Don't Bite the Newbies: How Reverts Affect the Quantity and Quality of Wikipedia Work

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

Aniket Kittur
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
nkittur@cs.cmu.edu

John Riedl
Grouplens Research
University of Minnesota
200 Union St. S.E.
Minneapolis, MN 55455
riedl@cs.umn.edu

ABSTRACT

Reverts are important to maintaining the quality of Wikipedia. They fix mistakes, repair vandalism, and help enforce policy. However, reverts can also be damaging, especially to the aspiring editor whose work they destroy. In this research we analyze 400,000 Wikipedia revisions to understand the effect that reverts had on editors. We seek to understand the extent to which they demotivate users, reducing the workforce of contributors, versus the extent to which they help users improve as encyclopedia editors. Overall we find that reverts are powerfully demotivating, but that their net influence is that more quality work is done in Wikipedia as a result of reverts than is lost by chasing editors away. However, we identify key conditions – most specifically new editors being reverted by much more experienced editors – under which reverts are particularly damaging. We propose that reducing the damage from reverts might be one effective path for Wikipedia to solve the newcomer retention problem.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems

Keywords

Wikipedia, WikiWork, Experience, Quality, Motivation, Productivity, Revert

1. INTRODUCTION

Wikipedia has become one of the largest and most successful examples of social production on the web. Hundreds of thousands of volunteers are working together to create an encyclopedia of human knowledge. One of the keys to Wikipedia's success has been the ability to gather contributions from a large, diverse community of volunteer editors. To obtain the participation of such a wide array of web users, Wikipedia eases the transition from reader to editor via the technical platform and supporting policies; for example, the MediaWiki software lowers participation barriers by letting editors contribute through their web browsers; and community policies enable editing of most pages without even registering as a member.

However, this same ease of editing can be a double-edged sword, making it easy not only for legitimate editors to contribute but also for malicious or biased contributors to degrade the quality of existing content. One way Wikipedia, like many other wikis, addresses this issue is through revert functionality. When an inappropriate edit is made to an article, any other editor can revert the article back to its previous state. The ability to easily revert changes alters the participation cost structure such that it costs less to fix an undesired edit than it costs to make the edit in the first place.

Unfortunately, past research has shown that modifying the work of Wikipedia editors can reduce their rate of contribution in the future [26]. Reverts are a particularly direct form of modification, so it seems likely that reverts will have similar negative consequences. According to our analyses, the total percentage of reverts has increased over time to approximately 10% of all edits as of 2010. In this paper, we focus on the effect that reverts, other than those labeled as vandalism or self-reverts[1], have on editor behavior. In just 2009, there were nearly 900,000 reverts of this type, an enormous amount of potentially lost effort and motivation.

1.1 Research Questions

The Collective Effort Model (CEM) provides a useful framework for categorizing the motivations of Wikipedia editors[11]. The CEM suggests a relationship between individual and group outcomes. For example, tasks that may produce high group value can increase individual motivation; furthermore, increasing group outcomes can increase individual motivation for other tasks by making the individual feel that overall group effort is producing more value.

Getting reverted may make individuals feel that their contributions are not valued by the group and are not leading to positive group outcomes, resulting in demotivating effects. In this case, being reverted might cause editors to produce less work after the revert.

RQ1: How does being reverted affect the quantity of editor work?

---

[1]Self-reverts are a specific type of revert where an editor reverts their own work – often when they've made a mistake.

One perspective that has not been examined much in the literature is that being reverted could be part of the learning process for editors. When an editors is reverted, they may reconsider the edits they make thus improving their work. Wikipedia's guideline for editing boldly[2] encourages editors to make changes as they see fit and let the collaborative process help them check the quality of their work. Therefore, we might expect to see that editors who are reverted learn something from the experience and produce higher quality work afterward.

RQ2: How does being reverted affect the quality of editor work?

One way that Wikipedia editors might avoid motivation loss when they are reverted is by discussing the situation. Commiseration with other editors might help preserve or even reinforce editors' motivations to continue participating in Wikipedia. Wikipedia's Bold, Revert, Discuss cycle[3] encourages editors to engage in discussion with other editors after their work has been reverted. However, if they are too strongly discouraged by having been reverted they may communicate less, losing the opportunity to learn about and improve Wikipedia.

RQ3: How does being reverted affect communication?

Reverts have been implicated as a factor in the dramatic participation decline Wikipedia has seen in the last few years. The source of this decline appears to be reduced newcomer retention. For example, while nearly 40% of new editors remained active for a year pre-2005, that number dropped to only 12-15% post-2007[24]. A significant challenge to recruiting and retaining newcomers is the difficulty they experience in understanding the vast history of prior contributions, decisions, policies, and standards that the community has evolved over time. Such factors have led to newcomers being reverted at higher rates than more experienced editors, a trend that has been getting exacerbated over time[21]. Could it be that early reverts are responsible for the reduction in newcomer activity?

The experience of editors who perform reverts is an interesting variable for analysis too. Editors who have been working in Wikipedia for a long time will have more experience interacting with other editors. This experience could give them insight into how to discard other editors' work without demotivating them. More experience could also make it more difficult to relate to newbies, especially in the changed context of the more crowded, more complete Wikipedia of today[21].

RQ4: How does experience moderate the effects of reverts on contribution?

1.2 Structure and Contributions

In this paper we analyze the impact that reverts have on editor behavior. Although prior studies have examined reverts, ours is the first study we are aware of to quantify the impact of reverts on editor behavior. Understanding these effects is crucial to understanding how emerging editorial policy in Wikipedia is creating a high-quality encyclopedia.

This paper considers hundreds of thousands of editors, 400,000 revisions, and a variety of techniques to analyze the performance of those editors. Lest the contributions get buried in the details, we summarize them here. Our most important findings are that (1) reverts do have a negative impact on editor contribution and survival, especially for newcomers; and (2) when editors do continue to contribute after a revert, the quality of their contributions increases. Taken together, our findings suggest a more nuanced view of reverts that both recognizes the benefits of reverts for learning while acknowledging their costs in editor motivation and contribution, especially for newcomers. This view has practical implications for system design as well as for the design of intelligent tools for supporting reverts that enhance their beneficial effects while minimizing costs.

2. RELATED WORK

Wikis have been studied in the contexts ranging from the classroom [3, 7] to enterprise [9]. The most successful example of a wiki, Wikipedia, is also the most studied (though see [13] and [17] for analyses of other wiki communities).

