Cumulative experience and recent behavior and their relation to content quality on Wikipedia

Michail Tsikerdekis

Western Washington University, Department of Computer Science[1]

Research highlights

- Recent behavior a more powerful predictor than cumulative experience as a user characteristic that impact on content quality
- Recent behavior in editing high quality articles impact content quality
- Integrated users in the community have a higher impact on content quality

ABSTRACT

Cumulative experience is often seen as a major factor for influencing content quality in collaborative projects such as Wikipedia. However, past studies often utilize cumulative experience based on the quantity of work rather than quality and context. Moreover, the perspective on cumulative experience assumes a final destination for user behavior, whereas much of the literature indicates that user behavior changes over time. This paper aims to address these two factors by providing better descriptions and context to determine their effect on content quality. The study rematerialized these factors based on

---

1 Communications Facility 495, 516 High Street, Bellingham, WA 98225. U.S.A. Telephone: 360-650-2401, Email: michael.tsikerdekis@wwu.edu

the quality of work and built a comprehensive model of how article quality materializes through groups of users. Regression models in this study indicate that recent behavior is a more powerful predictor of content quality change than cumulative experience. The implications of the findings impact the design of task-routing systems as well as designers and group managers' work in collaborative projects.

Cumulative experience and recent behavior and their relation to content quality on Wikipedia

Research highlights

- Recent behavior is a more powerful predictor than cumulative experience as a user characteristic that impacts content quality
- Recent behavior in editing high quality articles impacts content quality
- Integrated users in the community have a higher impact on content quality

ABSTRACT

Cumulative experience is often seen as a major factor for influencing content quality in collaborative projects such as Wikipedia. However, past studies often utilize cumulative experience based on the quantity of work rather than quality and context. Moreover, the perspective on cumulative experience assumes a final destination for user behavior, whereas much of the literature indicates that user behavior changes over time. This paper aims to address these two factors by providing better descriptions and context to determine their effect on content quality. The study rematerialized these factors based on the quality of work and built a comprehensive model of how article quality materializes through groups of users. Regression models in this study indicate that recent behavior is a more powerful predictor of content quality change than cumulative experience. The implications of the findings impact the design of task-routing systems as well as designers and group managers' work in collaborative projects.

1. INTRODUCTION

Collaborative article editing on Wikipedia has seen a number of studies examining the emergence of content quality (Mesgari, Okoli, Mehdi, Nielsen, & Lanamäki, 2015). The development of Wikipedia articles has been suggested to be the result of a complex, shared, stigmergic process (Elliott, 2006; Loveland & Reagle, 2013) of informal peer review where the actions of individuals (contributions and review) leads to the aggregation of high quality content (Stvilia, Twidale, Smith, & Gasser, 2008). Studies have indicated through different perspectives that this collaborative process has been found to be effective when conditions for implicit coordination involving few editors doing most of the work are met (Kittur & Kraut, 2008), while another perspective has shown that a group's composition in terms of experience diversity is far more important (Arazy & Nov, 2010). However, there has not been enough depth in these studies to demonstrate the types of cumulative experiences that could affect quality and how the experience could be contextualized for the purposes of understanding its effect on quality in such communities. Furthermore, the question of implicit coordination (i.e., few editors are required to do most of the work in order to affect quality) challenges the very notion of democratic production in terms of content on Wikipedia. Debatably, looking at implicit coordination and group composition, it is unclear which of the two has more weight in affecting quality and in what context within the lifecycle of an article. As such, both terms and their interaction offer different perspectives on collaborative editing that have not been thoroughly examined together. Group composition can be defined as the set of characteristics that users in a group may have (e.g., experience), while implicit coordination is how individuals put these skills together as a group towards accomplishing a particular task without explicit guidance.

Collaborative article editing on Wikipedia has seen a number of studies examining the emergence of content quality (Mesgari, Okoli, Mehdi, Nielsen, & Lanamäki, 2015). The development of Wikipedia articles has been seen as the result of a complex, stigmergic process (Elliott, 2006) of informal peer review where the actions of individuals (contributions and review) leads to the aggregation of high quality content (Stvilia, Twidale, Smith, & Gasser, 2008).

Past work has shown that in order to increase article quality, a single editor or a small group of editors generally took the lead – doing most of the contribution work – while a large group of editors contributed less (Kittur & Kraut, 2008). Based on these results, it was speculated that this pattern of *implicit coordination* was effective due to a reduction in coordination costs. This finding challenges previously conducted work by suggesting that it is indeed important for few individuals from a group to take on a leading position in order to affect an article's quality.

More recent work challenged the aforementioned conclusion. It showed how distributing work on an article among such participating individuals is not a top predictor of content quality change. Instead, *experience diversity* (originally referred to as global inequality), representing the distribution of work across the whole of Wikipedia among individuals in a group, is a more powerful predictor of content quality change (Arazy & Nov, 2010). The metric can be thought of as representation of a group's composition based on experience. This shifts the focus from how a group coordinates its work to a group's composition in terms of its members on Wikipedia; this is an easier aspect to engineer and influence from a sociotechnical design perspective. More recently, experience diversity among members of the group has been found to influence the likelihood of conflict in a group's coordination (Arazy, Nov, Patterson, & Yeo, 2011).

In this paper, I examine in more detail how a group's composition based on user characteristics can affect content quality. The aim is to harmonize different proposals into a coherent and comprehensive model that improves on the accounted variance of previous studies and is more descriptive of how users influence quality through their characteristics. The contributions of this work are itemized as follows:

- I expand models based on a group's cumulative experience as well as on the types of experience editors may bring to article work.

- I show that the amount, diversity and type of experience play a strong role in improving content quality on Wikipedia.

- I demonstrate that recent editor behavior can be a more powerful predictor of content quality change than a group's cumulative experience. I specifically show that contributors with editing experience in various spaces such as Wikipedia's policy are critical for the development of an article.

- I demonstrate the interaction between group composition and coordination within the context of an article's lifecycle. The results indicate in greater detail what a "successful" context in the article development process looks like and what constitutes an effective group composition in order to affect outcome quality.

I provide a discussion on the implications of the results highlighting the importance of process/norm literacy for community leaders and users that an evolving community needs to provide to its newcomers in order to sustain itself. In addition to the empirical findings, the study is one of a few instances where a Heckman 2-step selection model has been applied to social computing research.

2. RELATED WORK

In this study, I have examined how group composition and diversity of engagement affect the efficiency of the open production process that makes Wikipedia work. Four scholarly conversations revolve around the topics presented in this paper. These relate to how tasks are selected by users or allocated to users by various mechanisms, how editor roles are explicitly or implicitly assumed by users, the mechanisms of peer-review and damage control that maintain quality in articles, and how group processes can influence the emergence of quality.

2.1. Task Allocation

A rich body of literature covers the self-directed nature of open production systems. Newcomers to open production projects generally start by making minor contributions around content in which they are interested (Masli, Priedhorsky, & Terveen, 2011). Later, they transition to more substantial editing and coordination work (Bryant, Forte, & Bruckman, 2005). This pattern of social learning is referred to as legitimate peripheral participation *(Lave & Wenger, 1991)* and is well described across many different types of open online communities in various contexts (Preece & Shneiderman, 2009). Other work has shown that Wikipedia's editors coordinate their work with each other in hidden and nuanced, yet effective ways (Forte & Bruckman, 2008; Viegas, Wattenberg, Kriss, & Ham, 2007).

The processes by which editors discover the tasks to be performed has measureable effects on the quality and coverage of content produced. For example, editing around one's interests suggests a demand-driven production pattern common to peer production communities (i.e., the most important content is expanded first) (Benkler & Nissenbaum, 2006). This interest-driven process of work allocation also results in some concerning coverage disparities. In Wikipedia, coverage is generally more complete for topics of relevance to contributors (as opposed to readers) (Lehmann, Müller-Birn, Laniado, Lalmas, & Kaltenbrunner, 2014; Warncke-Wang, Ayukaev, Hecht, & Terveen, 2015), to men rather than women (Lam et al., 2011) and to western English-language-dominated cultures (Graham, Hogan, Straumann, & Medhat, 2014).

Not all work is self-directed on Wikipedia. WikiProjects, groups of editors who maintain a specific subject area or type of work, play a central role in work coordination on Wikipedia (Morgan, Halfaker, Taraborelli, Hwang, & Goggins, 2015). Within WikiProjects, some editors will focus primarily on coordinating work for others. This strategy brings attention to important articles that need work. Zhu et al. (2012) showed that WikiProjects that schedule a "Collaboration of the Week" task see both increased work towards the goal of improving the target article as well as an increase in activity peripheral to the explicit goal. Recent work has also shown that more experienced editors

will often coordinate their work in "WikiProjects" by building and iterating through alphabetized worklists (Wattenberg, Viégas, & Hollenbach, 2007).

Some work is even algorithmically directed. Researchers have shown that personalized recommender systems can be used to route contributors to tasks they are unaware of but will likely be interested in working on (Cosley, Frankowski, Terveen, & Riedl, 2007). This strategy can result in an increase in volunteer contributor activity and measurable benefits for the system as a whole (Priedhorsky, Masli, & Terveen, 2010).

2.2. Editor Roles

The behavioral change of community members as they gain experience does not lead to a terminal and singular role, but there are multiple roles from which users can choose and switch between. In Wikipedia, there are many roles (e.g., quality assurance, border patrol, admin) that are assigned to individuals based on their level of experience within the community. Promotions to rules are also dependent on several criteria as well as a consensus that needs to be reached by multiple votes that may also have conflicting opinions (Burke & Kraut, 2008). The transitional flow between these roles is toward roles characterized by a higher level of work and responsibility; there are few examples of users who regress (Arazy, Ortega, Nov, Yeo, & Balila, 2015). Large leaps across many levels are also observed. Users transition from content-focused work to process/policy-focused work.

