

Content is King, Leadership Lags: Effects of Prior Experience on Newcomer Retention and Productivity in Online Production Groups

Raghav Pavan Karumur, Bowen Yu, Haiyi Zhu, Joseph A. Konstan
GroupLens Research, University of Minnesota
Minneapolis, MN, USA
{raghav, bowen-yu, zhux0449, konstan}@umn.edu

ABSTRACT

Organizers of online groups often struggle to recruit members who can most effectively carry out the group's activities and remain part of the group over time. In a study of a sample of 30,000 new editors belonging to 1,054 English WikiProjects, we empirically examine the effects of generalized prior work-productivity experience (measured by overall prior article edits), prior leadership experience (measured by overall prior project edits), and localized prior work-productivity experience (measured by pre-joining article edits on a project) on early retention and productivity. We find that (1) generalized prior work-productivity experience is positively associated with retention, but negatively associated with productivity (2) prior leadership experience is negatively associated with both retention and productivity, and (3) localized prior work-productivity experience is positively associated with both retention and productivity within that focal project. We then discuss implications to inform the designs of early interventions aimed at group success.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Group and Organization Interfaces

Author Keywords

Wikipedia; WikiProjects; Prior Experience; Online Groups; Withdrawal; Retention; Newcomers; Online Communities; Productivity; Subgroups; Peer Production; Learning Transfer; Resocialization.

INTRODUCTION

Volunteer groups have existed for a long time in settings such as local non-profits, NGOs, and charity organizations. The explosive growth of computer technology and near-universal access to the Internet have enabled the growth of new forms

of volunteer contributions and groups at unprecedented scales. Examples include Wikipedia, OpenStreetMap, Open Source Software projects, product review forums, technical Q&A sites, citizen science projects, and online fund-raising, to name but a few. Bringing in people who actually do work and stick around long enough is a common problem to all these groups.

It is also common for volunteers in these groups to come in with varying levels of prior experience that shapes their activity in the group and perhaps, their success within the group. Prior work on new volunteer retention and productivity in online and offline groups, however, suggests that the effect of prior experience on newcomer success is complicated, with some studies showing positive effects [9, 10, 56, 57, 72] and others indicating negative ones [20, 26, 52]. Because prior experience is something that is usually visible and can be objectively measured, it is both theoretically and practically important to understand how what volunteers carry from their prior work experience affects their performance in the context of a new group. In this paper, we want to unpack prior experience and resolve conflicts in prior work. We, therefore, ask the following Research Question:

How does a new volunteer's prior experience affect their early retention and productivity in the group they join?

To answer this question, we review prior literature and identify three types of prior experience: (i) generalized prior work-productivity experience, which is prior experience associated with normal production activities (i.e., non-leadership activities) in other similar volunteer groups, (ii) prior leadership experience, which is the experience of organizing activities and managing people in other volunteer groups, and (iii) localized prior work-productivity experience, which is the amount of work a volunteer would invest in a group before joining the group (in other words, the "internship" experience). Because early identification of group failures can help community moderators intervene in a timely manner and shape the group for success [44], we identify two early outcomes: retention and productivity at the end of the first quarter after joining a group.

We explore the effects of prior experience in the specific context of WikiProjects. WikiProjects are subgroups in Wikipedia, which are intended to help organize volunteer effort around building and improving articles in specific topic areas.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2018, April 21–26, 2018, Montréal, QC, Canada

© 2018 ACM. ISBN 978-1-4503-5620-6/18/04...\$15.00

DOI: <https://doi.org/10.1145/3173574.3174080>

WikiProjects often share structure and volunteer membership. Additionally, the volunteers' editing records on Wikipedia are visible to the public, which gives us an opportunity to explore whether and how their prior record is predictive of future contributions to a group they join.

Our findings, indeed, show mixed effects of prior experience on retention and productivity in the group they join:

- Generalized prior work-productivity experience is positively associated with retention but negatively associated with productivity.
- Prior leadership experience is negatively associated with both retention and productivity.
- Localized prior work-productivity experience is positively associated with both retention and productivity within that focal project.

Using our findings, we hope to advance knowledge about the behavior of members in online production groups, the nuanced effects of prior experience, and inform the designs of early interventions aimed at shaping group success in online social collaborative knowledge systems.

The rest of the paper is organized as follows. In the next section, we discuss the theory and hypotheses concerning retention, productivity, and prior experience within online groups. We then discuss our platform, dataset, and study methodology. We state our research methods and describe our findings followed by a brief discussion of the results. We then conclude by discussing how prior experience may be more broadly effective in supporting the design and management of online social collaborative knowledge systems.

THEORY AND HYPOTHESES

Retention and Productivity

A number of online groups face the problems of lack of early retention and productivity from new users. 46% of members of guilds in World of Warcraft leave their group in less than a month, migrating to other groups within the game [73]. On MovieLens, 60% of new users do not come back after the first session [44]. In The Pearl Open Source Development Project, more than half of the newly registered developers never showed up after their first post [28]. In Usenet groups, 68% of newcomers did not return after their first post [2]. Half of the social, hobby, and work mailing lists had no traffic over a 130-day period [11] and even in active mailing lists, less than 50% of subscribers posted even a single message in a 4-month period [11].

Two outcome measures have been extensively studied in offline work groups as determinants of group success: the member retention to contribute to the group effort, and the quantity of work output [2, 16].

Prior organizational science literature views the lack of volunteer retention from three different perspectives [22, 15, 61]. Because the online community literature is generally in favor of sustaining a steady group of volunteers for continued production [51], the first and more dominant view is about the negative effects of low retention of volunteers. These include the

loss of productive volunteers [65], the loss of social capital [25, 41], the cost of training new, inexperienced volunteers [24], and the weakening of knowledge resources of the organization [41] - all of which deplete the available resources, disrupt the routines and established social ties, threaten the cognitive structures, and the eventual sustainability of the group. The second view sees the positive effect: helping screen out under-performing volunteers [50]. The third adopts a more neutral view that suggests that new volunteers with new skills and knowledge replace those who leave, maintaining the critical mass, and this may be optimal for organizational performance [25]. To reconcile the above arguments, Hausknecht and Holwerda [38] argue that the traditional, aggregated measures of volunteer dropouts such as turnover rates hide variation in the key causal factors that predict retention and performance and so, the specific details concerning who is being retained and who is not are more important than the level of turnover itself [39]. As an example, the loss of a productive manager may be more damaging than the loss of an under-performing employee. Depending on these details, the same level of retention could have different consequences [39, 76].

While it is ideal that volunteer groups should achieve both retention and productivity simultaneously, often, there may be a tension between the two [26, 70]. For instance, it is likely that volunteers may "free-ride" i.e., stay but not contribute, in some groups [1]. Or, core members who take the lion's share of workload may eventually "burnout" and leave the group [5, 21]. Or, core contributors might feel they have accomplished their mission by contributing everything they know and might stop contributing as further contribution may require more research and effort [67]. Or, the presence of multiple experienced core members with clashing interests might lead to conflicts that can erode each others' energy and enthusiasm causing them to leave the group. Accordingly, prior research has examined a number of factors influencing productivity and retention in online production communities. Some of them include members' personality [43, 45], socialization tactics used [18, 30, 31, 32, 47], members' ability to identify with and integrate into the group [14, 63, 75], the diversity of the subgroups they belong to [16], the availability of activities to perform [44], the leadership behaviors within the group [77], the feedback [79], and the type and amount of social support they receive [71] from other members of the group.