A number of studies have looked specifically at reverts in Wikipedia. One common subject of research has been quantifying revert activity. This has often focused on reverting "damaged" content, such as the effects of vandalism. Wikipedia editors have developed sophisticated semi-automated processes for detecting, escalating, and sanctioning vandalism [8], with the result that vandalism is typically reverted very quickly, often on the order of minutes [14, 15, 16, 23]. The prevalence of reverts has been growing over time [14], as has a trend towards reverting newcomers [21].

Although vandalism is an important challenge, reverts often signify more substantive disagreements. Significant prior research has also focused on detecting, visualizing, and understanding the editing dynamics that involve reverts and their relationship to conflict. Conflict is not unexpected in a user-generated knowledge base, in which hundreds of thousands of contributors each bring their own viewpoints and knowledge, often leading to factions, and territoriality [10, 14, 22]. Tools to visualize conflicts and viewpoint differences based on reverts or revisions have also been developed [2, 6, 14, 20, 23].

Relatedly, there have also been a number of studies developing explicit trust and quality metrics that leverage the longevity of an editors' contributions (e.g., [5] and [18]). For example, Adler and de Alfaro[1] derive author reputation from the survival of an author's edits over time and surface this information through an interface that color-codes the content they contribute. Zeng et al.[25] use dynamic Bayesian networks to calculate the evolution of trust in an article using the editing status of authors and inserted/deleted text as input. Explicit visualization of revert activity in a page's history (among other features) has also been shown to decrease readers' perceptions of the trustworthiness of a page [12].

Our study builds on this foundation of research, and adds to it by quantifying the impact of reverts on the behavior of the reverted editor. Although much is known about revert activity at a macro level, little is known about its impact on editor behavior. Some studies have looked at the effect of other kinds of feedback on contribution. Zhang & Zhu [26] conducted an analysis of editors who created new pages in Wikipedia, finding that editors are less likely to contribute when another editor edits the page they created, irrespective of whether they were adding or removing content. This effect was mitigated by previous collaborator experience. Choi et al.[4] examined the effect of socialization tactics on new-

comers to Wikiprojects, finding that negative (constructive) feedback but not positive feedback led to relative increases in subsequent contribution. These studies suggest that interacting with an editors' contributions – either directly or through explicit feedback – can have an impact on the editor's subsequent amount of contributions. However, neither Zhang & Zhu nor Choi et al. examine the impact of reverts, nor changes in the quality or boldness of contributions in addition to their quantity.

3. DATA AND METHODS

3.1 Dataset

To look for the effects of being reverted, we gathered a sample of 400,000 revisions created by registered editors from the January, 2010 data dump[4] of English Wikipedia by combining two randomly selected subsamples. The *reverted* subsample comprises 200,000 revisions that were reverted by another editor (no self-reverts) and not tagged as vandalism[5]. The *not-reverted* subsample comprises 200,000 revisions that were not reverted to be used as a control for comparison. These samples were gathered in such a way that no two revisions were performed by the same editor. This restriction ensures that the effects we observe are not dominated by a few prolific editors. For the rest of this paper, we refer to the reverted and not-reverted revisions in the combined sample as *sampled edits*.

We obtained word peristence metadata (described below) generated for the January 3rd, 2008 English Wikipedia data dump and intersected it with our sample to obtain a second dataset with 176,438 revisions (90,641 reverted, 85,797 not-reverted). Figure 1 notes which metrics are based on this smaller dataset.

3.2 Quantity of work

Editors perform work in Wikipedia by editing pages. Starting an article, contributing to an existing article, sending a message to another editor and asking a question are all represented within the system as an edit that creates a new revision of a page. For this paper, we are primarily interested in the quantity of work as it applies to the construction of Wikipedia's encyclopedia articles. We measure the quantity of article work in two ways: revisions/day and words added/day.

Revisions/day is the number of article revisions created by an editor per day. This metric quantifies work based on the rate than an editor makes changes to articles.

Words added/day is the number of non-stop-words added to articles by an editor per day. This metric quantifies work based on the rate than an editor adds content to articles.

3.3 Quality of work

We approximate the quality of an editor's contributions based on how other editors react to it. If Wikipedia's natural review mechanism of collaborative editing is effective in selecting for quality content (see [19] for more explanation), higher quality contributions should be more likely to be accepted by other editors in Wikipedia. We define two metrics

Figure 2: **Persistent Word Revisions**. The word persistence values for five revisions of a sample article about apples are presented with arrows that show how words persist between revisions. Stop-words are greyed out since they are not considered in the algorithm. Revisions #3 reverts back to revision #1 and restores the word "red".

for the quality of editor contributions: reverts per revision and PWR/word[10].

Reverts/revision is the proportion of an editor's revisions that have been reverted in a given timespan. Reverting a revision is an indication that the reverting editor does not consider the revision to be of acceptable quality.

PWR/word, based on the PWR metric (see figure 2 for an example application of PWR), is the average number of revisions that words added by an editor persist. Higher quality contributions should, on average, last longer.

3.4 Boldness of contributions

Both of the metrics we use as proxies for the quality of an editors work are measures of whether other editors revert their contributions. Under these metrics, an editor can appear to do work of higher quality by making "safer" edits that are less likely to be reverted by other editors. Since the "boldness" of an edit can change our interpretation of results related to our quality metrics, we watch for changes in boldness in response to reverts. To quantify the boldness of contributions, we measure two characteristics of edits: the number of words changed and the average establishment of words removed.

Words changed per revision is the number of not stop-words added or removed by an edit. Edits that change more article content are considered to be more bold than edits that change little or no content.

Establishment of words removed is the average PWR of words that an edit removes. Words with high PWR have survived many revisions, and therefore, the scrutiny of other editors who edit the article. The higher the PWR value, the more strongly the content has been established in this article, and the bolder an edit that removes it.

