Herding the Cats: The Influence of Groups in Coordinating Peer Production

Aniket Kittur, Bryan Pendleton, Robert E. Kraut
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213
{nkittur, bpendlet, robert.kraut}@cs.cmu.edu

ABSTRACT

Peer production systems rely on users to self-select appropriate tasks and "scratch their personal itch". However, many such systems require significant maintenance work, which also implies the need for collective action, that is, individuals following goals set by the group and performing good citizenship behaviors. How can this paradox be resolved? Here we examine one potential answer: the influence of social identification with the larger group on contributors' behavior. We examine Wikipedia, a highly successful peer production system, and find a significant and growing influence of group structure, with a prevalent example being the WikiProject. Comparison of editors who join projects with those who do not and comparisons of the joiners' behavior before and after they join a project suggest their identification with the group plays an important role in directing them towards group goals and good citizenship behaviors. Upon joining, Wikipedians are more likely to work on project-related content, to shift their contributions towards coordination rather than production work, and to perform maintenance work such as reverting vandalism. These results suggest that group influence can play an important role in maintaining the health of online communities, even when such communities are putatively self-directed peer production systems.

Categories and Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Management, Measurement, Performance, Design, Human Factors, Theory.

Keywords

Wikipedia, groups, self-identification, coordination, peer production, organizational citizenship behavior.

1. INTRODUCTION

Peer production systems harness the collaborative efforts of many

individuals in order to create artifacts of lasting value [2]. The successes of peer production include open source software (OSS) systems such as Linux and Apache, news and discussion sites such as Digg and Slashdot, and collaborative knowledge systems such as Wikipedia.

In peer production systems users self-select tasks that match their expertise and interests. People take on work in peer production systems to "scratch a personal itch" [24]. This self-assignment of work leads to benefits in both performance, when contributors select tasks that closely match their abilities, and in motivation, as they select tasks that match their interests. Some have argued this efficient self-selection of tasks is fundamental to the success of peer production [2]. This process is a variant of coordination through markets, in which people use their personal knowledge of what they are good at and interested in along with their knowledge of the tasks that need to be done to efficiently match work with people

However, there are many essential tasks in peer production systems that must be completed for the group as a whole to be successful, independent of whether individual community members find them interesting or rewarding. OSS systems must test the robustness of system components, ensure interoperability with other components, identify and fix bugs and produce documentation whether or not performing these tasks soothes any contributor's personal itch. In Wikipedia these essential tasks include maintenance work, such as reverting vandalism and categorizing articles, assessment work (e.g., rating articles and producing peer reviews), providing sources for articles by adding citations and verifying external links and 'wikifying' articles by adding appropriate markup and internal links to other articles.

To be successful, peer production systems must ensure that contributors perform critical tasks that are central to the organizations' mission and goals. In the case of OSS development systems, the community must produce modules or applications that the user-base most desires and fix critical bugs. In the case of a collaborative knowledge system such as Wikipedia, which strives to be the world's most complete and authoritative encyclopedia, some topics must be covered in sufficient depth for the product to be considered encyclopedic by its users. Even though the Wikipedia community of editors seems eager and capable of providing rich content on topics related to popular culture, such as rock musicians, TV shows and science fiction[1],

---

[1] In an acknowledgement of what might be misplaced priorities in an encyclopedia with aspirations to be authoritative, Wikipedia instituted *The Great LucasCruft Purge* in March of 2005, the informal name given to the deletion of many Star Wars-related articles.

Wikipedia would not be a credible encyclopedia if it did not have content at least equally rich on science, history and literature. However, if individuals choose tasks based solely on their personal interests then important but less popular topics may be neglected [10]. For example, the English Wikipedia's article on the 17th century philosopher, mathematician, scientist, and writer René Descartes is shorter and less complete than its article on the fictional character James T. Kirk, the original captain of the Starship Enterprise on the Star Trek TV show.

These points highlight one problem with relying on purely market-like processes for peer production. The labor market analogy fails for peer production systems for at least two reasons: 1) the goals of the producers may not be naturally aligned with those of the consumers; 2) there are difficulties with providing sufficient incentives to change those goals.

People contribute to peer production systems for a variety of reasons which may not be the same as either the goals of the project or the goals of the consumers. For example, programmers contribute to open source software not just for altruistic reasons but also to sharpen their skills, build reputation, and increase job opportunities [18][25]. In Wikipedia, people similarly have a variety of motivations for contributing [4][21]. However, these may not match the goals of the project to build a comprehensive and accurate encyclopedia of human knowledge, or with the goals of consumers, who may be looking to find the answer to a particular question.

Also, unlike markets, peer production systems are often based on the efforts of volunteers. While this can provide an enormous pool of potential workers, directing them can be difficult in the absence of monetary incentives to undertake work that is not intrinsically interesting. Furthermore, providing such incentives may undermine users' motivations to contribute, as they would then become focused on external rewards [8]. Finally, in a variant of the well-known public goods problem, people may not contribute what they believe would benefit all users of a site, if this contribution to the public good does not serve their personal self-interests.