There is also some work suggesting that prior experience is predictive of future retention and productivity in groups [10, 59, 60, 75]. However, a closer examination of the theory on retention of newcomers in offline and online groups suggests a more complicated relationship of prior experience with retention and productivity, with some types showing positive effects and others negative ones. And, to our knowledge, there is no work that either makes a clear distinction between the various kinds of prior experience that a volunteer can potentially possess or draws any conclusions about them either individually or together with other group and individual level factors. In this work, we treat *prior experience* more systematically and examine its effects on a new volunteer's early retention and productivity after they join a group.

Prior Experience

Because volunteers frequently move in and out of groups, it is useful to learn about the impact their prior experience in other groups has on their retention and productivity in the future groups they join.

Much of prior work in online communities suggests that prior experience has a positive impact on both the individuals and the communities as a whole. For example, the theory on Legitimate Peripheral Participation (LPP) widely used to describe the newcomer experience in online communities [9] suggests that newcomers' initial peripheral participation is important for them to be acquainted with the tasks, vocabulary, and organizing principles of the community. Experience gained in using editing tools, organizing activities, and communicating and collaborating with other members could positively affect their future performance. Also, prior experience is positively predictive of future productivity and administrative behaviors [10, 59, 60, 75].

On the other hand, research based on analysis of employee behavior in offline organizations suggests that prior experience might have a negative impact on people's performance in a new context. For example, experienced employees are also likely to leave due to mismatch in expectations [27, 54], the need to suppress their perspectives [8, 20], unfavorable group structures [54], or stress and exhaustion [21].

In this paper, we want to resolve the conflicts in prior work by unpacking prior experience. In online peer production groups, we identified three types of prior experience:

(i) **generalized prior work-productivity experience**, which is generalized prior experience associated with normal content production in all the other similar volunteer groups.

(ii) **prior leadership experience**, which is prior experience associated with the tasks of coordination and organization in other similar volunteer groups, and

(iii) **localized prior work-productivity experience**, which is the amount of work a volunteer would invest in a group, as they identify with it, before joining the group.

It is unusual for volunteers to have leadership experience on a group before joining it and so we do not consider the fourth kind of prior experience i.e., localized prior leadership experience.

Based on prior literature in online and offline groups, we now propose hypotheses about the primary effects of the different dimensions of prior experience.

Research Hypotheses

Effects of Generalized Prior Work-Productivity Experience

Prior work-productivity experience is usually positively associated with retention [10, 59, 60, 75] and the lack of it is associated with withdrawal [3, 4, 34]. However, prior work-productivity experience is also associated with a decrease in productivity. A majority of workers in offline work groups eventually reach a plateau in their contributions or decrease them [66]. The initial motivation to produce more could be the desire to learn or grow within the organization [48] and

such a motivation may not exist after they have accomplished their goals [67]. For instance, in university settings, faculty often shift their focus from research to administrative service work after promotion to full-professorship [64]. In online subgroups too, prior work has found that users' motivations change as they become more engaged in the community [9]. The initial motivation could be the desire to contribute what they know or to gain reputation. With an increase in contributions along with experience, they move into more caretaker roles. Accordingly, their contribution levels might change although they stick around. For instance, individuals in the GNOME project¹ increased their coordination work and decreased their technical contributions to specific projects after moving to more lateral authority roles such as board directors [23]. Members with such longer tenures tend to contribute less to subgroups and more to the larger community [70]. Some others who start strong, begin to decline in their contributions later due to a potential buildup of stress and exhaustion [21].

As we read these together, there is an interesting conflict. The more experienced someone is, the less likely they are to leave [10, 59, 60] but their contribution to an individual workgroup within the organization is likely to decrease with change in motivation or roles [70, 23, 66], buildup of stress and exhaustion [70, 21] or because they have contributed everything they know and accomplished their goals [67]. We, therefore, believe that a change in motivation or roles that comes with experience is likely to affect future productivity.

In order to test this, we frame the following two hypotheses regarding the effect of past work productivity experience on future productivity and retention in the new online groups they join within a larger online community:

Hypothesis 1a: *Higher prior generalized work-productivity experience is associated with greater retention in a focal group.*

Hypothesis 1b: *Higher prior generalized work-productivity experience is associated with lesser work-productivity in a focal group.*

Effects of Prior Leadership Experience

Broadly speaking, the prior literature suggests two perspectives to understanding the effects of prior leadership experience on retention and productivity within groups.

The first perspective suggests that when members gain more leadership experience, they are likely to be involved in many interactions outside of the group, and these are likely to pull them away from the focal subgroup [55] affecting both their performance as well as retention. As we have seen in the examples of faculty promotion to a full professorship and individuals moving to administrative roles [23, 64], increase in administrative activities and leadership behavior is strongly associated with a decrease in performance.

¹GNOME is a desktop environment composed of free and open-source software that runs on Linux and most BSD derivatives, and the GNOME project refers to the community behind it which consists of all the software developers, artists, writers, translators, other contributors, and active users of GNOME.

The second perspective suggests that leaders can find it challenging to adjust to a group for various reasons. Prior work in online groups found that users' perceptions of their roles change as they become more engaged in the community [9]. Those who are power users and administrators see themselves as caretakers, as leaders with an established reputation, identities, organizational perspectives, mental models, and existing modes of practice. According to the Social Cognitive Theory (SCT), their self-efficacy (belief in one's capabilities and the ability to complete various actions [6]) in tasks such as knowledge sharing [29, 40, 52] is high. And, the more familiar they are with a domain, the higher their self-efficacy is [17]. When they join a new group, they usually also carry their established reputation, mental models, organizational perspectives and modes of practice from their previous groups [7, 8, 13, 36, 80]. Often, existing members of a group vouch for native patterns and structures to protect native knowledge hierarchies and resist new, innovative ideas, differing practices, or past-reputation-based leadership of these experienced folks until they establish their identities independently in the new group [72]. As a result, for succeeding in their new role in the new group, they may need to modify or suppress their perspectives, innovations, practices or role identities [8, 20]. Sometimes, their performance in the new role may not match their prior performance, their own expectations, or the expectations of the new group [27, 54]. At other times, the layers of structures, bureaucratic requirements, and oppositional rigidities in the new group may serve as barriers for their contributions and practices and leave them frustrated [54]. Often, they themselves tend to make judgments about the level of disparity that exists between their old and new settings, colleagues, and practices [49]. Certain of their attributes or practices may be oppositional to the established knowledge and practice structures and frameworks in their new setting [36] and even generate counter-productive responses among new colleagues [42, 57]. For instance, volunteers tended to get bolder and increased the likelihood of having their work rejected [9, 37] in Wikipedia.