3.5 Productivity

In addition to measuring how many words a user adds to Wikipedia ("Quantity") and how long those words last on average ("Quality"), we are also interested in finding a way to combine these metrics to estimate how much of an impact a user is having on the encyclopedia. Such a measure should include the amount of output the editor produces as well as the quality of such output. On average, an editor who adds more words that last longer is affecting Wikipedia more than an editor who adds fewer words, or whose words are reverted more quickly. We therefore define the *productivity* of an editor to be:

---

[4]A software bug produced a small amount of incorrect data that should not have significantly affected our analysis. See http://blog.wikimedia.org/2010/05/29/xml-dumps-resumed/
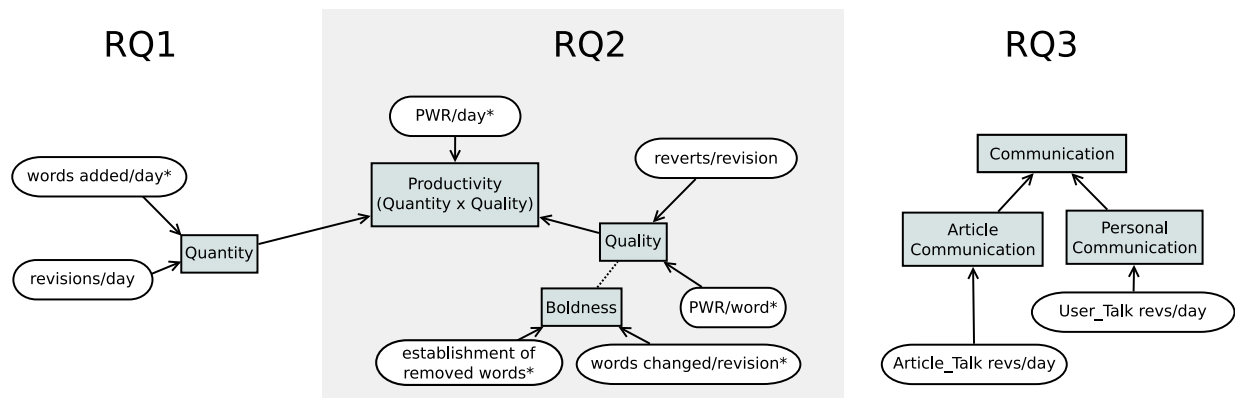[5]Vandalism was detected using the D_LOOSE approach from [16].

Figure 1: Hidden variables of editor activity are connected to the metrics that were used as proxies in the analysis and are divided by the research questions they are used to explore. Metrics obtained from 2008 data are signified with an "*". An arrow from A to B means "A is used as a proxy for B". The dotted line between boldness and quality represents the confounding effect described in Section 3.4. RQ4 does not rely on hidden variables.

PWR/day: the product of the number of not stop-words words added to articles and the number of subsequent revisions those words persist.

Of course, any such measure is an approximation of actual impact, since, at the extremes, an editor who focuses on improving Wikipedia policy might add significant value without ever editing a single word of an article. However, over a large number of edits, and a large and diverse set of editors, it is likely that measuring changes in how many words are contributed and in how long they last will serve as an effective proxy for productivity.

3.6 Measuring communication

For this work, we are interested in two types of communication: communication about article content and personal communication between editors. Wikipedia makes a clear distinction between these types of communication via the namespace[6] in which the communication occurs. Communication about article content occurs as edits to pages in the Article_Talk namespace and personal communication occurs as edits to pages in the User_Talk namespace.

Article_Talk revisions/day and *User_Talk revisions/day* are the number of revisions to pages in the Article_Talk and User_Talk namespaces created by an editor per day. Changes to the revisions/day in these namespaces should reflect changes in the amount of communication an editor participates in.

3.7 Detecting changes

To detect changes in the activities and characteristics of reverted editors, we directly compare measures of the activities of the editors before and after a sampled edit. To establish a pre-revert state, we analyze the edits an editor made in the week preceding the revert. We then compute the change in activity from this baseline level for each of the following four weeks of that editor's activity. We call the difference in each week the *activity delta* (Δ). We choose a week as the time frame for analysis due to the observation that many editors have weekly periodic edit profiles, perhaps in response to work and other "real life" activities.

When measuring activities that vary widely from day to day for editors, we divide the activity delta by the standard deviation of the activity/day in the pre-revert week to produce a *controlled activity delta* (Δ/σ). This ensures

---

[6]To differentiate between the different types of edit activity, Wikipedia has several "namespaces" devoted to different types of content.
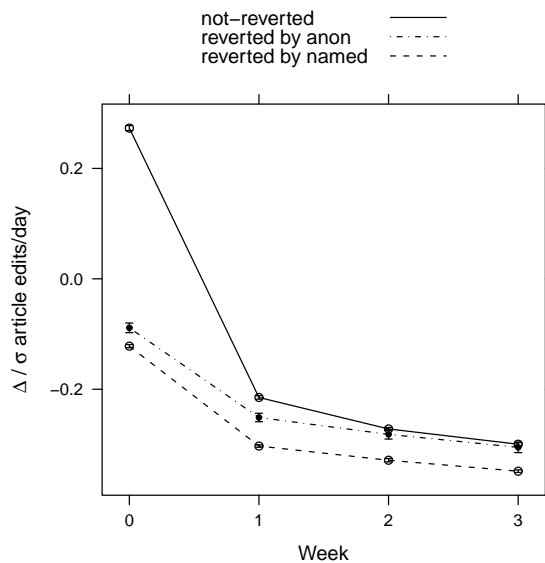

Figure 3: **Article activity** Δ/σ. For four weeks after a sampled edit, the change in article activity is reported. Reverted edits are split by whether the reverting editor was registered or anonymous. A control group of similar editors who were **not** reverted is included for comparison.

that all editors (regardless of overall amount of activity) are expressed in our results equally.

3.8 Notes on Figures and Tables

The plots in this paper include standard error bars. For many of the plots, the error bars may be too small to see due to the high number of observations sampled – but they are present. We use regressions many times throughout this paper. The most important regressions are summarized in Table 1, and discussed throughout Section 4. The results of other regressions are presented with coefficients and significance values where appropriate, but the regression tables are redacted to save space.

4. RESULTS AND DISCUSSION

4.1 RQ1: How does being reverted affect the quantity of editor work?

As discussed in Section 1, theory predicts that being reverted could demotivate editors, leading them to do less work in the future. To measure changes to the quantity

Table 1: Regressions over article activity Δ/σ, survival and PWR/day Δ/σ for four weeks after the sampled edits. Characteristics of the sampled edit's change to an article (words added, words removed, establishment of removed words) and whether it was a reverting edit itself or reverted back to by another editor are included in the regression to control for effects they could have on future work. For the discussion, statistical significance corresponds to $\alpha = .025$. Multicollinearity was checked for using correlation between explanatory variables. All correlation coefficients are below 0.5. PWR Δ/σ is scaled and logged to normalize it.

