Making Peripheral Participation Legitimate: Reader Engagement Experiments in Wikipedia

Aaron Halfaker
GroupLens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
halfaker@cs.umn.edu

Oliver Keyes
Wikimedia Foundation
149 New Montgomery St.
3rd Floor
San Francisco, CA 94105
okeyes@wikimedia.org

Dario Taraborelli
Wikimedia Foundation
149 New Montgomery St.
3rd Floor
San Francisco, CA 94105
dario@wikimedia.org

ABSTRACT

Open collaboration communities thrive when participation is plentiful. Recent research has shown that the English Wikipedia community has constructed a vast and accurate information resource primarily through the monumental effort of a relatively small number of active, volunteer editors. Beyond Wikipedia's active editor community is a substantially larger pool of *potential* participants: readers. In this paper we describe a set of field experiments using the *Article Feedback Tool*, a system designed to elicit lightweight contributions from Wikipedia's readers. Through the lens of social learning theory and comparisons to related work in open bug tracking software, we evaluate the costs and benefits of the expanded participation model and show both qualitatively and quantitatively that peripheral contributors add value to an open collaboration community as long as the cost of identifying low quality contributions remains low.

Author Keywords

Wikipedia; Social learning; Legitimate peripheral participation; quatitative; experiment; participation; open production

ACM Classification Keywords

H.5.3. Information Interfaces and Presentation (e.g. HCI): Group and Organization Interfaces

INTRODUCTION

Open collaboration systems like Wikipedia require a stable pool of volunteer contributors to remain productive. Without volunteers to occupy necessary roles, these systems would cease to function. The success of an open collaboration project appears to be highly correlated with the number of participants it maintains[7]. In order to maintain the pool of contributors, newcomers must be continually socialized into the organization[22, 18].

CSCW'13, February 23–27, 2013, San Antonio, Texas, USA.

Participation in peer production

Participation in online communities tends to manifest as a long-tail distribution[27]; a tiny, active minority produces most of the content while the majority of community members produces very little individually. Beyond the contributing population is an often overlooked population of users who do not contribute, commonly referred to as "lurkers". Despite the fact that lurkers do not contribute, their numbers tend to dwarf the rest of the community by orders of magnitude. The primary reason for lurking in online communities has been identified as the lack of a perceived "information benefit" in increasing their effort to contribute[17, 5, 12], yet in some communities, a substantial proportion of lurkers are simply not aware that they can participate[1, 16].

Recent work has shown that the English Wikipedia, an online encyclopedia often held as a prototypical example of open collaboration, has a particularly steep long-tail of participation[27]. Through an analysis of Wikipedia's historical edit logs, Priedhorsky et al. estimated that the most active 0.1% of contributors produce nearly half of the encyclopedia's value[19]. According to the Wikimedia Foundation's official statistics[1] report for March of 2012, the encyclopedia was edited by 113,304 editors and a comScore[2] report for the same month shows 1.47 billion unique visitors. These numbers suggest that the English Wikipedia's consumers outnumber producers by 10,000 to 1. Given what the current contributors have been able to achieve – a vast and highly accurate encyclopedia – the consumers of Wikipedia would represent a massive potential workforce if even a small percentage of them could be coerced to contribute productively.

Research in collaborative computing has explored effective mechanisms for boosting participation in online communities. For example, Beenen et al. used insights from the collective effort model (CEM) to increase participation in a movie recommender via email requests[2]. Rashid et al. extended their study by replacing the email requests with visual queues within the user interface[20]. Within Wikipedia, Cosley et al. showed the effectiveness of a task routing system designed to decrease the cost of finding *where* to contribute in order to boost contributions of current Wikipedians[6]. Wash and Lampe showed that directly asking users for comments on

[1] `http://stats.wikimedia.org/EN/TablesWikipediaEN.htm`
[2] `http://comscore.com`

news articles could temporarily increase participation without reducing the quality of comments[25]. However, none of these studies examined the process of transitioning from non-participant to participant.

Supporting peripheral participation

Social learning theory gives us a framework for understanding how newcomers to a community transition from non-participation to participation – i.e. from consumer to producer. In their highly cited work on how newcomers approach "communities of practice", Lave & Wenger describe the process by which initiates begin participation on the periphery of a community by performing simple and low risk, yet productive tasks that they refer to as legitimate peripheral participation (LPP)[15]. As newcomers gain experience, they become more familiar with the tasks, vocabulary and norms of the community, and by doing so, they are able to confidently expand their level of participation. Recent work by Preece & Shneiderman applies this framework to online communities by defining a set of roles that users adopt as they transition from the periphery of the community (reader) to its center (leader). Specifically, they recommend that online communities should be designed to "support legitimate peripheral participation so that readers can gradually edge into contributing"[18].

The open source software community has engineered a model for supporting a type peripheral participation in the form of open issue tracking software. Through such systems, users who do not have the knowledge, time or interest in directly contributing code to a project may submit bug reports and feature requests to the developers. Through the use of such systems, participation is expanded from the a relatively small number of producers (developers) to a much more numerous group of consumers. This contribution medium allows these consumers to learn about the norms and practices of a community by interacting at the periphery. In an analysis of the growth of the developer community around the Freenet project, an open source file sharing platform, Von Krogh et al. describes a "joining script" where some newcomers post bug reports and feature requests before attempting substantial participation. Newcomers who follow this sequence of activities are more likely to be granted privileged access to the source code repository[24].

Such expanded participation models allow prospective contributors to make simple and low risk, yet productive contributions, and by doing so, the workload for maintaining the community product is distributed more evenly among its stakeholders. When such participation is high in quality, it can reduce the workload of the small percentage of prolific contributors by spreading a time-consuming production activity (e.g. bug detection and reporting) over more individuals. This reduction in workload for the primary contributors is particularly valuable for volunteer communities where effort is the primary currency of progress.

When viewed this way, Wikipedia is not keeping pace with developments in collaborative computing that leverage peripheral participation. Although Bryant et al. uncovered, through a series of interviews with highly active Wikipedia editors, that social learning theory matches the way that editors view their integration into the Wikipedia community[4], the most peripheral type of activity they describe is making edits to fix mistakes. Antin and Cheshire argue that reading the encyclopedia should be viewed as a form a legitimate participation by showing evidence that experienced readers of Wikipedia know more details about the editing community[1], but they do not explain how reading is "productive", a condition of legitimate peripheral participation as described by Lave & Wenger[15]. MediaWiki, the software that runs Wikipedia, affords no half-step from reader to editor and recent work suggests newcomers are finding it increasingly difficult to transition into productive editors. In an analysis of the predictors of newcomer retention, Halfaker et al. showed that desirable new editors are likely to have their first few contributions rejected and that this rejection is a strong predictor that newcomers will give up on editing entirely[11, 9]. Given that lurkers in other communities have reported fear of strong, negative reactions[12] as their reason for not contributing, it seems likely that many Wikipedia readers don't contribute due to a reasonable fear of rejection.

