be interpreted as the effect of a unit increase in an independent variable (i.e., increasing concentration from completely even to completely concentrated, or doubling the number of editors) when other variables are at their mean level. All regression analyses had VIFs below 3, indicating no problems with multicolinearity.

The multiple regression analyses show a significant positive effect of the number of editors on machine-computed measures of coverage, but no significant results were found for the human ratings. This provides partial support for Hypothesis 1: having more editors helped for independent tasks such as adding information. In contrast, there was a marginal negative effect of the number of editors on readability. This provides strong support for the second part of Hypothesis 1: having more editors not only benefits independent tasks (coverage) more than interdependent tasks (readability), but further suggests that having more editors can be harmful for highly interdependent tasks.

Contrary to Hypothesis 2a, there were no benefits found for the main effect of editor concentration for readability. However, the interaction of concentration and the number of editors was positively related to readability. This provides support for Hypothesis 2b, that for interdependent tasks concentrating work indirectly effects quality by enabling a larger group of editors to contribute more effectively (i.e., the interaction). These benefits for concentration of editing were not seen for the coverage measures, also consistent with the second part of

Hypothesis 2: interdependent tasks with high coordination requirements benefit from concentration more than independent tasks (in which coordination requirements are already low).

Discussion
Together these results support both Hypotheses 1 and 2. An increased number of editors was associated with better coverage, which involves independent tasks, but was associated with lower readability, which involves interdependent tasks. In contrast, greater concentration of editing benefited interdependent work (readability) but not independent work (coverage), by enabling large numbers of editors to work more effectively.

However, caution is warranted in overgeneralizing from these results. The number of articles in the sample was small, largely due to the difficulty and time involved in coding each article. Furthermore, American rock singer articles did not strongly differ from each other on some of the quality metrics utilized here. This restriction of range in quality may have resulted from difficulty that the coders, who were not domain experts, had in identifying information that was missing or inaccurate or how the article compared to an optimal article structure and organization. In addition, the coders, who were not expert writers, had difficulty identifying subtle problems with the prose less obvious than spelling and grammar mistakes (which themselves were relatively rare). Finally, hand coding the articles was very time consuming, making this technique too expensive for analyses involving large numbers of articles, thereby reducing the power of statistical tests.

A lesson learned for researchers of Wikipedia and other collaborative knowledge bases is that having individual non-experts rate content may work for finding differences

Predictor	Coverage						Readability		
	Human-coded			Computer-coded			Human-coded		
	Coef	SE	p	Coef	SE	p	Coef	SE	p
# Editors (log2)	0.301	0.237		0.38	0.111	**	-0.276	0.137	
Editor concentration (gini)	0.598	3.545		1.914	1.66		-1.411	2.048	
# Editors X Concentration	1.608	2.241		1.013	1.049		2.869	1.295	*

Table 2. Regression analysis of the number of editors, concentration of editing, and interaction on measures of coverage and readability. * = p < .05

at the gross level (e.g., [18]), but may be less useful for making fine distinctions. Furthermore, the amount of time needed even for experts to carefully evaluate the quality of an article is non-trivial. This is consistent with previous research in which quality ratings of Wikipedia articles by expert Wikipedia administrators often took upwards of 30 minutes per article [17].

STUDY 2: PEER REVIEW

In Study 2 we use a novel approach to overcome some of these problems by taking advantage of Wikipedia's article review programs, in which expert Wikipedians review and rate the quality of articles. Using this approach, Study 2 aimed to extend the results of Study 1 to a larger set of articles and to include more detailed measures of quality.

Assessing Article Quality

To assess the quality of an article and problems remaining in it, Study 2 takes advantage of two formal quality assessment programs in Wikipedia, the peer review process and nominations for article quality levels. There have been thousands of peer reviews and article quality nominations in Wikipedia [16]. In both, an editor solicits reviews for an article. Other editors then review the article, providing detailed constructive feedback on the current state of the article and suggestions for improvement.

Compared to the quality nomination progress, the peer review program is relatively informal, used to elicit assessments when an article is still in a formative stage. The article quality nomination process is more formal, used for articles that are already at a relatively mature state. In this process, an editor proposes that the article be placed in a designated quality level (e.g., designated a "featured article"). Other editors, including experts in the topic domain, in writing and grammar, or in the rules of Wikipedia itself, then explain why they support or oppose moving the article to that level and provide suggestions for improvement. The highest level is a "featured article," which must meet stringent quality criteria. Articles can also be nominated for lower quality levels, including "A-Class" and "GA-Class" (Good Article). Past research on the external validity of these quality levels has shown strong correlations between community assessments and assessments by a more general audience (i.e., non-Wikipedians) [20].