Another study on Wikipedia attempted to identify roles for individuals who were not explicitly assigned to them but were implicitly assumed by individuals. The results identify rules that can differentiate editors between those who contribute content (named substantive editors), those who focus on process and policy (referred to as technical editors), to those who focus on quality control as well as rule enforcement (referred to as counter-vandalism editors) and those who spend most of their time in discussion spaces (referred to as social networker editors) (Welser et al., 2011). In this study, these roles were identified based on the distribution of work that users had across different spaces on Wikipedia; however, assigning work to single users may be problematic due to overlapping categories in the distribution of work. While these roles cannot be quantifiably identified for each user, the study provides a perspective that users exhibit *recent behavior* that can change over time and is separate from cumulative experience. A more recent study has also attempted to expand on the idea of roles that may influence quality (Yang, Halfaker, Kraut, & Hovy, 2016). For example, the role classified as the "substantive expert" (i.e., someone who adds substantial content as well as references) was found to influence article quality. However, the increase in accounted variance when editor roles were added to the baseline model was small (0.5%).

2.3. Mechanisms of Peer-Review and Damage Control

Since Wikipedia began attracting attention as an openly crowdsourced encyclopedia, researchers have been concerned about its quality. Earlier studies on Wikipedia articles found that their information was accurate overall (Chesney, 2006; Giles, 2005; K. West & Williamson, 2009). However, an early study around the same time discovered inaccuracies where a lack of sources and no attribution for quotations was a major issue for some articles (Rector, 2008). There are three general threads of inquiry around how Wikipedia attained such high quality: open quality process, damage mitigation and the effect of group composition on article quality improvement.

Past studies have examined how Wikipedia's open processes enable crowds of volunteer editors to produce and maintain high quality reference material. An early study on information quality on Wikipedia showed that both formal (e.g., discussion on article content) and informal (e.g., editors review and alter the previous work of others) mechanisms are used to ensure that quality emerges from the process (Stvilia, Twidale, Gasser, & Smith, 2005). The authors speculated that the efficiency of Wikipedia's peer-review mechanism could depend on the quality of the editorial group of an article. Stvilia et al. (2008) compared Wikipedia to traditional models of information quality and argue that Wikipedia's information creation and review process are different but still effective. A reader can become an editor and reverse the changes made by another editor, all without needing to register an account. This mechanism is built in the collaborative project by design. Additionally, there is little lag between product creation and delivery to an end user, as work coordination remains less formal on Wikipedia. Stvilia et al. (2008) argue that the combination of formal and informal methods used by editors on Wikipedia keeps the content current and accurate while contributing to quality improvement.

However, the open nature of Wikipedia and its rising popularity often make it a target for vandals. Wikipedia utilizes a set of tools and distributed processes to organize the community's work efficiently across the crowd of editors. While editors will generally review the articles that they read and edit for damage, most counter-vandalism occurs at the boundaries of Wikipedia (at the time of contribution) (Geiger & Halfaker, 2013; Geiger & Ribes, 2010). This quality control strategy focuses on the need to maintain quality in the face of continuous incoming damaging edits. Geiger and Ribes (2010) argue that the tools and processes that these boundary workers use enable efficient, coordinated, and decentralized action against vandals by editors.

Automated "bots" also play an important role in countering vandalism. These bots use a mixture of machine learning and natural language processing to identify damaging edits (Adler, de Alfaro, Mola-Velasco, Rosso, & West, 2011). Generally, these bots are only able to confidently identify and revert only the most egregious vandalism, but they

do so much faster than humans and minimize the human effort that is required to address the remaining more subtle damage (Geiger & Halfaker, 2013; Geiger & Ribes, 2010; Halfaker, Kittur, & Riedl, 2011).

2.4. Group Processes and Emergence of Quality

Beyond general processes and software that enforce rules to increase the quality of articles over time, some studies have examined group processes and patterns of effective collaboration.

Activity on an article's discussion page (referred to as explicit coordination), a space used for coordinating changes about an article, has a positive effect on quality (Kittur & Kraut 2008). This result was substantiated also by another study that showed that the article discussion space is used for strategic planning of edits and enforcement of standard guidelines and conventions (Viegas et al., 2007). Other studies have found talk pages (another name for the article discussion pages) to be the main source of coordination (Ehmann, Large, & Beheshti, 2008).

An article's age was also found to have a significant effect on quality change (Kittur & Kraut, 2008). However, as the article age is increased the rate of improvements has also been found to decrease (Kane & Ransbotham, 2016a). Lifecycle approaches have also been used to measure how articles evolve and change in quality over time (Wöhner & Peters, 2009). High quality articles tend to develop slowly, and as they mature, editing activity picks up compared to low quality articles that tend to receive high editing activity from their early stages.

Further, the idea of small uncoordinated article changes affecting quality has also been examined and demonstrated by Wilkinson and Huberman (2007). Another study attempted to identify how users come together to collectively achieve quality change in an article (Kittur & Kraut 2008). The authors of the study monitored the activity in a sample of articles over a six-month period and identified factors that are associated with an improvement in article quality that manifested as an increase in Wikipedia's article quality assessment scale. The results suggest that articles increase in quality most efficiently when most of the work in an article is performed by one or a few individuals from a larger group of editors.

Later studies have challenged the previous findings by demonstrating that it is not just how work in an article is distributed among editors (implicit coordination), but how these editors differ in terms of their cumulative experience working in Wikipedia (Arazy & Nov, 2010). Groups with a wide diversity in editor experience more efficiently increase

article quality than groups with less diversity. Similar work has also shown that crowd size and diversity can influence performance (Robert & Romero, 2015). The predictive power of this experience diversity was shown to overshadow that of the implicit coordination observed by Kittur & Kraut (2008). While the finding on experience diversity is impressive in predictive terms, it can lead to incorrect conclusions when investigating a proper group composition to affect quality due to the inability of distinguishing collective group experience. A group whose members are overall highly experienced and the distribution of experience among them is skewed could produce an equal diversity metric (Gini metric in the case of the aforementioned studies) compared to a group with overall low experience that has the same skewed distribution of experience among its members. As such, the practical implications of the finding are limited.

The importance of user experience has also been substantiated by other studies. Ortega (2009) quantitatively demonstrated that featured articles were mostly edited and developed by active and experienced contributors. Apart from the intuition of the need for experienced users, a mix of both new and long-term contributors also appeared to influence the attainment and maintenance of featured status (Ransbotham & Kane, 2011). Shared experience has also been found to have an effect on content quality along with group size (Carillo & Okoli, 2011). The studies have all used proxies for experience that considered experience as the amount of work conducted on articles rather than the quality of articles on which the work was conducted.

A study that sought to investigate quality over quantity of work examined experience in the context of the German-branch of Wikipedia. It linked experience in editing featured articles with an effect on quality change in other featured articles (Stein & Hess, 2007). However, the study was limited to an investigation of only featured articles, which have two levels in the German-branch of Wikipedia.

3. RESEARCH QUESTIONS

In this study, I attempt to identify the conditions and users within a group that can contribute to content quality change by expanding on previous work. This resulted in the creation of variables that provide further details into group collaboration, which can be compared with previously used variables in the literature in order to determine their predictive power regarding article quality.

This study builds on the initial work developed by Kittur & Kraut (2008), which was later expanded and challenged by Arazy and Nov (2010). Both of these studies uniquely formulated predictor models that studied the effect of several variables on article quality change. This is the foundational model expanded by this study.

It has been several years since the initial study by Kittur & Kraut (2008) discovered implicit coordination variables that affect article quality, and since then, Wikipedia has grown in size and evolved as a community. As such, I sought to establish whether the basis for these variables that measured explicit and implicit coordination were still important to article quality.

Q1: Does implicit coordination affect article quality change?

Q2: Does participation on an article's discussion page affect article quality change?

Many studies have demonstrated the significance of group experience for affecting quality (Arazy & Nov, 2010; Carillo & Okoli, 2011; Mesgari et al., 2015; Ortega, 2009; Rosenthal, Rosnow, & Rubin, 2000); however, experience was often used as a generalized measure. Experience was often reduced to measuring quantity of work rather than quality and treated all work as equal. However, work on Wikipedia involves more complex activities (e.g., writing articles or contributing to discussion pages). As such, *cumulative experience* was not properly contextualized in its nature. For example, is experience acquired solely with the passage of time, or is it the sum of all actions that a user has taken in the past? Are there any specific actions that should be given greater consideration as "experience" that can affect quality?

In this study, I expand on previous work such as the study conducted by Arazy & Nov (2010), but I also attempt to identify cumulative experience within the context of the actions of editors and by asserting that all experience is not equal. Moreover, I attempt to overcome the issues caused by experience diversity as a metric for predicting quality change since, as previously mentioned, it is not effective at representing what an ideal group composition should look like in terms of experience. Further investigation on how cumulative experience may interact with other variables in the model can also yield interesting results that will provide more details about the nature of any effect found to be significant with respect to content quality. For example, the degree of integration with the community could influence the dynamics of teams working on the article. The research question formulated here is expanded further in the methods section of this paper.

Q3: What types of cumulative experience affect content quality and to what degree?

Previous research (described in section 2) also suggests that there may be temporal qualities of an editor that can affect outcomes that go beyond what cumulative experience can explain. For example, we know that people can be qualitatively categorized into specific roles and that these roles may change over time (Welser et al., 2011) and such transitions are dependent on an editor's motivational orientation (Arazy et al., 2017). As such, while experience is an accumulating process, roles represent a transformational and

temporal process. Other studies (Arazy & Nov, 2010; Stein & Hess, 2007) have utilized metrics of experience, but they have never examined or compared them with recent behavior, which reflects changes in user attitudes. In this study, I use *recent behavior* to demonstrate that user attitudes can change over time as users may change the manner as well as the type of work they conduct on an online community. As such, recent behavior is perceived as a recent and temporary reflection of a user's cumulative experience. Further interaction effects are encapsulated in the formulated research question in order to establish in more detail the nature of the effect that recent behavior has on content quality. For example, addressing how the number of users and their recent work may influence content quality is an interaction that could potentially impact how tasks are being allocated in communities. The following research question was developed.