	Activity Δ/σ			Reverted activity Δ/σ			Survival (logistic)			log(PWR/day Δ/σ)		
	β	SE	$P(> \lvert z \rvert)$	β	SE	$P(> \lvert t \rvert)$	β	SE	$P(> \lvert t \rvert)$	β	SE	$P(> \lvert t \rvert)$
(Intercept)	.149	.004	< .001	-.149	.003	< .001	1.206	.006	< .001	.322	.013	< .001
Reverted(True)	-.292	.005	< .001				-.681	.008	< .001	.627	.021	< .001
Week since revert	-.178	.002	< .001							.021	.008	.013
Reverted editor tenure	.024	.002	< .001				1.594	.010	< .001	-.004	.010	.660
Words added	.062	.012	< .001				.004	.021	.844	.022	.019	.244
Words removed	-.003	.003	.309				-.017	.005	< .001	-.008	.009	.405
Est. of removed words	-.001	.003	.708				-.109	.006	< .001	.058	.014	< .001
Reverting(True)	.030	.012	.015				.326	.027	< .001	.399	.050	< .001
Reverted to(True)	-.037	.011	.001				.019	.021	.363	.022	.046	.633
Rvtd:Week since revert	.103	.003	< .001	-.074	.003	< .001				-.001	.012	.953
Rvtd:Reverted tenure	.043	.003	< .001	.064	.002	< .001	.373	.014	< .001	-.037	.014	.011
Rvtd:Reverting tenure				-.009	.002	< .001						
Rvtd:Words added	-.063	.012	< .001	-.000	.002	.872	-.047	.023	.041	-.035	.020	.084
Rvtd:Words removed	.002	.003	.746	-.002	.002	.253	.015	.006	.008	.031	.014	.029
Rvtd:Estab. of rm words	.003	.004	.367	.002	.002	.145	.045	.007	< .001	-.005	.016	.740
Rvtd:Reverting	.002	.014	.884	.026	.005	< .001	.334	.029	< .001	-.678	.057	< .001
Rvtd:Reverted to	.073	.013	< .001	.034	.006	< .001	.049	.025	.054	-.177	.057	.002
Rvtd:Reverting is anon				.012	.006	.032						

Table 2: Dependent variable characteristics. n for all dependent variables is $684,508$ after removing non-finite values.

	Reverted	Week	Rvtd tenure weeks	Words added	Words removed	Estab. of rm	Reverting	Reverted to
μ	0.50	1.00	21.01	41.28	73.70	30.95	0.09	0.07
σ	0.50	1.41	33.47	1082.78	2728.89	123.10	0.30	0.26
η	0.50	1.00	3.80	3.00	0.00	0.00	0.00	0.00

of work of an editor, we use revisions/day, the average number of revisions an editor completed per day. To check the validity of this metric, we also performed our analysis over words added to articles per day, and found no significant difference in the apparent effects.

To identify the effect of being reverted on the quantity of work perfomed by reverted editors, we performed a linear regression over controlled article activity delta[7]. The "Activity Δ/σ" column of Table 1 reports that being reverted is a significant, negative predictor($\beta = -.292, p < .001$) of controlled article activity delta. We also performed a regression over a controlled activity delta for words added per day Δ/σ and found that being reverted predicted a similar significant, negative effect ($\beta = -.934, p < .001$).

Figure 3 shows the controlled revisions/day delta of reverted editors split by whether the *reverting* editor was registered or anonymous beside the controlled revisions/day delta of not-reverted editors. In the first week after being reverted (week 0), not-reverted editors increase their activity about 0.3 standard deviations. By contrast, reverted editors decrease their activity by 0.1 standard deviations. By the second week after being reverted (week 1), the difference between reverted and not-reverted editor activity has decreased substantially. Although the average activity level of editors reverted by anonymous editors appears to converge by week 2, the activity of editors reverted by named editors

does not converge in the four observed weeks. Essentially, editors reverted by anonymous editors recover to the same average level of activity within a couple of weeks, but those reverted by named editors do not recover for at least one month (if ever).

Since the plot suggests that some editors do not recover in the month of time observed, we suspected that the long-term effect on reverted editors could have been due to editors being demotivated enough to leave Wikipedia entirely. To determine whether being reverted predicted a decreased survival rate in Wikipedia, we defined a simple metric for survival as a boolean variable that is `True` when an editor continues editing at least 8 weeks after the sampled edit and perform a logistic regression over it. The "Survival (logistic)" column of Table 1 reports that being reverted is a significant, negative predictor of survival ($\beta = -.681, p < .001$).

The regressions over article activity and survival have shown us that being reverted predicts both a decrease in activity and a reduction in the probability of survival. Can the reduction in activity be explained by the decreased percentage of surviving editors or do editors who will continue to edit decrease their activity as well? To answer this question, we performed another regression over the controlled article activity delta for only the surviving editors. Being reverted remained a significant, negative predictor ($\beta = -0.237, p < .001$). It appears that, even when editors will continue editing for at least two months, being reverted has a significant effect on the quantity of work they perform in the observed four weeks.

---

[7]To control for other factors that could confound the effects we observe in the regressions, we include editor tenure and characteristics of the reverted edit as independent variables in the regression.

To measure the amount of the difference between the reverted editors in Figure 3 that is independently due to the anonymous or named status of the reverting editor, we performed a regression. This regression included only the reverted sampled edits because those are the only ones with a reverting editor. Column "Reverted activity Δ/σ" in Table 1 shows that the reverting editor's anonymous status predicts a marginally significant, positive effect on future contributions($\beta = .012, p = .032$). There are two possible causes for this effect: a higher rate of contribution or a higher rate of survival among editors reverted by anonymous editors. To measure the independent effect of each of these two possible causes, we performed two additional regressions. We found that being reverted by an anonymous editor did not predict a higher rate of contribution for surviving editors($\beta = 0.011, p = .273$), but it did predict a higher rate of survival (logistic: $\beta = .153, p < .001$). This result suggests that the apparent demotivating effect of being reverted is significantly less severe when the reverting editor is anonymous and that the effect is largely due to an increased probability of survival.