The resulting feedback and suggestions are typically preserved on separate peer review or nomination pages, which provide a rich resource for gaining insights into the different types of problems with quality in an article. Some problems result from interdependent tasks (e.g., neutrality, style or consistency) others from independent tasks (e.g., missing information or spelling mistakes). In the following analyses we use these comments as outcome measures to predict the way an article is written (e.g., the number of editors, concentration of editing) influences the types of quality problems that remain (e.g., problems with independent versus interdependent tasks).

Sampling

We randomly sampled 230 articles that had at least two reviews; these could be peer reviews, A-class reviews, or Featured Article reviews. Ten articles were reserved as a training set for establishing common coding between judges. For each article we retrieved and separated all comments made in the reviews, resulting in a total of 10,002 individual comments. We also computed the number of editors and the Gini coefficient up to the date of review for each article.

Coding

We utilized the NWREL rubric [24] for judging article quality. This rubric was chosen for its popularity and ease of use. NWREL has trained 15,000 teachers in the use of the rubric since 1983 and has been the subject of several research papers regarding writing assessment [21].

The NWREL rubric provides guidelines for coding writing along seven dimensions: Ideas, Organization, Voice, Word Choice, Sentence Fluency, Conventions, and Presentation. We modified the NWREL rubric to develop a guide relevant to coding individual review suggestions[4]. Two authors independently judged whether each category involved primarily independent or interdependent tasks, with differences resolved by discussion. The new coding scheme is shown in Table 3 along with an example of each category. For example, the "Ideas" category was used when a comment talked about changing the information content of the page, e.g., adding, removing, or clarifying information; while the "Organization" category was used for merging, splitting, or moving information on a page or across pages. We also added Wikipedia-specific categories for links/categorization and for citations, which were the theme of many comments. Our modification does not include "Word choice," as coders initially found it too difficult to distinguish between comments dealing with word choice versus fluency. During the process of coding, the coders discovered that they were using the fluency metric for both interdependent problems, such as overall readability and flow, and independent problems, such as wording within sentences. Thus after the initial categorization was completed, two coders recoded the fluency comments as primarily interdependent or independent (Cohen's Kappa = 0.60). Prior to classifying the final data, two coders rated 10 articles' worth of training data to establish inter-rater reliability, with an average Cohen's Kappa across categories of 0.61. One coder then rated the remaining articles.

Out of the 10,002 comments, 4601 were categorized as suggestions for the article. The remaining consisted of various other types of comments, including: editors replying to suggestions (e.g, "Done. Restructured whole article"); follow-up and clarification (e.g., "That's not really

---

[4] The resulting coding guide is available upon email request to the first author.

Independent task problems

Category	Example	Ratings
Ideas	The Soundtrack section could do with some quotes from critics.	1532
Links / Categorization	Huge amount of redlinks need to be turned blue [target pages created] or removed	268
Presentation	Could do with a good shot of a modern locomotive in the current red livery.	418
Citations	Little or no references. If you wish to have this article featured, it must be referenced.	557
Fluency (Independent)	"His work has been adulated by eminent figures through the centuries" - very awkward passive	258

Interdependent task problems

Category	Example	Ratings
Organization	The article may flow better if history was the first section, then details of armament, etc, finishing with where they are now and reactivation potential	358
NPOV	This article definately [sic] has an Indian bias towards it and can be written in a more neutral tone. I suggest a collaborative effort with Pakistan-based editors.	130
Conventions	Improperly formatted dates [throughout] (e.g. "June 15th, 2007"; also, non-date enabled dates)	1039
Fluency (Interdependent)	One concern is readability. IMO, some the prose is jerky and does not flow, e.g the para on "Childhood".	578

Table 3. Categories used for coding review comments with examples and number of ratings for each.

the case here"); process issues (e.g., "I don't think one user should have final say over any article"); and simple statements of "Support" or "Oppose" for A-class and Featured Article reviews, in which the results determined the article's resulting class assessment. Most suggestions were for Featured Article reviews (2,734) and informal peer reviews (1,792), with very few A-class review suggestions (21). This small figure for A-class reviews was due both to fewer overall reviews and to the greater difficulty of finding and automatically extracting them in Wikipedia. All review types had relatively similar comment to suggestion ratios, with informal peer reviews eliciting more suggestions (58% of all comments) than A-class (48%) or FA (40%), as the latter two included the dual purposes of judging the suitability of the article for advancement in assessed class as well as providing specific feedback.

