Crowd Diversity and Performance in Wikipedia: The Mediating Effects of Task Conflict and Communication

Ruqin Ren
University of Southern California
Los Angeles, U.S.A
ruqinren@usc.edu

Bei Yan *
University of Southern California
Los Angeles, U.S.A
beiyan@usc.edu

ABSTRACT

Crowd diversity is a key attribute that impacts crowd performance in online collaboration systems. As a structural composition of a crowd, diversity is likely to influence crowd performance through communication processes during collaboration. This study examined how diversity influenced crowd performance under different conditions of task conflict and communication in Wikipedia article production. With a sample of 5,899 articles, we found that contribution diversity positively predicted crowd performance, whereas experience diversity was negatively related to performance. In addition, task communication and conflict partially mediated the relationship between crowd diversity and performance. Task communication positively predicted performance for both forms of diversity. Task conflict, on the other hand, was positively predicted by expertise diversity, but had negative associations with contribution diversity and performance. The findings help unpack the reasons for differential effects of diversity on crowd performance, and demonstrate the importance of including communication variables when studying online crowd collaboration.

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Computer-supported cooperative work, Theory and Models; K.4.3 Organizational Impacts: Computer-supported collaborative work; H.3.5 Online Information Services

Author Keywords

Wikipedia; crowd performance; diversity; conflict; communication

INTRODUCTION

Recent years have witnessed a proliferation of various forms of online collaboration, with prominent examples including open source software development and Wikipedia [5, 8, 24]. Different from traditional offline collaborations, these online

---

[1]The authors contributed equally to the article.

collaborations often rely on large, distributed crowds without formal organizing structures. Yet the crowds can be incredibly intelligent even when working on complex tasks [40, 14].

What makes a large, distributed group of individuals so intelligent? According to theory and research on crowd collaboration, a key factor that contributes to the "wisdom of the crowd" is the diversity contained in a crowd [32]. Large groups are likely to consist of diverse individuals with non-redundant information, perspectives and experiences, which may enhance collective performance. However, to harness the wisdom of the crowd, diversity in a crowd needs to be effectively integrated. The integration of diversity is often not only a mathematical aggregation process, but also a social process that requires communication and coordination within the crowd. Research on group collaboration has found that diversity may have paradoxical effects on group performance because, apart from the potential benefits, differences may also result in higher communication costs and even conflicts [44, 16].

Hitherto, research on Wikipedia has found that the diversity of editors who contribute to Wikipedia articles or projects is related to several group outcomes, including productivity, article quality, and member retention [20, 37, 36]. However, little existing research has directly investigated how crowd diversity influences crowd performance through processes of task conflict and communication.

To address this issue, our study examined task-related conflict and communication as mediators between crowd diversity and performance. The hypotheses were tested with a sample of 5,899 Wikipedia articles using bootstrapped mediation analysis. In the following sessions, we will start by defining and reviewing relevant concepts and theories. We will then propose our hypotheses, discuss analytical method and present our results. The paper will conclude with a discussion of key findings, limitations, and directions for future research.

LITERATURE AND THEORETICAL MODEL

Diversity in Wikipedia

Diversity is the distribution of differences among the members of a group regarding a common attribute [12]. Following this definition, diversity in Wikipedia is treated as a characteristic of the group of editors who contributed to the same article. The theorization distinguishes between two types of diversity: contribution diversity and experience diversity. Contribution

diversity is the difference in editors' within-article edits. Experience diversity is the difference among editors' previous edits within Wikipedia in general. We focus on differences of editors' editing behaviors rather than other commonly used traits such as demographics, because previous research on diversity in virtual groups suggests that the salience of detectable or surface-level demographic differences tends to be attenuated in computer-mediated contexts [3]. Therefore, studying observable behavioral differences among editors better captures the collaboration dynamics in online systems such as Wikipedia.

The study operationalizes contribution diversity as the coefficient of variation of the number of edits each editor contributed to a Wikipedia article. This operationalization closely follows Kittur and Kraut's [20] operationalization of "implicit coordination". In their study, implicit coordination was measured as the Gini coefficient of the editors' number of edits within an article. The same metric was used by Robert and Romero [38] to operationalize "local workload diversity". In the current study, coefficient of variation was chosen over Gini coefficient because we are more interested in the variation rather than the concentration of editors' edits within-articles.

Experience diversity reflects variation in editors' overall editing experience and commitment to Wikipedia. High experience diversity indicates a large discrepancy among editors' experience in Wikipedia. For articles with low experience diversity, the editors tend to be equally senior (or junior) in Wikipedia. The variable is similar to tenure diversity examined in previous research, which was measured as the coefficient of variation of editors' time in Wikipedia [36, 6]. In our study, experience diversity is operationalized as the coefficient of variation of editors' total edits, since actual volume of edits can better represent editors' editing experience within the system. This operationalization is similar to Robert and Romero's [37, 38] construct of "global workload diversity", although they measured the Gini coefficient rather than the coefficient of variation of editors' contributions to other Wikipedia articles.