However, opening up peripheral participation to a larger pool of users doesn't always come for free. The net value of expanding participation often depends on the associated cost of filtering and moderating a larger volume of contributions. A recent analyses of bug reports by casual users in large-scale open source software projects suggest that the mismatch between what users report and what developers find useful can undermine the value of broader participation [3]. A related study reports that core developers in software projects find higher value in smaller groups of highly committed bug reporters than in a larger group of unengaged contributors [13]. Given these concerns, it is essential to consider both a potential increase in participation and the moderation concerns when vetting extensions to participation.

Article feedback

Motivated by decreasing levels of newcomer retention in the English Wikipedia's contributor community[26, 21], the we worked with the Wikimedia Foundation to develop a half-step between reading and editing in the form of an extension to the Wikipedia's software called the *Article Feedback Tool* (AFT). This tool allows readers to submit feedback about encyclopedia articles to editors via the standard web interface. Like bug tracking software, AFT allows the community's consumers to communicate their concerns to the community's producers in a simple, low risk way. Submitting feedback to an article's editors can also be productive; like bug reports, feedback can be used by the producers as a mechanism for identifying problems and missed opportunities in articles. In this way, we view AFT as an extension of the contribution model of Wikipedia that supports a new mode of legitimate peripheral participation.

In this paper we extend the online participation literature and examine an aspect of LPP in Wikipedia through a set of field experiments (performed live on the English Wikipedia) designed to test AFT's effectiveness in eliciting participation from readers and encouraging the conversion from reader to

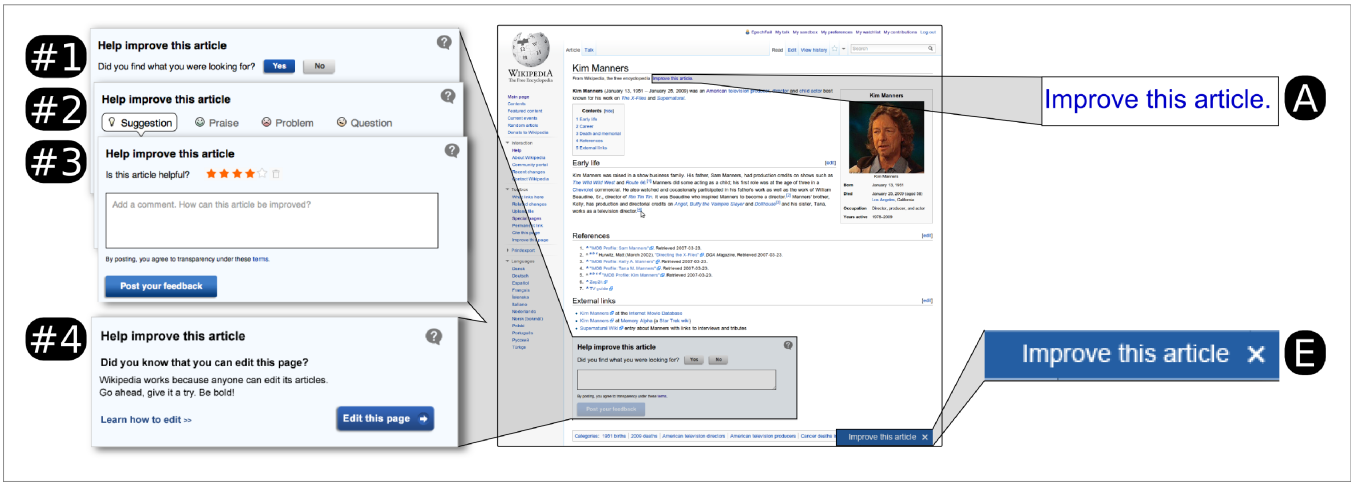


Figure 1: **The Article Feedback Tool's interface components.** The components of the AFT interface are called out from Wikipedia's article viewing interface. An article on Kim Manners, one of the randomly sampled articles, is loaded. **#1-3** represent different versions of the article feedback forms. **#4** represents the edit invitation form, a request for the reader to try editing the page. **A** and **E** represent links inserted into the page to direct readers to the feedback form.

editor. We also analyze the utility and productivity of the increased participation in the context of Wikipedia's quality control mechanisms.

THE ARTICLE FEEDBACK TOOL

The *Article Feedback Tool*[3] is an extension of Wikipedia's user interface that affords readers of an article the ability to submit feedback about that article to the editors. When a reader views an article, a small form appears after the content asking the reader to "Help improve this article". Figure 1 displays the different versions of this form that were tested as part of the present study. Each form consists of two components: a question prompt, which we refer to as an "elicitation", followed by a free-form text box. The feedback forms differ only by their elicitations:

- Form 1 asks the reader if she found what she was looking for: yes or no.
- Form 2 allows the reader to pick from 4 types of feedback she might give in the text box: suggestion, praise, problem or question.
- Form 3 asks the reader to rate the article on a 1-5 scale.

To evaluate the feedback form's effect on the conversion from reader to editor, we introduced a fourth type of form for comparison. Form 4 doesn't accept feedback; instead, the form invites the reader to make a contribution by editing the article using a UI of exactly the same size and location as the feedback forms.

Since many Wikipedia articles are very long, the feedback forms are often hidden beneath pages of scrolling. To allow for more straightforward access to the form and encourage more readers to leave feedback, we designed two prominent links (labelled A and E in figure 1) that act as shortcuts to the form. When a reader clicks on a prominent link, the form is loaded as an overlay on the page – allowing a reader to access the form without having to scroll. Although Link A appears at the top of the article, it is relatively hidden by the boilerplate statement "From Wikipedia, the free encyclopedia". Link E is intended to be much more likely to catch a reader's attention by (1) taking up more space, (2) making clear that it is a button and (3) remaining visible to the reader by staying fixed to the bottom of the window as she scrolls through the article.

When performing our experiments (described in Experiments and Results), we mixed and matched these interface components to form experimental conditions that test the effects that each component has on the quality and quantity of participation.

[3] `http://www.mediawiki.org/wiki/Article_feedback`

RESEARCH QUESTIONS

In this section we motivate three research questions that drive our analysis to understand the effectiveness of AFT as a mechanism for eliciting useful feedback and converting readers to Wikipedia editors.