Q4: What types of recent behavior can affect content quality, and to what degree?

4. METHODS

I evaluated all research questions by obtaining and analyzing Wikipedia's datasets offered by the Wikimedia Foundation. These consist largely of raw datasets that needed to be processed and converted into meaningful variables to evaluate and help address the research questions. In this section, I present how I accessed and processed these datasets and provide an overview of the methods that will be used to achieve the results in a later section.

4.1. Quality Assessment Classes in Wikipedia

To evaluate how editor collaboration on Wikipedia affects content quality, I utilized Wikipedia's article quality assessment scale as the output variable ("Wikipedia:Version 1.0 Editorial Team/Assessment," 2014). Quality scales have been seen positively for formalizing ratings but also for utilizing them as standardized measures in machine learning algorithms (Warncke-Wang, Cosley, & Riedl, 2013; Warncke-Wang, Ranjan, Terveen, & Hecht, 2015). These assessments are mainly conducted by Wikiprojects (groups of editors interested in a particular theme of articles) that initiate an evaluative voting process for an article under their theme. Quality assessments on Wikipedia have been tested and found to be in agreement with external reviewers in the past (Kittur & Kraut, 2008). A recent study has also shown that Wikipedia quality rating is a reliable measure of information quality, however, external reviewers in the study were only able to make a binary classification (high/low quality) for articles (Kane & Ransbotham, 2016b). Quality rating judgments on Wikipedia are made based on the community's criteria, which may differ from the criteria of an external panel measuring the quality of encyclopedic articles. However, the goal of this study was not to measure Wikipedia's reliability, which is discussed in many other studies (Mesgari et al., 2015), but to measure

how Wikipedia sees itself achieve the goals set forth by its community. Regardless, this limitation should be taken into account because there have been studies that have utilized our method (Kittur & Kraut, 2008) and others that have not (Arazy & Nov, 2010).

The article quality scale was formulated by the Wikipedia community in late 2005 and evolved to its current version later in 2006 (although the criteria for each category are constantly evolving as are standards for articles). Wikipedia's article quality rating scale consists of the following ratings: Featured Article (FA), A, Good Article (GA), B, C, Start, Stub. Featured Article signifies the highest rating in the scale while the Stub the lowest. Once an article quality rating is decided by a Wikiproject, it is delivered (sometimes using bots) to article discussion pages which at times can have multiple quality ratings by multiple Wikiprojects. An additional peer review process is applied for the highest article quality categories: FA and GA (Viégas, Wattenberg, & McKeon, 2007). The English branch of Wikipedia at the time of the study had 4,903,920 articles, some of which have quality ratings received by Wikiprojects ("Special:Statistics," 2015). Studies on Wikipedia have shown that most articles on Wikipedia belong to the lower level categories, while a few thousand make it to the top categories (Kittur & Kraut, 2008; Tsikerdekis, 2016).

4.2. Datasets

I obtained all article quality assessments that were ever delivered on an article's discussion page before July 1, 2011. A total of 7,864,682 quality ratings were gathered. Quality ratings were converted to a numeric corresponding scale ranging from 7 to 1, with 7 representing Featured Articles and 1 representing Stub article quality rating. The choice of numeric scale was made in order to make the results directly comparable to past work (Kittur & Kraut, 2008) that has also used this scale for regression models. Two additional quality ratings are used for lists (pages containing lists of articles) that were excluded from the dataset.

I developed observation periods in order to establish what factors affect change for a period of work on Wikipedia. This is similar to a procedure that was also utilized by Kittur & Kraut (2008). The two observation periods were July 1, 2010, through January 1, 2011, and January 1, 2011, through July 1, 2011. For each observation period, I obtained the rating at the beginning of a period (henceforth *Initial Rating*) and the rating at the end of the observation period (henceforth *Final Rating*). Ratings were calculated using the median of all ratings that appeared on an article's discussion page from Wikiprojects at the relevant time (beginning or end) of the observation period. Contrary to previous work that used the highest quality rating by all Wikiprojects (Kittur & Kraut, 2008) due to the early and sparser implementation of the rating system, I propose that the median reflects a more accurate quality of an article. Each article may have been

evaluated by multiple groups rather than one, and as such, utilizing only the highest rating misses important information. Many Wikiprojects may be more favorable and less objective towards certain articles. As such, multiple ratings increase the objectivity of results. Furthermore, the use of the median over the mean is suggested due to the rare occurrence of malicious ratings recorded during the observation period (e.g., a user changes the rating without approval and may require several days for the malicious change to be reverted back). Spearman correlations between the median and max Initial Rating and Final Rating scores were $r_s = .982$, $p < .001$ and $r_s = .982$, $p < .001$, respectively. Similarly, correlations between the median and mean Initial and Final rating scores were $r_s = .989$, $p < .001$ and $r_s = .989$, $p < .001$, respectively. Because the breakdown point of the median is more robust than that of the mean (Leys, Ley, Klein, Bernard, & Licata, 2013), the median was utilized to eliminate any probable anomalies (outliers) in the quality ratings.

Once quality ratings were obtained, they were rounded to zero decimal points (where the midpoint decimal values were rounded upward). Articles without an Initial Rating (no Wikiproject that evaluated them at the beginning of the observation period) were excluded. Similarly, articles with an Initial Rating of 7 at the beginning of the observation period were also excluded given that no higher rating exists in the scale and therefore no room for measurable improvement of an article. A small number of articles demonstrated a decrease in their quality rating and were also excluded. In most cases the decrease was due to an erroneous high initial quality rating being prematurely added to the article discussion page and the rating later being reverted. This was often a result of an ambitious editor who elevated the rating without coordination (approval by a project) or a vandal who had a vested interest in raising the quality of an article.

The two observation periods were merged into one dataset that totaled 4,729,128 articles. The distribution of final quality ratings ranging from 1 to 7 for articles in the sample is presented in Table 1. During the two observation periods, only 44,408 articles changed their article quality while the rest remained the same; this indicates a general tendency for articles to not change quality ratings.

[INSERT TABLE 1 HERE]

More than 20 million user accounts on Wikipedia have contributed to articles. Considering the time complexity for certain metrics required to be computed by the study, the dataset was reduced in size. I obtained a stratified sample due to the size of the dataset and the additional metrics that needed to be added. The sample was obtained without replacement and was based on the Final Ratings (seven possible values) and whether quality changed or did not change (two possible values). The resulting strata were a total of 12 given that there are no cases where an Initial Rating was not available (must be 1 or

higher) and that articles with a Final Rating of 7 do not have a stratum where no change occurs (because that would mean that the Initial Rating was 7). Random stratified samples are considered reliable in terms generalizing the results to the overall population of articles and usually are required to have at least 30 cases for each stratum (Somekh & Lewin, 2005). This generated a sample of 1,320 articles from the two observation periods with 12 unique strata. A stratified sample developed a more realistic picture of articles that changed as well as those that did not change during observation periods for all possible quality ratings.

4.3. Predictors of Quality Change

In order to examine editor characteristics for those editors who participated in an article within the sample, I developed variables for cumulative experience and recent behavior (based on the literature review and research questions). Cumulative experience is defined as all of the accumulated activity that an editor has performed since he or she joined Wikipedia, whereas recent behavior represents an editor's short-term activity.

Anonymous users and bots were excluded from the sample. While anonymous users play a large role in Wikipedia, it is difficult to discern unique persons through IP addresses to build the metrics developed in this study. Bots were excluded as they have absolute deterministic behavior, which is of no interest to the study. Additionally, many bots were also developed to deliver the human assigned article quality ratings in article discussion pages, which would have resulted in a signaling effect for the models developed.

For each user who edited an article in the sample during an article's observation period, I retrieved a record of all revisions made to other pages for all namespaces since their first registration and up until the last revision made during the observation period. Namespaces are spaces that Wikipedia is divided in based on different themes such as articles, discussion, policy, and user profiles. For each revision made to other articles, I additionally obtained each article's median quality rating at the time the revision was made (rounded to zero decimal points). All revisions used for the cumulative experience as well as recent behavior metrics were non-reverted revisions. As such, these metrics refer to revisions that survived subsequent reviews by other editors. The reasoning behind this was to measure only "quality" edits. For example, a user may contribute multiple revisions to high quality articles that are later reverted. If these reverted revisions were included in the sample, the aforementioned user would have appeared to be contributing to these articles and therefore assumed to be experienced in editing high quality articles, which is not the case.

After retrieving records for each individual user, I started developing variables for the group of editors who participated in the article and in the sample during the observation periods. For many of the metrics described below, I utilized a geometric mean instead of an arithmetic mean due to the presence of many outliers in the sample, as well as the fact that conceptually, user activity between editors in Wikipedia is not in its entirety independent of one another.

I developed the following cumulative experience metrics for the total participating users for each article in the sample during the respective observation period:

Cumulative experience in high quality articles: The geometric mean of revisions made by users to high quality articles, where high quality articles were considered those with ratings of 7, 6 and 5 on the quality assessment scale.

Cumulative experience in user and user talk namespaces: The geometric mean of revisions made by users to the user and user talk namespaces. The assumption is that these spaces are used for coordination and as such serve as a proxy for integration with the community. The two spaces may also display different activity by different users but as a way of summarizing they were used as one entity in this study.

Cumulative experience in Wikipedia and Wikipedia talk namespaces: The geometric mean of revisions made by users to the Wikipedia and Wikipedia talk namespaces. A similar assumption to user and user talk pages is held also for this metric with the intent that the two spaces serve as a proxy for community integration.