Task Communication in Wikipedia
Wikipedia articles have dedicated talk pages or discussion pages. These are local, low-cost venues that editors can use to communicate with each other. They provide a forum for editors to coordinate their editing work, discuss policies and procedures, resolve conflicts, and invite assistance from other editors [22]. A content analysis of Wikipedia talk pages found that talk pages are most often utilized to request coordinated editing, request other people to edit, request to join a discussion, request an opinion, request information, request tasks to be carried out, and FYI information [26]. Content that could not be categorized into this schema accounted for only 2% of total content. Hence, the amount of edits in article talk pages are good indicators of the amount of communication efforts devoted to article production. Thus, the volume of talk page discussion is used to reflect the level of task communication in Wikipedia article editing. It is the same metric used in [20].

Task Conflict in Wikipedia
Conflict in groups is defined as incompatibilities among group members [15]. Task conflict widely exists in Wikipedia ar-

ticle production process [22, 45]. One common behavioral manifestation of task conflict in Wikipedia is edit reverts [43, 22]. A revert refers to a situation in which an editor changes an article back to the previous version. Any work done on the article before the revert (including the revert itself) is lost [22]. A revert is likely to happen when one editor find the previous edits incompatible with her opinions. It suggests difficulty in reaching consensus, incompatibility among editors, and may even lead to editing wars [43, 22]. It is therefore a strong indicator of conflicts within the group of editors who work on the same article. Conflict manifested as reverts may sometimes benefit article production by protecting edits from vicious vandalism. However, if editors of an article spend too much time revoking each others' edits rather than contributing new contents to the article, the quality of the article is likely to suffer [1].

Related Work
To investigate the relationship among diversity, conflict, communication and crowd performance, the current study builds on the three bodies of work. The first body of work is rooted in theory and research on the wisdom of the crowd. Previous research on crowd collaboration generally holds a positive view of diversity, treating it as one of the most critical bases for the wisdom of the crowd [31, 40]. Diversity benefits the performance of the crowd in at least two ways [32]. First, as diversity increases in the crowd, the aggregated crowd performance gets closer to the mean. For example, in a problem solving task, diversity ensures that the average crowd performance can at least achieve the group mean. In collective estimation tasks, diversity cancels out errors in individual estimates, making the aggregated result closer to the real value [40]. Second, the marginal return of one category (one specific type of expertise) decreases as the category gets larger. Having a group of participants distributed across a number of categories, compared to having all group members concentrated in the same category, is likely to improve the performance of the whole group.

Nevertheless, this view implicitly assumes a mathematical aggregation process of human collaboration that is neutral and conflict free. It ignores the fact that synergy of differences is also a social process which may result in various collective dynamics. Simple aggregation may work well only when social interaction is completely absent in the collaboration process.

The second body of research, drawing from small group research, suggests that diversity in interactive groups can sometimes result in detrimental group dynamics, such as high cost of communication and increased conflicts [41, 42, 16, 44]. In fact, homogeneous groups tend to have greater levels of cohesion and more positive affect among members [23]. Differences in readily detectable attributes (such as race, ethnicity, language, gender, and age) are more likely to negatively influence group identification, psychological safety, and group satisfaction, particularly in the early formation stages of working teams [25].

Yet it may be difficult to directly apply research on small groups to the current context, since online crowd collaboration bears a number of features that are distinct from traditional

offline collaboration. To begin with, the crowd is much larger in size. Research on small groups typically studies groups or work teams with 3 to 20 members. In online collaborations, it is not uncommon to have hundreds or even thousands of individuals working together on the same task. The significant increase in size is likely to alter the internal dynamics in collective actions, influencing members' contributions and interactions with each other [29]. A second and related feature of crowd collaboration is that unlike small, face-to-face groups, the crowd in online collaborations are often geographically dispersed. They work together asynchronously via mediated communication, or sometimes, with no communication at all. Research on virtual teams has had inconsistent findings about diversity's effects on group performance. Some has found positive influences of diversity on group dynamics and performance [39], while others have suggested negative effects on group performance [33]. Overall, the research suggests that the effects of detectable demographic differences tend to be attenuated because crowd participants do not often interact face-to-face [3]. However, conflicts can happen in collaborations via deep-level, cognitive diversities (e.g. differences in behavior and expertise) [1]. The current study thus focuses on the latter type of diversity.

The third body of work comes from previous research on Wikipedia and other online collaborative communities. The research has, directly or in directly, examined the influence of several types of diversity on collective outcomes [20, 37, 6]. For example, in open source software communities, diversity in contribution and reputation was positively related to success of software projects[7]. Tenure diversity in Wikipedia had an inverted-U shaped curvilinear relationship with group productivity and member retention [36]. Yet in another study, diversity of editor experience was negative associated with crowd performance [37]. In addition, diversity of editors' within-article contribution was found to be positively related to crowd performance in Wikipedia [20, 36, 37].

Wikipedia is an online system that crowdsources encyclopedia writing tasks. Collaborative writing tasks on Wikipedia exhibit two structural attributes. On one hand, collaborative writing is a divisible task that can be broken into parts, which can then be completed independently by each editor. This enables contributions to be distributed and asynchronous, making it possible to harness the diversity of the crowd consisting of editors with various levels of expertise and availability. On the other hand, the task is synergistic rather than merely additive, because contributions made by each editor need to align with others' to make an article flow as a coherent whole. Thus, communication and coordination is necessary during collaboration to integrate the parts contributed by different editors. Conflicts may also happen during the process. These processes variables may mediate the relationship between diversity and performance. Nevertheless, these mechanisms still remain relatively under-studied. Our study aims to contribute to existing research by including two process variables – task communication and conflict, to unpack the mechanism through which diversity influences performance in online crowd collaboration.