Taken together, the results of this subsection suggest that being reverted both decreases the probability an editor will continue editing and decreases the motivation of those editors who do continue editing – at least temporarily.

The lesser demotivating effect in the case that the reverting editor is anonymous suggests that characteristics of the reverting editor could have an effect on the severity of the demotivation. Anonymous editors lack a persistent identity in Wikipedia and are likely to be perceived as outsiders with only a passing interest. In Wikipedia, allowing anonymous editors to contribute is controversial[8] and discrimination against anonymous editors has warranted community concern[9]. A possible explanation of the results is that reverted editors take the feedback of having their work discarded by anonymous editors less seriously.

4.2 RQ2: How does being reverted affect the quality of editor work?

Though editors reduce the quantity of their work in editing articles after being reverted, it is possible that reverts serve as an opportunity for feedback and learning. Here we explore the question of whether getting reverted tends to increase the quality of an editor's future work.

Quality. As described in Section 3.3, we first quantify the quality of editors contributions by measuring how likely they are to get reverted in the future. To ensure that the revert itself did not bias the results, that edit is dropped from the recent activity when establishing the pre-revert state for reverted editors. To check the validity of these results, we also performed our analysis over PWR/word, a word persistence based measure of quality, and found no significant difference in the apparent effects.

Figure 4 shows that editors who are reverted are less likely to get reverted in the future. Reverted editors immediately drop significantly in their probability of being reverted. The percentage drop of about 4% shown in the figure is substantial since it represents more than a third of the underlying
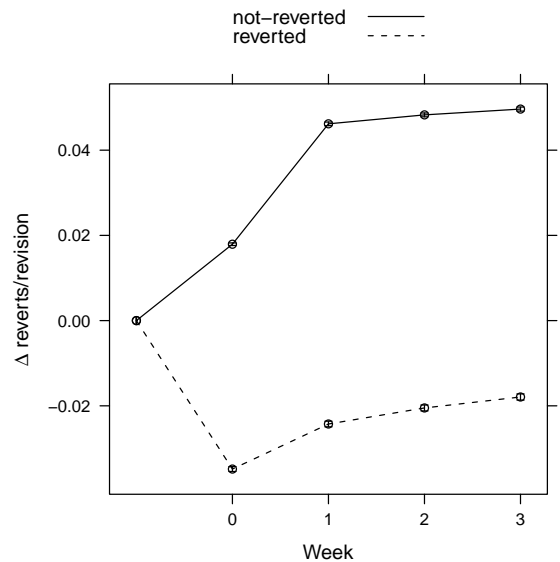
Figure 4: **Reverts per revision** Δ. For the four weeks immediately after a sampled edit, the change in quality of work as reverts/revision is reported. A control group of similar editors who were **not** reverted is included for comparison.

revert probability (11%). By contrast, not-reverted editors become more and more likely to be reverted each week.

It is interesting to note that a week after the sampled edit, both reverted and not-reverted editors make edits that are increasingly likely to be reverted each week of continued editing – and at a similar rate. One interpretation is that the being reverted is a learning experience for the reverted editor and acceptability of their work to other editors after being reverted represents an increase in quality. Another interpretation is that editors decrease their boldness in order to make their work more acceptable after being reverted, since edits that are less bold (eg. copy-edits and other minor changes) should be less likely to be rejected by other editors.

Boldness. To find out whether editors change the boldness of their contributions to articles, we measured two characteristics of the edits they made: words changed/revision and the establishment of removed words.

Figures 5a shows that reverted editors will change fewer words per revision and remove less established words after being reverted. The number of words changed per revision delta for reverted editors is consistently $75 - 100$ below the delta for not-reverted editors. Interestingly, the difference in the establishment of removed words only diverges a week after being reverted. These results suggests that being reverted could be a check on the boldness of the type of edits being made and reverted editors could be improving their revert percentage by playing it safe.

Productivity. Does the change in boldness explain the change in reverts/revision or are editors also increasing the quality of their work? In order to determine whether editors only become less bold after being reverted or whether they increase in quality of their work, we measure how their productivity changes based on PWR/day, a measure of both the quantity and quality their work described in Section 3.5.

Figure 6 plots the controlled PWR/day delta for reverted and not-reverted editors. Reverted editors increase in their productivity more than not-reverted editors immediately after being reverted and the difference stays rel-
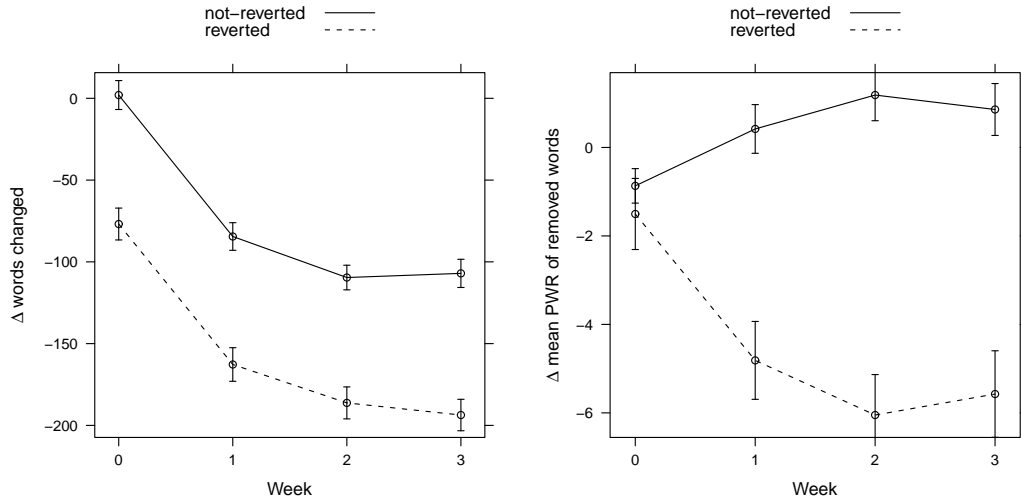
(a) **Words changed/revision** Δ.

(b) **Establishment of removed words** Δ.