RQ1: How do different elicitations affect the volume and utility of feedback?

A larger pool of contributors isn't necessarily better. Although Wash and Lampe found that increasing participation by asking readers for comments on news articles did not decrease comment quality[25], these comments were not judged for their *usefulness* to the journalists. When Bettenburg et al. measured the usefulness of bug reports to developers, they found that reports by casual users were less useful due to mismatches between user and developer concerns[3]. We hypothesize that focusing readers toward the concerns of Wikipedia editors when eliciting feedback will encourage submissions that are more useful to editors. Conversely, we expect designing the feedback form around the concerns of Wikipedia readers may encourage more readers to submit feedback at the cost of decreased utility to editors.

We expect to find a tradeoff between the quantity and quality of participation whereby elicitations that receive more feedback submissions will do so by encouraging submissions that are less useful to editors, or more precisely:

- Asking readers whether they found what they were looking for (form 1) should increase participation since the question directly addresses readers' concerns but decrease utility since many of those concerns will not be shared by editors.

- Asking readers to categorize their feedback (form 2) should elicit more useful feedback by giving cues to readers about what types of feedback are expected but decrease participation due to the increased transaction cost of requiring the reader to provide meta-information.

- Asking readers to rate the article (form 3) should increase the quality of comments by encouraging readers to consider the the article's quality within the context of the encyclopedia (editors' concern) but decrease participation by forcing readers into the role of experts.

H1.1: Form elicitations that increase participation will do so at the cost of the decreased utility of submissions.

RQ2: How does the prominence of the elicitation affect the volume and utility of feedback?

The design decision to place the feedback forms after an article was strategic. Given that most articles in Wikipedia contain enough content to require scrolling the browser window, feedback forms positioned at the bottom of the article are hidden such that only readers who have scrolled through the article will ever see them appear on the screen. Presumably, these readers are particularly suited to leaving feedback because they should be more likely to have examined some content from the article before seeing the form. However, there may be readers who would have submitted valuable feedback, but never realized they could because they didn't scroll through the entire article.

By presenting a prominent link to the feedback form, we provide an alternate route to leaving feedback that does not require the reader to scroll through the article and we expect that the prominent placement will capture the attention of more readers. We expect that by making feedback forms more prominent, we will increase the number of submissions at the cost of decreasing the overall usefulness of submissions.

H2.1: The volume of feedback submissions will increase with prominence.

H2.2: The utility of feedback submissions will decrease with prominence.

RQ3: How does the presence of the feedback interface affect new editor conversion?

A prominent invitation to leave feedback as opposed to making a productive edit to an article may undermine the "sofixit"[4] and "be bold"[5] culture of Wikipedia that encourages individuals who feel that an article should be changed to boldy edit the article themselves. Further, even for a reader whose interest in making an edit herself was held constant despite submitting feedback, the act of submitting feedback should "cannibalize" some of her finite time and effort that could have been put towards more direct participation in the form of an edit. If this is true, the presence of AFT may actually be counterproductive.

H3.1: The presence of AFT will decrease the rate of new editor conversions.

As a stopgap measure, presenting the reader with an invitation to edit (form 4 in figure 1) after submitting feedback should encourage potential new editors to continue through the normal process of "boldly" making an edit themselves. However, such an invitation to edit might also short circuit the normal process of peripheral participation by encouraging a reader to make the transition to editing before they've lurked for long enough to understand the norms and goals of the community.

H3.2: New editors who are invited to start editing articles will be less productive than editors who start by their own volition.

METHODS

The article sample

To explore our three research questions, we performed three distinct experiments on the English Wikipedia. For each experiment, the AFT interface components were loaded on a 0.6% random sample of encyclopedia articles. All readers who viewed these articles were randomly assigned to experimental conditions (described in Experiments and Results). We used browser cookies to extend the continuity of experimental groups between sessions and internet connections (IP addresses) so that readers would remain within the same experimental condition through the duration of an experiment.

To minimize the potentially disruptive impact of testing AFT on a live website, the same random sample of articles was used for all three experiments. However all users were rebucketed between experiments to control for a potential ordering bias.

Feedback utility

In order to address the research questions postulated in Research Questions, we required a way to determine which of the feedback submissions were useful to editors. The judgement of experienced Wikipedia editors to perform this evaluation was invaluable in determining the usefulness of feedback since, by design, they will be responsible for making use of feedback after it has been submitted, so they will know best which feedback submissions are useful. In other words, we consider their judgement to be a direct measure of utility. We

[4] `http://en.wikipedia.org/wiki/Template:sofixit`
[5] `http://en.wikipedia.org/wiki/WP:BOLD`

organized a group of Wikipedian volunteers interested in determining the overall utility of AFT via requests posted on the documentation page[6] and several IRC sessions.

To support these Wikipedian evaluators, we built a user interface to present a feedback submission in the context in which it was submitted while hiding details about the experimental condition from which feedback was submitted. For each feedback submission loaded into this user interface, the version of the article at the time feedback was submitted is loaded into an article preview pane, and the feedback text that was submitted is loaded into the feedback text pane. The rater was instructed to use an evaluation form to answer two questions:

Is this useful? Raters were instructed to mark a feedback submission as "useful" if they could imagine making use of this feedback to edit the article[7].

What's the intent? Raters were instructed to categorize the intent of the feedback. They could select zero to many options from: *suggestion*, *praise*, *question*, *issue*, *irrelevant* and *abuse*. These categories were decided upon through a pilot run in cooperation between the researchers and a subset of the raters.

20 Wikipedians participated in this hand-coding process. We randomly assigned work such that each feedback submission was rated by exactly two different Wikipedians. Each Wikipedian rated between 50 and 350 feedback submissions.

New editor productivity

To measure the productivity of newly converted editors, we examine their contributions to encyclopedia articles during their first week of tenure. We assume that an edit is "productive" if it is not reverted by another editor within 48 hours.

We identify reverted edits by looking for subsequent revisions that completely discard the changes of a new editors' revisions using the approach for determining "identity reverts" described by Halfaker et al. Their work suggests that the proportion of reverted edits is a useful proxy for the quality of an editor's work in most cases [10]. Given Wikipedia's efficient vandal-fighting system (see [8] for an overview), most unproductive edits should be reverted within moments, so 48 hours should be ample time to capture a reverting edit. We consider an editor to be "productive newcomer" if she makes at least one productive edit to an encyclopedia article within a week of her conversion.