Theory and Hypotheses
Diversity and Crowd Performance
When editors work independently on different parts of a Wikipedia article, crowd diversity is likely to have a direct effect on the contents they contribute, even without any direct collaboration among them. Yet different forms of diversity may have distinct effects on online collaboration outcomes. Contribution diversity indicates the differences in within-article contribution among the group of editors who worked on the same article. High contribution diversity means that the distribution of contribution across all editors working on one article is uneven. In other words, a small number of editors made most of the article edits while others contributed only a small portion of the total edits. On the contrary, low contribution diversity indicates that editors of an article have made even contributions. Previous research on online contribution has suggested that having an active core group of participants who outline the task structures can benefit from task coordination and motivate contribution from peripheral members [7, 30]. Diverse contribution among editors within a Wikipedia article can be considered as a form of implicit coordination among the editors, which helps to improve collaboration efficiency and outcome [20]. Therefore, it is hypothesized that:

H1: Contribution diversity is positively related to crowd performance.

On the other hand, experience diversity represents differences in editors' overall editing experience in Wikipedia. Editors of a particular Wikipedia article often come with different levels of editing experience, commitment to the task, and familiarity with community policies and rules [11]. These differences are likely to result in varied attitudes and opinions about how and what to contribute to a particular article. When experience diversity is high in an article, the article is likely to have a higher diversity of perspectives, writing styles, and levels of clarity. When these distinctive edits are put together into one piece, they are less likely to flow well. More coordinating and structuring work needs to be done to fix an article written by editors with diverse experiences, but strong top-down coordination is often lacking in online systems like Wikipedia. Previous research on Wikipedia has found a negative effect of experience diversity on collaboration outcome [37, 38]. Thus, we hypothesize:

H2: Experience diversity is negatively related to crowd performance.

The Mediating Effects of Task Conflict and Communication
The synergistic feature of collaborative writing in Wikipedia inevitably leads to communication (direct or indirect) during collaboration process, since some editors have to work on re-structuring the article and reconcile the inconsistencies generated by different editors. Conflicts are more likely to happen when a highly diverse group of individuals interact with each other [41, 44]. Based on existing literature, it is argued that the effects of diversity on crowd performance in interactive groups are at least partially mediated by the conflict and communication processes during collaboration. Also, the two types of diversities – contribution diversity and experience diversity, are likely to have distinct effects on internal conflict

and communication processes due to the features of crowd collaboration in Wikipedia.

The divisibility of collaborative editing in Wikipedia allows editors to sometimes work independently on one part of the article. Independence prevents the direct clash of different perspectives and opinions, limiting potential conflicts among the editors working on the same article. Therefore, contribution diversity, which decreases the chance of direct encounters among participants, is likely to reduce task conflicts. In addition, when contribution diversity is high, editors of the same article contribute unevenly in the article, implicitly agreeing to take on different roles in the collaboration process [20]. This is also likely to lower the probability of disagreements and task conflicts during collaboration.

H3: Contribution diversity is negatively related to task conflict within a crowd.

Moreover, high contribution diversity should lead to higher task-related communication because editors who are doing the majority of the work need to inform other collaborators about the work structure to avoid overlapping efforts and to facilitate coordination. It is common practice for Wikipedia editors to claim their responsibilities in article talk pages. We thus argue that while contribution diversity should be negatively associated with task conflict within Wikipedia, it should be positively related to task communication in article talk pages:

H4: Contribution diversity is positively related to task communication within a crowd.

On the contrary, experience diversity, which indicates difference in overall editing experience, may impede the synthesis of different edits made by editors due to a lack of common ground and mutual understanding [4]. In a group of editors with high experience diversity, senior editors are more likely to come across contributions made by newcomers. When the newcomers' edits do not meet the requirements of the Wikipedia community (which is very likely), senior members are likely to modify or revert those contributions. This unwelcoming behavior from senior editors is likely to generate negative feelings from the junior editors, decreasing their continued interest and commitment to Wikipedia [11]. Furthermore, if junior editors modify the edits of senior members, another round of reverting from senior editors is likely to occur. Therefore, conflict is more likely to occur when experience diversity is high.

H5: Experience diversity is positively related to task conflict within a crowd.

However, these disagreements in article editing may also lead to increased communication with their collaborators, as editors may choose to express their different opinions and dissatisfaction in the talk page. In addition, when experience diversity is high, interactions among highly diverse participants during article synthesis is more likely to occur. Yet the difference in their experiences may lead to very different writing styles and perspectives. More task-related communication is thus needed to reconcile and resolve these differences and make the article flow as a consistent whole. Hence:

H6: Experience diversity is positively related to task communication within a crowd.

Previous research on small groups has generated inconsistent findings about the effects of conflict on group performance [15]. While earlier findings have suggested that conflict within groups may lead to detrimental group outcomes due to reduced member productivity and satisfaction [44], later studies have found that conflict can be beneficial to group creativity and innovation if disagreements can promote deeper discussion and cognitive processing of group members [28, 1]. Jehn [15] suggested that the inconsistency may be resolved by specifying what type of conflict a study investigates. In our study, the focus is on behavioral task conflict among editors – a process in which editors revoke each others' edits. This focus is different from previous studies that have examined conflicts as communication efforts to resolve disagreements on talk pages [1] or member's perception of disagreement in groups [16]. This is because perception of disagreements and verbal disputes of different views may not necessarily lead to direct detrimental behaviors to the group product. In fact, if diverse views can be fully discussed during group communication, diversity could benefit group outcome and creativity [28, 1].