The aforementioned metrics relate to Q3. A metric similar to cumulative experience in high quality articles was produced by Stein and Hess (2007) for the German-language version of Wikipedia and was based on the revisions made on featured articles (two levels of featured articles in German Wikipedia). The overall experience (or reputation of editors, as it was called) in an article was measured in various forms based on the total revisions on an article, or limited to just the first few revisions ever made on an article. The selection of the particular namespaces, as opposed to others, was made based on past literature suggesting that Wikipedia's personal communication networks (in user talk pages) have an impact on quality (Tsikerdekis, 2016) and that some users are better integrated with community activities (by assuming different roles) (Welser et al., 2011), which could potentially impact quality. The latter was obtained through a proxy, which was participation on Wikipedia's policy pages. The additional article discussion namespace that was found to have an impact on quality by a previous study (Kittur & Kraut, 2008) was also included and is described later in this section.

Experience metrics fall short of taking into account the fact that being a user in a collaborative project is a transformative experience as much as an accumulating one. Transformation results in temporal changes in behavior. Using the same dataset, I also derived the following *recent behavior* metrics based on the most recent 250 revisions for each user during their last revision on each article in the sample within the observation period. In the sample of articles, 71.2 percent of editors who contributed to any of these articles had more than 250 revisions (the median revisions per editor was 2,659). The study recognizes that multiple values could qualify as metrics for recent behavior. For example, time could also qualify as a factor in measuring recent behavior; however, the purpose of this study was to measure recent behavior as a "change in attitudes", which would not be feasible if time was used as a factor (e.g., user A appears to have had zero activity during the last 6 months, but his or her recent 250 revisions out of a total of 10,000 revisions reflects a major change in editing attitude). Furthermore, a previous study that relates to a theme of this paper also utilized 250 revisions as a distinction metric since the number is important for gaining access to counter-vandalism tools as well as AutoWikiBrowser access (Panciera, Halfaker, & Terveen, 2009).

Recent behavior in high quality articles: The geometric mean of revisions made by users to high quality articles, where high quality articles were considered those with ratings of 7, 6, and 5.

Recent behavior in user and user talk namespaces: The geometric mean of revisions made by users to the user and user talk namespaces (2 and 3 numerical identifiers).

Recent behavior in Wikipedia and Wikipedia talk namespaces: The geometric mean of revisions made by users to the Wikipedia and Wikipedia talk (4 and 5 numerical identifiers) namespaces.

These metrics related to Q4 and can be thought of as a user's preference in participating in various namespaces and articles on Wikipedia. This can also be thought of as the short-term profile of a user. For example, a user may have 1,000 revisions on Wikipedia as a whole with most of them applied on the article namespace. However, when looking at their most recent 250 revisions, they may have been contributed most on Wikipedia policy namespaces. Furthermore, many users have fewer than 250 revisions on Wikipedia; however, because the study is based on how a group operated on an article as a whole, there is variation between cumulative experience and recent behavior metrics, which in turn reveals the differences between recent behavior (or attitude) and cumulative experience.

Additionally, I calculated two metrics that were utilized and demonstrated a high effect on content quality in past studies (Arazy & Nov, 2010; Kittur & Kraut, 2008).

These were *experience diversity* (global inequality), as the Gini distribution of revisions made throughout the lifetime of a user's account that participated in an article during an observation period, and *implicit coordination* (editor concentration), as the Gini distributions of revisions made by users on the article of interest during an observation period. The latter was also referred to as local inequality by Arazy and Nov (2010). Both of these metrics utilized the Gini coefficient (a single number bound between 0 and 1) to represent the aforementioned distributions of revisions across users. The choice of the Gini coefficient was made in order for the results to be directly comparable to past studies.

For the article sample during the observation period, I also derived the following variables, some of which are used for comparisons to previous literature-derived models:

Number of editors: The total number of editors who made changes to an article during the observation period.

Article age: An article's age in months since its first revision (which created the article).

Number of revisions on article discussion: The total number of revisions made on an article's discussion page during the observation period.

Article discussion change: The change in an article's discussion page during the observation period measured in bytes.

4.4. Approach

Because the aim was to evaluate how article quality changes based on a group's composition as well as collaborative patterns, I utilized a lagged multiple regression. Instead of using the quality rating of an article at the end of the observation period (or at the beginning of that period), I utilized the *article quality change* for the ratings before and after the observation period as the outcome variable. The *initial article rating* at the beginning of the observation period is held constant in the prediction model. This procedure removes the potential influences of the initial rating on predictors as well as controls for other unobserved variables that can influence quality rating change. The procedure is intentionally identical to the one used by the foundational model of this study (Kittur & Kraut, 2008).

I further controlled for selection, biases caused by the limited amount of articles on Wikipedia that are being evaluated, as well as the fact that a rating evaluation did not necessarily occur for every article in the sample within their respective observation periods even though articles may change. This control was provided using a Heckman 2-

step selection model (Heckman, 1979), which consists of a probit regression model predicting whether an article will receive an evaluation within the observation period based on a set of variables. The outcome variable of the logistic model is then used as a control variable in the multiple regression model. To review for selection bias, the inverse Mill's ratio was utilized to establish that all Heckman models were consistent.

To establish an article's likelihood for quality assessment, I utilized several variables that were found to have an impact on the likelihood of an article receiving a rating by Kittur and Kraut (2008). These were the number of revisions an article had prior to the beginning of the observation period, the age of the article, the number of editors working on an article during the observation period, the distribution of edits and the number of revisions made in an article's discussion page during the observation period.

5. RESULTS

5.1. Descriptive Statistics and Model Summaries

Descriptive statistics for outcome and predictor variables are presented in Table 2. The mean implicit coordination is closer to zero, which is suggestive that work is more evenly distributed among editors. Groups of editors in the sample did not perform much work on high quality articles based on *Cumulative experience in high quality articles* and *Recent Behavior in high quality articles*. All metrics were normalized between zero and one in subsequent tests and regression models because comparisons between them may have been difficult due to standard deviation differences.

[TABLE 2 HERE]

Pearson correlations for all variables that were used in the Heckman 2-step selection models (presented later in this section) are shown in Table 3. There is negative correlation between initial quality rating and quality rating change. Implicit coordination is also correlated (positively) with article quality change. This means that a skew distribution of how users work on an article (an individual does most of the work) appears to coincide with article quality. Experience in editing high quality articles also correlates with recent behavior in editing high quality articles. The same connection between experience and recent behavior can also be observed with regard to Wikipedia's policy and process pages. If cumulative experience was equal or almost identical to recent behavior for users, we would expect these numbers to be higher. This demonstrates that while cumulative experience affects recent behavior to a degree, the two metrics have substantial differences.

[TABLE 3 HERE]

The variables described in the previous section were gradually fitted into models. Table 4 illustrates the model used by Kittur and Kraut (2008) that describes implicit coordination, as well as explicit coordination patterns (participation on article discussion pages). The models indicate that article quality and implicit coordination are predictors of quality change. Additional interactions exist between implicit coordination and quality change, which are discussed later in the paper. One of the findings that does agree with previous studies is the lack of a significant coefficient for neither the article discussion variable nor subsequent interaction effects related to it. These were highlighted in previous work as explicit coordination (Kittur & Kraut, 2008). The lack of a significant result may be attributed to a change in attitudes over time on how discussion pages are used as well as a change in the name for this namespace.

[TABLE 4 HERE]

Table 5 illustrates the model used by Arazy & Nov (2010) comparing experience diversity to implicit coordination along with article discussion page change. While the last variable was originally measured in number of words, for the purposes of this model I used the change in number of bytes. Overall, many variables that were found in the original study to be significant are also significant on Table 5. The results indicate a large effect for implicit coordination, while experience diversity is statistically significant but less influential on quality change. The only result that contradicts the original study relates to the discussion pages. These do not indicate an impact on article quality change while the original study found this variable to be significant.

[TABLE 5 HERE]

Table 6 builds on the findings of Tables 4 and 5. The approach of building nested models utilized in the table enables a direct comparison of model effectiveness (Faraway, 2014; Harrell, 2015). The table illustrates all variables that describe cumulative experience and recent behavior. Interaction effects were included in Table 7. All models were examined for multicollinearity and had a Variance Inflation Factor of less than 2. Results are discussed in the sections that follow.

[TABLE 6 HERE]

[TABLE 7 HERE]

5.2. Baseline Models

Models 1a and 1b served as a baseline model in order to help assess the increase in accounted variance as the rest of the metrics are included in subsequent model. These

models address Q1 (implicit coordination and article quality change). The results appear to agree with the original studies. These show that when implicit coordination is skewed, it reduces coordination costs and increases the likelihood for quality change. In particular, two variables were found to have the most impact based on article quality: *experience diversity* on Wikipedia among editors of an article and *implicit coordination*. These two variables are more powerful predictors than the total number of editors who edit an article during the observation period. In other words, having a large number of editors is not necessarily beneficial to an article because coordination costs increase, as described by Kittur and Kraut (2008). It was additionally speculated that this form of implicit coordination may be a proxy that represents a shared mental model existing among experienced editors who edit an article.

5.3. Cumulative Experience Models

In these models, I included all cumulative experience variables that measured the quality of articles that editors edit, as well as the namespaces that they have already edited throughout their lifetime. Models 2a and 2b address Q3 (cumulative experience and content quality).

Model 2a demonstrates a statistically significant ($F(3) = 6.076$, $p < .001$) small improvement compared to the baseline model 1a (1.3% increase in the accounted variance). The only significant effect appears to exist for those participating in Wikipedia namespaces. There are several factors that contribute to this effect. Familiarity with Wikipedia namespaces increases knowledge of article quality criteria, contribution to Wikiprojects that target articles as well as nominations for higher quality articles (e.g., Featured articles).