To understand the cost of the increasing non-productive edits, we examined *how* these newcomer edits were reverted. The Wikipedia community's approaches to reverting damaging contributions can be organized into three categories:

- *manual* - Human editors revert edits directly via the web interface. Reverts performed manually require the most human effort.
- *semi-automated* - Wikipedians have developed a suite of tools to make the process of identifying and reverting damage require less human effort. Contributions revert by wikipedians using semi-automated tools require less effort than manual reverts.
- *automated* - Autonomous computer programs commonly referred to as "bots" monitor recent changes and perform reverts without human input. Contributions reverted by bots require essentially no human effort.

Advantageously, the automated and semi-automated tools identify themselves by leaving structured comments along with the revisions they make. Through the use of a set of regular expressions on these edit comments, we are able to categorize reverts into the above three categories. Reverts performed by user accounts with a bot "flag" were classified as *automated*. Reverts performed using Huggle[8], Twinkle[9], Popups[10], Rollback[11] or STiki[12] were classified as *semi-automated*. All other reverts were classified as *manual*. We examined the proportion and raw number of reverts that are caught by each tool to understand the cost of non-productive newcomer contributions.

EXPERIMENTS AND RESULTS

RQ1: How do different elicitations affect the volume and utility of feedback?

To explore this research question, we randomly divided readers into three exprimental conditions – one for each of the three feedback forms (1-3 from figure 1). For each condition, the appropriate feedback form appeared at the end of the encyclopedia article. Note that neither of the prominent links were used for this experiment.

Results

We considered different designs of the AFT as treatments applied to the same population of articles and measured the amount of daily feedback submissions generated in response of each treatment for the duration of the experiment (Dec. 27th - Jan 24th, 2012). Figure 2 plots the median submissions per day over the last two weeks of the experimental period. The first two weeks were discarded to limit the effect of novelty for the new feedback interface. Forms 1, 2 and 3 generated a total of 1666, 1539 and 1148 feedback submissions respectively. We used a one-way analysis of variance to test the null hypothesis that the means of these treatments are equal. The test allows us to reject the null hypothesis at $\alpha = 0.01$ ($F(2, 42) = 6.57, p = 0.003$). Post-hoc comparisons using the Tukey range test indicates that the means for form 1 ($M = 111, SD = 32.3$) and form 2 ($M = 102.6, SD = 27.4$) were significantly higher than form 3 ($M = 76.5, SD = 20.5$) at $\alpha = 0.05$. However, the

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

[7] Raters were allowed to select a checkbox named "unsure" if they didn't feel confident about their rating.

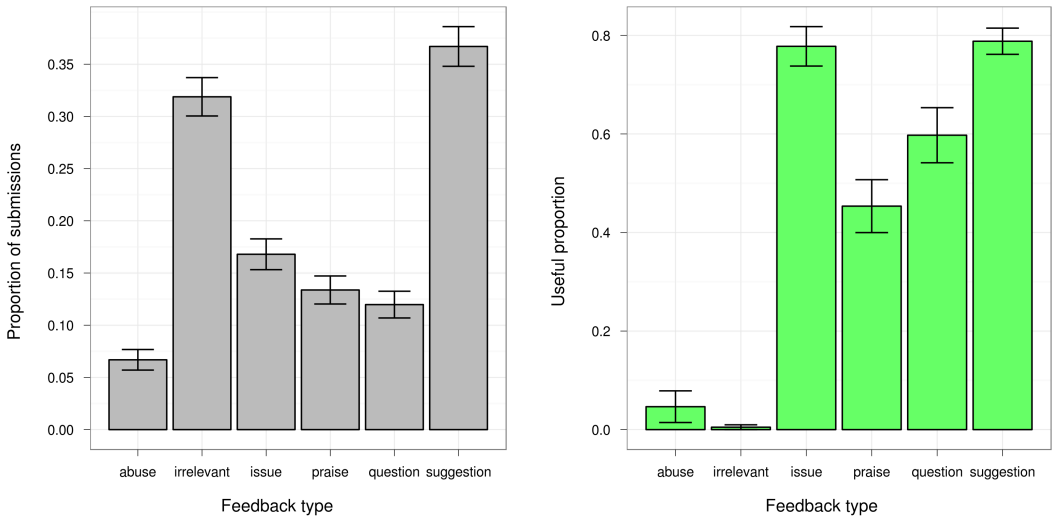
[8] `http://enwp.org/WP:HG`

[9] `http://enwp.org/WP:TW`

[10] `http://enwp.org/WP:Popups`

[11] `http://enwp.org/WP:Rollback`

[12] `http://enwp.org/WP:STiki`



(a) **Proportion of feedback types**.

(b) **Utility of feedback types**.

Figure 3: (left) The proportion of feedback submitted is plotted by intention as determined by at least one Wikipedian. (right) The proportion of useful feedback is plotted for each type. These plots draw aggregate proportions from feedback submitted through all three experimental conditions.

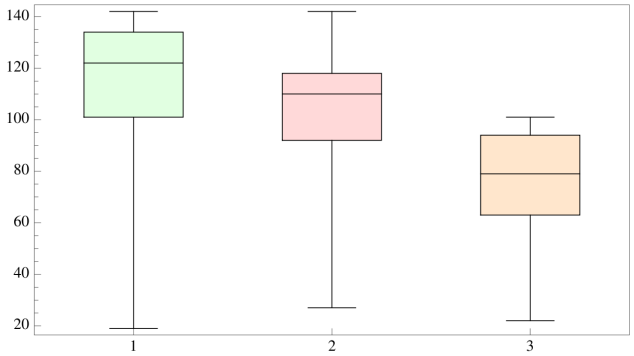


Figure 2: **Quantity of feedback by experiment.** The median daily feedback submissions per day is plotted for the last two weeks of the experiment with box limits at the 25% - 75% quantiles and bar limits at the most extreme observed values.

rate of feedback submissions for form 1 did not significantly differ from form 2.

Next, we looked at the utility of feedback submissions via the three forms as determined by our Wikipedian evaluaters. A random sample of up to 250 feedback submissions was gathered for each condition such that no two submissions came from the same article over the entire experimental period[13]. This sampling approach was used to produce statistics that reflect the expected utility of feedback submissions per article by controlling for the overrepresentation of popular articles. To measure the quality of feedback submitted, we examined the proportion of useful feedback submissions as determined via hand coding by multiple Wikipedians, as described in Methods.

To combine the two ratings per feedback submission into a single assessment, we employed three aggregation strategies:

- **someone**: useful to at least one Wikipedian
- **both**: useful to both Wikipedians
- **strict**: useful to both Wikipedians and neither was unsure

We found the three rating aggregation strategies to compare similarly between experimental conditions, so we opt to only report the proportion of feedback determined useful by both Wikipedians ("both" strategy).

