Defense Mechanism or Socialization Tactic? Improving Wikipedia's Notifications to Rejected Contributors

R. Stuart Geiger[1], Aaron Halfaker[2], Maryana Pinchuk[3], Steven Walling[3]

[1] School of Information, University of California, Berkeley

[2]Grouplens Research, University of Minnesota

[3]Wikimedia Foundation

rsg@berkeley.edu; halfak@cs.umn.edu;{mpinchuk, swalling}@wikimedia.org

Abstract

Unlike traditional firms, open collaborative systems rely on volunteers to operate, and many communities struggle to maintain enough contributors to ensure the quality and quantity of content. However, Wikipedia has historically faced the exact opposite problem: too much participation, particularly from users who, knowingly or not, do not share the same norms as veteran Wikipedians. During its period of exponential growth, the Wikipedian community developed specialized socio technical defense mechanisms to protect itself from the negatives of massive participation: spam, vandalism, falsehoods, and other damage. Yet recently, Wikipedia has faced a number of high profile issues with recruiting and retaining new contributors.

In this paper, we first illustrate and describe the various de fense mechanisms at work in Wikipedia, which we hypothe size are inhibiting newcomer retention. Next, we present re sults from an experiment aimed at increasing both the quan tity and quality of editors by altering various elements of these defense mechanisms, specifically pre scripted warn ings and notifications that are sent to new editors upon re verting or rejecting contributions. Using regression models of new user activity, we show which tactics work best for different populations of users based on their motivations when joining Wikipedia. In particular, we found that per sonalized messages in which Wikipedians identified them selves in active voice and took direct responsibility for re jecting an editor's contributions were much more successful across a variety of outcome metrics than the current mes sages, which typically use an institutional and passive voice.

Introduction

A substantial amount of research has investigated how new members to organizations are recruited, retained, and socialized. A longstanding framework (Van Maanen & Schein, 1979) for understanding socialization tactics distinguishes between institutionalized and individualized tactics: institutionalized tactics are highly-structured and formalized, while individualized tactics are more ad-hoc and situated. Jones (1986) argues that institutionalized techniques work to reduce task uncertainty, which is critical for reducing the anxiety newcomers face. Gruman et al (2006) found that in firms with institutionalized tactics, such as explicit mentorship programs or classroom-type training sessions, newcomers are more likely to engage in proactive behaviors with regards to their tasks, most notably information seeking from established members. The uncertainty reduction theory or URT (Berger & Calabrese, 1975; Lester, 1987) has been widely used to explain why explicit guidence and task directives are critically important feedback mechanisms.

However, these studies have been done in traditional workplaces or formal mentorship programs in which employees are more or less bound to managers and firms. These findings and theories might not apply in voluntary peer production communities, due to the self-directed nature of participation and work practices. Kraut et al (2010) argue a similar point in a substantial review of the socialization literature as it pertains to on-line communities, noting that highly-institutionalized socialization practices are quite successful in bringing newcomers into an organization. They rightly point out that on-line communities like Wikipedia do not function like most institutions in terms of the formality of their various social mechanisms. Socialization in Wikipedia and open source software communities is highly individualized, with few formal mentoring programs or spaces dedicated to in-depth training. New members are thrown into the project and left to fend for themselves, often learning the various rules of the project by breaking them. In Wikipedia, the project's longstanding policies invite new users to "be bold" in editing and discourage Wikipedians from "biting the newbies" when they make mistakes.

Yet the longstanding distinction between institutionalized and individualized tactics does not speak to a core difference between how socialization currently occurs in Wikipedia as compared to other organizations. As we show in our first study, socialization of new members increasingly takes place through the project's highly-automated defense mechanisms, with new members' first interaction

with another Wikipedian taking the form of having one's contributions reverted and being sent a warning message. While the specific tactics used in these communications to new users vary, the overall *regime of socialization* in which these tactics are deployed has remained constant since its inception in 2006-07 -- the period in which the community experienced exponential growth in terms of both users and content. In our second study, we experimentally test different socialization tactics used by veterans when interacting with new users, and we do so within Wikipedia's dominant regime of socialization: the fast-paced 'revert-and-warn' mode that is made possible by the proliferation of bots and semi-automated tools.

Choi et al. (2010) performed a similar study regarding socialization tactics in Wikipedia, focusing on sending newly active editors different kinds of invitations to join a WikiProject – a group of editors dedicated to improving a specific topic area. They found that more personalized messages were significantly more effective at recruiting and retaining new members than boilerplate messages. Our study expands on this research, testing the hypothesis that personalized messages are better at not only recruiting and retaining new members, but increasing the amount of communication between new and veteran members.