Nevertheless, task conflict manifested in editing behaviors in Wikipedia not only demonstrates disagreements and incompatibility among editors, but also generates process loss in article production. Frequent reverts at least suggest inefficiency in the group collaboration process. In extreme cases, it may lead to editing wars that interrupt article production and severely harm article quality [43, 22]. Task conflict, beyond its contribution to creative abrasion, may impede collaboration and have negative consequences [1]. Crowds in online collaboration communities may suffer more from behavioral task conflict, since the coordination of conflict resolution is more difficult virtually than in face-to-face groups[27]. Thus, it is expected that:

H7: Task conflict is negatively related to crowd performance.

In addition, Wikipedia talk pages are often a place for editors to coordinate their work, discuss their views and resolve disagreements. Previous study on Wikipedia has found that the level of within article communication and conflict level is positively associated [21, 22]. Yet the causal direction remains unclear since previous studies were correlational. In the current study, we argue that the collaboration process starts with diverse inputs from article editors, yet differences may lead to conflicts. When conflicts happen, editors are likely to turn to article talk pages to inquire about the situation, coordinate their editing efforts and discuss solutions.

H8: Task conflict is positively related to task communication within a crowd.

Lastly, task-related communication provides an interactive channel for the discussion of task-related issues, which helps coordinate editors' contributions and synthesize their ideas. It could also help resolve disagreements and conflicts, thus mitigating the negative effects caused by differences and helping the crowd harness its diverse resources. Empirical study in Wikipedia has found that task communication was a plausible
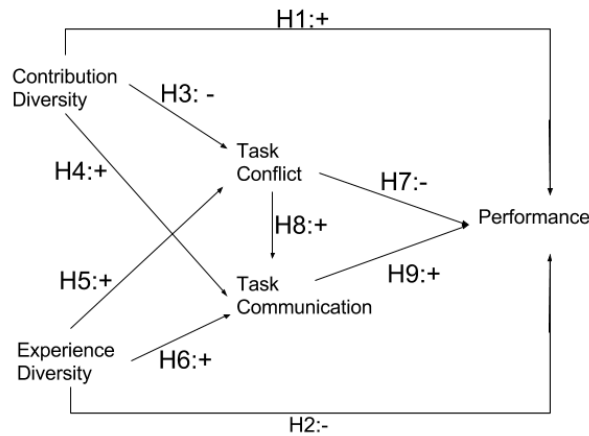
Figure 1. Theoretical Model

Geography	2603
History	1183
Philosophy	571
Religion	484
Mathematics	482
Sociology	205
Technology	191
Science	97
Culture	83

Table 1. Number of articles sampled from each topic category

way of conflict resolution [45] and was positively associated with article quality in Wikipedia [20]. Particularly, research on diversity in Wikipedia has showed that efforts to resolve disagreements on talk pages benefited collective performance when the crowd was diverse [1]. Therefore:

H9: Task communication is positively related to crowd performance.

The summary of this proposed theoretical model is shown in Figure 1.

METHOD AND DATA

The data collection was conducted by accessing Wikipedia API and online query service. A stratified sampling was conducted on nine WikiProjects, which are: History, Sociology, Geography, Culture, Technology, Mathematics, Science, Philosophy, and Religion. Since topic controversy may influence within-article conflict, these nine Wikiprojects were chosen to cover topics of various conflict levels [19]. A similar approach was used by Arazy and colleages [1]. For each article in the data set, its complete editing history was collected between the article's creation time to April, 2016. The initial sampling resulted in 6,000 articles. Among them, 101 articles had only one editor – meaning there's no potential social interaction at all. These were excluded from the sample, leaving a final data set of 5,899 articles.

The unit of analysis, a crowd, is defined as all the editors who have contributed to a particular article. For each article, we collected the complete list of editors who had contributed to the article, as well as their previous editing history. The data set included 142,908 unique editors. An average editor contributed to about 4.16 articles. Bots were excluded from the editor list. Table 1 presents the number of articles sampled from each topic category.

Dependent Variable

The dependent variable, crowd performance, was measured as an article's *quality score*. We collected each article's peer-reviewed quality score (the 5-point scale from high to low

is: Featured Article, Good Article, A-quality, B-quality, C-quality, and Start-quality. A-level articles are relatively few so we collapsed A-level with Good Article). This quality score has been commonly used in previous works and has been validated by external reliability tests [18, 20, 35]. Then, each article was assigned a quality score from 5 to 1 corresponding to letter grade *Featured, Good/A, B, C, Start*, respectively.

Independent Variables

Contribution diversity: Contribution diversity was measured as the coefficient of variation of the number of edits each editor made to an article (log base-10 transformed).

Experience diversity: Experience diversity was measured as the coefficient of variation of each editor's previous edits in all Wikipedia articles (log base-10 transformed).