The differences between the proportion of useful feedback submitted between the three interfaces varied insignificantly around 0.45. Of the three conditions, the most substantial difference was observed between form 2 (0.436) and 3 (0.469), however a χ^2 test found that difference to be insignificant ($p = 0.604$).

We also found a relatively strong consistency between the apparent intentions of feedback submitted via the three interfaces with a couple of exceptions: a χ^2 test showed that form 2 elicited a significantly more useful issues than form 1 ($\Delta = 0.225, p = 0.035$) and form 1 elicited a significantly higher proportion of questions than form 3 ($\Delta = 0.100, p = 0.005$). Although these differences are statistically significant, the difference in useful issues could be considered insubstantial and may have come about by chance given the number of statistical tests performed to identify differences (6 types of feedback * 2 tests * 2 conditions = 24 tests).

[13]Form 3 didnt elicit feedback submissions across enough distrinct articles during this period, so we were only able to sample 143 submissions.

To explore a potential mismatch between the types of feedback readers were likely to submit and the types of feedback Wikipedians found useful, we merged the feedback submitted via the three interfaces to compare the aggregate rate of feedback types and utility. Figures 3a and 3b show some substantial differences between the feedback submitted by readers and the feedback that Wikipedians found useful. As an example, the proportion of feedback determined irrelevant by Wikipedians was the second most frequent type (31.9%) while the proportion of irrelevant feedback that was marked useful was vanishingly small (0.5%) and could be attributed to rater error given that it represents a single submission labelled both useful and irrelevant.

Figures 3a and 3b also shows some similarities. For example, suggestions appear to be both the most prevalent (36.7%) and useful (78.9%) submission type.

The relatively small proportion of abuse is also worth noting. Only 6.6% of feedback was determined to be submitted in bad-faith. To put this into context, 10.0% of anonymous edits to encyclopedia articles are explicitly labelled as vandalism when they are reverted[11].

Discussion

Although the elicitation directed towards readers' concerns (form 1) did elicit more participation than the form directed toward editors' concerns (form 3), there was not a signficiant difference in the proportion of useful feedback. This result refutes hypothesis 1.1 and suggests that the strategy of directing readers towards editor interests – at least in this case – did not affect the usefulness of their participation.

However, we did see a slight effect on the intentions of submitted feedback. We observed a significantly lower proportion of questions submitted via form 3 which asks readers to rate an article. We suspect that, to evaluate the quality of an article, readers must put themselves in the role of an expert – someone qualified to perform an evaluation. Assuming readers who have legitimate questions would be less likely to adopt this role, it makes sense to see both a lower proportion of questions submitted and fewer overall submissions for form 3. Given the relatively high overall utility of questions (59.7% useful), this suggests that a substantial number of useful questions may not have been submitted by readers in this condition.

RQ2: How does the prominence of the elicitation affect the volume and utility of feedback?

Given that the results reported in the last section suggest that the different forms had no substantial effect on the utility of submissions and our Wikipedian hand-coders preferred form 1, we continue our experimentation using only that form. To look for changes in the utility of feedback submitted based on the amount of reader participation, we performed another experimental run using links "A" and "E" described in figure 1 in an attempt to increase the prominence of the feedback form. We re-shuffled readers into three new experimental conditions:

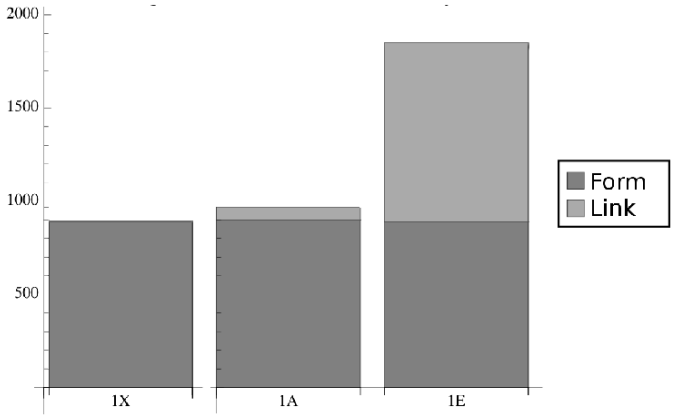


Figure 4: **Volume of feedback by experiment.** The raw amount of feedback submitted via the three experimental conditions is plotted by whether the feedback was submitted via the prominent link or via the form at end of articles.

- **1X**: Form 1 loads only on the bottom of the article.
- **1A**: Same as 1X but includes link A
- **1E**: Same as 1X but includes link E

Results

To ensure that the prominent links A and E were serving their intended function, we measured the total amount of feedback submitted over the period between April 5 and April 19 (15 days).

As figure 4 shows, the prominent link appears to have been effective in eliciting more feedback. While all cases elicited approximately the same amount of feedback directly via the form at the bottom of the article as the non-prominent condition (1X), the feedback submitted via the prominent link itself appears to purely supplement the total amount of feedback submitted (1X: 892, 1A: 967 and 1E: 1851). In the case of 1E, the prominent link apparently increased the rate at which feedback was submitted by about 91%.

To look for differences in the utility of feedback submitted via each of the experimental conditions, we randomly sampled 300 feedback submissions per condition and asked our Wikipedian coders to rate their usefulness. The differences between the proportion of useful feedback submitted between the three interfaces varied insignificantly around 0.42. A χ^2 of the largest observed difference between the conditions 1X (no link) and 1A (less prominent link) was insignificant ($\Delta = 0.065, p = 0.124$). If the prominence of the interface caused a decline in the utility of feedback, we would expect to see a more substantial dip in the proportion of useful feedback in the 1E condition given the amount of additional feedback submitted.

However, we did observe a significant difference in utility based on the whether the feedback originated via the form at the bottom of the article or the prominent link E. Figure 5 plots the proportion of useful feedback submitted for each experimental interface by whether it originated via the form

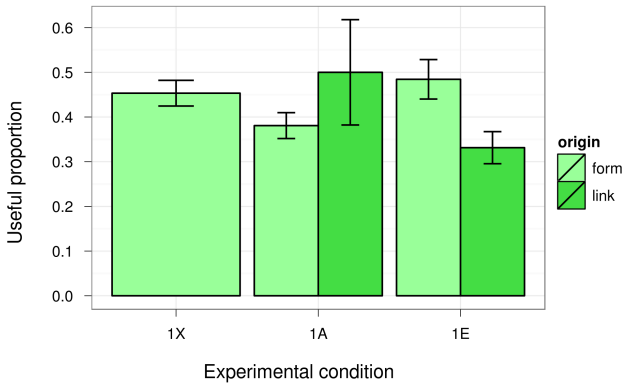


Figure 5: **Utility of feedback by origin.** The proportion of useful feedback is plotted for each condition by the origin from which it was submitted with standard error bars. Note that the large error bars around the proportion of useful feedback submitted via the link in the 1A condition is due to the small amount of feedback sampled in that condition (n=18).