Model 2b also shows the interaction effects where participation in the user namespace becomes significant when it interacts with implicit coordination. The model is a statistically significant ($F(4) = 6.216$, $p < .001$) improvement compared to the baseline model 1b (1.7% increase in the accounted variance). The interaction effect included in the model is plotted in Figure 1. The high and low cumulative experience was determined based on the median of the sample of articles. Low cumulative experience was defined as values lower than the median cumulative experience value, and high cumulative experience was defined as higher or equal values to the median cumulative experience. In general, skewed implicit coordination (high Gini coefficient) appears to be effective only under a specific condition. Such a condition requires that the group of users has an overall low experience in editing user namespaces (user and user talk pages). In such cases, having fewer editors doing most of the work seems to be more effective in terms of changing article quality. On the other hand, when there are editors who have a high level of experience in editing user pages, implicit coordination does not seem to affect quality

change. A possible explanation for this is probably due to the nature of experienced editors being able to better coordinate with others but also understanding the requirements for article quality change. This may also be in line with other studies showing that an editor may repeatedly support an article over time (Keegan, Gergle, & Contractor, 2013). Editors who maintain active user and user talk pages are likely to be more involved in the community.

[FIGURE 1 HERE]
Figure 1: Interaction effects between cumulative experience and implicit coordination. Abbreviations are as follows: Cumulative experience on user namespace and implicit coordination.

5.4. Recent Behavior Models

Models 3a and 3b incorporate recent behavior metrics and several interactions with other variables. These models address Q4 (recent behavior and content quality).

Model 3a demonstrates a statistically significant ($F(3) = 23.541$, $p < .001$) substantial improvement) compared to the baseline model 1a (5.3% increase in the accounted variance. In particular, substantial recent participation on high quality articles is a strong predictor of quality change. The coefficient remains at 3.395 in the absence of all interaction effects (model 3a), and it is one of the highest coefficients in the model aside from the implicit coordination. Recent participation in Wikipedia policy pages appears to also predict quality change.

Model 3b (which includes interaction effects) demonstrates a statistically significant ($F(6) = 22.436$, $p < .001$) substantial improvement compared to the baseline model 1b (11% increase in the accounted variance). Significant interaction effects exist between recent behavior on high quality articles and initial quality rating, implicit coordination and total number of users participating during the observation period on an article. Figure 2 shows these interactions.

[FIGURE 2 HERE]
Figure 2: Interactions effects between recent behavior in high quality articles variables and initial rating, implicit coordination and total number of users.

The interactions show that recent activity in high quality articles can have a more radical impact on quality change as the initial quality of an article rises. On the other hand, groups that are less experienced in editing high quality articles can be effective in lower quality articles but not in articles that have a higher initial quality. This may be interpreted as signaling or as a result of learning. In the first interpretation, quality editors impact the quality of articles. As quality rises, so does the need for more competent

editors. However, the alternative way of looking at this is that there is a learned behavior for editors who edit high quality articles and participate in Wikipedia's policy and process pages. The outcome of this learning process not only makes these editors less effective compared to their "untrained" counterparts in editing low quality articles but also makes them highly efficient in editing high quality articles. Regardless of the interpretation, the effect is considerable.

An additional finding of interest is revealed by the interaction of recent behavior in editing high quality articles and implicit coordination. Articles benefit more when few individuals are doing most of the work, and of those individuals, most have a low recent behavior of editing high quality articles. For example, an article that has a group of editors who have not recently edited high quality articles would benefit from a coordination pattern in which few individuals (perhaps more experienced) end up doing most of the work. On the other hand, the benefit of implicit coordination on quality is less pronounced when individuals working on an article have a high recent behavior of editing high quality articles.

[FIGURE 2 HERE]

Figure 2: Interactions effects between recent behavior in high quality articles variables and initial rating, implicit coordination and total number of users.

6. DISCUSSION

Further discussion points based on this study's findings are elaborated in separate sections below.

6.1. Contextualized Cumulative Experience

While experience in previous studies (Arazy & Nov, 2010; Arazy et al., 2011; Halfaker, Kittur, Kraut, & Riedl, 2009) has been measured singularly (e.g., quantity of work), this study demonstrates the need for a more detailed and contextualized expression of experience. Results for Q3 ("What types of cumulative experience affect content quality, and to what degree?") suggest that all experience is not equal in terms of impacting content quality. An example can be found in the interaction effect between cumulative experience on user pages and implicit coordination that substantially impacts content quality. In other words, all revisions made in Wikipedia are not equal, with some namespaces appearing to affect an individual's potential to impact the quality of articles on which they eventually work. While it needs to be affirmed by future work that is beyond the scope of this paper, a possible explanation is the presence of a learning process that allows for users who work on other namespaces to develop practices that can improve their impact on articles. Understanding the context from which experience is

gained will result in a better understanding of what group composition is ideal for bringing about change.

6.2. Adaptive Implicit Coordination Patterns based on Cumulative Experience

In previous work on implicit coordination (Arazy & Nov, 2010; Kittur & Kraut, 2008), there has been an emphasis on effective collaborative patterns. Background theory suggests that a successful pattern for implicit coordination occurs when few editors from a group do most of the work on an article during an observation window. This in turn will affect content quality due to a reduction in coordination costs. This effect was confirmed by our results addressing Q1 ("Does implicit coordination affect article quality change?"). However, implicit coordination did produce an incomplete picture as a predictor for quality change. This is demonstrated by the accounted variance of baseline models 1a and 1b. However, once cumulative experience is incorporated into the model, other patterns emerge that call for a re-evaluation of what was previously perceived as an effective pattern of coordination.

The perspective that a skewed distribution of work between editors can reduce coordination costs and positively affect quality change that is argued by Kittur and Kraut (2008) cannot be generalized to all cases. It is rather conditioned upon the cumulative experience on user pages in a group's participating editors. Put simply, reduction in coordination costs by having few editors doing most of the work is an effective strategy among integrated editors (e.g., those that frequently edit user and user talk pages) but not for those who are not well integrated. Under this lens, Wikipedia's democratic process is not challenged but the pattern observed is rather a necessary configuration to optimize a group's performance due to "inadequate" group composition in terms of cumulative experience among group members.

6.3. Explicit Coordination

Considerable discussion has focused on the effectiveness of the article discussion space where editors get to coordinate the work as well as solve disputes involving articles. In addressing Q2 ("Does participation in an article's discussion page affect article quality change?"), I discovered that contrary to past studies that found the discussion namespace to have an effect on article quality change, the same effect was not observed in this study. It is likely that this is due to the nature of the discussion namespace, which is mainly geared toward dispute resolution or writing corrections. Recent work also seems to point to this conclusion in terms of how the article discussion namespace is used (Ferschke, Gurevych, & Chebotar, 2012). As such, if it is the only variable in a model, it is likely to have some predictive power in detecting article quality change. However, once more specific variables are added that aid in measuring group composition as well as implicit

coordination, the effect of the variable, and as such, the importance of the article discussion page as a predictor of quality change, is diminished. Future studies should investigate in depth how different types of activity on an article discussion space could influence article quality.

6.4. Recent Behavior and Cumulative Experience

Experience (albeit not contextualized in detail) in the prior literature has been found to affect content quality (Arazy & Nov, 2010; Arazy et al., 2011; Halfaker et al., 2011; Mesgari et al., 2015). However, a considerable amount of unaccounted variance has been left as speculation to infer other user characteristics that may affect content quality. The literature on user roles (Welser et al., 2011) has displayed the potential for identifying roles that are assumed by users over their time working in an online community. I addressed Q4 ("What types of recent behavior can affect content quality and to what degree?") by identifying a more tangible way for measuring this recent behavior as the recent temporal extension of a user's experience (a user's recent profile). The results suggest that not only is recent behavior important for content quality but that it is substantially more important than cumulative experience. Users who have worked on high quality articles in the past and display a recent preference for working on high quality articles can positively affect content quality.

The results also relate to a study that examined whether active editors in Wikipedia are "born" or "made" (Panciera et al., 2009). Editors with just a few revisions over the lifetime of an account were found to contribute less in terms of quality. The current paper complements this study while noting that cumulative experience does matter but that recent behavior matters even more. Put simply, active editors can further be distinguished in terms of their activity, which in turn impacts quality. An argument can be made (although further examination of the topic is necessary) that impactful editors are both "born" and "made" over time. That is, in agreement with past work (Panciera et al., 2009), some traits (e.g., editing high quality articles) are picked up by individuals who integrated more with the community.

7. RECOMMENDATIONS

Based on the aforementioned findings, I provide recommendations in a generalized manner that could be implemented not only in a Wikipedia environment but also in other collaborative projects. As this is a Wikipedia-specific study, such implementations should be taken with caution because different environments may yield different results depending on a system's design as well as policies for users.

7.1. For Designers/Developers

The findings have direct implications for designers and developers who need to utilize their users' resources and time in the most effective way in order to improve the quality of outcomes for collaborative projects. The results suggest that group composition when contextualized within a content's lifecycle will impact content quality. The efforts of more experienced members could be redirected to higher quality content leaving room for training in lower quality content for inexperienced users. From a design standpoint, enabling the creation of training groups could also be of benefit to a community. This is a likely effect that can be found on Wikipedia's Wikiprojects (Zhu, Kraut, & Kittur, 2016). Finally, developers involved in setting up task routing software (Cosley et al., 2007) can take into account both cumulative experience and recent behavior and redirect users to complete tasks. This way, weaker groups can be enhanced by the addition of more experienced members.

7.2. For Community Users

In conjunction with legitimate peripheral participation (Lave & Wenger, 1991), there appears to be a path for inexperienced users to reach levels that can influence higher quality articles positively. Users can be made aware of their performance and as a result they can seek to develop their skills but also understand their role in the community. For example, they can choose to enhance their recent skills by integrating more within the community. Users could then assume a more active role with an elevated understanding on how their performance will in turn affect articles and their community as a whole.