Figure 5: For the four weeks immediately after a sampled edit, the change in the boldness of work is reported via two metrics: words changed/revision and establishment of removed words. A control group of similar editors who were **not** reverted is included for comparison.
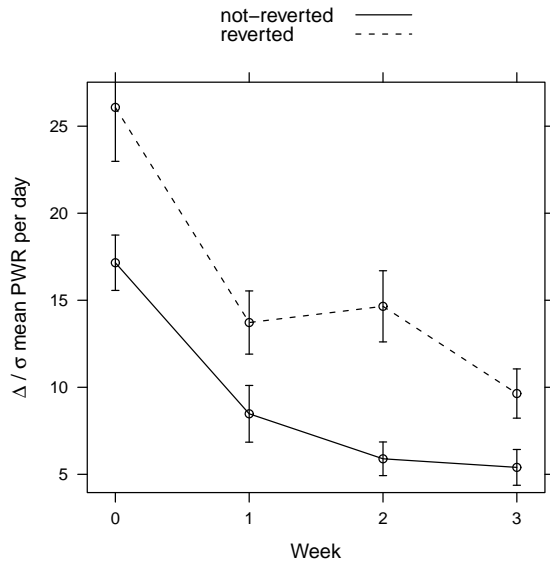

Figure 6: **PWR/day** Δ/σ. For the four weeks immediately after a sampled edit, the change in productivity is reported as the controlled PWR/day delta. A control group of similar editors who were **not** reverted is included for comparison.

atively consistent through the four observed weeks. The "log(PWR/day Δ/σ)" column in Table 1 confirms that being reverted is a significant, positive predictor of future productivity ($\beta = .627, p < .001$) and the interaction between being reverted and week has a small, insignificant effect ($\beta = -.001, p = .953$). The pre-revert PWR/day (not shown in Table 1 for formatting reasons) was included in the regression as a control for how productive an editor was before being reverted. The interaction between being reverted and recent productivity was a significant, negative predictor ($\beta = -.412, p < .001$). This effect suggests that editors who are producing high PWR/day before being reverted do not gain as much PWR/day after being reverted.

Although the results reported in this section show that reverted editors were more likely to leave Wikipedia and that the ones who stay will become less active and less bold in their work, this result shows that they will increase their quality, and therefore their productivity enough to more than make up for the difference. However, highly productive editors do not benefit in the same way that less productive editors do from being reverted. This result suggests that the learning effect from being reverted is only experienced by editors who are not already very productive to begin with.

4.3 RQ3: How does being reverted affect communication?

When editors reduce their effort toward editing Wikipedia articles, do they move that effort toward other Wikipedia activities, such as communication? To answer this question, we applied the same approach that we used in Section 4.1 for article activity to produce variance controlled activity deltas (Δ/σ) for Article_Talk and User_Talk. We limited our analysis to surviving editors since editors who do not continue work on Wikipedia are predisposed to not make edits to Article_Talk and User_Talk.

To determine whether being reverted was a significant, independent predictor of future communication, we performed two regressions over controlled activity deltas for of Arti-
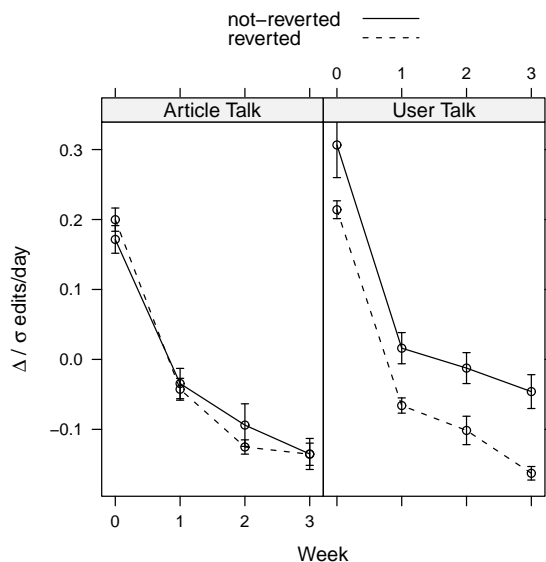
Figure 7: **Communication activity** Δ/σ - For the four weeks immediately after a sampled edit made by surviving editors (defined in Section 4.1), the change in Article_Talk and User_Talk communication activity is reported. A control group of editors with similarly distributed tenure who were **not** reverted is included for comparison.

cle_Talk and User_Talk Δ/σ. The regressions report a significant negative effect of being reverted on User_Talk activity ($\beta = -.079, p < .001$), but an insignificant effect of being reverted on Article_Talk activity ($\beta = -0.023, p = 0.663$).

It's interesting to note that reverted editor tenure was a powerful predictor of the change in Article_Talk activity ($\beta = .123, p < .001$). Since, reverted editor tenure is significantly different between the reverted and not-reverted subsamples(T-test: diff=76.29, $p < .001$), we controlled the not-reverted sample to be distributed similarly by tenure[10] to the reverted subsample before plotting values for Figure 7.

Figure 7 plots the controlled activity deltas for surviving editors after controlling for reverted editor tenure. For Article_Talk, the difference between reverted and not-reverted editors appears insignificant, but for User_Talk, the difference is substantial (0.1 standard deviations) and reverted editors did not recover to the communication activity levels of not-reverted editors in the four weeks observed.

Overall we have seen that when editors were reverted they reduced their personal communications, but did not reduce their communication over article content significantly. For Wikipedia, this result may be good news, because it suggests that the reverted editors are not being entirely demotivated from participating in Wikipedia.

4.4 RQ4: How does experience moderate the effects of reverts on contribution?

We now know that although being reverted reduces an editor's contributions, it also increases the quality of their work. To understand how the effects of being reverted are moderated by experience, we repeated the analysis from sections

---

[10]To control for tenure differences, we matched proportions in quantiles between the two subsamples.
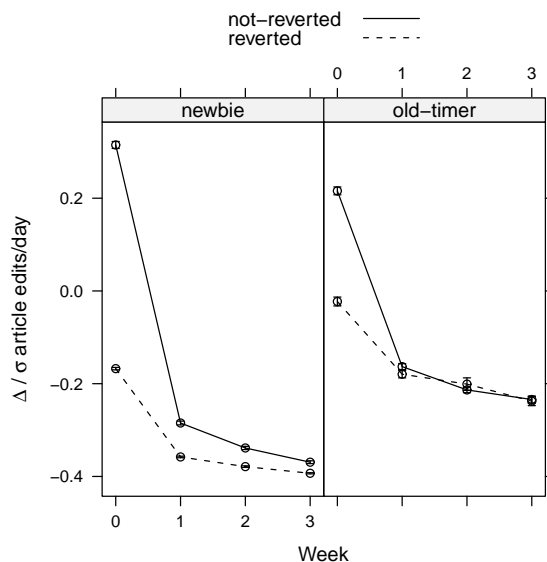

Figure 8: **Article activity** Δ/σ **by reverted editor tenure**. For the four weeks immediately after a sampled edit, the change in article activity is reported for newbies and old-timers. A control group of similar editors who were **not** reverted is included in each graph for comparison.