Yet as much as we agree with the recommendations of Choi and colleages that veteran users ought to send messages to new users that are tailored to the specific kinds of activities that they performed – e.g. "thanks for fixing that typo!" – the socio-technical defense mechanisms that have come to overwhelmingly dominate interactions between veteran and new users makes this difficult. In a sense, Choi et al. similarly tested different socialization tactics, but did so within an entirely different regime of socialization compared to how new members are typically treated in Wikipedia. The participants in their study were editors who had been specifically identified by veteran contributors as potentially valuble contributors to a particular topic area -- typically based on the kinds of articles they had previously edited. This kind of highly-situated and contextual mode of identifying potentially valuble new contributors and bringing them into a small, focused, topic-specific group is a promising way of introducing less experienced users to others in the community. However, the tactics employed in this regime might function quite differently when newcomers are 'welcomed' by having their contributions reverted by a semi- or fully-automated tool, who then sends a boilerplate message telling them not to make any more 'unconstructive' edits.

Regimes of Socialization in Wikipedia

Defense Mechanism or Socialization Tactic?

Despite constant comparison between Wikipedia and other peer production systems – most notably, open source software development – Wikipedia faces a unique challenge in that it has an astounding lack of formal gatekeeping mechanisms. With a few exceptions for controversial articles and blocked IP addresses, almost any Internet user has the technical ability to edit almost every encyclopedia article in whatever manner they see fit. This model is a foundational principle of the Wikipedian community, and both communal wisdom and academic research holds that these lower barriers to entry do make the encyclopedia vulnerable to error and vandalism, but dramatically increase rates of participation (Wilkinson, 2008). Yet most peer production systems use a privileged contribution system, such as commit access in a software project. For example, any coder can submit a patch to the Linux kernel or Apache, but the contribution will only be made part of the codebase if it is approved. Such systems also have a stronger and implicit regimes of ownership that also serve to filter for quality. In a case study of Apache software projects, Mockus et al. found that developers who had created or maintained a specific portion of code extensively were given greater say in what changes would be made to it. (Mockus, Fielding, & Herbsleb, 2000)

Because of this, Wikipedia is generally assumed to not employ a quality control system since any contribution can be made and saved instantly; however, several studies have shown that the system employs effective mechanisms for dealing with damaging edits after they are made. Stvilia et al. argued that Wikipedia's open editing system constitutes an informal peer review that moderates the quality of articles. (Stvilia, Twidale, Smith, & Gasser, 2005) Halfaker et al. modeled the features of editors and the quality of their contributions and found that the majority of the rejection of contributions (through the mechanism of a revert) is can be predicted by independent measures of quality, though a substantial amount of rejection is related to editor bias. (Halfaker, Kittur, Kraut, & Riedl, 2009) Furthermore, these mechanisms are assumed to function quite well, given that numerous studies have demonstrated that damage in Wikipedia is, in general, reverted very quickly. (Priedhorsky et al., 2007; Viegas, Wattenberg, & Kushal, 2004)

Much of the work of patrolling new contributions, however, is performed by not by humans reading through full-length encyclopedia articles, but instead by highly-sophisticated, fully-autonomous bots, which are reverting vandalism, spam, and other malicious contributions. (Geiger, 2009) In addition, there are a number of highly-dedicated "vandal fighters" (Geiger & Ribes, 2010) who use specialized tools and scripts to monitor changes to articles in near real-time, reverting the editors who make less obviously malicious contributions. One notable aspect about these bots and tools is that in addition to supporting the high-speed reversion of unwanted contributions, they also send a pre-written message to the offending editor. Some of these messages are more generic and aim at generally improving the quality of the offending editor, while others are highly specific to the kind of damage or mistakes made. They collectively constitute what we term a regime of socialization, which is defined not by the particular strategies used to interact with newcomers but the overarching organizational context in which those tactics are deployed. Wikipedia's regime of socialization thus

differs from both traditional firms and smaller peer production communities in that socialization tactics are predominantly deployed by a small number of highly-active bots and tool-assisted humans who are typically focused on seeking out and reverting poor contributions.

~~Who~~ What is Interacting with New Editors?

Our decision to experimentally test tool-initiated messages to new users arose from the following descriptive analysis of how new editors were being socialized. Most notably, we found that vandal fighters and their tools have come to dominate not only editing, but also interpersonal communication. For readers unfamilar with interpersonal communication in Wikipedia, it is important to understand that the MediaWiki software upon which Wikipedia runs has no built-in private messaging system. Instead, each user has their own 'user page' that often serves as a profile, and Wikipedians leave messages for each other by editing the attached discussion or 'user talk' page. A prominent yellow banner stating "You have new messages" appears on every page if another user has changed an editor's user talk page since the last time the editor viewed it.