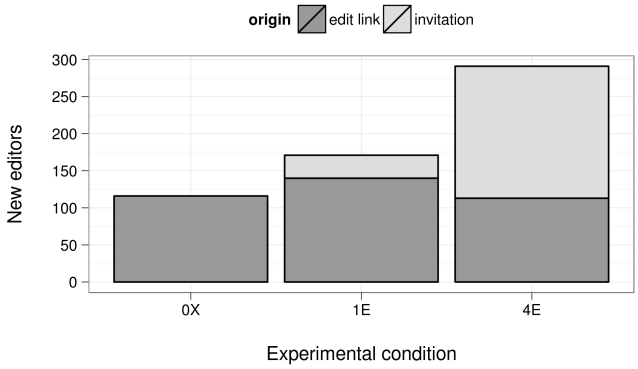


Figure 6: **New users by origin.** The number of new editors is plotted and stacked by the origin of their first edit. "edit link" refers to the standard vector for accessing the edit pane. "invitation" refers to form 4 which invites the user to make an edit.

at the bottom of the article or the prominent link. The proportion of useful feedback submitted via the most prominent link (1E) was significantly lower than the feedback submitted via the corresponding form ($\Delta = 0.153, p = 0.010$). This appears to be partially due to complaints about the prominent link submitted via the prominent link. After the experiment, we received messages from some concerned users that felt the prominent link was distracting. Out of the 8 feedback comments submitted via 1E's link by registered editors, only 1 feedback submission was determined to be useful while the rest were useless protests of the presence of the feedback interface.

Discussion

Increasing the prominence of the feedback form increased the volume of contributions in the expected way, affirming hypothesis 2.1. The results above suggest that the rate at which feedback is submitted can be nearly doubled by introducing prominent link E into the interface.

Surprisingly, we also found that expanding participation by making the interface more prominent did not affect the overall proportion of useful submissions, thus refuting hypothesis 2.2. We had hypothesized that increasing participation arbitrarily via a more obvious interface would elicit participation from readers who were increasingly less interested in contributing constructively, and therefore, utility would fall. However, we found no significant change in the proportion of useful feedback submitted despite more than doubling the amount of submissions. This result suggests that there are many potentially productive contributors available who may simply not be aware of their ability to contribute and that hiding the means to contribute is not an effective mechanism for improving the utility of contribution.

However, we did observe that feedback submitted via the most prominent link was of lower utility than that submitted directly via the form. Although it is tempting to conclude that the prominent link itself elicits lower utility submissions, it is important to note that the overall utility of submissions through the most prominent condition (1E) was not significantly lower than the 1X condition despite the fact that the prominent link nearly doubled the rate at which feedback was submitted. Given the slightly larger proportion of useful feedback submitted directly via the 1E form than the 1X form, we suspect that the underlying cause of this disparity is a rerouting of less useful feedback (that would have been submitted anyway) through the prominent link.

RQ3: How does the presence of the feedback interface affect new editor conversion?

To look for a potential cannibalization effect of AFT on new editor conversions, we re-shuffled readers into three new experimental conditions:

- **0X**: Control condition (no feedback form or link)
- **1E**: Feedback form 1 displayed at the bottom of the article, with prominent link E. Users were presented with edit invitation form 4 after successfully submitting feedback (indirect invitation)
- **4E**: Edit invitation form 4 displayed at the bottom of the article, with prominent link E (direct invitation)

Results

To measure new editor conversions, we observed user activity for the period between April 27th and May 7th 2012. During this observation period, both the direct invitation (4E) and indirect invitation (1E) conditions saw more new editors conversions than the control case. Figure 6 shows that a similar number of new editors originated via the standard vector (edit link) in the treatment conditions (1E: $n = 140$, 4E: $n = 113$) as in the control condition (0X: $n = 116$). The new editors whose first edit originated via the invitations to edit (1E: $n = 31$, 4E: $n = 178$) only appear to supplement these numbers.

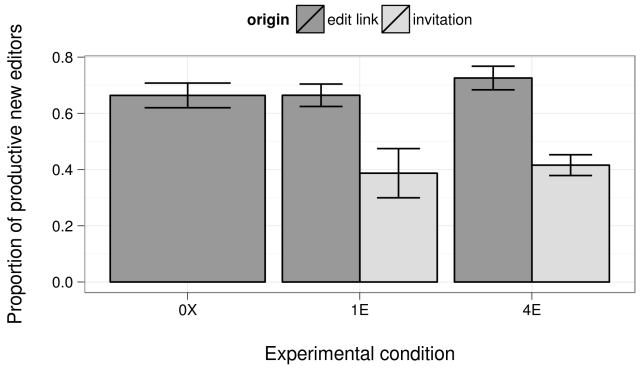


Figure 7: **Proportion of productive new users by origin.** The proportion of new editors who made at least one productive contribution in their first week is plotted by the origin of their first edit for the three experimental conditions.

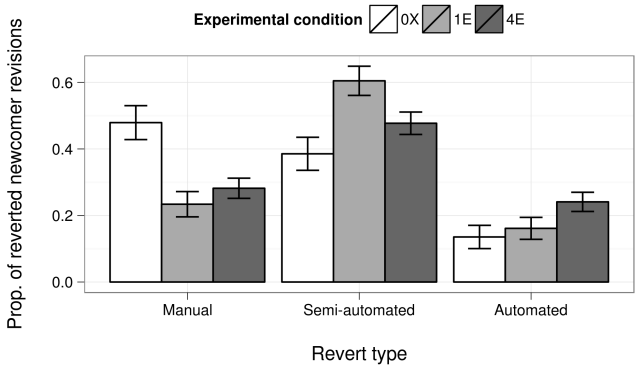


Figure 8: **The cost of non-productivity**. The proportion of newcomer revisons reverted for each experimental condition by how the revision was reverted.