1.1 Managerial control

Managerial control is one solution to the coordination problems of allocating contributors to important tasks and topics. In government agencies, businesses and other conventional organizations, the institution of management is the standard coordination technique used to match people to the tasks that need to be accomplished. Managers assign tasks to employees and can reward or punish them, e.g., through a raise or loss of job, respectively.

Some degree of managerial control can exist in peer production systems, just as it does in many other organizations with heavy volunteer participation, such as the Red Cross or the American Cancer Association. In Linux, for example, a single individual (Linus Torvalds) had putative control over all changes to the system. Similarly, Wikipedia founder Jimmy Wales' role in the community has been described as "benevolent dictator for life" [14]. In less extreme cases of managerial control in peer production systems, the online Encyclopedia of Life uses a system of 'curators' who are responsible for maintaining the quality of the encyclopedia's vetted content; similarly, Wikipedia has 'administrators' with special responsibilities and powers, including the power to ban other editors.

However, the control exerted by the authorities in each of the cases just described is limited and tends to be permissive rather than directive. While the authorities shape key policy decisions, they don't make the low-level decisions about which volunteer should work on which task. This may be because the leaders may not have enough time or attention to devote to these personnel issue. Alternatively, the limits on managerial control in peer production systems may stem from the volunteer character of these communities. If control is too great, the volunteers may simply leave, with fewer consequences for their economic welfare and their personal identities than if they had quit a wage-paying job.

1.2 Group self-management in Wikipedia

Although many peer production communities do not exercise strong managerial control, they do exercise self control though informal social pressures, formalized rules of operation and bureaucratic social structures. Wikipedia, for example, as of September 2007 had 44 wiki pages labeled as "Wikipedia Official Policy" another 248 pages labeled as Wikipedia guidelines, and an additional 45 pending proposals for policies or guidelines [5]. While regular editors are responsible for acting according to the policies and guidelines and enforcing them through discussion and informal social pressure, Wikipedia also has a cadre of administrators who have special technical powers to enforce policy and to adjudicate in the case of conflict over them.

In addition to policies and guidelines, which provide a framework that individual Wikipedians could use to regulate their own behavior, Wikipedia strongly relies upon group self-management as a coordination device. Wikipedia has grown a collection of almost 2,000 subgroups known as WikiProjects to coordinate the work of individual editors. WikiProjects are collections of editors interested in improving the coverage and quality of articles in a particular domain. Topics for WikiProjects range from military history to the Beatles. Projects designate their scope by placing a tag similar to the one shown in Figure 1 on the articles that they encompass. Multiple projects can assume responsibility for a single article. For example, the US history, military history and the Pennsylvania projects have all incorporated the Battle of Gettysburg article into their scope.

WikiProjects can also be defined by task rather than content, such as copyediting[2] or improving the quality of article images. Each WikiProject has dedicated pages (which exist in a separate namespace from regular article content) on which project members can discuss issues, coordinate, set group goals, and request help. Editors can join a project simply by adding their name to the member list, though some projects move members who have not been active for a certain period of time to an inactive list.

There are often further divisions of group structure within WikiProjects. For example, the military history project, one of the largest WikiProjects, has separate departments for assessing articles, contests, logistics, outreach, review, and even a stress
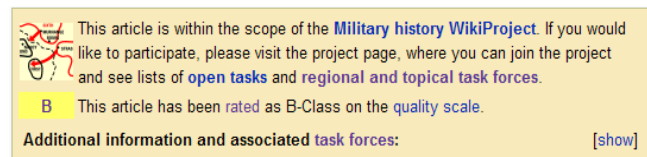
Figure 1. Example of a template for the Military History WikiProject. Adding the template to an article places it in the scope of the WikiProject, and the template can be used to display assessments of the quality of the article.
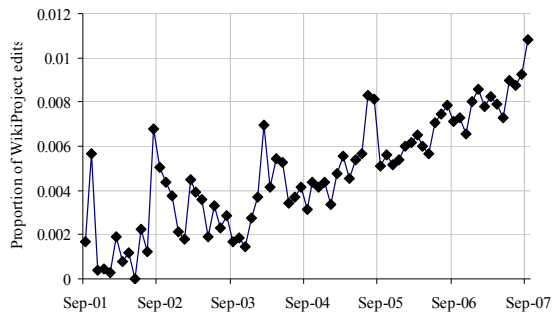
Figure 2: Growth in activity in WikiProjects

hotline for editors experiencing "wikistress" (stress caused by Wikipedia vandals, trolls, edit wars, and incivility[3]). Furthermore, many large projects subdivide themselves into task forces, smaller groups that focus on more specific topics (e.g., African military history or the Napoleonic era). Task forces may have their own member lists, unique templates and standards, and track articles in their scope separately from their parent WikiProject.