Analysis and Results

The dependent measure in this analysis was the proportion of problems cited in reviews that referenced interdependent rather than independent work tasks[5]. A positive coefficient

---

[5] Unlike Study 1, the measures of independent and interdependent quality here are not independent, as they are both based on the same pool of problems cited in reviews. Thus we are using a single measure which is the relative proportion of interdependent versus dependent problems cited. The drawback to this approach is that it makes it difficult to interpret whether an effect is uniquely driven by

thus indicates worse interdependent quality (i.e., more interdependent problems than independent problems), while a negative coefficient indicates better interdependent quality (i.e., fewer interdependent problems relative to independent ones). As in Study 1, we used a multiple regression approach with the number of editors and the concentration of editing as predictors.

Since each article included at least two reviews and multiple comments, we used hierarchical linear modeling with the article as a random effect to deal with the non-independence in the data [8]. Variables were centered as in the previous analysis, and thus coefficients should be interpreted as the effect of a unit increase in a predictor when other variables are at their mean level. VIFs for the analysis were below 2, indicating low multicolinearity between variables.

The regression coefficients are shown in Table 4. The coefficient for the number of editors is significantly positive, indicating that articles with many editors had more problems with interdependent tasks relative to independent ones. This is consistent with both Hypothesis 1 and data from the first study, which showed a positive impact of the number of editors on independent tasks but a negative impact on interdependent tasks.
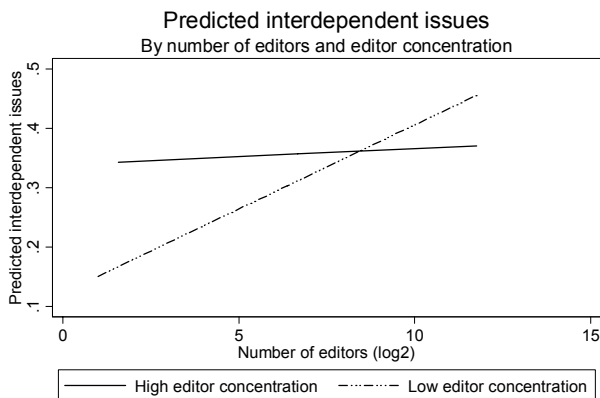
---

the influence of coordination techniques on interdependent or independent work.

Also consistent with the first study, there was no significant main effect of editor concentration on the proportion of interdependent issues, contrary to Hypothesis 2a. However, the coefficient for the interaction of the number of editors and their concentration is significantly negative, indicating that a high concentration of editing reduces problems of having many editors when working on interdependent tasks. This provides support for Hypothesis 2b and is consistent with results from the first study, which showed a positive impact for coordination factors such as concentrating the work in fewer editors for interdependent tasks but not independent tasks. As shown in Figure 2, the proportion of interdependent problems grows with the number of editors when concentration is low, but remains flat when concentration is high. This suggests that concentration of work in a smaller group of editors is an effective means of harnessing the efforts of the crowds for interdependent tasks.

GENERAL DISCUSSION

Tasks in Wikipedia vary in the degree of coordination required between editors. In two studies we examined how increasing the number of editors differentially benefits independent versus interdependent tasks. The first study examined a small set of articles rated by the researchers on their coverage (independent work) and readability (interdependent work). The results supported Hypothesis 1: more editors were associated with better coverage, involving low coordination tasks, but did not help and sometimes hurt readability, involving high coordination tasks. Results also suggested that concentrating the work in fewer editors benefited high coordination tasks but did not benefit low coordination tasks.

The second study took advantage of the large article review system already in place in Wikipedia to determine how the number and concentration of editors were associated with the types of problems cited in reviews. Consistent with the first study, Study 2 found that a greater number of editors was associated with more interdependent problems.

Predictor	Interdependent issues		
	Coef	SE	p
# Editors (log2)	0.016	0.007	*
Editor concentration (gini)	0.16	0.109	
# Editors X Concentration	-0.127	0.059	*

Table 4. Regression analysis of the number of editors, concentration of editing, and their interaction on the proportion of interdependent problems cited in article reviews. * = p < .05

However, concentrating the editing within a subset of all editors reduced the interdependent problems associated with large numbers of editors.

In both studies, the benefits from concentrating the work were due to enabling more effective contributions by a large number of editors rather than the direct effects of the core editing group, consistent with Hypothesis 2b but not 2a. This suggests that core editors may be creating a structure or environment for peripheral editors to flesh out and promoting a shared mental model of the article rather than simply doing better work themselves.