Using records from a backup copy of Wikipedia's mySQL database, we first analyzed all edits made to the encyclopedia project from 11 July to 11 September 2011. Out of the 922,773 total edits to the User talk namespace – which is the designated space for messages from users to users – 190,732 (or 20.7%) were from fully-automated bots; 64,364 (or 6.98%) were from a semi-automated counter-vandalism tool called Huggle; and 152,034 (or 16.5%) were from a semi-automated multi-purpose tool called Twinkle. In all, this means that 44%, almost half of all activity on User talk: pages, originates from highly-automated human and bot users – and many but not all of whom are engaged in counter-vandalism activity. It should be noted that even the semi-automated tools like Huggle and Twinkle are template-based: with the click of a button, a Wikipedian can instantly leave one of many pre-written message templates. While in this paper we only focus on the warnings and notifications to new users whose edits have been reverted, these tools also pre-script a variety of other organizationally-significant actions in Wikipedia.

In order to better understand the ways in which vandal fighters interact with new users, we then examined the first edit made to a registered or anonymous user's talk page. We classified 6.82 million initial contacts with 'new' users, out of a total of 30.44 million registered and anonymous users who have ever edited Wikipedia. Of course, some of these users were contacted within seconds of their first edit, while others were longstanding editors who had contributed for months, even years without receiving a single message on their user talk page. We found that the proportion of first messages sent by automated bots and semi-automated vandal fighting tools has grown substantially since 2006 (Figure 1). At present, at least 80% of initial contacts to users are made from a tool or bot, and it should be noted that this method actually underestimates these proportions, given that some users can choose to disable the specific traces that make such tools easily identifiable. From one perspective, such findings reveal the very necessary algorithmic response to the substantial number of new contributors who come to Wikipedia each day. From another perspective, it speaks to the primary orientation that

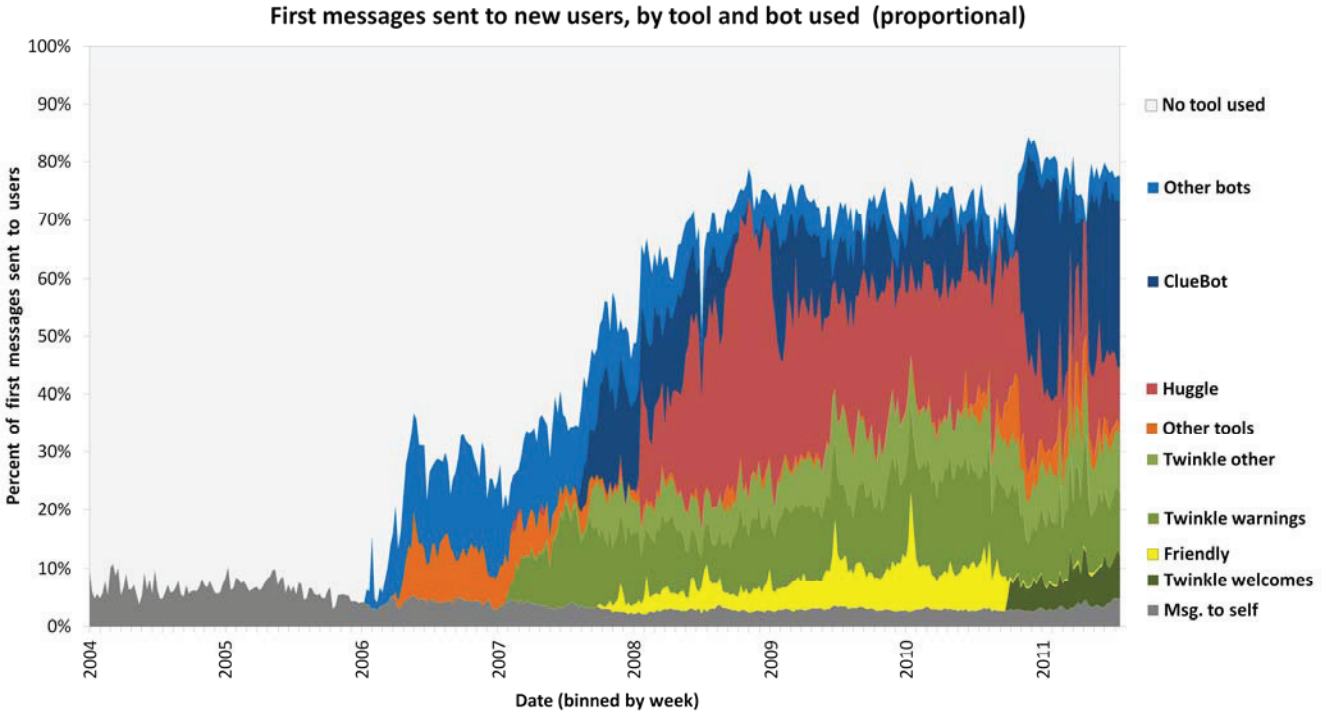

Figure 1: Bots and tools have come to dominate new users' first interactions with Wikipedians

Wikipedia, as a decentralized socio-technical system, has adopted towards its contributors, particularly non-registered and new contributors.