8. LIMITATIONS

Experience as defined in this study involves only activity within the bounds of Wikipedia (editing articles and subsequent namespaces). However, a recent study has shown that editor activity and as such experience accumulated over time span multiple platforms (R. West, Weber, & Castillo, 2012). Many editors browse relevant pages around a topic on the web and will eventually edit related articles. They also make use of social networking to become more engaged with the topics they edit. These activities and subsequent writing activities (e.g., blog writing) make up the sum of an editor's experience. As such, experience metrics reflected in this study can be seen as an estimate and only part of an editor's actual experience. The age of a user's account is also an additional factor that could influence experience, and future studies should investigate this and the aforementioned aspects further.

Not measuring reverted revisions also may influence the results in the event that some revisions are valid but a more experienced editor exercised authority over a newer user. While this is not a common occurrence on Wikipedia, it has been documented (Halfaker et al., 2011). Additionally, some content that is the result of vandalism may be removed

by an additional revision that deletes it rather than a revert action. Furthermore, differences in the window of observation (larger or smaller) could also influence metrics such as the Gini coefficient and the activity observed for a page. While it does not change the statistical significance of the results presented in this paper, it is important to note such a limitation. Additionally, many experienced users may also hold privileged positions that can help them "push" towards the promotion of articles to a higher quality rating. An increase in rating requires voting by members of a Wikiproject, and many Wikiprojects often initiate their own voting for article ratings. Regardless, some of the effects observed in this paper may be influenced to a certain degree by how experienced editors can also hold influential positions (e.g., administrators). Finally, recent behavior as specified in this study includes a subset of 250 recent revisions out of an entire set. However, a subset could have also been defined based on another number or a percentage that would influence the results. Although the effect of recent behavior is significant, a larger (or smaller) effect could be identified using another definition that constitutes "recent" behavior.

9. CONCLUSION

Ever since January 15, 2001, when Wikipedia went live, content quality and the group processes that govern it have been a pertinent issue for multiple research studies. A recent literature review includes more than 98 articles focused on understanding content quality on Wikipedia (Mesgari et al., 2015). The importance of this work can help developers and managers of online communities and collaborative projects achieve better qualities of outcomes. I show that not only implicit coordination and group composition is important but that both of these conditions have to be contextualized based on an article's lifecycle, the experience of the group and the temporal behaviors that a group may have. These findings demonstrate how work on Wikipedia is much more complex than previously thought. While popular culture often views Wikipedia's work as an uncoordinated wisdom of the crowd effort, the reality is different. Collaborative projects such as Wikipedia provide their users with the unique ability to not only voluntarily contribute but to also voluntarily choose how they will develop as community members. This ability for self-actualization provided by Wikipedia's "be who you want to be" design helps demonstrate that quality is not only an attribute of content but of the community that contributes and manages it. Much like ancient texts revealed the reality of their authors, Wikipedia's content reveals the qualities and culture of its editors.

REFERENCES

Adler, B. T., de Alfaro, L., Mola-Velasco, S., Rosso, P., & West, A. (2011). Wikipedia Vandalism Detection: Combining Natural Language, Metadata, and Reputation

Features. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing SE - 23* (Vol. 6609, pp. 277–288). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-19437-5_23

Arazy, O., Liifshitz-Assaf, H., Nov, O., Daxenberger, J., Balestra, M., & Cheshire, C. (2017). On the "How" and "Why" of Emergent Role Behaviors in Wikipedia. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 2039–2051). New York, NY, USA: ACM. http://doi.org/10.1145/2998181.2998317

Arazy, O., & Nov, O. (2010). Determinants of wikipedia quality. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 233–236). New York, New York, USA: ACM Press. http://doi.org/10.1145/1718918.1718963

Arazy, O., Nov, O., Patterson, R., & Yeo, L. (2011). Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict. *Journal of Management Information Systems*, *27*(4), 71–98. http://doi.org/10.2753/MIS0742-1222270403

Arazy, O., Ortega, F., Nov, O., Yeo, L., & Balila, A. (2015). Functional Roles and Career Paths in Wikipedia. In *Proceedings of the 2015 ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1092–1105). New York, NY, USA: ACM. http://doi.org/10.1145/2675133.2675257

Benkler, Y., & Nissenbaum, H. (2006). Commons-based Peer Production and Virtue*. *Journal of Political Philosophy*, *14*(4), 394–419. http://doi.org/10.1111/j.1467-9760.2006.00235.x

Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (pp. 1–10). New York, NY, USA: ACM. http://doi.org/10.1145/1099203.1099205

Burke, M., & Kraut, R. (2008). Mopping Up: Modeling Wikipedia Promotion Decisions. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (pp. 27–36). New York, NY, USA: ACM Press. http://doi.org/10.1145/1460563.1460571

Carillo, K., & Okoli, C. (2011). Generating quality open content: A functional group perspective based on the time, interaction, and performance theory. *Information & Management*, *48*(6), 208–219. http://doi.org/10.1016/j.im.2011.04.004

Chesney, T. (2006). An empirical examination of Wikipedia's credibility. *First Monday*, *11*(11). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/1413

Cosley, D., Frankowski, D., Terveen, L., & Riedl, J. (2007). SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces* (pp. 32–41). New York, NY, USA: ACM. http://doi.org/10.1145/1216295.1216309

Ehmann, K., Large, A., & Beheshti, J. (2008). Collaboration in context: Comparing article evolution among subject disciplines in Wikipedia. *First Monday*, *13*(10).

http://doi.org/10.5210/fm.v13i10.2217

Elliott, M. (2006). Stigmergic Collaboration: The Evolution of Group Work. *M/C Journal*, *9*(2). Retrieved from http://journal.media-culture.org.au/0605/03-elliott.php

Faraway, J. J. (2014). *Linear Models with R, Second Edition*. Taylor & Francis.

Ferschke, O., Gurevych, I., & Chebotar, Y. (2012). Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 777–786). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2380816.2380911

Forte, A., & Bruckman, A. (2008). Scaling Consensus: Increasing Decentralization in Wikipedia Governance. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)* (pp. 157–157). IEEE. http://doi.org/10.1109/HICSS.2008.383

Geiger, R. S., & Halfaker, A. (2013). When the levee breaks. In *Proceedings of the 9th International Symposium on Open Collaboration - WikiSym '13* (pp. 1–6). New York, NY, USA: ACM Press. http://doi.org/10.1145/2491055.2491061

Geiger, R. S., & Ribes, D. (2010). The Work of Sustaining Order in Wikipedia: The Banning of a Vandal. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (pp. 117–126). New York, NY, USA: ACM. http://doi.org/10.1145/1718918.1718941

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*(7070), 900–901. Retrieved from http://dx.doi.org/10.1038/438900a

Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven Geographies of User-Generated Information: Patterns of Increasing Informational Poverty. *Annals of the Association of American Geographers*, *104*(4), 746–764. http://doi.org/10.1080/00045608.2014.910087

Halfaker, A., Kittur, A., Kraut, R., & Riedl, J. (2009). A Jury of Your Peers: Quality, Experience and Ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (p. 15:1--15:10). New York, NY, USA: ACM. http://doi.org/10.1145/1641309.1641332

Halfaker, A., Kittur, A., & Riedl, J. (2011). Don't bite the newbies. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration - WikiSym '11* (p. 163). New York, New York, USA: ACM Press. http://doi.org/10.1145/2038558.2038585

Harrell, F. (2015). *Regression Modeling Strategies*. Berlin, Heidelberg: Springer International Publishing. http://doi.org/10.1007/978-3-319-19425-7

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153–161. Retrieved from http://www.jstor.org/stable/1912352

Kane, G. C., & Ransbotham, S. (2016a). Content as Community Regulator: The Recursive Relationship Between Consumption and Contribution in Open

Collaboration Communities. *Organization Science*, *27*(5), 1258–1274. http://doi.org/10.1287/orsc.2016.1075

Kane, G. C., & Ransbotham, S. (2016b). Research Note—Content and Collaboration: An Affiliation Network Approach to Information Quality in Online Peer Production Communities. *Information Systems Research*, *27*(2), 424–439. http://doi.org/10.1287/isre.2016.0622

Keegan, B., Gergle, D., & Contractor, N. (2013). Hot Off the Wiki: Structures and Dynamics of Wikipedia's Coverage of Breaking News Events. *American Behavioral Scientist*, *57*(5), 595–622. http://doi.org/10.1177/0002764212469367

Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08* (pp. 37–46). New York, NY, USA: ACM Press. http://doi.org/10.1145/1460563.1460572

Lam, S. (Tony) K., Uduwage, A., Dong, Z., Sen, S., Musicant, D. R., Terveen, L., & Riedl, J. (2011). WP:Clubhouse?: An Exploration of Wikipedia's Gender Imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (pp. 1–10). New York, NY, USA: ACM. http://doi.org/10.1145/2038558.2038560

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

Lehmann, J., Müller-Birn, C., Laniado, D., Lalmas, M., & Kaltenbrunner, A. (2014). Reader Preferences and Behavior on Wikipedia. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media* (pp. 88–97). New York, NY, USA: ACM. http://doi.org/10.1145/2631775.2631805

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. http://doi.org/http://dx.doi.org/10.1016/j.jesp.2013.03.013

Loveland, J., & Reagle, J. (2013). Wikipedia and encyclopedic production. *New Media & Society*, *15*(8), 1294–1311. http://doi.org/10.1177/1461444812470428

Masli, M., Priedhorsky, R., & Terveen, L. (2011). Task Specialization in Social Production Communities: The Case of Geographic Volunteer Work. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 217–224). Palo Alto, California: AAAI Press.