4.1 and 4.2 in the context of the tenure[11] of both reverted and reverting editors.

4.4.1 Reverted editor tenure

Activity and survival. Figure 8 represents activity changes for two interesting subsets of editors: *newbies* are editors with less than one month of tenure and *old-timers* are editors with more than one year of tenure.

After a revert, old-timer editors do experience a temporary reduction to their article activity, but they return to the level of activity of their not-reverted counterparts within two weeks of being reverted. For newbie editors, the difference in the activity delta is both stronger and longer-lasting. Reverted newbies take more than four weeks to return to the activity levels of not-reverted newbies.

The "Activity Δ/σ" column of Table 1 reports a significant, positive effect of reverted editor tenure ($\beta = .043, p < .001$) on the controlled activity delta when an editor is reverted. This positive effect suggests that the more experience an editor has in the system the less their article activity will decline after being reverted. The "Survival (logistic)" column in Table 1 also reports that reverted editor tenure is a significant, positive predictor of editor survival ($\beta = .373, p < .001$). In summary, newbies are less likely to continue editing after being reverted than old-timers and the ones who do continue to edit reduce their activity more than old-timers.

Quality and productivity. We now consider how the changes to quality and productivity seen after a revert are moderated by the experience of the reverted editor. The "log(PWR per day Δ/σ)" column of Table 1 shows a significant, negative effect on controlled PWR/day delta for reverted editor tenure ($\beta = -.037, p = .011$). That is, the more experience reverted editors have in Wikipedia,

---

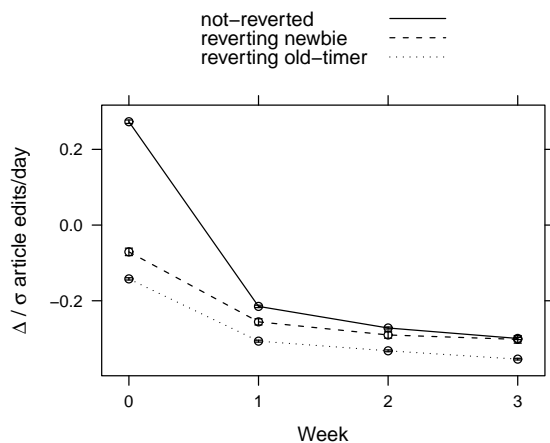[11]The amount of time since an editor's first edit.

Figure 9: **Article activity Δ/σ by reverting editor tenure**. For the four weeks immediately after a sampled edit, the change in article activity is reported. Reverted edits are split by whether the reverting editor was an newbie or old-timer. A control group of similar editors who were **not** reverted is included for comparison.

the lower the productivity boost they see after being reverted. This result is suprising given the *Activity and survival* analysis above, that more experienced editors see less of a drop in contribution after being reverted than less experienced editors. A plausible explanation is that more experienced editors increase the quality of their article contributions less. A linear regression over PWR/word delta, finds a marginally significant, negative effect of reverted editor tenure ($\beta = -1.145, p = .180$), confirming that the smaller quality increase for editors with more experience can be explained by the lower productivity boost they experience.

4.4.2 Reverting editor tenure

Activity and survival. We now examine the impact of the tenure of the editor who performs the revert. Figure 9 compares the controlled activity deltas of editors reverted by newbie and old-timer *reverting* editors. Here, we see the opposite effect of reverted editor tenure. While the activity of editors reverted by newbies will recover within two weeks, the activity of editors reverted by old-timers did not recover in the four observed weeks.

The "Reverted activity Δ/σ" column in Table 1 shows a significant, negative effect for reverting editor tenure on the controlled activity delta ($\beta = -.009, p < .001$). A logistic regression over the survival of reverted editors also showed a significant, negative effect for reverting editor tenure ($\beta = -.075, p < .001$). This suggests that the longer the reverting editor has been participating in Wikipedia, the lower the likelihood that the reverted editor will continue to make edits. Even for editors who will survive, more reverting editor experience means a larger reduction in the quantity of work the reverted editor will perform after being reverted.

This result suggests that editors with a high level of experience are not as effective in maintaining the motivation of the editors they revert than editors who are less familiar with wiki-work. Another interpretation is that old-timers have an enhanced ability to identify unconstructive editors and chase them away.

Quality and productivity. Linear regressions over our metrics for quality and productivity of reverted editors did not report significance for the tenure of the reverting editor. This suggests that quality and productivity changes in the reverted editor are not significantly affected by the experience of the reverting editor.

5. CONCLUSIONS

The key question driving this research is to understand what effects of reverts are on the contributions of editors. Since it is difficult to measure motivation directly, we instead measure indirect effects that might be caused by changes in motivation, including reducing the amount of contribution, or the ultimate reduction, withdrawal from Wikipedia.

We find that editors are more likely to withdraw from Wikipedia after being reverted and that the ones who stay decrease their quantity of work. These effects are especially strong for reverted newbies and when the reverting editor has more experience in Wikipedia. This finding is especially relevant since Wikipedia has an increasing problem with newcomer retention [21, 24].

On the other hand, editors who continue to do work in Wikipedia after being reverted increase the quality of their work. This effect is especially true for newbies and less productive editors. Being reverted appears to be a learning experience, helping the editors who need it most learn to be more effective Wikipedians. Under our measure of productivity (PWR/day), the net effect of reverts on Wikipedia is positive: on average, an editor who is reverted produces more persistent words per day – even if we include those editors who withdraw from Wikipedia in the calculation!

Editors also change their communication patterns after being reverted. Although we observed an overall decrease in the communication activities of reverted editors, editors do not appear to decrease their communication about article content after being reverted. The sustained article communication activity is a positive sign. Wikipedia's Bold, Revert, Discuss cycle encourages article discussion as a reaction to being reverted, and discussion should help reverted editors learn how to improve their work. However, editors did decrease their communications to other editors after being reverted. This reduction may be a sign of withdrawal risk: personal communications with others in a community can reinforce incentives to participate and reinforcement may be exactly what is needed for editors whose revisions were just reverted.