Thus, prior leadership experience can create barriers to fit, adaptation and integration [27, 35]. Consequently, they often experience lesser satisfaction and high degrees of frustration and conflict in their attempts to connect with others in a way similar to their previous setting for needing support for their performance [49]. In the online groups of Wikipedia and del.icio.us, researchers find that there was a dramatic shift in workload from power users to the common user [46]. We, therefore, posit that:

Hypothesis 2a: *Higher prior leadership experience is associated with lesser retention in a focal group.*

Hypothesis 2b: *Higher prior leadership experience is associated with lesser work productivity in a focal group.*

Effects of Localized Prior Work-Productivity Experience

Prior research concerning the transition of potential members from outsiders to organizational members shows that volunteers who strongly identify with a topic area or a group tend to more positively evaluate it, are willing to become more active, and exert more effort than those who don't. And, as they become more active, they tend to contribute more [62,

58]. Also, during the evaluation period, those who see that the group fits their needs join it and remain in it longer, whereas those who don't see it as a fit leave (see [44] for a review). Thus, those who join after preliminary experience with a group are likely to remain longer and contribute more. Similar research examining the hypotheses concerning the effects of college internships on individuals shows a strong support for future employment with the organization for individuals with internships [12, 68]. We therefore hypothesize:

Hypothesis 3a: *Higher localized prior work-productivity experience is associated with greater retention in the focal group.*

Hypothesis 3b: *Higher localized prior work-productivity experience is associated with greater work productivity in the focal group.*

It is unusual for volunteers to have leadership experience on a group before joining it and so we do not consider the fourth kind of prior experience i.e., localized prior leadership experience.

METHODS

Study Platform

We study membership and editing contributions in Wikipedia through WikiProjects². Wikipedia is best known for its *articles* – community-edited pages devoted to specific topics and collectively forming an encyclopedia – but it also has other pages devoted to collaboration (*talk pages* and *project pages*), to people (*editor pages*), and to policies and guidelines. Individual units of contribution are called *edits*, and such edits can be made on any of types of page. Any internet user can contribute content to Wikipedia's pages and is called an *editor*.

WikiProjects are subgroups within Wikipedia where editors come together to improve Wikipedia's coverage of a particular topic. Usually, this is done by organizing a group of related articles under one heading. A typical organization effort might include gathering all pages related to a particular topic under one heading, expanding the content of these articles, aligning articles to the same style of writing, and ensuring the articles meet certain quality standards. A typical main page of a project called the *project page* includes a brief description of the project and its scope, a list of members volunteering to contribute to the project, the list of tasks to be done, and guidelines and policies adhering to which members should work toward content production. Discussions regarding project maintenance and resolution of issues within the broader scope of the project are done in dedicated pages called *project talk pages*.

We choose WikiProjects as our research platform for three reasons. First, prior work identified WikiProjects as an example of Ostrom's *nested organizational structures* with clear goals [33]. Second, there is rich historical data available about editor activities in Wikipedia as a whole as well as various WikiProjects which helps us explore the concept of prior experience and design various metrics around it. Third, because WikiProjects span a large topical scope, we feel conclusions drawn from WikiProjects are more likely to be generalizable

² <https://en.wikipedia.org/wiki/WikiProject>

than those drawn from narrower communities such as health and technical forums.

Dataset

We use the English Wikipedia data dump of June 2, 2015, downloaded from the site³ hosted by the Wikimedia Foundation. The dump data contains the complete revision history of all the pages in English Wikipedia. We use an open source Python package⁴ to pre-process the dump files and extract the revision information stored in the HTML format. To construct the WikiProjects for our analysis, we parse the project templates on articles' talk pages which included the information about which WikiProjects an article belongs to. We include articles that belonged to multiple WikiProjects in all those WikiProjects. This resulted in an initial set of 1,949 WikiProjects. From these, we exclude projects that never grew to more than three members (which is the minimum size of a group) as we want to understand this in a collaborative context. Further, we exclude projects that are not related to specific topics such as *WikiProject: Articles for creation*. This resulted in a final dataset of 1,054 WikiProjects.

Many editors edit the pages without being aware of any projects. So, it would not make sense to look at edits to any page in the scope of the project randomly. Also, we want to explore the notion of pre-joining contributions for which we want to explicitly identify volunteer membership in the groups. Two approaches to identifying volunteer membership in projects are common in prior literature: declared membership, based on voluntary sign-up on the project page, and participatory membership, where an editor is considered to have joined a project when they made their first edit to either the project page or the project talk page. Morgan et al. compared the two approaches and found no significant difference [56]. In this work, we choose the participatory approach. This yielded a total of 88,427 members of the projects in our sample (excluding the bots) who contributed a total of 44,135,006 edits over 14 years.

Operationalization

Definition of Joining: In this work, we operationalize *joining* as the first explicit *project or project talk-page edit*. This definition is not original to our work; it is used in prior work by others including [69, 75, 78].

Independent Variables

Generalized Prior Work-productivity Experience: We count the total number of edits an editor made on the article (and the corresponding talk) pages in Wikipedia before joining the focal project except edits on the focal project as their generalized prior work-productivity experience as these represent efforts in individual article content production.

Prior Leadership Experience: We count the total number of edits an editor made on the project pages and the corresponding talk pages in Wikipedia before joining the focal project as their prior leadership experience as these represent organizational behavior at the project level.

³ <https://dumps.wikimedia.org/enwiki/20150602/>

⁴ <https://github.com/earwig/mwparserfromhell>

Localized Prior Work-productivity Experience: We count the total number of edits an editor made on the main article pages and the corresponding talk pages of articles within the scope of the project before joining it as the localized prior work-productivity experience as these represent efforts on content production for individual articles for a specific project.

The explicit joining action indicates the editor's first point of awareness of a larger community of members and of a collection of pages beyond the page (or pages) they are editing. This is the point where they *begin* documented project-level collaborations and *begin* exhibiting different behaviors with group members compared to non-group members [56]. Prior research shows that those that explicitly join groups share a strong sense of group identity [63], establish group norms and common repertoires [53], may exhibit in-group favoritism [28, 74], which non-members may not. We find that even employees who join a company, despite interning many times, are considered *new* and go through new-employee training. For all these reasons, we consider those with localized experience also as *newcomers*.

Dependent Variables

Early identification of group failures can help community moderators intervene in a timely manner and shape the group for success [44]. And prior work that studied WikiProjects longitudinally collected project-related measures for quarters i.e., 90-day periods, as a quarter captures a regular editing cycle on Wikipedia [61, 70]. Since we are interested in studying the effects of prior experience on early retention and early productivity, we measure our response variables at the end of the first quarter for all the volunteers.

Early Retention: We measure this as a binary variable. Consistent with prior work [75], we regard an editor as having withdrawn from a project if they have not made any edits for a continuous six-month period at the end of the first quarter in any of the article, the article talk pages, the project or the project talk pages.

Early Productivity: We measure early productivity in terms of the number of edits made [16, 46, 47, 70, 75] on all articles within the scope of the project during the first quarter after joining.

Control Variables

Prior work shows that a number of other factors are likely to influence outcomes of members' successful collaborations in WikiProjects [16, 70, 75]. We, therefore, explore our three dimensions of prior experience along with all of these factors to see if prior experience measures provide an additive value over these in determining early retention and productivity of new volunteers in the focal project. We have operationalized many of these in ways consistent with prior work:

Project Scope: This is a count of the number of articles within the scope of the project [16, 70, 75].

Project Size: This is a count of the number of editors who participated in the focal project before the focal editor joined [75].