To test if this increase in conversion was significant, we approximated the view-to-editor conversion rate by drawing 95% confidence intervals around the approximate views that each experimental condition received based on the total number of views that the sample received during the observation period (9,424,041). The random function which assigned readers to experimental conditions can be represented as a binomial proportion where the underlying probability of assignment to any one condition is $\frac{1}{3}$. Therefore, we can divide the total views and draw confidence intervals using the binomial approximation to a normal distribution($\frac{9424041}{3} = 3141347 \pm 945.47$). We used this confidence interval to perform a conservative χ^2 test that appropriately reduces the likelihood of a type I statistical error. Let u_1 and u_2 be set of new users such that $|u_1| > |u_2|$ we performed the following test:
$\chi^2(\frac{|u_1|}{3141347+945.47}, \frac{|u_2|}{3141347-945.47})$

The conservative χ^2 test found the difference in the rate of new editor conversions to be significant for all three cases (0X $<$ 1E: $p = 0.002$, 1E $<$ 4E: $p < 0.001$). This suggests that the indirect invitation condition (1E) converted significantly more views to new editors than control condition and the direct invitation condition (4E) converted significantly more views to new editors than both 1E and the control.

To explore the productivity of new editors conversions via the experimental conditions, we performed an analysis of the first week of contributions made by each new editor to look for productive contributions as described in Methods. We define a "productive editor" as a new editor who made at least one edit to an article that was not reverted within 48 hours by another editor.

As figure 7 shows, editors who originated via the invitation to edit through both the direct and indirect call to action were significantly less likely to make productive contributions in their first week than those who originated through the edit links within the same condition (1E: $\Delta = 0.277, p = 0.008$; 4E: $\Delta = 0.310, p < 0.001$). When comparing only those new editors who originated via the edit link between experimental conditions, the proportion of productive new editors varied insignificantly around 0.68. These results suggest that new editors who originated via an invitation to edit are about half as likely to contribute productively in their first week as those editors who originated via the standard vectors.

To examine the cost of introducing more unproductive newcomer contributions to Wikipedia, we classified the reverted newcomer editors in each condition into *manual*, *semi-automated* and *automated* as described in Methods. As figure 8 suggests, a χ^2 test found that a significantly smaller proportion of unproductive newcomer edits were reverted manually in both the 1E($\Delta = -0.25, p < 0.001$), and 4E($\Delta = -0.19, p = 0.001$) conditions. We also see that a significantly larger proportion of 1E revisions were reverted with semi-automated tools($\Delta = 0.219, p = 0.002$) and that a significantly larger proportion of 4E revisions were reverted with fully automated tools($\Delta = 0.134, p = 0.019$) when compared to the control. These differences suggest that Wikipedia's damage filtering systems (automated and semi-automated tools) are making up for some of the increased cost of damage due to the invitation to edit.

Discussion

The results described in this section are a solid refutation of hypothesis 3.1. Even without including those editors who saw the edit invitation, (insignificantly) more new editors made their first contribution in the 1E condition through the standard vectors to the edit pane (edit link) than in the control case. The indirect call to action appeared to only supplement new editor conversions. Further, this class of new editors appears to be just as likely to be productive in their first week of tenure as those who started editing via the control case.

However, there appears to be a cost to boosting conversions via invitations to edit – at least initially. A substantially smaller proportion of editors originating via the invitation made a productive contribution in their first week of edit activity.

RQ1: How do different elicitations affect the volume and utility of feedback?		
H1.1	unsupported	Although form 1 and form 2 elicited more feedback submissions than form 3, the quality of submissions was insignificantly different.
RQ2: How does the prominence of the elicitation affect the volume and utility of feedback?		
H2.1	supported	The prominent link conditions elicited significantly more feedback than the control.
H2.2	unsupported	Although feedback submitted via the prominent links was less useful, the overall differences were insignificant.
RQ3: How does the presence of the feedback interface affect new editor conversion?		
H3.1	unsupported	The presence or absence of the feedback interface had no observed effect on the rate of new editor conversions through the standard vectors.
H3.2	supported	New editors whose first edits originated via an invitation were substantially less likely to contribute productively in their first week.

Table 1: A summary of our hypotheses and findings.

As we suspected, the edit invitation strategy is a double-edged sword. Although the invitation appears to have effectively convinced substantially more readers to try their hand at editing, these invited editors were less likely to make productive contributions in their early wiki-career than those who started editing on their own. This result supports hypothesis 3.2.

Social learning theory suggests that newcomers to a community tend to go through a natural process by which they learn about a community and build confidence before expanding their contribution pattern[18]. It seems likely that explicitly inviting new editors to contribute would artificially inflate their confidence and speed up the process of introduction to the community which would naturally lead to more mistakes – at least initially.

Our analysis suggests that, at least in part, Wikipedia's efficient damage filtering mechanisms reduce the cost of the increase in unproductive newcomer contributions. Although we can assume that the cost in human effort of a few more automated reverts is essentially zero, the cost of operating the semi-automated reverting tools is unclear. Without such understanding, it's difficult to reason about just how much Wikipedia's damage filtering mechanisms make up for the increase in unproductive newcomer contributions.

CONCLUSIONS

In this paper, we examined a strategy for increasing participation from Wikipedia's consumers (readers) through a set of experiments performed using the Article Feedback Tool, an extension of the software that allows encyclopedia readers to make simple and low risk, yet productive, contributions – thereby supporting legitimate peripheral participation. Table 1 summarizes the results of our three experiments in the context of the research questions and hypotheses stated in Research Questions.

Although we hypothesized a trade-off between the quality and quanity of participation, our observations of such an effect were inconsistent. While we were able to boost the rate of feedback submissions both by aligning AFT's feedback forms to readers concerns and by making the interface more prominent to readers, we did not observe a decrease in the quality ("useful"ness as determined by Wikipedian raters) of feedback submitted. This result reflects the findings of Wash and Lampe[25] and suggests that there are many potentially productive non-contributors available who may simply be unaware of their ability to contribute and that obfuscating the mechanism for contribution is not an effective mechanism for improving the quality of contribution. In other words, increasing transactional costs either through requiring the reader to occupy a certain role or through obfuscating the ability to contribute do not appear to be effective strategies for increasing contribution quality for this low investment type of contribution. On the contrary, both our results and those of Wash and Lampe[25] suggest that participation can be increased by making the means of contribution more prominent without sacrificing quality with simple and straightforward contribution types like comments and feedback.