Mediators

Task Conflict: Task conflict was measured as a function of number of reverts in an article. To identify reverts, we computed a unique identifier of every revision made to every article using the SHA-1 hashing scheme (commonly used to check if text files are identical), and identified it when a later revision exactly matched the hash of a previous article. This is consistent with previous research [22, 45] that used the MD5 hashing scheme to detect reverts. To operationalize conflict, revert counts were normalized with the article length. The rationale was that the same number of reverts to a short article should represent higher level of conflict than a longer article. The Pearson correlation between article length and total edits in an article is 0.7 ($p < 0.001$) in our sample.

Task Communication: Task communication was measured as the number of edits in an article's talk page (log base-10 transformed).

Control Variables

Article Size: Total number of editors who contributed to an article (log base-10 transformed).

Article Age: The time difference between an article's creation time and April, 2016, the data collection time (log base-10 transformed).

Average experience: The average number of edits contributed by an article's editors in Wikipedia (log base-10 transformed).

Method

The hypotheses were tested using bootstrapped mediation analysis [13] using the PROCESS macro in SPSS. This macro

provides a tool to run mediation analysis proposed by Preacher and Hayes [34]. The analysis uses an ordinary least squares-based path analysis to estimate direct and indirect effects. Bootstrap and Monte Carlo confidence intervals are used by the model to estimate indirect effects. This method was chosen because one of the main goals of the study was to introduce conflict and communication as two mediating variables to help explain the relationship between crowd diversity and performance. The VIF test results of the variables were all below 10 and the tolerance values were all above 0.1, suggesting that there were no serious multicollinearity issues within our data. Table 2 and 3 present the summary statistics and correlation matrix of variables. The variables were log transformed because preliminary analysis of the data set revealed that some of the independent variables were positively skewed.

Statistic	Mean	St. Dev.
Article Quality	1.42	0.78
Contribution Diversity	1.61	1.16
Experience Diversity	2.39	0.88
Task Conflict	0.004	0.04
Communication	63.01	565.69
Article Size	99.21	196.10
Article Age	7.97	0.71
Average Experience	418775.73	328628.53

Table 2. Summary Statistics

RESULTS

Figure 2 graphically summarizes the results of the mediation analysis. The number of bootstrapping was set to 10,000. The coefficients were unstandardized. The control variables were entered into the model but not shown in Figure 2. Their coefficients can be found in the specific model equations below.

$$ArticleQuality = i_1 + 2.01 * ContributionDiversity - 0.84$$
$$* ExperienceDiversity + 0.43 * Size - 1.53$$
$$* Age - 0.10 * AverageExperience + e_1$$
(1)

$$TaskConflict = i_2 - 0.88 * ContributionDiversity$$
$$+ 0.52 * ExperienceDiversity$$
$$+ 0.68 * Size - 0.04 * Age(nonsig.)$$
$$- 0.30 * AverageExperience + e_2$$
(2)

$$TaskCommunication = i_3 + 0.91 * ContributionDiversity$$
$$+ 0.55 * ExperienceDiversity$$
$$+ 0.20 * TaskConflict + 0.74$$
$$* Size - 2.21 * Age - 0.19$$
$$* AverageExperience + e_3$$
(3)

$$ArticleQuality = i_4 + 1.70 * ContributionDiversity$$
$$- 0.80 * ExperienceDiversity$$
$$- 0.25 * TaskConflict + 0.13$$
$$* TaskCommunication + 0.49 * Size - 1.27$$
$$* Age - 0.14 * AverageExperience + e_4$$
(4)

Model (1) regressed article quality on contribution diversity and experience diversity without mediators. H1 predicted that contribution diversity is positively related to crowd performance, and this was supported ($\beta = 2.01, p < .001$). H2 hypothesized a negative relationship between experience diversity and crowd performance. This was also supported ($\beta = -.84, p < .001$). The R^2 of this model was 0.54.
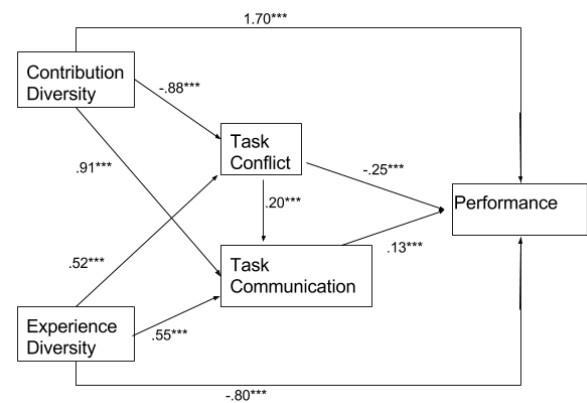

Figure 2. Model results. ***p < 0.001,**p < 0.01,*p < 0.1

To test H3 to H9, the following models added the two mediators, task conflict and task communication. As demonstrated in Model (2), contribution diversity negatively predicted task conflict ($\beta = -.88, p < .001$), whereas experience diversity was positively associated with conflict ($\beta = .52, p < .001$). Therefore, H3 and H5 were supported. Model (2) had an R^2 of 0.69.

H4 and H6 predicted that both types of crowd diversity would be positively related to task communication. Model (3) supported the two hypotheses by showing that both contribution diversity ($\beta = .91, p < .001$) and experience diversity ($\beta = .55, p < .001$) were positively related to task communication. In addition, task conflict was also positively associated with task communication ($\beta = .20, p < .001$), supporting H8. Model (3) had an R^2 of 0.72.