WikiProjects may also incorporate a hierarchical management structure. Some projects have coordinators and even a lead coordinator, and may have formal elections to fill these posts. Although coordinators do not receive additional technical capabilities or executive privileges, they are expected to play an important role in the administration and maintenance of the project. They are also influential in guiding the development of project-related policies and procedures[4].

As Figure 2 shows, WikiProject activity has been growing steadily in both absolute number and relative influence over the years. By October 2007 the proportion of edits to WikiProject pages increased to over 1% of all edits made in Wikipedia. This is remarkable given that this number does not include edits to any of the actual content articles that are the subject of the WikiProjects, but only reflects communication and coordination work relating to the WikiProject itself.

WikiProjects provide a mechanism to influence the decision by individual editors about their choice of tasks to perform. Techniques used to achieve this influence fall into at least two distinct groups: (1) directing editors' attention to high priority tasks and (2) setting group goals.

As an example of identifying high priority tasks, the military history project as of April 2008 identified over 66 thousand articles under its scope that needed some type of improvement. Project can also encode this priority information in its templates. Templates can include information on the importance of an article in the context of the project, as shown in Figure 1. High priority articles can be aggregated and surfaced, providing a way to identify important areas that need work.

In terms of goals, many projects sponsor both project-wide contests and target collaborations of the week. For example, the military history project has a regular, monthly contest in which project coordinators awards points to members for improving an

article in its domain. At the end of the month, a scoreboard displays contestants' rank, the number of articles they edited in a particular that month, and the total improvement points they earned.

One of the important tasks that projects encourage their members to perform is assessing article quality. Even though Wikipedia as a whole has a quality guide and a quality assessment editorial team to selecting articles to appear in a DVD version of the encyclopedia[5], in practice the assessment of particular articles is typically done by the subject matter experts in WikiProject, who grade articles according to the Wikipedia wide guidelines.

The same project template that lays claim to an article is also used to provide assessments of the quality of the article. For example, the template in Figure 1 indicates that this is a B-class quality article. Individual WikiProjects often track the quality of the articles in their scope as they improve and prominently showcase those of outstanding quality. For example, the homepage of Military History project shows that the project encompasses 350 featured articles and 168 A-class articles.[6]

WikiProjects also provide mechanisms for members to self-identify and to acknowledge each other. Members can place project banners on their user pages, identifying their online personas with the group. Members can also recognize other members by nominating them for or directly giving them awards or "barnstars" for important contributions.

In summary, WikiProjects incorporate many characteristics of traditional groups. They provide mechanisms for members to coordinate, set goals, self-organize, self-identify, and reward each other.

1.3 Group identification

Joining a group often causes people to feel that their identity is tied to that of a group. As a result, their view of their self-interest may expand to include the group and perceive that helping the group serves their own self-interest in a way that identical behavior would not if they had not identified with the group [13]. By identifying with the group they should be willing to take on tasks that the group considers important or that they think is needed for the group's success. Supporting this idea, there is a large body of literature in offline groups on how identification with a group or organization can significantly change people's behavior and lead to improved outcomes for the group. These outcomes include increased effort and participation, better task performance, reduced turnover, and greater intrinsic motivation and job satisfaction [1][17][26][16]. In addition to direct task-related behaviors, identification increases information sharing, coordinated action, and cooperation [6][12]. Identification leads people to assist the group in other ways, such as helping other group members in times of stress, providing social support to them, defending the organization, and engaging in good citizenship behaviors [11][28]. When they identify with the group, they are also more likely to be influenced by organizational norms and values, causing them to make decisions favorable to the organization, hold positive evaluations of the organization, and be

---

[3] http://meta.wikimedia.org/wiki/Wikistress

[4]
http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Military_history/Coordinators#Responsibilities

[5]
http://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team#Wikipedia_1.0_projects

[6]
http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Military_history

more influenced by the organization [6][13][26]. Thus explicitly joining a group may have a number of beneficial effects on both the individual and the group.

1.4   Research hypotheses
The discussion to this point proposes that WikiProjects are important mechanism through which Wikipedia influences editors to work on tasks that are important to the encyclopedia as a whole and especially to the portion of Wikipedia that the projects encompass. Based on prior research in social psychology and organizational behavior, there is substantial reason to think that when editors join a WikiProject they will become more influenced by its specific coordination recommendations (e.g., its lists of important tasks), since most of these tasks require editors to edit specified articles encompassed by the project. In addition, when editors join a WikiProject they should be more willing to work hard in the project's behalf, to be willing to cooperate with other members, and exhibit the diffuse helpfulness known as organizational citizenship behavior. This should be reflected in their devoting more of their work in Wikipedia to articles that the project encompasses, to direct more of their effort to coordination with other project members, and perform organizationally useful chores, such as reverting vandalistic edits, that reflect good project citizenship.

Increasing activity. If joining a WikiProject increases overall motivation to contribute to the cause, this increased motivation should be reflected in the editors' overall *activity*, including their direct production work, in editing Wikipedia articles, and their coordination work, in planning, evaluating and discussing articles.