However, even if semi- and fully-automated messages pose problems for new users, simply disallowing them is not the optimal strategy. This would not only be a near-impossible decision to enforce on the Wikipedian community, but also fails to grasp the reasons why these tools have come to dominate post-hoc gatekeeping and new member socialization in Wikipedia. While the most common invocation of Wikipedia's inequality in contribution rates – that 10% of editors are responsible for a majority of the content (Ortega, Gonzalez-Barahona, & Robles, 2008) – is one of criticism, this inequality is an organizational problem as well. In order to review the contributions of millions of newcomers and socialize them to the Wikipedian community, highly-automated tools and tactics are critical. For example, in 2008, 211 distinct Wikipedians (0.019% of all editors) were responsible for sending over half of the first messages to project's new editors that year.

Experimentally Testing Socialization Tactics

Our second study tests the strategy of improving these boilerplate messages by infusing them with automated yet personalized introductions. Such a tactic requires no extra work on the part of the veterans who are interacting with new users, and only minimal effort on the part of those who maintain the tools and bots which are used in these kinds of activities.

The default templated messages Wikipedians send to new users upon making misakes deploy a highly institutionalized and passive voice, as well as institutionalized tactics. One widely-used message informs users that "one of your recent edits ... has been reverted, as it appears to be unconstructive." Task directives are highly prevalent in such messages, instructing new editors to do a number of tasks *except* editing encyclopedia articles again: these include reading an introduction or tutorial, reviewing the project's policies, or making edits in the Sandbox – a special page outside of the encyclopedia that is intended for testing.

We tested three different messages, which we referred to as personal with directives, personal without directives, and the default, which contained directives but was not personalized. We used three cases (Figure 2) in order to test both the effects of the personalization as well as the specific directives present, both of which we hypothesized were problematic in the default message. The directives which were included in two of the three messages are longstanding elements of Wikipedia's warning messages, and urge the user to edit the much-championed "Sandbox", provide "an informative edit summary", and finally read the community's introduction to editing. The personalized messages include a friendlier introduction that identifies the sender by their username and uses active voice to take responsibility for reverting the recipient's contribution. The personalized message invites the user to ask the sender questions and includes a link to the sender's user talk page.

This experimental design was designed to test four hypotheses regarding the design of socialization tactics:

H1. Both the personalized messages will increase the communication between recipients and senders, given the personal introduction and invitation to ask questions.

H2. The personalized message without directives will increase communication even more, given that the only actionable link was to the sender's talk page.

H3. Personalized messages will also increase the number

- default

  Welcome to Wikipedia. Although everyone is welcome to contribute to Wikipedia, at least one of your recent edits, such as the edit you made to Science has been reverted, as it appears to be unconstructive. Use the sandbox for testing; if you believe the edit was constructive, ensure that you provide an informative edit summary. You may also wish to read the introduction to editing. Thank you. EpochFail(talk|work) 14:44, 21 September 2011 (UTC)

- personal with directives

  Hello, and welcome to Wikipedia! I edit Wikipedia too, under the username EpochFail. I noticed that one of your recent edits, such as the one you made to Science with this edit, appeared to be unconstructive, and I've reverted it. In the future, please use the sandbox for testing and be sure to provide an informative edit summary. You may also wish to read the introduction to editing. Please feel free to ask me questions about editing Wikipedia (or anything else) on my talk page. EpochFail(talk|work) 14:44, 21 September 2011 (UTC)

- personal without directives

  Hello and welcome! I edit Wikipedia too, under the username EpochFail. Wikipedia is written by people like you and me, so thank you for taking the time to participate. I wanted to let you know that I undid one of your recent contributions, such as the one you made to Science with this edit, because it didn't appear constructive to me. If you think I made a mistake, or if you have any questions about editing, you can leave me a message on my talk page. Thanks! EpochFail(talk|work) 14:44, 21 September 2011 (UTC)

Figure 2: The three messages tested in the experiment

of future edits to articles as well as overall retention.

H4. Users who received more personalized messages will be more likely to be warned again, as the less-institutional voice could serve as less of a deterrent.