Model (4) is the full model including the independent variables and the mediators. The model shows that task conflict was negatively associated with crowd performance ($\beta = -.25, p < .001$), supporting H7. In contrast, task communication positively predicted performance ($\beta = .13, p < .001$), providing support for H9. After including the mediators, the effects of crowd diversity were still significant, suggesting a partial mediation effect of task conflict and communication on the relationship between crowd diversity and performance. The R^2 of the final model was 0.57.

	1	2	3	4	5	6	7
1. Quality							
2. Contribution Diversity	0.53						
3. Experience Diversity	0.33	0.43					
4. Task Conflict	-0.00	0.03	0.10				
5. Communication	0.19	0.26	0.25	0.03			
6. Article Size	0.38	0.45	0.69	0.09	0.49		
7. Article Age	0.14	0.22	0.58	0.05	0.08	0.31	
8. Average Experience	-0.23	-0.28	-0.55	-0.07	-0.11	-0.35	-0.23

Table 3. Correlation Matrix.

We also tested the robustness of our findings following Baron and Kenny's [2] procedure, applying ordinal logistic regression when the dependent variable is ordinal. The results corroborated the findings observed in the above bootstrapped mediation analysis.

DISCUSSION

This study examined the relationship between crowd diversity and performance in Wikipedia article production under different conditions of task conflict and communication. With a sample of 5,899 articles, the study found that contribution diversity was positively related to crowd performance, whereas experience diversity negatively predicted performance. The bootstrapped mediation analysis revealed that task communication and conflict partially mediated this relationship: task communication was positively predicted by both forms of diversity, and positively predicted performance for crowd diversity as well. Task conflict, on the other hand, was positively predicted by expertise diversity, but had a negative association with contribution diversity and performance. Taken together, the findings suggest the importance of including communication variables in the study of online crowd collaboration.

The results demonstrated that the relationship between crowd diversity and performance was partially mediated by communication processes. While previous work on crowd collaboration has examined the effects of diversity on crowd performance, few studies have directly included communication variables to further investigate the relationship. Our results suggest that the positive effect of contribution diversity on performance is partly due to its positive association with task communication and negative association with task conflict. The mechanism by which experience diversity depressed performance is partially due to task conflict associated with it. In addition, task communication helped to mitigate the negative effects of experience diversity and conflict on crowd performance.

These findings help reconcile the mixed findings about the effects of diversity on group performance [12] by showing that different types of diversity within a crowd may have differential effects on a crowd's communication processes and collaboration outcome. Contribution diversity, which indicates the difference among the editors' contribution level within an article, was positively associated with crowd performance. Highly uneven workload distribution across the crowd is likely to suggest an implicit division of labor among the article's editors: with a group of core editors who set up the scope and structure of an article, and the rest of the workers voluntarily (also less actively) contribute to the article as outlined by the core editors. This diversity makes collaborative editing easier to accomplish and could improve coordination efficiency [20].

In contrast, experience diversity was negatively associated with conflict and performance. High experience diversity reflects greater difference in editors' previous contributions and commitment to Wikipedia. Existing research on online communities suggests that online communities suffer from high turnover, and retaining newcomers can be important to keep a community lively [17]. However, this finding suggests possible gate-keeping behavior of old-timers during collaboration. It helps to explain the finding that newcomers are hard to retain in Wikipedia [10]. Distributed collaborative communities often experience conflicts about group norms. When experience diversity is high, conflicts are more likely to happen because newcomers who are not familiar with existing community norms are working with experienced contributors who may find their contributions unsatisfying. This negative experience in newcomers' initial interaction with the community may keep the newcomers away from the community. Explicit communication on talk pages, however, seems to be a helpful way to mitigate this negative effect. Newcomers should therefore be encouraged to explicitly communicate their task-related concerns and coordinate with other editors on talk pages to reduce misunderstandings among the collaborators.

One note of caution is that existing research on groups has generated mixed findings about the effects of conflict on performance [15]. Conflict can sometimes be detrimental to group task performance, but may also stimulate innovation and creativity within groups [28]. The mixed findings about the effects of conflict may have to do with type of conflicts researchers investigated [16]. The negative effect of conflict found in the current study partially has to do with the operationalization of our construct. Conflict here was operationalized as number of reverts normalized by article length. It directly measures the density of editing reverts in a Wikipedia article, but does not tap into other forms of conflict in editor discussion and interaction. Previous research on Wikipedia has found that conflict, as measured by efforts to resolve disagreements in talk pages, can benefit group outcome when the crowd is more diverse [1]. So we are cautious in claiming that all task conflicts have negative effects on performance in online communities.

We are also cautious about making causal claims over the association between conflict and communication. Kittur and colleagues [21] found that greater communication was associated with greater task conflict. Our finding is consistent with

theirs, but the study does not claim causal relationship between the two variables since the data is correlational. In fact, Kittur et al [21] took a similar stance and emphasized in their work that "...the proportion of article talk was positively associated with conflict. It is difficult to distinguish whether article talk is a cause of conflict or a marker of conflict, since much conflict happens during the discussion of an article (p.222)."