H1. Joining a WikiProject should lead editors to increase their overall activity in Wikipedia.

Directing behavior. If joining a WikiProject causes editors to enlarge their view of their self-interest to also include the interests of the group they joined, this shift in self-identification should be reflected in where these editors work and the type of work they do. Rather than merely increasing their overall activity, joining a group may direct attention towards tasks that are part of the group's goals and away from those that are not. As a result, joining a WikiProject may lead editors to devote more of their work to in-project tasks, that is to tasks relevant to the articles that are encompassed by the WikiProjects they joined.

H2. Joining a WikiProject should lead editors to shift their work towards articles that are considered within the project's scope and away from articles that are not.

Furthermore, joining may influence the *types* of tasks editors work on. Specifically, upon joining a project, editors may shift away from the direct production work of editing articles and towards work involving cooperation and coordination with others. In Wikipedia much of the coordination about articles happens on the articles' corresponding talk or discussion page; thus one measure of influence is the change in the distribution of editing behavior to article vs. discussion pages.

H3. Joining a WikiProject should lead editors to shift work towards discussions with other editors on the talk pages associated with the project and its articles and away from direct production work, editing the project's articles themselves.

WikiProjects and Wikipedia need good organizational citizenship to be successful. Organizational citizenship behavior (OCB) has been defined by Organ [22] as "individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the effective
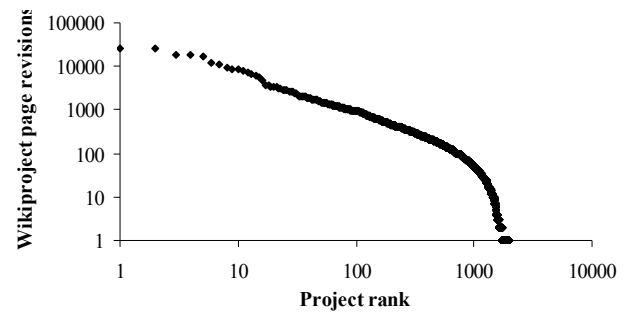
Figure 3. WikiProject activity ranked by total number of revisions to pages and subpages.

functioning of the organization". Such behavior is vital to the continued functioning of many online communities, which are often maintained and supported by volunteer efforts. Thus although related to cooperative activity, here we include OCB separately in a first attempt to explore the factors which can influence it. Joining a project may cause editors to increase working on "citizenship" tasks that benefit the project as a whole with little direct personal benefit to the individual who completes them. For example, reverting vandalism is an important task in Wikipedia, which maintains the integrity of the encyclopedia, but takes time away from direct production or even coordination work.

H4. Joining a WikiProject should lead editors to increase their good citizenship behaviors such as fighting vandalism

2.   Methods
2.1   Sample
In the following analyses we use a download provided by the MediaWiki foundation which included all edits from Wikipedia's inception up to October 2007 (approximately 144 million edits). To handle this data volume, we used the Yahoo! M45 computing cluster running Hadoop to prepare data for analysis. As of October 2007 there were approximately 2000 WikiProjects in Wikipedia. To measure activity differences between WikiProjects we summed the number of revisions to project pages and subpages (i.e., pages where users coordinated about the project). The distribution of project activity was highly skewed, following an approximate power-law distribution (see Figure 3). The tail of the distribution drops sharply, perhaps due to the start-up commitment of a WikiProject: if a user puts in enough effort to start a project, it is likely that they will make at least some edits to it.

We sampled WikiProjects, including 22 of the top 40 most active by total edits, and 51 from the US States project (the 50 states plus D.C.), resulting in a final set of 73 different WikiProjects. For each project, we identified the project templates used to label an encyclopedia article as a member of that project, and any project management pages that contained the list of project members. We identified all editors who were ever listed on the membership lists for the 73 sampled projects. We defined their date of joining the project as the date in which they first edited any of the pages listing project members, presumably adding themselves to the list. Although editing the membership list pages might not always reflect editors' true join date, editors rarely edited project pages unless participating in the project. To identify the list of pages associated with a project, we searched for pages that had ever contained the project's template in their content. We expanded this set of pages to include both the "article" and the "article talk"

pages, as much of the non-production work occurs against "talk" pages.

To complete the sampling, we then randomly sampled up to 125 editors who had explicitly joined each project. For each project member, we selected a matched control, non-member editor who never explicitly joined the project, but who made at least one edit to a page within the scope of the project during the month that their matched member joined. For each editor in the sample we calculated monthly editing activity for both pages within the scope of the project (in-project pages) and those not within the scope (out-project pages). These pages included both article and talk pages. Overall, this resulted in a sampling of 7973 distinct editors including 133,861 monthly observations.

2.2 Independent Variables

Eventual Member. Whether an editor eventually becomes a member of the project (Member = 1) or not (Member = 0). Note that all editors sampled made at least one edit to a page considered in the scope of the project.