Variables	Descriptive Statistics						Correlations								
	Min	25%ile	50%ile	Mean	75%ile	Max	1	2	3	4	5	6	7	8	9
1. Project Scope	0	1016	5179	37455	18119	1143441	1.00								
2. Project Size	0	65	194	482	517	5248	0.41	1.00							
3. Project Age	0	22	47	52	77	167	0.16	0.52	1.00						
4. Editor Tenure	0	3	16	26	39	165	-0.03	0.05	0.36	1.00					
5. Interest Match	0	0.1	0.1	0.1	0.2	1.0	0.01	-0.04	-0.04	0.01	1.00				
6. No. of Simul Projects	0	1	5	39	22	1251	-0.04	-0.09	0.01	0.18	-0.03	1.00			
7. Prior Gen. Work Exp.	0	97	1664	17948	12418	1285322	-0.03	-0.04	0.11	0.37	-0.01	0.29	1.00		
8. Prior Leadership Exp.	0	0	6	209	94	23451	-0.04	-0.06	0.04	0.25	-0.02	0.38	0.35	1.00	
9. Prior Loc. Work Exp.	0	0	6	128	50	134810	0.17	0.09	0.06	0.10	0.05	0.01	0.14	0.02	1.00

Table 1. Descriptive Statistics and Correlations of Variables.

Project Age: This is a count of the number of months from the project’s creation until the focal editor joined [75]. This variable is used to control for the project maturity, which may affect the ease with which new members could integrate into and contribute to the project.

Editor Tenure: This is a count of the number of months from the registration of the editor in Wikipedia to the time they joined the focal project.

Interest Match: This measures the interest match between an editor and the focal project. Following prior work [75], we create a topic vector for the editor based on their prior edits on articles, another topic vector for the project based on the articles within the scope, and compute the cosine similarity between the two vectors.

Number of Simultaneous Projects: This is a simple count of the number of projects the editor has any edits in during the time he is a member of the focal project.

ANALYSIS AND RESULTS

Analysis Strategy

We present the descriptive statistics and correlations among all our variables in Table 1.

Table 1 suggests that most of the variables have a heavily right-skewed distribution. We, therefore, log-transform all the above variables (except *Interest match*, which is between 0 and 1) to stabilize the variance and improve the fit of the models in which we will use them as predictors. We also standardize all of them (i.e., normalize to mean zero and unit standard deviation) for ease of comparing their relative importance (i.e., the coefficients across the predictors in the models we build). Most of the correlations between the variables are low. Nonetheless, in order to examine and remove any potential multi-collinearity between the individual predictors, we compute the VIFs (Variance Inflation Factors) for all the variables included in the model and find that removing the variables *Project Size* and *Editor Tenure* from the set of predictors achieves a set with all individual VIFs sufficiently below 5, the recommended maximum for behavioral sciences data [19] (including these two gave at least two values very close to 5). The VIFs for all predictors used in our models are shown in Table 2.

Variable	VIF
Project Scope	1.51
Project Age	1.11
Interest Match	1.12
No. of Simul Projects	3.04
Prior Gen. Work Exp.	3.16
Prior Leadership Exp.	3.68
Prior Loc. Work Exp.	1.87
Mean VIF	2.21

Table 2. Collinearity diagnostics on all the Independent Variables after log-transforming and standardizing.

The standard errors are small and we have seen that the predictor variables do not change signs when we try to remove variables further from the remaining set of predictors here, indicating that this set of predictors do not pose problems of multicollinearity.

Each project can have multiple editors and an editor can belong to multiple projects. Our data, therefore, is cross-nested between WikiProjects and individual editors. We, therefore, use random-effects regression models to take care of potential correlations across observations that are nested within a level (e.g., editors nested under projects). For our first outcome measure, i.e., determining the early retention, we use a binary response variable that measures whether or not an individual volunteer remains in the project by the end of first quarter. For our second outcome measure, i.e., determining the early productivity, we see that our dependent variable is the total number of edits made in the first quarter after joining which is a count variable with over-dispersion (i.e., the variance is much higher than the mean). We, therefore, use a negative binomial regression model to handle this scenario.

We control for the effect of variables examined in prior literature (namely, Project Scope, Project Age, Interest Match, and Number of Simultaneous Projects) while examining the additive effects of the three Prior Experience variables (namely, Generalized Prior Work-productivity Experience, Prior Leadership Experience, and Localized Prior Work-productivity Experience).

Variables	Models for Early Retention						Models for Early Productivity					
	Model I		Model II		Model III		Model IV		Model V		Model VI	
	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.
Project Scope	1.012***	0.024	0.576***	0.021	0.518***	0.02	1.659***	0.045	0.805***	0.034	0.774***	0.034
Project Age	-0.511***	0.01	-0.447***	0.011	-0.41***	0.011	-0.297***	0.009	-0.173***	0.008	-0.146***	0.008
Interest Match	0.61***	0.010	0.305***	0.011	0.301***	0.011	1.103***	0.013	0.61***	0.01	0.583***	0.01
No. of Simul Projects	0.912***	0.01	1.605***	0.019	1.706***	0.02	0.246***	0.010	0.637***	0.013	0.686***	0.013
Prior Gen. Work Exp.			0.449***	0.016	0.091***	0.029			-0.276***	0.013	-0.431***	0.015
Prior Leadership Exp.			-1.381***	0.02	-1.169***	0.02			-0.808***	0.013	-0.771***	0.014
Prior Loc. Work Exp			1.046***	0.014	1.142***	0.015			1.496***	0.009	1.584***	0.010
Gen. Work \times Leadership					-0.294***	0.013					-0.159***	0.011
Gen. Work \times Loc. Work					-0.390***	0.019					-0.26***	0.011
Leadership \times Loc. Work					0.378***	0.017					0.274***	0.001
AIC	100434		83859		82936		579852		548927		548091	
χ^2			16581.42***		929.25***				30931.46***		841.93***	

Table 3. Results of the effects of prior experience on Early Retention (Models I through III) and Early Productivity (Models IV through VI). We use the following notation in tables for p-value significance *: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, ns: $p > 0.05$**

Results

Out of our initial dataset, we use a sample of 30,000 editors along with all their edits in all the WikiProjects they participated. To examine whether the prior experience variables have additive value over and above the variables we are controlling for, we build three separate models each for retention and productivity: base models (I and IV) containing just the control variables, the models (II and V) containing the control variables as well as the prior experience variables, and the models (III and VI) that also include potential 2-way interactions among the prior experience variables. We do not include 3-variable and higher interactions for they not only make interpretation considerably more complex but also do not significantly improve our understanding of interactions between the variables. The results of the random effects logistic regression for retention (Models I, II, and III) at the end of the first quarter and those of the random effects negative binomial regression for productivity during the first quarter are shown in Table 3.

Choosing the best model

The model's fitness to the data can be determined either by comparing actual values with the predicted values using the model or by comparing the model with other competing models. Comparison with competing models seems more appropriate in this context since the outcomes are over-dispersed counts. We use the Akaike Information Criterion (AIC) to evaluate the goodness of fit for each of these models. AIC rewards goodness of fit of the model to the data while penalizing complexity (i.e., more number of predictors). AICs are always compared with each other and individual AIC magnitudes are not interpreted by themselves as they are affected greatly by sample size. In general, the smaller the AIC among a set of candidate competing models, the better the model. Using the AIC, we note that the models including both the prior experience variables and their interactions (Model III for early retention and Model VI for early productivity) are

better. We find also that the difference in log likelihoods of the base model with Model II is statistically significant ($p < 0.001$) with $\chi^2 = 16581.42$ and of Model II with Model III is statistically significant ($p < 0.001$) with $\chi^2 = 929.25$, indicating again that Model III is better than Model II and Model I. Similarly, we find Model VI is better than Models V and IV. We also find that prior experience variables have coefficients that are comparable in magnitude to the control variables.