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. Å., & Lanamäki, A. (2015). "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia. *Journal of the Association for Information Science and Technology*, *66*(2), 219–245. http://doi.org/10.1002/asi.23172

Morgan, J. T., Halfaker, A., Taraborelli, D., Hwang, T., & Goggins, S. (2015). Advancing an Industry/Academic Partnership Model for Open Collaboration Research. In

Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (pp. 293–296). New York, NY, USA: ACM. http://doi.org/10.1145/2685553.2685559

Ortega, F. (2009). *Wikipedia: A quantitative analysis*. King Juan Carlos University. Retrieved from https://www.researchgate.net/profile/Felipe_Ortega2/publication/200773248_Wikipedia_A_quantitative_analysis/links/00463519b697045820000000.pdf

Panciera, K., Halfaker, A., & Terveen, L. (2009). Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work* (pp. 51–60). New York, NY, USA: ACM Press. http://doi.org/10.1145/1531674.1531682

Preece, J., & Shneiderman, B. (2009). The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. *AIS Transactions on Human-Computer Interaction*, *1*(1), 13–32.

Priedhorsky, R., Masli, M., & Terveen, L. (2010). Eliciting and Focusing Geographic Volunteer Work. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (pp. 61–70). New York, NY, USA: ACM. http://doi.org/10.1145/1718918.1718931

Ransbotham, S., & Kane, G. C. (2011). Membership Turnover and Collaboration Success in Online Communities: Explaining Rises. *MIS Quarterly*, *35*(3), 613–627.

Rector, L. H. (2008). Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, *36*(1), 7–22. http://doi.org/10.1108/00907320810851998

Robert, L., & Romero, D. M. (2015). Crowd Size, Diversity and Performance. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1379–1382). New York, NY, USA: ACM. http://doi.org/10.1145/2702123.2702469

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and Effect Sizes in Behavioral Research: A Correlational Approach*. Cambridge: Cambridge University Press.

Somekh, B., & Lewin, C. (2005). *Research Methods in the Social Sciences*. London: Sage. http://doi.org/10.1111/j.1467-8535.2005.00515_7.x

Special:Statistics. (2015). Retrieved from http://en.wikipedia.org/wiki/Special:Statistics

Stein, K., & Hess, C. (2007). Does it matter who contributes: a study on featured articles in the german wikipedia. In *Proceedings of the 18th conference on Hypertext and hypermedia HT 07* (pp. 171–174). New York, New York, USA: ACM. http://doi.org/10.1145/1286240.1286290

Stvilia, B., Twidale, M. B., Gasser, L., & Smith, L. C. (2005). Information quality in a community-based encyclopedia. In *Knowledge Management: Nurturing Culture, Innovation, and Technology-Proceedings of the 2005 International Conference on*

Knowledge Management (pp. 101–113). Charlotte, NC: World Scientific Publishing Company.

Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. (2008). Information quality work organization in wikipedia. *Journal of the American Society for Information Science and Technology*, *59*(6), 983–1001. http://doi.org/10.1002/asi.20813

Tsikerdekis, M. (2016). Personal communication networks and their positive effects on online collaboration and outcome quality on Wikipedia. *Journal of the Association for Information Science and Technology*, *67*(4), 812–823. http://doi.org/10.1002/asi.23429

Viégas, F. B., Wattenberg, M., & McKeon, M. M. (2007). The Hidden Order of Wikipedia. In D. Schuler (Ed.), *Online Communities and Social Computing: Second International Conference, OCSC 2007, Held as Part of HCI International 2007, Beijing, China, July 22-27, 2007. Proceedings* (pp. 445–454). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-73257-0_49

Viegas, F., Wattenberg, M., Kriss, J., & Ham, F. (2007). Talk Before You Type: Coordination in Wikipedia. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07)* (pp. 78–78). Washington, DC, USA: IEEE. http://doi.org/10.1109/HICSS.2007.511

Warncke-Wang, M., Ayukaev, V. R., Hecht, B., & Terveen, L. G. (2015). The Success and Failure of Quality Improvement Projects in Peer Production Communities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 743–756). New York, NY, USA: ACM. http://doi.org/10.1145/2675133.2675241

Warncke-Wang, M., Cosley, D., & Riedl, J. (2013). Tell Me More: An Actionable Quality Model for Wikipedia. In *Proceedings of the 9th International Symposium on Open Collaboration* (p. 8:1--8:10). New York, NY, USA: ACM. http://doi.org/10.1145/2491055.2491063

Warncke-Wang, M., Ranjan, V., Terveen, L., & Hecht, B. (2015). Misalignment Between Supply and Demand of Quality Content in Peer Production Communities. In *Proceedings of the 9th International Conference on Web and Social Media - ICWSM - 15*.

Wattenberg, M., Viégas, F., & Hollenbach, K. (2007). Visualizing Activity on Wikipedia with Chromograms. In C. Baranauskas, P. Palanque, J. Abascal, & S. Barbosa (Eds.), *Human-Computer Interaction – INTERACT 2007 SE - 23* (Vol. 4663, pp. 272–287). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-74800-7_23

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., & Smith, M. (2011). Finding social roles in Wikipedia. In *Proceedings of the 2011 iConference on - iConference '11* (pp. 122–129). New York, New York, USA: ACM Press. http://doi.org/10.1145/1940761.1940778

West, K., & Williamson, J. (2009). Wikipedia: friend or foe? *Reference Services Review*, *37*(3), 260–271. http://doi.org/10.1108/00907320910982758

West, R., Weber, I., & Castillo, C. (2012). Drawing a Data-driven Portrait of Wikipedia Editors. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (p. 3:1--3:10). New York, NY, USA: ACM. http://doi.org/10.1145/2462932.2462937

Wilkinson, D., & Huberman, B. (2007). Assessing the value of cooperation in Wikipedia. *First Monday*, *12*(4). http://doi.org/10.5210/fm.v12i4.1763

Wöhner, T., & Peters, R. (2009). Assessing the Quality of Wikipedia Articles with Lifecycle Based Metrics. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (p. 16:1--16:10). New York, NY, USA: ACM. http://doi.org/10.1145/1641309.1641333

Yang, D., Halfaker, A., Kraut, R., & Hovy, E. (2016). Who Did What: Editor Role Identification in Wikipedia. In *International AAAI Conference on Web and Social Media*. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13077

Zhu, H., Kraut, R. E., & Kittur, A. (2016). Contingency View of Transferring and Adapting Best Practices within Online Communities. In *In CSCW '16: 19th ACM Conference on Computer Supported Cooperative Work & Social Computing Proceedings*. New York, NY, USA: ACM.

Zhu, H., Kraut, R., & Kittur, A. (2012). Effectiveness of Shared Leadership in Online Communities. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 407–416). New York, NY, USA: ACM. http://doi.org/10.1145/2145204.2145269
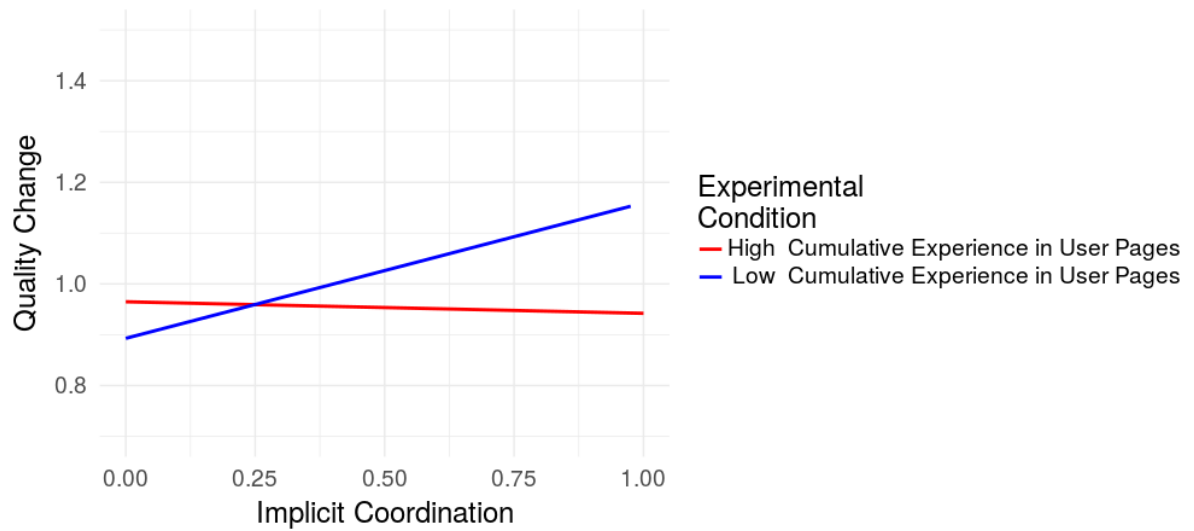
Figure 1: Interaction effects between cumulative experience and implicit coordination. Abbreviations are as follows: Cumulative experience on user namespace and implicit coordination
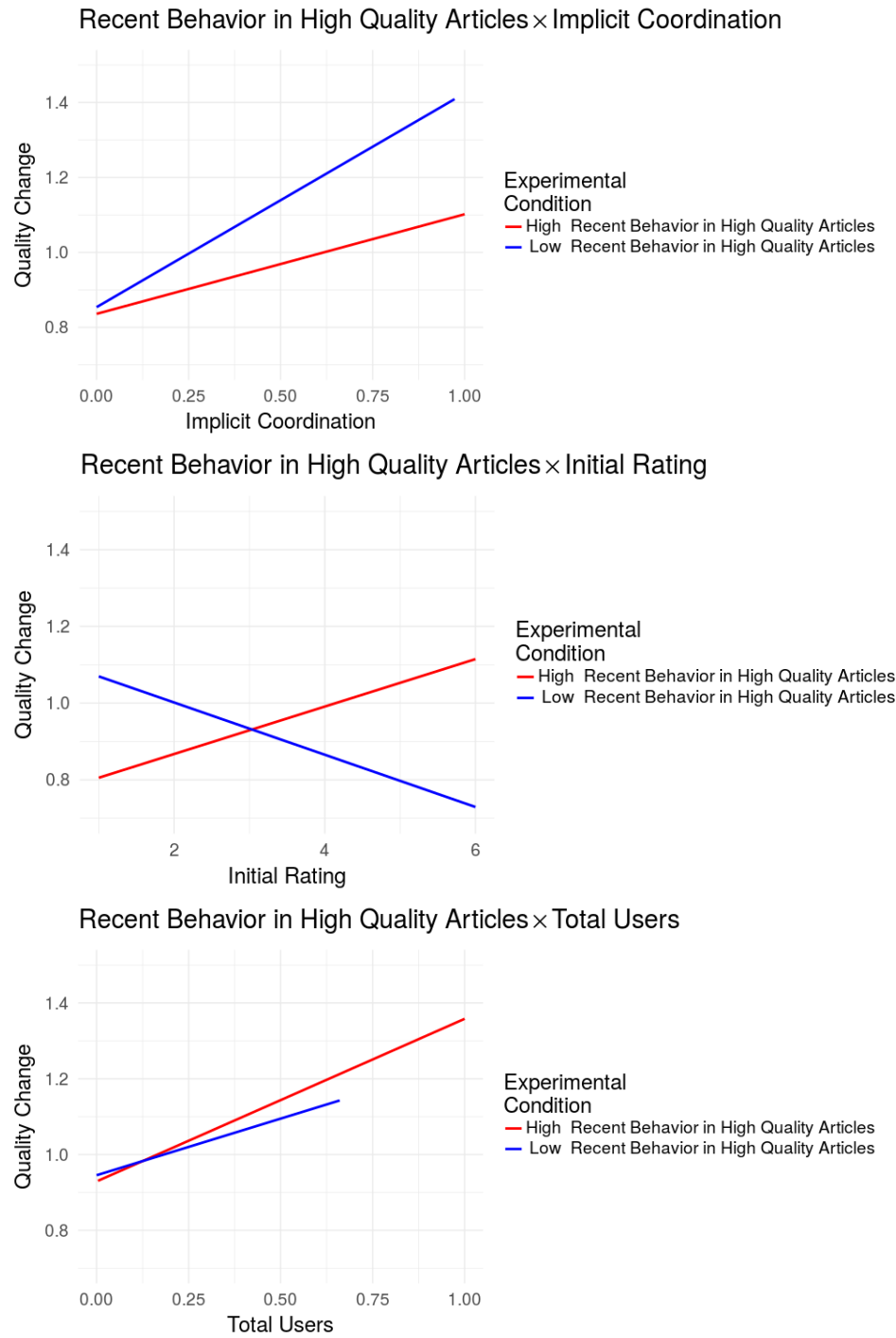
Figure 2: Interactions effects between recent behavior in high quality articles variables and initial rating, implicit coordination and total number of users.