These results suggest a paradox of quality for systems of large-scale distributed collaboration. On the one hand, having many editors increases the pool of available workers and the amount of work that can be done. Additional potential benefits include reducing bias, catching errors, and increasing breadth of covered knowledge. However, increasing the number of editors also increases coordination costs for tasks with high coordination requirements, potentially leading to process losses, lower motivation, social loafing, and conflict. Here we demonstrate both the benefits and the drawbacks of having many editors on different kinds of tasks within a single system of distributed collaboration.

We also find that structuring the work to reduce coordination requirements by concentrating editing in fewer editors can benefit highly interdependent tasks that otherwise would suffer from having many editors. These results are consistent with prior research on coordination in Wikipedia [20], demonstrating that concentrating the work in a smaller group of editors is a strategy that scales well to large numbers of editors, whereas coordination through communication does not. In this paper we extend that work by showing that concentrating the work is an effective strategy for tasks that are highly interdependent, but not for independent tasks.

Implications for Design

These studies indicate that the benefits of harnessing the power of the crowds critically depends on the type of tasks involved. Since most real world systems involve many different types of tasks, it will be important for designers to take into account task-level coordination dependencies when designing large-scale social collaboration systems.

Predicted interdependent issues
By number of editors and editor concentration

Figure 1. Joint influence of number and concentration of editors on number of issues cited in reviews relating to interdependent work tasks.

Below we outline a set of design recommendations based on this implication.

Our results indicate that for tasks that are highly interdependent, concentrating edits is especially useful. This suggests that support for more centralized document control for interdependent tasks such as consolidation or restructuring could be beneficial. Although formal centralized control methods may not be consistent with the ideology of wiki systems and could even have a detrimental effect of driving away users who would otherwise contribute, informal methods may still be useful. For example, editors could take on an "article guide" title for periods of time in which the article needs restructuring, such as often happens in featured article review. The "article guide" may have privileged tools, such as temporarily locking parts of the page or semi-blocking them so that different sets of users do not conflict with each other while the article undergoes significant change. Even without special tools the presence of such a role could reduce conflict and coordination costs by helping identify a person who is invested in improving the page and who will coordinate the efforts of other users and work to build consensus and reduce conflict.

Another recommendation is to develop work procedures that allocate different types of tasks at different times to different types of people. While new users or those with low commitment levels may be helpful for completing tasks with low coordination requirements, they may be less useful or even harmful for tasks where coordination requirements are high. It would be possible to direct more experienced and committed users to high-coordination tasks, either through targeted recruitment or by displaying those tasks in areas that low commitment users rarely visit. This approach could be enhanced by automated systems aimed at improving contribution through routing user attention, i.e., "intelligent task routing" [9]. However, our results suggest a potential problem with such systems: channeling attention may not be useful and may even be harmful to articles when the work tasks are highly interdependent. Such systems could be modified to take into account the coordination requirements of the tasks that need to be accomplished and correspondingly alter their recommendations.

More generally, one may be able to better assign individuals to tasks using profiles of the skills of individuals and the coordination properties of tasks. This system would be analogous to classical job matching algorithms used in organizations, e.g., vocational inventories that identify what tasks people are good at and help human resource managers decide where to assign them. For example, having interpersonal skills may be especially important for the small set of people who participate in most of the discussion on articles [20] but less useful for spell-checking. Although in the systems examined here it is unlikely that people would explicitly fill out inventories, it may be possible to learn which tasks people are good at based on their contribution history, such as their expertise in independent tasks such as creating hyperlinks or improving coverage, or in interdependent tasks such as restructuring an article or improving its readability. More research is needed to determine how to classify people and tasks; however, we hope the research and ideas here may provide an initial step towards achieving this goal.

Our results also suggest that mechanisms for structuring work to reduce coordination dependencies could have a significant impact on harnessing the power of the crowd. Artifacts that embody community norms, such as templates that allow similar content to follow a common format, FAQs addressing common issues, and manuals of style to guide formatting, may reduce coordination dependencies and thus allow more editors to effectively contribute. However, communicating these norms to contributors, especially newcomers who may be overwhelmed when beginning to contribute, is a challenge that merits further investigation.

CONCLUSION

Social collaboration systems often involve tasks with a variety of coordination dependencies. Here we demonstrate that these coordination dependencies predict whether a task will benefit from additional contributors, or from concentrating the work in fewer editors. These findings have important implications for the design of large-scale social collaboration systems that require significant coordination between contributors.