We interpret Models III and VI to understand the impact of various kinds of variables. Note that the above variables are log-transformed (with e as the base of the logarithm) and normalized to mean 0 and a standard deviation of 1. This makes it easier to understand the impact of different predictors with respect to each other. First, we note that all the predictors are significant and the effects of control variables are largely consistent with prior work. We, therefore, focus on interpreting only the variables of interest (i.e., the prior experience variables) on the linear scale to understand the actual impact of prior experience.

Overall Effects of Prior Experience Variables

Based on Model III, we find that holding all the other variables constant, an e -fold (i.e., roughly 2.7 times) increase in generalized prior work-productivity experience (in terms of number of prior article and article talk page edits) is roughly associated with an overall 3% increase in the odds of retention, whereas an e -fold increase in prior leadership experience (in terms of number of prior project and project talk page edits) is roughly associated with an overall 62% decrease in the odds of retention, and an e -fold increase in localized prior work-productivity experience (in terms of pre-joining article edits to the focal project) is roughly associated with an overall 70% increase in the odds of retention. And based on Model VI, we find that holding all the other variables constant, an e -fold increase in generalized prior work-productivity experience is roughly associated with a 17% decrease in productivity

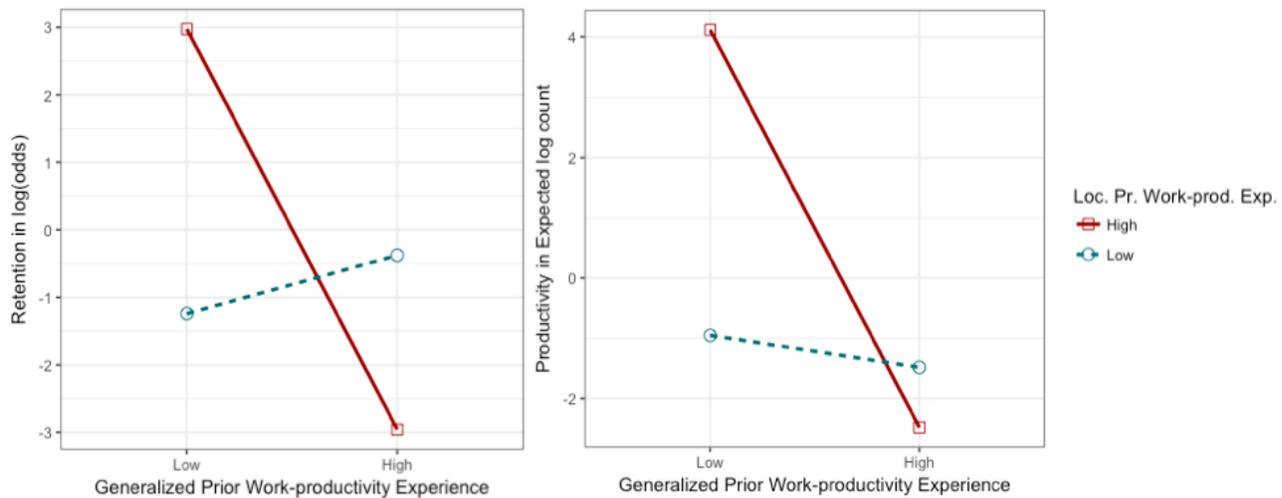


Figure 1. Figure showing the interactions between low and high values of the various prior experience variables. The plots for retention are on the left and the plots for productivity are on the right.

(i.e., the expected count of number of edits made) during the first quarter, an e -fold increase in prior leadership experience is associated with a 37% decrease in the expected count of number of edits and an e -fold increase in localized prior work-productivity experience is associated with a 108% increase in the expected count of number of edits.

The above overall percentages include the effects of interactions within them. In order to tease out the effects of individual interactions, we plot the interaction plots for the two response variables for low and high values of various prior experience variables. Below, we present and discuss a couple that are interesting.

Interaction Effects of Prior Experience

Figure 1 shows how generalized prior work productivity experience interacts with localized prior work-productivity experience. We see that the retention and the productivity are the lowest when both generalized prior work-productivity and localized prior work-productivity experiences are high. One potential scenario for high localized work-productivity is when the volunteers have already contributed everything they know and contributing more would require much more research and effort. A high generalized prior work-productivity experience might be indicating a potential burnout effect due to stress or exhaustion - the combination of which is possibly associated with the low retention and productivity in the focal project. On the other hand, we see that the retention and productivity are the highest when generalized prior work-productivity experience is low - an example of this is a situation where a potential burnout has not yet happened and the high level of attachment associated with the high level of localized prior work-productivity experience is potentially responsible for high retention and high productivity in the focal project.

DISCUSSION

First, our findings show that generalized prior work-productivity experience is positively related to retention and

negatively related to the productivity confirming our hypotheses **1a** and **1b**, prior leadership experience is negatively related to both the retention as well as the productivity confirming our hypotheses **2a** and **2b**, and localized prior work-productivity experience is positively associated with both the retention and productivity confirming our hypotheses **3a** and **3b**.

Second, while prior work shows only a positive relationship between metrics based on prior experience and future productivity and administrative behaviors [10, 59, 60, 75], our work confirms that the relationship is, indeed, much more complicated, with some types showing positive effects and others negative ones. Even with the caveat that we are talking about productivity in its simplest form i.e., edit count, our work shows that prior experience, in general, is worse for productivity although better for retention.

Third, it is interesting to observe some of these interactions. Consider the interactions between localized prior work-productivity experience and prior leadership experience. This could be understood in two ways: (1) Localized prior work-productivity in a specific topic area has a huge positive effect that it dampens any of the negative effects of prior leadership experience. OR (2) The benefits of localized work-productivity get cut down, the more someone has overall prior leadership experience. However, the net effect of localized work-productivity still remains positive (see Table 3. Hence, *content is king, and leadership lags*. The effect of generalized prior work experience in the presence of interactions is pretty small (e.g., compare models II and III). This means that generalized prior work experience is useful and positively predictive of retention only when the volunteers do not have localized or leadership experiences. The benefits of localized work productivity experience which are substantial get cut down significantly if you have too much overall experience (see Figures 5 and 6). Future experiments along these directions could reveal interesting insights about causal relationships.

Theoretical Implications

While the notion of *prior experience* has been explored before, in this work, we show for the first time, that *different kinds* of prior experience reveal and predict more interesting and nuanced effects in volunteering groups. We show, for instance, the importance of identifying and distinguishing between prior work productivity experience and prior leadership experience, and between generalized and localized experiences. We think these concepts might generalize not just to other online peer production communities but to volunteer or organizational management more generally. Given our findings, it would be useful to study other domains (and revisit prior studies of newcomer contribution/retention) through the lens of different types of prior experience.

Practical Implications

The practical implication we have for WikiProjects is that we can improve the success of recruiting and retaining productive contributors. Looking only at the primary effects, the first implication is that the localized prior work-productivity experience is the most effective indicator of whether the new volunteer will become a productive and dedicated member. Our findings suggest that WikiProjects looking to recruit and retain productive workers should focus their recruitment foremost on those who have already demonstrated a commitment to the specific work or cause of the project.