Practical Implications

The results also provided practical implications for designers of crowd collaboration systems. First, diversity is a collective level attribute and is relational in nature, since a single editor cannot be "different" from a nonexistent other. The positive effect of contribution diversity on crowd performance implies that crowd collaboration systems should allow and encourage participants to make different levels of contributions to an article in order to reach the best collective outcomes. This can be achieved by rewarding various forms and levels of contribution within the same article. Currently, several Wikipedia barnstars (such as Typo, Minor, Copyeditor) reward editors who contribute high quality, minor edits. Yet these barnstars tend to reward consistent individual efforts rather than contribution in relation to other editors' behaviors. New barnstars such as "Coordinator" can be created to award editors who make minor edits when the overall structure is already laid out or editors who proactively take a leadership role when most other editors are making minor contributions.

On the other hand, the negative consequences of experience diversity suggest that online collaboration systems should help newcomers to adapt to the community faster and encourage senior editors to be more friendly to new comers' contributions. For example, Wikipedia can identify edits that are contributed by new editors, and the system can remind other editors to be friendly to new newbies' contributions when editors try to revert newbies' edits. In the meantime, the community can cultivate welcoming and friendly behaviors from existing members to new contributors by rewarding barnstars such as "Newbie Adviser" to experienced editors who spend time tutoring newcomers' editing behavior.

The tutorial page for newcomers should not only list rules of the community, but also encourage new members to express their concerns, learn about community rules and coordinate with other editors using the talk page, since our findings have demonstrated that communication in the talk page could help mitigate the negative effect of experience diversity. Senior editors should also be encouraged to use the talk page to communicate with other collaborators before reverting others' contributions directly. A "Communicator" barnstar can be created to reward editors who use talk pages frequently to coordinate with other editors.

Limitations and Future Directions

The study has a number of limitations. To begin with, the data used was cross-sectional which prevents us from making causal claims based on these statistical correlations. Nevertheless, since crowd diversity is a compositional attribute, the mediating effects of group processes (for example, communication and conflict) is a logical argument based on theory and empirical evidence. Second, previous research has found that crowd size may moderate the relationship among crowd diversity, communication, and crowd performance [37, 38, 20]. In the current study, since the main purpose was to identify the mediating effects of conflict and communication, crowd size was included as a control variable rather than a moderator in the analysis. Future studies could further investigate whether the relationship in the proposed model is also contingent upon crowd size.

In addition, this study only sampled articles from Wikipedia projects. Some other collaboration patterns may exist in articles that were not included in Wikipedia projects. These distinctive patterns may prevent us from generalizing this finding to all articles in Wikipedia. Lastly, previous literature has pointed out that various types of conflicts exist in online collaboration systems [9]. Exploring the effects of other forms of conflict, such as affective conflict or relational conflict, is a very interesting possibility for future investigation.

It is worth noting that the mediating effect of task conflict and communication was partial rather than full, suggesting that crowd diversity could directly influence performance without going through processes of conflict and communication. This could be caused by Wikipedia's task structure and collaboration structure. The core task in Wikipedia is collaborative editing, which is a divisible task. However, the integration of the parts requires synergy rather than mere addition. As a result, editors in Wikipedia can work on selected parts of an article independently most of the time, but their individual work needs to be combined to become a consistent and coherent article. This then requires the editors to coordinate with each other through direct or indirect interaction. Therefore, crowd diversity can directly influence crowd performance even without communication among the editors through their independent work, but it also impacts the final product indirectly via the communication and conflict processes during article integration.

We hypothesize that conflict and communication would have stronger mediating effects between crowd diversity and performance as the degree of social interaction within the crowd increases. In tasks where individuals give independent solutions without communicating with each other, the effects of conflict and communication will be negligible. However, when groups work on a complex, conjunctive task that needs every member's input, the social interaction to facilitate coordination is highly essential, and crowd diversity needs to be integrated and reconciled in the interaction processes in order to achieve good performance at the collective level. In the latter case, the mediating effects of communication and conflict should be much stronger. Future studies can test this hypothesis by categorizing crowd collaboration systems based on the level of task interaction, and compare the effects of conflict and communication on the relationship between crowd diversity and performance across different task types.

REFERENCES

1. Ofer Arazy, Oded Nov, Raymond Patterson, and Lisa Yeo. 2011. Information quality in Wikipedia: The effects of group composition and task conflict. *Journal of Management Information Systems* 27, 4 (2011), 71–98.

2. Reuben M Baron and David A Kenny. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* 51, 6 (1986), 1173.

3. Zoe I Barsness, Kristina A Diekmann, and Marc-David L Seidel. 2005. Motivation and opportunity: The role of remote work, demographic dissimilarity, and social network centrality in impression management. *Academy of Management Journal* 48, 3 (2005), 401–419.

4. Beth A Bechky. 2003. Sharing meaning across occupational communities: The transformation of understanding on a production floor. *Organization science* 14, 3 (2003), 312–330.

5. Yochai Benkler. 2006. *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

6. Jilin Chen, Yuqing Ren, and John Riedl. 2010. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 821–830.

7. Sherae Daniel, Ritu Agarwal, and Katherine J Stewart. 2013. The effects of diversity in global, distributed collectives: A study of open source project success. *Information Systems Research* 24, 2 (2013), 312–333.

8. Samer Faraj, Sirkka L Jarvenpaa, and Ann Majchrzak. 2011. Knowledge collaboration in online communities. *Organization science* 22, 5 (2011), 1224–1239.

9. Anna Filippova and Hichang Cho. 2015. Mudslinging and Manners: Unpacking Conflict in Free and Open Source Software. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1393–1403.