We made use of the existing user warning infrastructure in order to experimentally test new warnings. With the assistance of a number of Wikipedians actively involved with the development and use of these tools, we modified the templates to randomly insert one of 3 warnings. Because of the way in which tools like Huggle and Twinkle use templates, modifying the template at the server-level automatically changes the behavior of all clients. Specialized hidden identifiers (known as "z-templates") inserted in each type of message enabled us to track which users received which messages, and were designed such that only users who received a randomized message were tracked. In order to determine when editors actually read their messages (as opposed to relying on a potentially-faulty metric of when the message was posted), we made use of a specialized table in the mySQL database that kept track of when editors visited their own user talk pages. With assistance from the Wikimedia Foundation, we were able to query this table in near real-time to determine which users visited this page after the experimental messages were left. The experiment was live for approximately two weeks, between September 25th and October 10th, 2011. Of note for future research on the effectiveness of the messaging system is that out of the 4,512 users who received our message, only 2,451 actually read the message (that is, viewed their user talk page) within one month of receiving it. Only the users who did read their message were included in our analysis.

To identify the effects of the experimental messages on newcomers reverted by veteran Wikipedians, we employ logistic regressions (for boolean outcomes – e.g. continues to edit) and linear regressions (for scalar outcomes – e.g. # of edits to encyclopedia articles). Regression models are useful since they allow us to explicitly control for random environmental features and characteristics of the reverted user while providing a mechanism for identifing relationships between these features/characteristics and our experimental variables in the form of interaction terms. When the regression table reports a statistically significant effect for a variable, that means the variable has a significant effect independent of the other parameters of the model. In the case of our experimental variables, we interpret such effects as a causal relationship.

Many of our outcome metrics measure activity within a 72 hour window. We were wary of extending this window further given that the majority of our editors were unregistered and therefore only identified by their IP address. Because an IP address can be dynamic, we felt that 72 hours was the maximium time in which we could confidently be assured that the editor making contributions was the same individual who received and read the experimental message. We also controlled for both unregistered editors and unregistered editors from institutions or other shared networks, as explained below.

Predictor variables:

unregistered: MediaWiki tracks this field, which is 0 if the user has an account with a user name and 1 if the user is identifiable only through their IP address.

from shared IP: Wikipedians place specialized templates on the user talk pages of IP addresses that are shared among many users, such as libraries and corporations. The talk pages of all editors were scanned to see if these templates were present, recording a 1 if they were found and a 0 if they were not.

# of prior article edits: The number of edits the account made to the main namespace – also called namespace 0, where encyclopedia articles reside – 72 hours before the message was read. Determined by querying the user's contribution history and filtering by namespace.

# of prior talk edits: The number of edits the account made to the article or user talk namespace – also called namespaces 1 and 3, where discussion about individual encyclopedia articles and user-to-user communication takes place – 72 hours before the message was read.

warning was first message: Whether or not the message left was the first edit to the user's talk page, determined by retrieving the number of previous revisions from the database.

Results

Contributions to encyclopedia articles: For edits to the article namespace, that is, edits to encyclopedia articles, the most significant predictor was the number of previous edits made to articles. In addition, the personalized message with directives had an independent, negative effect on the number of future edits that recipients made to encyclopedia articles. However, this effect was negated for users who had previously made more edits to both encyclopedia articles and the user talk namespace. Such an independent effect was not found for the personalized message without directives, but there was a similar interaction such that users with more edits to encyclopedia articles prior to receiving the message were likely to make more edits to articles afterwards, compared to the default. Of marginal significance in this model was that users were less likely to make edits if they were anonymous, and less likely to make edits if the warning was their first message.

Discussion about encyclopedia articles: For edits to the article talk namespace, the designated place for Wikipedians to discuss specific encyclopedia articles, the number of previous edits to talk namespaces was the strongest predictor of future edits. In addition, prior edits to encyclopedia articles had a small but significant positive effect. With the messages, one interesting effect was that users who received both personalized messages were less likely to contribute to the article talk namespace compared to the default, but only if they had previously made edits to the talk namespaces. However, for users with more edits to the talk namespaces, the message with directives increased communication in the article talk namespace almost twice the amount as the personalized message without directives.

Regression Tables

Bold indicates statistical significance at a p-value below 0.05, italics indicate marginal signifiance at a p-value below 0.1.