However, projects that are smaller or in their initial phases may not have a lot of candidates with a demonstrated commitment to the project to recruit. In these cases, project recruiters can recruit those with generalized prior work-productivity experience who are likely to stay in the project longer but not necessarily be very productive. Recruiting experienced leaders is much trickier, as leadership experience is generally associated with low productivity and retention. When leaders are sought (e.g., to bolster leadership in a new group), WikiProject organizers may want to both verify commitment to the cause/topic and consider specific re-orientation/transition plans to help the leader better integrate and be successful in the WikiProject.

Generalization

The framework, metrics, and hypotheses we provide in this paper apply not just to WikiProjects, but also to other peer production groups such as OSS projects in GitHub, communities in StackExchange, groups in GoodReads or projects in OpenStreetMap. Considering prior experience can reveal more interesting and nuanced effects, one might even consider broader kinds of prior experience available in these specific volunteer groups. For instance, in the case of projects in OpenStreetMap, these results might suggest using their prior communication patterns with project members within projects along with prior map creation and leadership activities to predict new members' retention and productivity. We encourage future research in these communities.

We believe similar effects of prior experience may appear in offline groups. Consider a mosque looking to start a homeless shelter. As it recruits a collection of volunteers to staff the shelter, how much should it draw on top volunteers in other efforts (interfaith outreach, study sessions), how much on

leaders of other efforts, and how much on volunteers who have worked in other homeless shelters and similar projects, even outside the mosque? Of course, much additional research is needed to validate the generalizability of our findings.

CONCLUSION

In this study, we explore the effects of prior experience of new volunteers on their early retention and productivity in the group they join with the understanding that early identification of group failures can help community moderators intervene in a timely manner and craft the group for success. We found that certain kinds of prior experience have positive effects on newcomer retention and productivity whereas other kinds of prior experience have negative effects.

Specifically, we carried out the study on a sample of 30,000 new editors to 1,054 WikiProjects, which are groups dedicated to building content around specific topic areas. This platform allowed us to measure prior experience in multiple dimensions and potential interactions between them which could generalize to other communities with similar structures. Also, WikiProjects have been ideal for such an exploration owing to their well-established and shared structures, shared membership and publicly available historical data about each volunteer. Through our analysis, we found that (i) generalized prior work-productivity experience (measured by overall prior article and article talk page edits) is positively associated with retention, but negatively associated with productivity within the focal group, (ii) prior leadership experience (measured by overall prior project and project talk page edits) is negatively associated with both retention and productivity within the focal group, and (iii) localized prior work-productivity experience (measured by pre-joining article edits on a focal group) is positively associated with both retention and productivity within the focal group.

Limitations, Future Work and Potential Impact

In this study, we made a preliminary investigation of the effects of prior experience on early retention and productivity within the subgroups of a larger community in order to understand if examining prior experience has any value and we found that considering the prior experience of a member adds value over other group-level metrics such as composition and structure, and, even within prior experience, some kinds of prior experience have positive effects on group outcomes whereas other kinds have negative effects. However, prior experience might also vary with factors such as project age and the number of simultaneous projects. Future work should look at these variations in order to gain a deeper insight into the effect of prior experience.

Our data analysis, although providing us with key insights into the interactions of various dimensions of prior experience and their effects on retention and productivity, provides only limited support for understanding *why* the association between prior experience and outcomes exist and in what ways they are causal. In addition, we see interactions that at this point we don't have data to explain. Further qualitative studies could reveal more insights into this which we leave for future work to explore. Based on our findings, future studies could also

run field experiments with varying on-boarding processes for volunteers with different kinds and/or levels of experience. For instance, in groups with no opportunities for pre-joining contributions, volunteers in one condition might require going through a probationary period where they achieve a certain level of contributions before they become members and be compared with volunteers in another condition where there is no such requirement and both may be measured on their retention and productivity.

We do not have information regarding the amount of workload of these editors in their personal lives or in other online communities with similar skillset and we believe high contributors online are also potentially very knowledgeable in their respective fields which might also affect their performance once they undertake too many activities online. Again, conducting qualitative studies might help us gather this information and further insights into the interplay. We leave this also for future research to explore.

Consistent with much of prior work, our study used productivity and retention as measures of group success. Future research could extend this by incorporating more nuanced measures such as the quality of the artifacts produced or rate or amount of progress toward group/community goals.

In this study, we examine the phenomenon of near transfer, i.e., how prior experience is associated with group outcomes in case of groups having similar structures. However, knowledge, usability experience, and human capital may be easier to transfer across groups with similar structures than they would be across groups with dissimilar structures within the same community or within different communities. Future work should consider extending these findings to more heterogeneous environments with different structures and affordances.

The *potential impact* of this work lies in three areas. First, we have demonstrated the importance of considering diverse types of prior experience in predicting the longevity and productivity of experienced newcomers. This result makes a theoretical contribution to our understanding of newcomer behaviors in online groups. Second, this work is directly applicable to WikiProjects where it can be used to identify individuals to recruit and to plan pre-joining activities to test and/or build commitment to a project. Third, while our results have not yet been tested outside Wikipedia, we provide a framework for extending this research into new domains, including generalized hypotheses and research methods that can be used for systematic research and exploration.

ACKNOWLEDGMENTS

We thank all the reviewers, and Andrew Hall, Zach Levonian, and Taavi Tajjala of GroupLens Research for their helpful feedback. We are grateful to NSF for funding our research through the grant IIS-13-19382.

REFERENCES

1. R. Albanese and D.D. Van Fleet. 1985. The Free-Riding Tendency. *Academy of Management Review* 10, 2 (1985), 244–255.
2. Jaime Arguello, Brian S. Butler, Elisabeth Joyce, Robert Kraut, Kimberly S. Ling, Carolyn Rosé, and Xiaoqing Wang. 2006. Talk to Me: Foundations for Successful Individual-group Interactions in Online Communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 959–968. DOI : <http://dx.doi.org/10.1145/1124772.1124916>
3. Hugh J Arnold and Daniel C Feldman. 1982. A multivariate analysis of the determinants of job turnover. *Journal of applied psychology* 67, 3 (1982), 350.
4. Jeffrey B Arthur. 1994. Effects of human resource systems on manufacturing performance and turnover. *Academy of Management journal* 37, 3 (1994), 670–687.
5. A.B. Bakker, K.I. Van Der Zee, K.A. Lewig, and M.F. Dollard. 2007. The Relationship Between the Big Five Personality Factors and Burnout: A Study Among Volunteer Counselors. *The Journal of Social Psychology* 146, 1 (2007), 31–50.
6. Albert Bandura. 1997. *Self-efficacy: The exercise of control*. Macmillan.
7. Janice M. Beyer and David R. Hannah. 2002. Building on the Past: Enacting Established Personal Identities in a New Work Setting. *Organization Science* 13, 6 (Nov. 2002), 636–652. DOI : <http://dx.doi.org/10.1287/orsc.13.6.636.495>
8. Andrew D. Brown. 2015. Identities and Identity Work in Organizations. *International Journal of Management Reviews* 17, 1 (2015), 20–40. DOI : <http://dx.doi.org/10.1111/ijmr.12035>
9. SL Bryant and A Forte. 2005. Bruckman. A. 2005. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. *Proceedings of GROUP'05* (2005), 1–10.
10. Moira Burke and Robert Kraut. 2008. Mopping up: modeling wikipedia promotion decisions. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, 27–36.
11. Brian S. Butler. 1999. When is a group not a group: An empirical examination of metaphors for online social structure. In *Carnegie Mellon University*.
12. Gerard Callanan and Cynthia Benzing. 2004. Assessing the role of internships in the career-oriented employment of graduating college students. *Education+ Training* 46, 2 (2004), 82–89.
13. Jon C. Carr, Allison W. Pearson, Michael J. Vest, and Scott L. Boyar. 2006. Prior Occupational Experience, Anticipatory Socialization, and Employee Retention. *Journal of Management* 32, 3 (2006), 343–359.
14. Luis V Casaló, Carlos Flavián, and Miguel Guinalú. 2013. New members' integration: Key factor of success in online travel communities. *Journal of Business Research* 66, 6 (2013), 706 – 710. DOI : <http://dx.doi.org/10.1016/j.jbusres.2011.09.007> International Tourism Behavior in Turbulent Times.