Joined Project. For editors that eventually become members of the project, this indicates for a given month whether they had joined (Joined = 1) or not (Joined = 0).

In-project work. Indicates whether edits were made to pages within the scope of the project (in-project = 1) or outside the scope of the project (in-project = 0).

Coordination Work (Work type). Whether edits were made to article pages directly (work type = 0) or o the discussion pages (work type = 1).

Age. Months since the editor made their first edit in Wikipedia. We included age as a control variable in some analyses.

2.3 Dependent Variables

Edits. The number of edits made to a page in a given month.

Vandal reverts. Edits made explicitly to combat vandalism. To quantify such behavior we identified revisions which had been annotated with common vandalism-fighting comments, such as "rvv" (revert vandalism). Although this approach is not perfect, as there may be vandalism reversions that are not marked, or marked reversions that are not actually fighting vandalism, prior research has shown that it is a reasonable proxy [23][15].

2.4 Analysis strategy

We examine the influence of project membership on the amount and type of work that editors do through two complementary statistical approaches. In the first, to assess how membership

compare the editing trajectories of people who eventually become project members with people who also edited articles in the project but who never become members. On average, editors joined a project about 9 months after they first start editing in Wikipedia (mean=9.4 months, median=7 months, std=9.37). If our hypotheses are correct, editors who eventually become members should increase their contributions to Wikipedia, direct their contributions to the projects they joined and disproportionately increase their willingness to do coordination and maintenance work later in their histories in Wikipedia, when as they are more likely to be project members. In contrast, non-member control editors should not have as steep increases in these variables, because they never become project members.

The second approach compares the editing behavior of eventual project members before and after they join a project. Evidence consistent with our hypotheses would be that upon joining a project these editors will increase their overall contributions to Wikipedia, direct their contributions to the projects they joined and disproportionately increase their willingness to do coordination and maintenance work.

Each of these techniques is imperfect. By contrasting the editing trajectory of eventual members with that of non-members, the first analysis does not control for the enthusiasm or other personal characteristics that differentiate the joiners from the non-joiners. The second analysis overcomes this limitation by comparing editing behavior by the same people before and after they joined a project. However, this analysis may confound the effects of joining with unrelated history or environmental effects (e.g., a campaign to increase contribution). Because the first analysis compares different people at the same time, it overcomes this limitation.

Both analyses use negative binomial regression as implemented in Stata (xtnbreg) as the dependent variables (number of edits or vandal reverts) are over-dispersed. Since each editor had multiple months of editing, we used hierarchical linear modeling with the editor as a random effect to deal with non-independence in the data [7].

3. Results

We report the results for both analyses together for each hypothesis because results from the two analyses tell a consistent story. Table 3 shows the results from a negative binomial regression comparing editors who eventually become project members with matched editors who did not. Table 4 compares the eventual project members before and after they joined the project. In both tables, effects are reported as Incidence Rate Ratios (IRR),

Factor	Organizational identification factors	Wikipedia measures
Increasing activity	Effort; participation; motivation; task performance	Number of edits
Directing - in-project vs. out-project	Control by the organization; organizationally beneficial decision making	Whether edits are to in- vs. out-project pages
Directing - article vs. discussion	Cooperation; information sharing; coordinated action	Whether edits are to direct (article) work or coordination (discussion) work
Directing - good citizenship behavior	Organizational citizenship behaviors; Defense of the organization	Number of edits reverting vandalism

Table 1. General factors examined in the study, with corresponding factors from the organizational identification literature and measures as operationalized in Wikipedia.

influences the amount and type of materials editors contribute we   the ratio by which increasing an independent variable by a unit

		Edits		Vandal reverts	
		Pre-join	Post-join	Pre-join	Post-join
Article	Out-project	256.5	234	1.63	1.73
	In-project	13.9	24.2	0.057	0.141
Talk	Out-project	95.4	106.3		
	In-project	2.8	8		

Table 2. Mean number of edits and vandalism reverts made by project members pre- and post-joining, to pages either in or out of the project, and to article or talk pages.

Figure 4. Percentage change of edits pre- vs. post-joining a project to article, talk, and vandalism reverts, separated by whether the page was outside or inside the scope of the project.

changes the dependent variable (e.g., the IRR for coordination work of .231 in Table 3 shows there are only 23% talk page edits compared to article edits).