Our research suggests that overall reverting activity in Wikipedia is healthy and valuable, with the training effects dominating the demotivating effects. However, there are specific cases in which reverting activity might be managed better, to dampen the negative effects and amplify the positive. We offer the following advice to designers of social production communities like Wikipedia:

Support rebuffed users. We found that being reverted often precedes a reduction in participation. Furthermore, the reverted editors decrease their communication with other editors, at a time when they are vulnerable to leaving the community. Perhaps having other users reach out to them could help reinforce their connections to the community. Support should be personal since other research has suggested that impersonal socialization tactics can do more harm then help [4].

Encourage the learning effect. We found that being reverted predicts an increase to the quality and productivity

of an editor's article work. The negative feedback appears to be an opportunity for users to improve the quality of their participation. Viewing feedback like reverts as an opportunity to teach should be both encouraged and supported. While performing a revert, the reverting editor should be encouraged to provide clear feedback to help the reverted editor grow as a member of the community.

Focus on newcomers. We found that newcomers are particularly likely to decrease their contributions after they are reverted. We also saw some evidence that they can learn the most from being reverted. Newcomers should be reached out to actively to help them become socialized into Wikipedia. Furthermore, the more curmudgeonly old-timers should be kept away from newcomers until they have gained some experience in the system.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] B. T. Adler and L. de Alfaro. A content-driven reputation system for the Wikipedia. In *WWW '07*, pages 261–270, New York, NY, USA, 2007. ACM.

[2] U. Brandes and J. Lerner. Visual analysis of controversy in user-generated encyclopedias. *InfoVis*, 7:34–48, March 2008.

[3] A. Bruns and S. Humphreys. Wikis in teaching and assessment: The M/Cyclopedia project. In *WikiSym '05*, pages 25–32, New York, NY, USA, 2005. ACM.

[4] B. Choi, K. Alexander, R. E. Kraut, and J. M. Levine. Socialization tactics in Wikipedia and their effects. In *CSCW '10*, pages 107–116, New York, NY, USA, 2010. ACM.

[5] P. Dondio, S. Barrett, S. Weber, and J. Seigneur. Extracting trust from domain analysis: A case study on the Wikipedia project. In L. Yang, H. Jin, J. Ma, and T. Ungerer, editors, *Autonomic and Trusted Computing*, volume 4158, pages 362–373. Springer Berlin, 2006.

[6] M. D. Ekstrand and J. T. Riedl. rv you're dumb: Identifying discarded work in wiki article history. In *WikiSym '09*, pages 4:1–4:10, New York, NY, USA, 2009. ACM.

[7] A. Forte and A. Bruckman. From Wikipedia to the classroom: Exploring online publication and learning. In *ICLS '06*, pages 182–188. International Society of the Learning Sciences, 2006.

[8] R. S. Geiger and D. Ribes. The work of sustaining order in Wikipedia: The banning of a vandal. In *CSCW '10*, pages 117–126, New York, NY, USA, 2010. ACM.

[9] J. Grudin and E. S. Poole. Wikis at work: Success factors and challenges for sustainability of enterprise wikis. In *WikiSym '10*, pages 5:1–5:8, New York, NY, USA, 2010. ACM.

[10] A. Halfaker, A. Kittur, R. Kraut, and J. Riedl. A jury of your peers: Quality, Experience and Ownership in Wikipedia. In *WikiSym '09*, pages 1–10, New York, NY, USA, 2009. ACM.

[11] S. J. Karau and K. D. Williams. *Groups at work: Theory and research*, chapter Understanding Individual Motivation in Groups: The Collective Effort Model, pages 113–141. Lawrence Erlbaum Associates, Inc., 2001.

[12] A. Kittur and R. E. Kraut. Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. In *CSCW '08*. ACM, 2008.

[13] A. Kittur and R. E. Kraut. Beyond Wikipedia: Coordination and conflict in online production groups. In *CSCW '10*, pages 215–224, New York, NY, USA, 2010. ACM.

[14] A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi. He says, she says: Conflict and coordination in Wikipedia. In *CHI '07*, pages 453–462, New York, NY, USA, 2007. ACM.

[15] P. Magnus. Early response to false claims in Wikipedia. *First Monday*, 13(9), 2008.

[16] R. Priedhorsky, J. Chen, S. Lam, K. Panciera, L. Terveen, and J. Riedl. Creating, destroying, and restoring value in Wikipedia. In *GROUP'07*, Sanibel Island, FLorida, USA, 2007.

[17] C. Roth, D. Taraborelli, and N. Gilbert. Measuring wiki viability: an empirical assessment of the social dynamics of a large sample of wikis. In *WikiSym '08*, pages 27:1–27:5, New York, NY, USA, 2008. ACM.

[18] B. Stvilia, M. B. Twidale, and L. C. Smith. Information quality: Discussions in Wikipedia. 2005.

[19] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser. Assessing information quality of a community-based encyclopedia. ACM Press, 2005.

[20] B. Suh, E. Chi, B. Pendleton, and A. Kittur. Us vs. them: Understanding social dynamics in Wikipedia with revert graph visualizations. In *VAST '07*, pages 163 –170, Nov. 2007.

[21] B. Suh, G. Convertino, E. H. Chi, and P. Pirolli. The singularity is not near: Slowing growth of Wikipedia. In *WikiSym '09*, pages 1–10, New York, NY, USA, 2009. ACM.

[22] J. Thom-Santelli, D. R. Cosley, and G. Gay. What's mine is mine: Territoriality in collaborative authoring. In *CHI '09*, pages 1481–1484, New York, NY, USA, 2009. ACM.

[23] F. B. Viégas, M. Wattenberg, and K. Dave. Studying cooperation and conflict between authors with history flow visualizations. In *CHI'04*, pages 575–582. ACM, 2004.

[24] Wikipedia. Editor trends study. http://strategy.wikimedia.org/?oldid=80283, March 2011.

[25] H. Zeng, M. A. Alhossaini, R. Fikes, and D. L. McGuinness. Mining revision history to assess trustworthiness of article fragments. In *CollaborateCom '06*, pages 1 –10, 2006.

[26] X. Zhang and F. Zhu. Intrinsic motivation of open content contributors: The case of Wikipedia. *Workshop on Information Systems and Economics*, 2006.