	Future contributions to encyclopedia articles:			**Future user-to-user communication:**			**Contacting the reverting user:**		
	$R^2 = 0.13$			**$R^2 = 0.11$**			**$R^2 = 0.04$**		
	coef	**error**	**p-val**	**coef**	**error**	**p-val**	**coef**	**error**	**p-val**
(Intercept)	**1.245**	**0.273**	**< .001**	**0.258**	**0.076**	**< .001**	0.014	0.036	0.694
personal-directives	**-1.047**	**0.41**	**0.011**	*-0.191*	*0.114*	*0.093*	0.047	0.054	0.381
personal-nodirectives	0.454	0.416	0.275	**0.312**	**0.116**	**0.007**	**0.278**	**0.055**	**< .001**
warning was first message	*-0.312*	*0.188*	*0.097*	-0.034	0.052	0.519	0.018	0.025	0.474
unregistered	*-0.492*	*0.258*	*0.057*	**-0.174**	**0.072**	**0.015**	-0.004	0.034	0.916
from shared IP	0.084	0.255	0.742	-0.105	0.071	0.14	-0.01	0.034	0.768
# of prior article edits	**0.079**	**0.023**	**< .001**	**0.014**	**0.006**	**0.024**	-0.001	0.003	0.8
# of prior talk edits	-0.14	0.101	0.167	-0.001	0.028	0.965	*0.024*	*0.013*	*0.073*
personal-directives interactions :									
warning was first message	0.324	0.27	0.231	0.048	0.075	0.52	0.032	0.035	0.368
unregistered	0.293	0.382	0.443	0.091	0.106	0.389	-0.061	0.05	0.221
from shared IP	0.202	0.359	0.574	0.056	0.1	0.578	0.013	0.047	0.79
# of prior article edits	**0.301**	**0.035**	**< .001**	0.013	0.01	0.177	-0.003	0.005	0.576
# of prior talk edits	**0.749**	**0.195**	**< .001**	**0.452**	**0.054**	**< .001**	**0.081**	**0.026**	**0.002**
personal-nodirectives interactions:									
warning was first message	-0.384	0.271	0.156	-0.018	0.075	0.812	0.003	0.036	0.938
unregistered	-0.228	0.389	0.558	**-0.297**	**0.108**	**0.006**	**-0.269**	**0.051**	**< .001**
from shared IP	0.133	0.356	0.71	0.015	0.099	0.882	-0.023	0.047	0.618
# of prior article edits	**0.103**	**0.033**	**0.002**	-0.013	0.009	0.161	0.005	0.004	0.235
# of prior talk edits	0.092	0.211	0.662	**0.478**	**0.059**	**< .001**	0.004	0.028	0.873

	Future discussions about encyclopedia articles:			**Short-term retention:**			**Future malicious activity:**		
	$R^2 = 0.36$			**AIC = 3090.43**			**AIC = 2738.63**		
	coef	**error**	**p-val**	**coef**	**error**	**p-val**	**coef**	**error**	**p-val**
(Intercept)	-0.023	0.043	0.598	**0.564**	**0.26**	**0.03**	**-0.642**	**0.253**	**0.011**
personal-directives	-0.033	0.065	0.609	**-0.826**	**0.387**	**0.033**	-0.185	0.385	0.63
personal-nodirectives	0.05	0.066	0.451	0.088	0.393	0.823	0.192	0.381	0.614
warning was first message	-0.032	0.03	0.29	**-0.966**	**0.17**	**< .001**	0.002	0.194	0.993
unregistered	0.054	0.041	0.189	**-0.742**	**0.236**	**0.002**	**-0.775**	**0.237**	**0.001**
from shared IP	-0.058	0.041	0.151	0.367	0.23	0.111	*0.433*	*0.251*	*0.085*
# of prior article edits	**0.008**	**0.004**	**0.031**	**0.161**	**0.038**	**< .001**	0.035	0.022	0.112
# of prior talk edits	**0.506**	**0.016**	**< .001**	0.506	0.329	0.123	0.082	0.147	0.578
personal-directives interactions :									
warning was first message	0.031	0.043	0.476	*0.452*	*0.244*	*0.063*	-0.006	0.276	0.983
unregistered	0.003	0.06	0.959	0.369	0.347	0.288	0.204	0.351	0.562
from shared IP	0.021	0.057	0.718	0.374	0.324	0.248	0.09	0.348	0.797
# of prior article edits	0	0.006	0.962	0.053	0.059	0.374	0.053	0.037	0.156
# of prior talk edits	**-0.227**	**0.031**	**< .001**	-0.048	0.422	0.909	0.134	0.215	0.535
personal-nodirectives interactions:									
warning was first message	0.02	0.043	0.645	**0.555**	**0.243**	**0.022**	-0.009	0.275	0.975
unregistered	-0.06	0.062	0.327	-0.315	0.359	0.379	-0.088	0.352	0.802
from shared IP	0.042	0.057	0.462	0.489	0.324	0.131	0.341	0.342	0.319
# of prior article edits	-0.008	0.005	0.146	-0.014	0.053	0.795	0.016	0.033	0.637
# of prior talk edits	**-0.417**	**0.033**	**< .001**	-0.225	0.39	0.564	-0.084	0.237	0.722

Tables 1 and 2: Outcome metrics for regressions

This suggests that the personalized messages are only more effective than non-personalized messages at increasing this form of communication for recipients who already have experience communicating directly with other users.