3.1 H1: Increasing activity

We hypothesized that joining a project should lead editors to contribute more to Wikipedia overall. For the member vs. control analysis, this hypothesis was only weakly supported: over time members increased in their editing only slightly relative to non-members (see the Member X Age interaction in Table 3). The pre-post member analysis also shows significant but very weak effects: joining a project leads to only a 1.6% increase in total edits (see the Joined project effect in Table 4). Overall, despite the statistical significance of the results, the influence of project membership on total editing activity appears practically negligible.
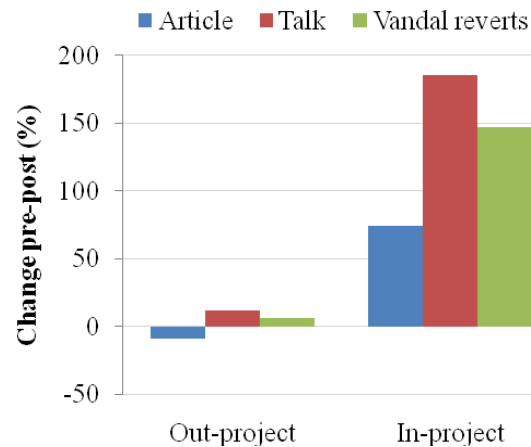
3.2 H2: Directing attention – In- vs. out-project

We predicted that group membership would shift editors' work towards the goals of the project, and specifically towards articles considered to be within the scope of the project. For the members vs. controls analysis, there was a significant interaction such that members were much more likely to edit in-project pages (Member X In project), with the effect becoming even more pronounced over time (Member X In project X Age). Similarly, the pre-post member analysis shows that after members join the project they are more likely to edit in-project pages (Joined project X In project). These results are evident in the raw means (Table 2) and summarized in Figure 4, which shows that joining a WikiProject is associated with large percentage increases in editing in-project pages but not out-project ones. Overall, this is strong evidence that joining a project shifts an editor's activity towards pages that are important to the group.

3.3 H3: Directing attention – Article vs. discussion

In addition to shifting where editors perform their work, H3 predicted a shift in the type of work they perform upon joining a project. Specifically, we hypothesized that they would engage in more coordination work, reflected by an increase in discussion page activity relative to article page production work.

The member vs. control analysis shows support for this hypothesis. Members are much more likely to make discussion edits than non-members (Member Coordination work). However, there is little evidence that this changes over time: there is a significant but negligible change in the interaction with age (Member X Work type X Age). In contrast, the pre-post member analysis provides stronger support for the hypothesis, with members changing their behavior towards discussion activity upon joining the project (Joined project X Work type).

We also examined the hypothesis resulting from the combination of H2 and H3; that is, that members would shift in their editing activity towards more coordinative work, and that this shift should be more pronounced for articles within the scope of the project. Both analyses show significant positive support for this prediction (Member X In project X Work type; Joined X In project X Work type).

To summarize, joining a project appears to change the behavior of an editor from simply editing articles to other, often more community-oriented behaviors such as coordination; and, this shift is especially powerful for work that is in the domain of the joined project.

3.4 H4: Good citizenship behaviors

We predicted that joining a group would increase good citizenship behaviors, such as fighting vandalism. While members did not appear to increase in overall vandal fighting over time more than non-members (Member X In project in Table 3), they did appear to shift their vandal fighting efforts towards articles in the project (Member X In project), an effect that grew stronger with age (Member X In project X Age). The pre-post member analysis is largely consistent with these results, showing a shift in vandal reverting towards in-project pages upon joining (Joined project X In project in Table 4); however, we see a substantial main effect of joining the project on overall vandalism reverting as well (Joined project). Overall, these results support the hypothesis that joining a group increases good citizenship and maintenance behaviors, especially for work that is within the group's domain.

4. DISCUSSION

Although peer production systems like Wikipedia are typically thought to be driven by the interests of individual contributors, these organizations have developed mechanisms to encourage contribution and coordinate the work of individuals. In the case of Wikipedia, we found that WikiProjects play a major coordinative role. We identified a number of significant changes in editor behavior upon joining a project group. Contrary to our initial hypothesis, these changes did not include substantial increases in

Hyp	Variables	Edits			Vandalism Reverts		
		IRR	SE	P	IRR	SE	P
	Age	1.001	.0004		1.019	.001	***
	Member	.883	.0081	***	1.172	.024	***
	In project work	.073	.0012	***	.033	.002	***
2, 4	Eventual Member X In project	3.193	.0605	***	6.002	.443	***
1, 4	Member X Age	1.002	.0005	***	.999	.001	
	In project X Age	.967	.0012	***	.974	.005	***
2, 4	Member X In project X Age	1.026	.0013	***	1.020	.005	***
	Coordination work	.231	.0025	***			
	Member X Coordination work	1.397	.0187	***			
3	In project X Coordination work	.669	.0205	***			
	Age X Coordination work	1.007	.0006	***			
	Member X In project X Coordination	1.804	.0565	***			
2&3	Member X Coordination X Age	.996	.0007	***			
3	In project X Coordination X Age	1.003	.0007	***			

Table 3. Regressions comparing editors who became project members versus matched editors who did not.
Note: *** $p<.001$

the amount of editing taking place. However, we did find large shifts in both where work was taking place and what type of work was performed, with more work directed at pages relevant to the project and an increase in coordination and discussion work, especially for pages within the project scope. This pattern of results also held true for good citizenship behaviors such as reverting vandalism, although we did find some evidence for an overall increase in vandal reverting upon explicitly joining a project as well as the shifts in the place and character of work found above.