15. Galyen N. Chandler, Benson Honig, and Johan Wiklund. 2005. Antecedents, moderators, and performance consequences of membership change in new venture teams. *Journal of Business Venturing* 20, 5 (2005), 705 – 725.
16. Jilin Chen, Yuqing Ren, and John Riedl. 2010. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 821–830.
17. Jengchung Victor Chen, Chinho Lin, David C Yen, and Kyaw-Phyo Linn. 2011. The interaction effects of familiarity, breadth and media usage on web browsing experience. *Computers in Human Behavior* 27, 6 (2011), 2141–2152.
18. Boreum Choi, Kira Alexander, Robert E Kraut, and John M Levine. 2010. Socialization tactics in wikipedia and their effects. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM, 107–116.
19. Jacob Cohen, Patricia Cohen, Stephen G West, and Leona S Aiken. 2013. *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.
20. Samantha A Conroy and Anne M O’Leary-Kelly. 2014. Letting go and moving on: Work-related identity loss and recovery. *Academy of Management Review* 39, 1 (2014), 67–87.
21. C.L. Cordes and T.W. Dougherty. 1993. A Review and An Integration of Research on Job Burnout. *Academy of Management Review* 18, 4 (1993), 621–656.
22. John L. Cotton and Jeffrey M. Tuttle. 1986. Employee turnover: A meta-analysis and review with implications for research. *Academy of management Review* 11, 1 (1986), 55 – 70.
23. Linus Dahlander and Siobhan O’Mahony. 2011. Progressing to the center: Coordinating project work. *Organization science* 22, 4 (2011), 961–979.
24. Y. Rene Darmon. 1990. Identifying sources of turnover costs: A segmental approach. *The Journal of Marketing* (1990), 46–56.
25. Gregory G. Dess and Jason D. Shaw. 2001. Voluntary turnover, social capital, and organizational performance. *Academy of management review* 26, 3 (2001), 446–456.
26. Morton Deutsch. 1959. Some factors affecting membership motivation and achievement motivation in a group. *Human Relations* 12, 1 (1959), 81–95.
27. Gina Dokko, Steffanie L. Wilk, and Nancy P. Rothbard. 2009. Unpacking Prior Experience: How Career History Affects Job Performance. *Organization Science* 20, 1 (2009), 51–68. DOI : <http://dx.doi.org/10.1287/orsc.1080.0357>
28. Nicolas Ducheneaut. 2005. Socialization in an open source software community: A socio-technical analysis. *Computer Supported Cooperative Work (CSCW)* 14, 4 (2005), 323–368.
29. Matthew S Eastin and Robert LaRose. 2000. Internet self-efficacy and the psychology of the digital divide. *Journal of Computer-Mediated Communication* 6, 1 (2000), 0–0.
30. Rosta Farzan, Laura A Dabbish, Robert E Kraut, and Tom Postmes. 2011. Increasing commitment to online communities by designing for social presence. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 321–330.
31. Rosta Farzan, Robert Kraut, Aditya Pal, and Joseph Konstan. 2012. Socializing Volunteers in an Online Community: A Field Experiment. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW ’12)*. ACM, New York, NY, USA, 325–334. DOI : <http://dx.doi.org/10.1145/2145204.2145256>
32. Andrea Forte, Niki Kittur, Vanessa Larco, Haiyi Zhu, Amy Bruckman, and Robert E Kraut. 2012. Coordination and beyond: social functions of groups in open content production. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 417–426.
33. Andrea Forte, Vanesa Larco, and Amy Bruckman. 2009. Decentralization in Wikipedia governance. *Journal of Management Information Systems* 26, 1 (2009), 49–72.
34. Rodger W Griffeth, Peter W Hom, and Stefan Gaertner. 2000. A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of management* 26, 3 (2000), 463–488.
35. Boris Groysberg. 2010. *Chasing stars: The myth of talent and the portability of performance*. Princeton University Press, Princeton, NJ, USA.
36. Boris Groysberg, Linda Eling-Lee, and Robin Abrahams. 2010. What it takes to make ‘star’ hires pay off. *MIT Sloan Management Review* 51, 2 (2010), 57–61.
37. Aaron Halfaker, Aniket Kittur, Robert Kraut, and John Riedl. 2009. A jury of your peers: quality, experience and ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*. ACM, 15.
38. John P. Hausknecht and Jacob A. Holwerda. 2013. When does employee turnover matter? Dynamic member configurations, productive capacity, and collective performance. *Organizational Science* 24, 1 (2013), 210–225.
39. John P. Hausknecht and Charlie O. Trevor. 2011. Collective turnover at the group, unit, and organizational levels: Evidence, issues, and implications. *Journal of Management* 37, 1 (2011), 352–388.