Future user-to-user communication: Examining edits to the user talk namespace, the personalized message without directives was independently more effective at increasing one-to-one communication between users. However, this independent effect was almost entirely negated for anonymous users, indicating that a lack of directives only increases user-to-user communication for registered users. For users who had previously made edits to the user talk namespace, both personalized messages were even more effective (with similar-sized effects) than the non-personalized message. There was also a marginally significant effect ($p = .093$) suggesting that the personal message with directives has an independent negative effect on the amount of future user talk edits. In short, both personalized messages increase communication in talk namespaces, but the default message with directives is better at increasing communication in the article talk namespace while the personal messages – particularly the one without directives – is more effective at increasing communication in the user talk namespace.

Contacting the reverting user: Looking specifically at whether the recipient contacted the user who reverted their edits and left them the message within 72 hours of reading the original message, the personalized message without directives was independently the largest predictor of such an outcome. However, we see a similar, negative effect for anonymous editors who receive the personal-nodirectives message, which suggests that the benefit is primarily observed for registered editors. Interestingly, the number of prior edits to the user talk namespace was only a marginally significant independent predictor of whether the recipient contacted the sender. Only for the personalized message with directives did prior one-to-one communication have a positive interaction with this outcome metric.

Short-term retention: Examining short-term retention, one of the largest and most important predictors was whether or not the message received was the first message that user had ever received. Furthermore, the personal message without directives did not have a statistically significant effect on retention compared to the default, unless it was the first message received. If it was the first message received, the overall effect was still negative, but the personalized message without directives was more effective compared both to the default and the personalized message with directives. In contrast, the personalized message with directives had an independently negative effect on retention, which was somewhat lessened if the message was the first message received. A user editing without an account (i.e. an anonymous user) was also independently a strong negative predictor of retention.

Future malicious activity: We examined whether or not the user received additional warnings within 72 hours of reading the message, and the only significant predictor was whether or not they were registered. Anonymous users were less likely to receive future warnings, a marginally significant effect ($p = 0.085$) which was lessened but not negated for users editing from a shared IP address. None of the messages sent had independent or interacting effects on this outcome metric.

Editing the sandbox: Finally, we examined if each new user made an edit to the sandbox -- the designated space for test edits. The sandbox is prominently mentioned in both of the messages with directives, and new users are strongly encouraged to use the sandbox instead of the article space. We found that only 16 users out of our entire pool of 2,445 users edited the sandbox after receiving a message. 1.08% of editors who received the default-directives message edited the sandbox, compared to 0.62% of users who received personal-directives and 0.25% of users who received personal-nodirectives. These numbers were too low to find statistical significance between the test groups, but their overall rate indicates the ineffectiveness of that particular directive.

Limitations

The three messages which were experimentally tested were designed based on existing warnings, and in order to limit the number of experimental cases and maximize our sample size, each of the messages contains a number of different links and phrases. However, this means we were not able to independently test the effect of directives compared to personalization, and a study that included a non-personalized message without directives could provide more explanation of this phenomenon. In the present study, it is also difficult to know which of the elements in the three messages produced the effects observed. For example, each of the three directives removed may have an independent effect on new users. A future study would be able to understand which directives are helpful and which are not: the sandbox does not appear to be a helpful directive, but we do not know what effect the introduction to editing has outside of simply being a directive. Similarly, the personalized messages use a number of strateiges to humanize the message, and it is unclear if the invitation to contact the editor has an effect independent from the "I edit Wikipedia too" opening. Finally, the personalized message without directives differs from the personalized message with directives not only in a lack of tasks and instructions, but also adds the phrase "Wikipedia is written by people like you and me." It is therefore possible that the effects which we attribute to a lack of directives could alternatively have resulted from a sense of group solidarity instilled by such a phrase.

We also remain wary of generalizing our findings to socialization tactics in collaborative communities at large, considering that we did not experimentally test the entire regime of socialization in Wikipedia – only the presentation of one of the first aspects of socialization was tested. However, future research may place new users into cohorts and longituinally test an enitre suite of institutional vs. individualized tactics.

Conclusion and Future Research

In addition to testing socialization tactics, this paper shows the necessity of taking into account the regimes of socialization in which those tactics are deployed. Given the broader context in which users are recruited and move from peripheral to full participants, different tactics may have wildly different effects. Messages with task directives were shown in our study to inhibit both communication and further contributions compared to those without directives. While this seems to contradict longstanding research from a number of fields, we explain this phenomeon by refernce to the way in which initial contact with new members predominantly takes place by reverting a user's contributions and sending them a warning message.