Quality	Number of articles
Stub	3,178,308
Start	1,284,018
C	127,110
B	116,575
A	21,686
Good Articles	1,012
Featured Articles	419

Table 1: Summary of distribution of articles and quality ratings during the observation period of the study.

	Mean	Median	SD
Initial Rating	3.05	3.00	1.65
Quality Change	0.96	1.00	1.23
Total Users	13.21	5.00	23.96
Implicit Coordination	0.28	0.24	0.27
Cumulative Experience in High Quality Articles	573.22	149.64	1662.05
Cumulative Experience in User Pages	21381.09	3833.07	162872.81
Cumulative Experience in Wikipedia Pages	978.14	264.04	2455.58
Recent Behavior in High Quality Pages	8.32	5.38	10.93
Recent Behavior in User Pages	113.04	114.41	69.86
Recent Behavior in Wikipedia Pages	10.05	7.18	12.47
Experience Diversity	0.42	0.49	0.31
Age	54.15	54.00	28.22
Number of Revisions in Article Discussion	8.12	1.00	38.22
Article Discussion Change	8753.99	45.00	90995.87

Table 2: Descriptive statistics for all variables used in this study.

	Q_c	Q_i	N	IC	CE_{hqa}	CE_u	CE_w	RB_{hqa}	RB_u	RB_w	ED	Age	$N_{rev\_discus.}$
Q_c													
Q_i	-0.27 ***												
N	0.16 ***	0.15 ***											
IC	0.59 ***	0.10 ***	0.36 ***										
CE_{hqa}	0.04	0.07 *	-0.10 ***	0.00									
CE_u	0.02	0.00	-0.07 **	-0.03	0.30 ***								
CE_w	0.03	0.03	-0.10 ***	-0.03	0.43 ***	0.50 ***							
RB_{hqa}	0.26 ***	0.23 ***	-0.02	0.26 ***	0.27 ***	0.04	0.11 ***						
RB_u	0.16 ***	0.10 ***	0.16 ***	0.21 ***	-0.02	0.19 ***	0.05	0.15 ***					
RB_w	0.24 ***	0.13 ***	0.01	0.24 ***	0.11 ***	0.17 ***	0.44 ***	0.35 ***	0.27 ***				
ED	0.28 ***	0.16 ***	0.52 ***	0.61 ***	-0.11 ***	-0.10 ***	-0.11 ***	0.07 *	0.21 ***	0.11 ***			
Age	0.00	0.14 ***	0.31 ***	0.15 ***	-0.10 ***	-0.10 ***	-0.10 ***	-0.07 *	0.06 *	-0.05	0.34 ***		
$N_{rev\_discussion}$	0.16 ***	0.06 *	0.31 ***	0.28 ***	-0.02	-0.03	-0.03	0.04	0.04	0.03	0.22 ***	0.12 ***	
$B_{discussion\_change}$	0.08 **	0.05	0.16 ***	0.14 ***	-0.01	-0.02	-0.02	0.02	0.02	0.01	0.12 ***	0.10 ***	0.71 ***

	Without Interactions		With Interactions	
	Coefficient	SE	Coefficient	SE
Intercept	0.574	0.362	0.358	0.373
Initial Rating	-0.031 ***	0.025	-0.094 *	0.039
Age	-0.335	0.202	-0.593	0.331
Total Users	0.191	0.493	-2.041	1.215
Implicit Coordination	3.514 ***	0.485	4.26 ***	0.569
Implicit Coordination x Total Users			4.30	2.26
Implicit Coordination x Initial Rating			-0.493 ***	0.08
Implicit Coordination x Age			0.551	0.549
Number of revisions in article discussion	-0.079	0.476	-1.827	1.558
Number of revisions in article discussion x Total Users			-0.332	3.022
Number of revisions in article discussion x Initial Rating			0.111	0.387
Number of revisions in article discussion x Age			2.123	1.768

Table 4: All variables used by Kittur and Kraut (2008) to describe implicit and explicit coordination effects on article quality change. Note: model without interactions: adjusted R^2 = .352, model with interactions: adjusted R^2 = .385, * p< .05, ** p<.01, *** p<.001.

	Coefficient	SE
Intercept	0.915 **	0.314
Initial Rating	-0.306 ***	0.025
Experience Diversity	-0.515 ***	0.143
Implicit Coordination	3.355***	0.394
Article Discussion Change	0.442	0.962

Table 5: Variables used by Arazy and Nov (2010) to compare distribution of experience with distribution of effort. Note: Adjusted R^2 = .361, * p< .05, ** p<.01, *** p<.001.

	Model 1a		Model 2a		Model 3a	
	Combined Table 3 & 4		Cumulative Experience		Recent Behavior	
	Coef.	SE	Coef.	SE	Coef.	SE
Intercept	0.813 *	0.326	0.621	0.332	0.184	0.341
Initial Rating	-0.311 ***	0.025	-0.325 ***	0.025	-0.378 ***	0.025
Total users	0.516	0.448	0.710	0.456	1.025 *	0.481
Implicit Coordination	3.51 ***	0.412	3.556 ***	0.413	3.652 ***	0.421
Experience Diversity	-0.599 ***	0.155	-0.477 **	0.155	-0.388 **	0.145
Cumulative Experience in high quality articles			1.37	0.859		
Cumulative Experience in user pages			-1.1	0.606		
Cumulative Experience in Wikipedia pages			2.22 **	0.816		
Recent Behavior in high quality articles					3.395 ***	0.523
Recent Behavior in user pages					0.8	0.485
Recent Behavior in Wikipedia pages					1.305 **	0.411
Adjusted R^2	.362		.375		0.415	
Mean Squared Errors	0.926		0.85		0.845	

Table 6: Summary of models built for addressing the research questions. Note: * p< .05, ** p<.01, *** p<.001.

	Model 1b		Model 2b		Model 3b	
	Combined Table 3 & 4		**Cumulative Experience**		**Recent Behavior**	
	Coef.	**SE**	**Coef.**	**SE**	**Coef.**	**SE**
Intercept	0.276	0.330	0.093	0.336	0.110	0.314
Initial Rating	-0.094 *	0.038	-0.109 **	0.037	-0.086 *	0.041
Total users	0.582	0.442	0.869	0.456	-3.317 ***	0.879
Implicit Coordination	4.94 ***	0.468	4.803 ***	0.472	4.015 ***	0.449
Implicit Coordination x Initial Rating	-0.492 ***	0.07	-0.488 ***	0.076	-0.508 ***	0.074
Experience Diversity	-0.721 ***	0.152	-0.594 ***	0.151	- 0.416 **	0.143
Cumulative Experience in high quality articles			0.904	0.836		
Cumulative Experience in user pages			-1.511 *	0.572		
Cumulative Experience in Wikipedia pages			1.541	0.802		
Cumulative Experience in user pages x Implicit Coordination			8.547 **	2.986		
Recent Behavior in high quality articles					3.128 **	1.096
Recent Behavior in user pages					0.925	0.48
Recent Behavior in Wikipedia pages					1.264 **	0.398
Recent Behavior in high quality articles x Implicit Coordination					3.05	1.841
Recent Behavior in high quality articles x Initial Rating					-0.91 **	0.279
Recent Behavior in high quality articles x Total users					111.045 ***	19.839
Adjusted R²	.394		.411		.484	
Mean Squared Errors	0.878		0.849		0.742	