40. Meng-Hsiang Hsu, Teresa L Ju, Chia-Hui Yen, and Chun-Ming Chang. 2007. Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectations. *International journal of human-computer studies* 65, 2 (2007), 153–169.
41. M.A. Huselid. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *AMJ* 38, 3 (1995), 635–672.
42. John Kammeyer-Mueller, Connie Wanberg, Alex Rubenstein, and Zhaoli Song. 2013. Support, undermining, and newcomer socialization: Fitting in during the first 90 days. *Academy of Management Journal* 56, 4 (2013), 1104–1124.
43. Raghav Pavan Karumur and Joseph A. Konstan. 2016. Relating Newcomer Personality to Survival and Activity in Recommender Systems. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization (UMAP '16)*. ACM, New York, NY, USA, 195–205. DOI: <http://dx.doi.org/10.1145/2930238.2930246>
44. Raghav Pavan Karumur, Tien T. Nguyen, and Joseph A. Konstan. 2016. Early Activity Diversity: Assessing Newcomer Retention from First-Session Activity. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 595–608. DOI: <http://dx.doi.org/10.1145/2818048.2820009>
45. Raghav Pavan Karumur, Tien T. Nguyen, and Joseph A. Konstan. 2017. Personality, User Preferences and Behavior in Recommender Systems. *Information Systems Frontiers* (2017).
46. Aniket Kittur, Ed Chi, Bryan A Pendleton, Bongwon Suh, and Todd Mytkowicz. 2007. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World wide web* 1, 2 (2007), 19.
47. Aniket Kittur, Bryan Pendleton, and Robert E Kraut. 2009. Herding the cats: the influence of groups in coordinating peer production. In *Proceedings of the 5th international Symposium on Wikis and Open Collaboration*. ACM, 7.
48. Peter Kollock. 1999. The economies of online cooperation. *Communities in cyberspace* 220 (1999).
49. Russell F Korte. 2009. How newcomers learn the social norms of an organization: A case study of the socialization of newly hired engineers. *Human Resource Development Quarterly* 20, 3 (2009), 285–306.
50. David Krackhardt and Lyman W. Porter. 1985. When friends leave: A structural analysis of the relationship between turnover and stayers' attitudes. *Administrative science quarterly* (1985), 242–261.
51. Robert E. Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph A. Konstan, Yuding Ren, and John Riedl. 2012. *Building Successful Online communities: Evidence-based social design*. MIT Press, Cambridge, MA, USA.
52. Robert LaRose. 2009. Social cognitive theories of media selection. *Media choice: A theoretical and empirical overview* (2009), 10–31.
53. Jean Lave and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge university press.
54. Claire M. Gardiner. 2016. Legitimizing processes: Barriers and facilitators for experienced Newcomers' entry transitions to knowledge practices. (06 2016).
55. J Miller McPherson, Pamela A Popielarz, and Sonja Drobnic. 1992. Social networks and organizational dynamics. *American sociological review* (1992), 153–170.
56. Jonathan T Morgan, Michael Gilbert, David W McDonald, and Mark Zachry. 2013. Project talk: Coordination work and group membership in WikiProjects. In *Proceedings of the 9th International Symposium on Open Collaboration*. ACM, 3.
57. Bjørn Erik Mørk, Thomas Hoholm, Gunnar Ellingsen, Bjørn Edwin, and Margunn Aanestad. 2010. Challenging expertise: On power relations within and across communities of practice in medical innovation. *Management Learning* 41, 5 (2010), 575–592.
58. Charles A O'Reilly III, David F Caldwell, and William P Barnett. 1989. Work group demography, social integration, and turnover. *Administrative science quarterly* (1989), 21–37.
59. Katherine Panciera, Aaron Halfaker, and Loren Terveen. 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*. ACM, 51–60.
60. Katherine Panciera, Reid Priedhorsky, Thomas Erickson, and Loren Terveen. 2010. Lurking? cyclopaths?: a quantitative lifecycle analysis of user behavior in a geowiki. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1917–1926.
61. Xiangju Qin, Michael Salter-Townshend, and Pádraig Cunningham. 2014. Exploring the Relationship between Membership Turnover and Productivity in Online Communities. In *Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM '14)*. AAAI.
62. Amy E Randel and Kimberly S Jaussi. 2003. Functional background identity, diversity, and individual performance in cross-functional teams. *Academy of Management Journal* 46, 6 (2003), 763–774.

63. Yuqing Ren, Robert Kraut, and Sara Kiesler. 2007. Applying Common Identity and Bond Theory to Design of Online Communities. *Organization Studies* 28, 3 (2007), 377–408. DOI: <http://dx.doi.org/10.1177/0170840607076007>
64. Yuqing Ren and John Riedl. *Helping Wikipedia versus Helping a WikiProject: Subgroup Dynamics, Member Contribution and Turnover in Online Production Communities*. Technical Report. <https://pdfs.semanticscholar.org/0601/4b5cde0c70ff0bb33151e80ba5ede53385ab.pdf>
65. Myra H. Strober. 1990. Human capital theory: Implications for HR managers. *Industrial Relations* 29, 2 (1990), 214–239.
66. Michael C Sturman. 2003. Searching for the inverted U-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships. *Journal of Management* 29, 5 (2003), 609–640.
67. Lesley .C Taylor. 2009. Thousands of editors Leaving Wikipedia. Article. (23 November 2009). Retrieved September 4, 2017 from https://www.thestar.com/news/2009/11/23/thousands_of_editors_leaving_wikipedia.html.
68. M Susan Taylor. 1988. Effects of college internships on individual participants. *Journal of Applied Psychology* 73, 3 (1988), 393.
69. Hang Ung and Jean-Michel Dalle. 2010. Project Management in the Wikipedia Community. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration (WikiSym '10)*. ACM, New York, NY, USA, Article 13, 4 pages. DOI: <http://dx.doi.org/10.1145/1832772.1832790>
70. Loxley Sijia Wang, Jilin Chen, Yuqing Ren, and John Riedl. 2012a. Searching for the Goldilocks Zone: Trade-offs in Managing Online Volunteer Groups. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 989–998. DOI: <http://dx.doi.org/10.1145/2145204.2145351>
71. Yi-Chia Wang, Robert Kraut, and John M. Levine. 2012b. To Stay or Leave?: The Relationship of Emotional and Informational Support to Commitment in Online Health Support Groups. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 833–842. DOI: <http://dx.doi.org/10.1145/2145204.2145329>
72. Etienne .C. Wenger-Trayner and Beverly Wenger-Trayner. 2015. *Learning in landscapes of practice: Boundaries, Identity, And Knowledgeability in Practice-Based Learning* (1st ed.). Routledge, New York, NY, Chapter Learning in a landscape of practice: A framework., 13–29.
73. Dmitri Williams, Nicolas Ducheneaut, Li Xiong, Yuanyuan Zhang, Nick Yee, and Eric Nickell. 2006. From tree house to barracks: The social life of guilds in World of Warcraft. *Games and Culture* 1 (2006), 338–361.
74. KY Williams and CA O'Reilly III. 1998. Demography and Diversity in Organisations: A review of 40 years of research in BM Staw and LL Cummings (eds) *Research in Organisational Behaviour* Vol. 20. *Jai Pres, Connecticut* (1998).
75. Bowen Yu, Yuqing Ren, Loren Terveen, and Haiyi Zhu. 2017a. Predicting Member Productivity and Withdrawal from Pre-Joining Attachments in Online Production Groups. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1775–1784. DOI: <http://dx.doi.org/10.1145/2998181.2998227>
76. Bowen Yu, Xinyi Wang, Allen Yilun Lin, Yuqing Ren, Loren Terveen, and Haiyi Zhu. 2017b. Out With The Old, In With The New?: Unpacking Member Turnover in Online Production Groups. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 117 (Dec. 2017), 19 pages. DOI: <http://dx.doi.org/10.1145/3134752>
77. Haiyi Zhu, Robert Kraut, and Aniket Kittur. 2012a. Effectiveness of Shared Leadership in Online Communities. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 407–416. DOI: <http://dx.doi.org/10.1145/2145204.2145269>
78. Haiyi Zhu, Robert Kraut, and Aniket Kittur. 2012b. Organizing Without Formal Organization: Group Identification, Goal Setting and Social Modeling in Directing Online Production. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 935–944. DOI: <http://dx.doi.org/10.1145/2145204.2145344>
79. Haiyi Zhu, Amy Zhang, Jiping He, Robert E. Kraut, and Aniket Kittur. 2013. Effects of Peer Feedback on Contribution: A Field Experiment in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2253–2262. DOI: <http://dx.doi.org/10.1145/2470654.2481311>
80. Robert Zinko, Gerald R. Ferris, Fred R. Blass, and Mary D. Laird. 2009. Toward a Theory of Reputation in Organizations. *Research in Personnel and Human Resources Management* 26, 1 (2009), 163–204. DOI: [http://dx.doi.org/10.1016/S0742-7301\(07\)26004-9](http://dx.doi.org/10.1016/S0742-7301(07)26004-9)