These findings also demonstrate the potential for making substantial improvement in the area of new user retention with relatively small changes to the messages which are sent to newcomers. Personalization was shown to be quite effective in increasing the amount of communication new users performed, while directives did not appear to have as many positive effects. However, many of these positive effects only apply for registered users or users who have already shown some amount of encyclopedia editing or interpersonal communication, indicating that they are best suited not as welcome messages, but as socialization tactics for new users who have already shown themselves to be somewhat active. Finally, even the messages without directives were not shown to have adverse effects – a longstanding concern among Wikipedian editors who work tirelessly to patrol the encyclopedia against error, spam, and vandalism.

Based on these findings, we aim to further test which elements of these messages work and do not work across a variety of outcome metrics. This is because this study does not cleanly illustrate, for example, which directives have which effects, or whether using active voice is more important than identifying one's own username. In future studies, we also aim to experimentally test not only the default warning message used by Wikipedians to notify users that their contributions were rejected, but a wide variety of templated messages. These include welcome templates that simply introduce an editor to Wikipedia, informational templates that alert a user to various events in Wikipedia, notifications that their articles or images are nominated for deletion, and more. With a number of these templates, the target recipients are registered users and not anonymous contributors, which is helpful in tracking long-term retention. In future studies, we also aim to identify with more granularity the different potential populations of users who are receiving such messages, given that certain individuals are likely to react differently to different kinds of rhetorical techniques.

Acknowledgments

This work was funded by NSF grant IIS 09-68483 and the Wikimedia Foundation. The members of WikiProject User Warnings, who assisted us in running these experiments and provided critical feedback. We would also like to thank Adam Shorland, Ryan Faulkner, Diederik van Liere, Melanie Kill, Jonathan Morgan, and Dario Taraborelli for their assistance in this research project. We would like to especially thank the reviewers, whose comments assisted us in substantially improving this paper.

References

Berger, C. R., & Calabrese, R. J. 1975. Some Explorations in Initial Interaction and Beyond: Toward a Developmental Theory of Interpersonal Communication. *Human Communication Re search*, 1(2): 99 112.

Choi, B., Alexander, K., Kraut, R. E., & Levine, J. M. 2010. So cialization Tactics in Wikipedia and Their Effects. In *Proc CSCW 2010.*

Geiger, R.S. 2009. The social roles of bots and assisted editing programs. In *Proc Wikisym 2009.*

Geiger, R.S., & Ribes, D. 2010. The work of sustaining order in wikipedia: the banning of a vandal. In *Proc CSCW 2010.*

Gruman, J. A., Saks, A. M., & Zweig, D. I. 2006. Organizational socialization tactics and newcomer proactive behaviors: An inte grative study. *Journal of Vocational Behavior*, 69(1): 90 104.

Halfaker, A., Kittur, A., Kraut, R., & Riedl, John. 2009. A Jury of Your Peers: Quality, Experience and Ownership in Wikipedia. In *Proc Wikisym 2009.*

Jones, G. R. 1986. Socialization Tactics, Self Efficacy, and New comers' Adjustments to Organizations. *Academy of Management Journal*, 29(2): 262 279.

Kraut, R., Burke, M., & Riedl, John. 2010. Dealing with New comers. In R. E. Kraut & P. Resnick (Eds.), *Evidencebased Social Design Mining the Social Sciences to Build Online Communities*: 1 42. MIT Press.

Lester, R. E. 1987. Organizational culture, uncertainty reduction and the socialization of new organizational members. In S. Thomas (Ed.), *Culture and communication: Methodology, behav ior, artifacts and institutions*: 105 113. Norwood, NJ: Ablex.

Van Maanen, J., & Schein, E. H. 1979. Toward a theory of organ izational socialization. *Research In Organizational Behavior*, 1(1): 209 264.

Mockus, A., Fielding, R. T., & Herbsleb, J. 2000. A case study of open source software development: the Apache server. In *Proc ICSE 2000.*

Ortega, F., Gonzalez Barahona, J., & Robles, G. 2008. On the Inequality of Contributions to Wikipedia. In *Proc HICSS 2008.*

Priedhorsky, R., Riedl, J., Lam, S. T. K., Panciera, K., Chen, J., & Terveen, L. 2007. Creating, destroying, and restoring value in wikipedia. In *Proc GROUP 2005.*

Stvilia, B., Twidale, M., Smith, L., & Gasser, L. 2005. Assessing information quality of a community based encyclopedia. In *Proc ICIQ 2005.*

Viegas, F., Wattenberg, M., & Kushal, D. 2004. Studying Coop eration and Conflict between Authors with history flow Visuali zations. In *Proc CHI 2004*.

Wilkinson, D. M. 2008. Strong regularities in online peer produc tion. In *Proc EC 2008.*