10. Aaron Halfaker, R Stuart Geiger, Jonathan T Morgan, and John Riedl. 2012. The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist* (2012), 0002764212469365.

11. Aaron Halfaker, Aniket Kittur, and John Riedl. 2011. Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work. In *Proceedings of the 7th international symposium on wikis and open collaboration*. ACM, 163–172.

12. David A Harrison and Katherine J Klein. 2007. What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of management review* 32, 4 (2007), 1199–1228.

13. Andrew F Hayes. 2009. Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication monographs* 76, 4 (2009), 408–420.

14. Jeff Howe. 2008. *Crowdsourcing: How the power of the crowd is driving the future of business*. Random House.

15. Karen A Jehn. 1995. A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative science quarterly* (1995), 256–282.

16. Karen A Jehn, Gregory B Northcraft, and Margaret A Neale. 1999. Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative science quarterly* 44, 4 (1999), 741–763.

17. Elisabeth Joyce and Robert E Kraut. 2006. Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication* 11, 3 (2006), 723–747.

18. Gerald C Kane and Sam Ransbotham. 2016. Content and Collaboration: An Affiliation Network Approach to Information Quality in Online Peer Production Communities. *Information Systems Research* (2016).

19. Aniket Kittur, Ed H Chi, and Bongwon Suh. 2009. What's in Wikipedia?: mapping topics and conflict using socially annotated category structure. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1509–1512.

20. Aniket Kittur and Robert E Kraut. 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, 37–46.

21. Aniket Kittur and Robert E Kraut. 2010. Beyond Wikipedia: coordination and conflict in online production groups. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM, 215–224.

22. Aniket Kittur, Bongwon Suh, Bryan A Pendleton, and Ed H Chi. 2007. He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 453–462.

23. Terri R Kurtzberg. 2005. Feeling creative, being creative: An empirical study of diversity and creativity in teams. *Creativity Research Journal* 17, 1 (2005), 51–65.

24. Karim R Lakhani and Eric Von Hippel. 2003. How open source software works:'free' user-to-user assistance. *Research policy* 32, 6 (2003), 923–943.

25. Frances J Milliken, Caroline A Bartel, and Terri R Kurtzberg. 2003. Diversity and creativity in work groups: A dynamic perspective on the affective and cognitive processes that link diversity and performance. *Group creativity: Innovation through collaboration* (2003), 32–62.

26. Jonathan T Morgan, Michael Gilbert, Mark Zachry, and David McDonald. 2013. A content analysis of Wikiproject discussions: Toward a typology of coordination language used by virtual teams. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*. ACM, 231–234.

27. Mark Mortensen and Pamela J Hinds. 2001. Conflict and shared identity in geographically distributed teams. *International Journal of Conflict Management* 12, 3 (2001), 212–238.

28. Charlan J Nemeth, Bernard Personnaz, Marie Personnaz, and Jack A Goncalo. 2004. The liberating role of conflict in group creativity: A study in two countries. *European Journal of Social Psychology* 34, 4 (2004), 365–374.

29. Mancur Olson. 1965. The logic of collective action: Public goods and the theory of groups. *Cambridge, Mass* (1965).

30. Siobhán O'Mahony and Fabrizio Ferraro. 2007. The emergence of governance in an open source community. *Academy of Management Journal* 50, 5 (2007), 1079–1106.

31. Scott E Page. 2008. *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.

32. Scott E Page. 2010. *Diversity and complexity*. Princeton University Press.

33. Linda Peters and Ronald J Karren. 2009. An examination of the roles of trust and functional diversity on virtual team performance ratings. *Group & Organization Management* (2009).

34. Kristopher J Preacher and Andrew F Hayes. 2004. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior research methods, instruments, & computers* 36, 4 (2004), 717–731.

35. Sam Ransbotham and Gerald C Kane. 2011. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quarterly* 35, 3 (2011), 613.

36. Yuqing Ren, Jilin Chen, and John Riedl. 2015. The Impact and Evolution of Group Diversity in Online Open Collaboration. *Management Science* (2015).

37. Lionel Robert and Daniel M Romero. 2015. Crowd Size, Diversity and Performance. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1379–1382.

38. Lionel P Robert and Daniel M Romero. 2016. The influence of diversity and experience on the effects of crowd size. *Journal of the Association for Information Science and Technology* (2016).

39. Pnina Shachaf. 2008. Cultural diversity and information and communication technology impacts on global virtual teams: An exploratory study. *Information & Management* 45, 2 (2008), 131–142.

40. James Surowiecki. 2005. *The wisdom of crowds*. Anchor.

41. Daan Van Knippenberg, Carsten KW De Dreu, and Astrid C Homan. 2004. Work group diversity and group performance: an integrative model and research agenda. *Journal of applied psychology* 89, 6 (2004), 1008.

42. Daan Van Knippenberg and Michaela C Schippers. 2007. Work group diversity. *Annu. Rev. Psychol.* 58 (2007), 515–541.

43. Fernanda B Viégas, Martin Wattenberg, and Kushal Dave. 2004. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 575–582.

44. K Williams and Charles O'Reilly. 1998. The complexity of diversity: A review of forty years of research. *Research in organizational behavior* 21 (1998), 77–140.

45. Taha Yasseri, Robert Sumi, András Rung, András Kornai, and János Kertész. 2012. Dynamics of conflicts in Wikipedia. *PloS one* 7, 6 (2012), e38869.