ACKNOWLEDGMENTS

REFERENCES
1. Aiken, L. & West, S. *Multiple regression: Testing and interpreting interactions.* Sage, CA (1991).

2. Argote, L. Input uncertainty and organizational coordination in hospital emergency rooms. *Administrative Science Quarterly 27* (1982), 420-434.

3. Atkinson, A.B. On the measurement of inequality. *Journal of Economic Theory 2* (1970), 244-263.

4. Baecker, R.M., Nastos, D., Posner, I.R. & Mawby, K.L. The user-centered iterative design of collaborative writing software. *Proc. CHI,* ACM Press (1993). 399-405.

5. Benkler, Y. Coase's penguin, or Linux and the nature of the firm. *Yale Law Journal 112* (2002), 367-445.

6. Brooks, F.P. *The mythical man month.* Addison-Wesley Longman Publishing Co., Inc. Boston, MA (1975).

7. Cataldo, M., Wagstrom, P., Herbsleb, J.D., & Carley, K. *Identification of coordination requirements:*

Implications for the design of collaboration and awareness tools. ACM Press, New York, 2006.

8. 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).

9. Cosley, D., Frankowski, D., Terveen, L., Riedl, J. SuggestBot: Using intelligent task routing to help people find work in Wikipedia. *Proc. IUI,* ACM Press *(2007)*, 32-41.

10. Galton, F. Vox populi. *Nature 75* (1907), 7.

11. Giles, G. Internet encyclopaedias go head to head. *Nature 438* (2005), 900-901.

12. Gittell, J.H. Coordinating mechanisms in care provider groups: Relational coordination as a mediator and input uncertainty as a moderator of performance effects. *Management Science 48* (2002), 1408-1426.

13. Hill, G.W. Group versus individual performance. Are N = 1 heads better than one? *Psychological Bulletin 91*, 3 (1982), 517-539.

14. Wikipedia Database Download.

    http://en.wikipedia.org/wiki/Wikipedia:Database_download

15. Wikipedia Featured Article Criteria.

    http://en.wikipedia.org/wiki/Wikipedia:Featured_article_criteria

16. Wikipedia Featured Article Statistics.

    http://en.wikipedia.org/wiki/Wikipedia:Featured_article_statistics

17. Kittur, A., Suh, B., Chi, E., & Pendleton, B. A. He says, she says: Conflict and coordination in Wikipedia. *Proc. CHI 2007,* ACM Press (2007), 453-462.

18. Kittur, A., Chi, E., & Suh, B. Crowdsourcing user studies with Mechanical Turk. *Proc. CHI 2008*, ACM Press (2008), 453-456.

19. 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).

20. Kittur, A*.,* & Kraut, R. E. (2008). Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. *CSCW 2008,* San Diego, CA (2008).

21. Kozlow, M., & Bellamy, P. Experimental study on the impact of the 6+1 trait writing model on student achievement in writing. *Association for Supervision and Curriculum Development Conference 2005*, Orlando, FL (2005).

22. Malone, T.W., & Crowston, K. The interdisciplinary study of coordination. *ACM Computing Surveys (CSUR) 26*, 1 (1994), 87-119.

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. Northwest Regional Educational Laboratory. 6+1 Trait Writing: Scoring Continuum. (2001). Available online http://www.nwrel.org/assessment/pdfRubrics/6plus1traits.PDF

25. O'Reilly, T. (2005, Sep 30). What is Web 2.0? Design patterns and business models for the next generation of software. Retrieved Aug 15, 2008, from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/whatis-web-20.html

26. 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.

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

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. Steiner, I. D. *Group process and productivity*. New York: Academic Press (1972).

30. Stewart, G. L. A meta-analytic review of relationships between team design features and team performance. *Journal of Management 32*, 1 (2006), 29.

31. Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. Assessing information quality of a community-based encyclopedia. *Proc. ICIQ 2005*, MIT Press (2005), 442-454.

32. Surowiecki, J. *The wisdom of crowds*. New York: Doubleday (2005).

33. Tushman, M. Work characteristics and subunit communication structure: A contingency analysis. *Administrative Science Quarterly 24* (1979), 82-98.

34. Van de Ven, A., Delbecq, A., & Koenig, R. Determinants of coordination modes within organizations. *American Sociological Review 41* (1976), 322-338.

35. Wilkinson, D.M., & Huberman, B.A. Assessing the value of cooperation in Wikipedia. *First Monday 12*, 4 (2007).