These results are largely consistent with the literature in offline groups, which have shown effects of group identification on both activity-centric metrics such as effort, participation, motivation, and task performance as well as group goal-centric metrics such as cooperation, OCB, defense of the organization, organizationally beneficial decision making, and other behaviors which demonstrate that individuals have internalized the goals of the group in addition to their own (see [3][27] for reviews). Surprisingly, we did not see a substantial increase in overall editing upon group joining or over time for members, although increases in overall activity have been consistently found in the literature. One reason for this may be that members are already established within Wikipedia by the time they join a project – on average, members join 9 months after their first edit in Wikipedia – and thus may have already maxed their available time to contribute, and instead shift their efforts to align them more closely to the group's goals.

These results are important for a number of reasons. First, few studies have examined the impact of group identification in online groups (though see [20]), and those that have, have focused on changes in total activity. Here we show a qualitative shift in the kind of activities engaged in upon joining a group, as well as a shift in focus towards in-group tasks vs. out-group tasks. We believe both of these are novel contributions and merit further investigation. Furthermore, our finding of no substantial increase in total activity upon group joining is inconsistent with the findings of [20] and others, and suggests an interesting counter case that could be useful in zeroing in on the mechanisms of group influence in online communities.

Another reason these results are important is because they demonstrate the influence of group goals in a system which is generally thought to be driven by individual interests. As noted by Benkler [2], the success of peer production systems is at least in part due to the efficient matching of editor interests and expertise with tasks that need to be accomplished. Some previous research in Wikipedia has aimed at improving the matching process, including SuggestBot [8] which recommends articles to editors, as well as a task surfacing mechanism for making task matching easier [18]. However, many tasks in the system may remain undesirable or have low priority yet still be essential to the continued health of the community. Here we show that even within an individualistic peer production system, group goals can influence behavior through self-identification with a sub-group.

One particularly intriguing result is that self-identification can

Hyp	Variables	Edits			Vandalism Reverts		
		IRR	SE	P	IRR	SE	P
1, 4	Joined project	1.016	.0068	*	1.484	.0186	***
	In project work	.179	.0017	***	.141	.0033	***
2, 4	Joined project X In project	1.300	.0157	***	1.433	.0404	***
	Coordination work	.312	.0026	***			
3	Joined project X Coordination	1.152	.0123	***			
	In project X Coordination	1.091	.0183	***			
2&3	Joined X In project X Coordination	1.207	.0254	***			

Table 4. Regressions comparing before and after a member joined a project on predicted variables.
Note: *** $p<.001$, ** $p<.01$, * $p<.05$

have a positive effect on tasks related to good citizenship behavior. Such tasks are often critical to group survival, but because individuals rarely have incentives to accomplish them they can go undone. Although we only looked at one kind of citizenship behavior in this study (reverting vandalism), our results highlight the potential importance of identification with a group and suggest that it is a rich topic for further research.

4.1 Limitations

The correlational nature of our data involves assumptions that can be challenged. Our analyses take into account differences between members and non-members for one analysis, and take into account individual differences among editors for another. However, the use of correlational analysis rests upon an assumption that all the relevant variables have been measured. It is possible that some unmeasured variable that co-varies with joining a project may account for what appears to be a direct relationship between group joining and the influence of group goals on individual behavior.

4.2 Design recommendations

Our results have significant implications for the design of online communities.

Explicit subgroups. Joining subgroups (WikiProjects) had a significant influence on behavior above and beyond joining Wikipedia itself. This suggests that having multiple explicit subgroups in a community can provide an environment for self-identification.

Mechanisms for self-identification. In addition to having groups, editors must have some method for self-identifying with them. In the WikiProjects studied here editors could become members by placing their names on a member roll. Other ways of self-identifying in Wikipedia include placing a banner of the group on one's editor page, or placing oneself in a category of members. Since non-members who edited similar pages as members did not show the effects that members did, having an explicit self-identification method appears to be a necessary prerequisite for spreading group influence.

Defining group scope. We found that members were more likely to edit and fight vandalism on pages that were within the scope of the group than pages outside the scope of the group. Thus providing a mechanism to define the scope of a group is an important precursor to harnessing this effect. Within Wikipedia one way scope is defined is by editors placing a template on a page that results in the article being included in a group-specific category. In other communities scope may be defined by more central rather than distributed methods (e.g., group leaders deciding on what should be included) or by algorithmic means (e.g., through measures of content similarity or other recommendation methods).

Defining group goals and norms. For a member to be influenced by group goals and norms they must be in a form that can be transmitted and learned by the member. In Wikipedia goals are transmitted in a variety of ways, some of which we discussed earlier, such as "collaboration of the week" or "articles needing attention". Group norms may be explicitly transmitted through project-specific guideline pages, or may be implicitly transmitted through newcomers observing or conversing with old timers. More generally, providing mechanisms for transmission of group goals and norms to new members is necessary for the behavior of those newcomers to be influenced by the group.