However, when we invited editors to edit articles via an explicit invitation, those supplemental new editors were less likely to be productive in their first week of tenure in Wikipedia. Editing an encyclopedia is a more demanding type of contribution than leaving a comment or submitting feedback, because editors must both contribute novel encyclopedic information and work within Wikipedia's complicated set of policies and guidelines – a requirement that new editors tend to struggle to meet[9]. Under the framing of legitimate peripheral participation, we might expect potential new editors to naturally go through a lurking process by which they build a situated understanding of the Wikipedia community and their place within it before making their first contribution – an assertion supported by [1]. When viewed this way, the invitation to contribute could be short circuiting the natural "joining script"[24] of new Wikipedia editors. If this is true, other systems with non-trivial contribution difficulty should see a similar effect.

It is useful to understand the value of AFT's invitation to new editors to the Wikipedia community's goals from a system-level perspective. In order to draw a data-based recommendation on whether AFT's invitation to edit should be released at a larger scale or not, both the value of new productive contributions and the cost of unproductive contributions must be accounted for. Systems with a high value on new contributions/contributors and/or efficient filters for dealing with un-

wanted contributions will be more likely to benefit from invitations despite the increase in unproductive edits.

Related work by Geiger et al. suggests that the English Wikipedia's automated and semi-automated anti-vandal tools are a efficient and scalable solution to the problem of moderation[8]. Our analysis suggests that these tools are effectively reducing the increased cost of unproductive contributions via the invitation to edit. In this context, it seems likely that the invitation to edit represent a net benefit to Wikipedia despite the larger proportion of unproductive edits. Related works examining Slashdot's distributed moderation system [14] and YouTube's copyright infringement detection algorithms[23] suggest a trend toward efficient, distributed and automated curation mechanisms in Web 2.0 systems that may also reduce the cost of such unwanted contributions in other systems.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the National Science Foundation under grants IIS 09-68483 and IIS 11-11201. We would also like to thank the Wikimedia Foundation and the following Wikipedians for their work evaluating feedback submissions: Salim Fadhley, Fluffernutter, GorillaWarfare, Adam Hyland, Jamietw, John Minthorne, Tom Morris, Osarius, Rohit Patnaik, Utar and Whiteguru.

REFERENCES

1. Antin, J., and Cheshire, C. Readers are not free-riders: reading as a form of participation on wikipedia. In *CSCW '10*, ACM (New York, NY, USA, 2010), 127–130.
2. Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., and Kraut, R. E. Using social psychology to motivate contributions to online communities. In *CSCW '04*, ACM (New York, NY, USA, 2004), 212–221.
3. Bettenburg, N., Just, S., Schröter, A., Weiss, C., Premraj, R., and Zimmermann, T. What makes a good bug report? In *SIGSOFT '08*, ACM (New York, NY, USA, 2008), 308–318.
4. Bryant, S. L., Forte, A., and Bruckman, A. Becoming wikipedian: Transformation of participation in a collaborative online encyclopedia. In *GROUP '05*, ACM (New York, NY, USA, 2005), 1–10.
5. Butler, B., Sproull, L., Kiesler, S., and Kraut, R. E. Community effort in online groups: Who does the work and why? *Human-Computer Interaction Institute* (2007), 90.
6. Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. Suggestbot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*, IUI '07, ACM (New York, NY, USA, 2007), 32–41.
7. Ducheneaut, N. Socialization in an open source software community: A socio-technical analysis. In *CSCW '05*, ACM (New York, NY, USA, 2005), 323–368.
8. Geiger, R. S., and Ribes, D. The work of sustaining order in Wikipedia: The banning of a vandal. In *CSCW '10*, ACM (New York, NY, USA, 2010), 117–126.
9. Halfaker, A., Geiger, S., Jonathan, M., and Riedl, J. The rise and decline of an open collaboration system: How Wikipedia's reaction to sudden popularity is causing its decline. *American Behavioral Scientist* (in-press).
10. Halfaker, A., Kittur, A., Kraut, R., and Riedl, J. A jury of your peers: Quality, Experience and Ownership in Wikipedia. In *WikiSym '09*, ACM (New York, NY, USA, 2009), 1–10.
11. Halfaker, A., Kittur, A., and Riedl, J. Don't bite the newbies: How reverts affect the quantity and quality of Wikipedia work. In *WikiSym '11*, ACM (2011), 163–172.
12. Katz. Luring the lurkers. http://news.slashdot.org/story/98/12/28/1745252/luring-the-lurkers, Dec 1998.
13. Ko, A. J., and Chilana, P. K. How power users help and hinder open bug reporting. In *CHI '10*, ACM (New York, NY, USA, 2010), 1665–1674.
14. Lampe, C., and Resnick, P. Slash(dot) and burn: distributed moderation in a large online conversation space. In *CHI '04*, ACM (New York, NY, USA, 2004), 543–550.
15. Lave, J., and Wenger, E. *Situated learning: Legitimate peripheral participation*. Cambridge University Press, 1991.
16. Nonnecke, B., and Preece, J. Why lurkers lurk. In *Americas Conference on Information Systems* (2001).
17. Nonnecke, B., and Preece, J. Silent participants: Getting to know lurkers better. *From usenet to CoWebs* (2003), 110–132.
18. Preece, J., and Shneiderman, B. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Trans HCI 1 1*, 1 (2009), 13–32.
19. Priedhorsky, R., Chen, J., Lam, S., Panciera, K., Terveen, L., and Riedl, J. Creating, destroying, and restoring value in Wikipedia. In *GROUP'07* (Sanibel Island, FLorida, USA, 2007).
20. Rashid, A. M., Ling, K., Tassone, R. D., Resnick, P., Kraut, R., and Riedl, J. Motivating participation by displaying the value of contribution. In *CHI'06*, ACM (New York, NY, USA, 2006), 955–958.
21. Suh, B., Convertino, G., Chi, E. H., and Pirolli, P. The singularity is not near: Slowing growth of Wikipedia. In *WikiSym '09*, ACM (New York, NY, USA, 2009), 1–10.
22. Taraborelli, D., and Roth, C. Viable Web communities: Two case studies. In *Viability and Resilience of Complex Systems*, G. Deffuant and N. Gilbert, Eds. Springer, 2011, 75–105.

23. Tartelton, G. The politics of platforms. *New Media & Society 12*, 3 (2010), 347–364.

24. Von Krogh, G., Spaeth, S., and Lakhani, K. Community, joining, and specialization in open source software innovation: a case study. *Research Policy 32*, 7 (2003), 1217–1241.

25. Wash, R., and Lampe, C. The power of the ask in social media. In *CSCW '12*, ACM (New York, NY, USA, 2012), 1187–1190.

26. Wikimedia. Editor trends study. **http://strategy.wikimedia.org/wiki/Editor_Trends_Study**, March 2011.

27. Wilkinson, D. Strong regularities in online peer production. In *Ecommerce '08*, ACM (2008), 302–309.