4.3 Future work

There may be many other ways of quantifying individual behavior than those used here. For example, we use vandalism fighting as a proxy for good citizenship behaviors. However, there may be better proxies available which may require additional work to mine, such as editor assessment behavior, mediation of conflict, or mentoring of newcomers. More generally, it may be possible to determine the degree to which a behavior is needed by the community by the efforts to recruit editors to engage in those behaviors. Understanding the factors that increase editor engagement in such tasks could be of significant theoretical and practical value.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Bartel, C. A. 2001. Social comparisons in boundary-spanning work: Effects of community outreach on members' organizational identity and identification. *Administrative Science Quarterly*, 46: 379-413.

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

[3] Blake E. Ashforth, Spencer H. Harrison and Kevin G. Corley (2008) Identification in Organizations: An Examination of Four Fundamental Questions, *Journal of Management* 2008; 34; 325

[4] Bryant, S., Forte, A., Bruckman, A. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Enciclopedia. *GROUP 2005*, ACM Press (2005).

[5] Butler, B., Joyce, E., and J. Pike, "Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia," *CHI 2008*, ACM Press (2008), pp. 1101-1110.

[6] Cheney, G. 1983a. On the various and changing meanings of organizational membership: A field study of organizational identification. *Communication Monographs,* 50: 342-362.

[7] Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. *Applied multiple regression/correlation analysis for the behavioral sciences.* Lawrence Erlbaum Associates, Mahwah, New Jersey (2003).

[8] Cosley, D., Frankowski, D., Terveen, L., & Riedl, J. SuggestBot: using intelligent task routing to help people find work in wikipedia. *IUI 2007*. New York: ACM Press.

[9] Deci, E.L., Koestner, R. & Ryan, R.M. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125 (6). 627-668

[10] Denning, P., Horning, J., Parnas, D., and Weinstein, L. Wikipedia risks. *C ACM* 48 (2005), 152.

[11] Dukerich, J. M., Golden, B. R., & Shortell, S. M. 2002. Beauty is in the eye of the beholder: The impact of organizational identification, identity, and image on the cooperative behaviors of physicians. *Administrative Science Quarterly*, 47: 507-533.

[12] Grice, T. A., Gallois, C., Jones, E., Paulsen, N., & Callan, V. J. 2006. "We do it, but they don't": Multiple categorizations

and work team communication. *Journal of Applied Communication Research*, 34: 331-348.

[13] Haslam, S. A., & Ellemers, N. 2005. Social identity in industrial and organizational psychology: Concepts, controversies and contributions. *International Review of Industrial and Organizational Psychology*, 20: 39-118.

[14] Reagle, J. M. Do as I do: Authorial leadership in Wikipedia. *WikiSym 2007*, ACM Press (2007), 143-156.

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

[16] Kogut, B., & Zander, U. 1996. What firms do? Coordination, identity, and learning. *Organization Science*, 7: 502-518.

[17] Kramer, R. M. 2006. Social capital and cooperative behavior in the workplace: A social identity perspective. *Advances in Group Processes*, 23: 1-30.

[18] Krieger, M., Stark, E. M., Klemmer, S. R. Coordinating tasks on the commons: designing for personal goals, expertise and serendipity. *CHI 2009*, ACM Press (2009).

[19] Lakhani, K.R., and Hippel, E. Von, "How open source software works:"free" user-to-user assistance," *Research policy*, vol. 32, 2003, pp. 923-943.

[20] Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frankowski, D., Terveen, L., and Rashid, A.M. "Using social psychology to motivate contributions to online communities," *Journal of Computer-Mediated Communication*, vol. 10, 2005.

[21] Nov, O. 2007. "What motivates Wikipedians?," *CACM,* vol. 50, 11 (2007), 60-64.

[22] Organ, D.W. *Organizational citizenship behavior: The good soldier syndrome*, Free Press, 1988.

[23] Priedhorsky, R., Chen, J., Lam, S., Panciera, K., Terveen, L., & Riedl, J. Creating, destroying, and restoring value in Wikipedia. *GROUP 2007*, ACM Press (2007), 259-268.

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

[25] Roberts, J., Hann, I.-H. & Slaughter, S. Understanding the Motivations, Participation and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects. *Management Science*, 52 (7), 984 - 999.

[26] Simon, H. A. 1976. *Administrative behavior: A study of decision-making processes in administrative organization*, 3rd ed. New York: Free Press.

[27] van Dick, R. 2004. My job is my castle: Identification in organizational contexts. *International Review of Industrial and Organizational Psychology*, 19: 171-203.

[28] van Dick, R., Grojean, M. W., Christ, O., & Wieseke, J. 2006. Identity and the extra mile: Relationships between organizational identification and organizational citizenship behaviour. *British Journal of Management,* 17: 283-301.