Effectiveness of Shared Leadership in Online Communities

Haiyi Zhu, Robert Kraut, Aniket Kittur
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA15213
{haiyiz, robert.kraut, nkittur}@cs.cmu.edu

ABSTRACT
Traditional research on leadership in online communities has consistently focused on the small set of people occupying leadership roles. In this paper, we use a model of shared leadership, which posits that leadership behaviors come from members at all levels, not simply from people in high-level leadership positions. Although every member can exhibit some leadership behavior, different types of leadership behavior performed by different types of leaders may not be equally effective. This paper investigates how distinct types of leadership behaviors (transactional, aversive, directive and person-focused) and the legitimacy of the people who deliver them (people in formal leadership positions or not) influence the contributions that other participants make in the context of Wikipedia. After using propensity score matching to control for potential pre-existing differences among those who were and were not targets of leadership behaviors, we found that 1) leadership behaviors performed by members at all levels significantly influenced other members' motivation; 2) transactional leadership and person-focused leadership were effective in motivating others to contribute more, whereas aversive leadership decreased other contributors' motivations; and 3) legitimate leaders were in general more influential than regular peer leaders. We discuss the theoretical and practical implication of our work.

Author Keywords
Shared leadership, online communities, Wikipedia, motivation

ACM Classification Keywords
H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work, Web-based interaction; K.4.3 [Computers and Society]: Organizational Impacts – Computer supported collaborative work.

CSCW 2012, February 11–15, 2012, Seattle, Washington, USA.

INTRODUCTION
Who are the leaders in online communities? Previous researchers tend to use traditional vertical leadership models [2,15] to explain leadership in online communities. They suggest that the leadership role is a specialized one. People who are appointed or elected to perform this role are designated as "leaders" [29, 6, 20]. According to this view, some of the responsibilities and functions associated with leadership cannot be shared too widely without jeopardizing the effectiveness of the group.

In contrast, we suggest a shared leadership framework to explain leadership in online communities. The shared leadership framework was originally proposed by researchers investigating offline leaderless groups such as self-managing teams, volunteer organizations and employee-managed companies [24, 23, 30]. They argue that leadership—involving persuading and influencing other people to pursue a common goal—emanates from members at all levels, not simply from elites in formal leadership roles. Any member of the group can exhibit some level of leadership at any time, and there is no clear distinction between leaders and followers. Members mutually influence each other on what tasks are to be done, how tasks should be done, and the way they relate to each other. Rather than leadership being invested in specialized roles, leadership is viewed as a shared influence process.

Although every member can enact some level of leadership behaviors, the shared leadership model does not assume the same effectiveness of leadership behaviors across individuals. This paper investigates how distinct types of leadership behaviors and the legitimacy of the people who deliver them (people in formal leadership positions or not) influence their effectiveness. Here, we operationalize the effectiveness of leadership in terms of the extent to which those exhibiting leadership behavior can motivate others to invest effort in the community and contribution to it.

In this article, we first review prior theorizing about types of leadership behaviors and leader legitimacy. Based on path-goal theory, we explain the process by which leaders influence others' motivation. We offer several hypotheses regarding the effectiveness of different types of leadership behavior and leaders. Subsequently, we describe our measurement and methods, present our results, and discuss research and design implications.

TYPES OF LEADERSHIP BEHAVIOR

Researchers investigating traditional vertical leadership have identified a range of effective leadership behaviors [2, 3, 30]. In a shared leadership context, these strategies continue to be relevant [23]. Rather than prejudge results with terms like "leader" and "follower", in the following sections we use the term "influencer" to indicate the influence exercisers and "targets" to indicate those they are trying to influence.

The distinction between task-based leadership behaviors (those dealing with task accomplishment) and person-based (those facilitating team interaction and development) is common in nearly every taxonomy of leadership behaviors. Similar dichotomies include initiating structure versus consideration in Ohio State's program on leadership research [12], task-oriented versus relationship-oriented behaviors in the University of Michigan research program [17], and task-focused versus person-focused behaviors in Burke et al's recent paper reviewing leadership behaviors [3]. We follow this general classification and then specifically differentiate three subcategories of task-based leadership behaviors, using Pearce and Sims's classification [23].

Task-based leadership

Transactional leadership

With transactional leadership, leadership behavior is considered a transaction or exchange between the influencer and the target. Transactional influencers provide praise and rewards or withhold punishment from targets who comply with role expectations. Representative transactional leadership behaviors include (1) providing personal rewards, (2) providing material rewards, (3) managing by exception (active), (4) managing by exception (passive) [23]. Sample questionnaire items measuring this type of leadership include 1) "X will recommend that I am compensated well if I perform well"; 2) "X gives me positive feedback or special recognition when I perform well"; 3) "X tracks mistakes"; 4) "X delays taking action until problems become serious".

Aversive leadership

In contrast to transactional leadership, aversive leadership uses intimidation and reprimands to decrease undesired behaviors from targets [23]. Sample questionnaire items measuring aversive leadership include 1) "X tries to influence me through threat and intimidation" and 2) "X lets me know about it when I perform poorly".

Directive leadership

Directive leadership emphasizes the need to provide direction to targets and specify their roles and responsibilities. Directive behaviors include 1) issuing instructions and commands and 2) assignment goals [23]. Sample questionnaire items measuring directive leadership include 1) "when it comes to work, X gives me instructions on how to carry it out" and 2) "X establishes the goals for my work".

Person-based leadership

In contrast to task-focused leadership behaviors, which directly focus on task accomplishment, person-based leadership behaviors emphasize the target as a person and personal relationships between the influencer and the target [3].

Consideration was first proposed as a type of person-based leadership behavior in 1950s in the Ohio State leadership research program [12]. Consideration is the degree to which an influencer acts in a friendly and supportive manner, showing concern for targets, helping them to develop, supporting group cohesion, and maintaining close social relationship with them [30]. In general, dyadic relationships characterized by consideration reflect two-way open communication, mutual respect and trust, and an emphasis on satisfying employees' needs. The concept of relationship-oriented leadership in the University of Michigan research on leadership [17] is similar.

Later, researchers developed and elaborated the concept of person-based leadership by proposing ideas such as transformational leadership, which highlights encouragement, inspiration and intellectual stimulation [4, 2, 23], and empowering leadership, which focuses on self-management skills and team work [23].

In general, person-focused leadership behaviors are friendly and supportive, aimed at maintaining close social relationships supporting group cohesion, and developing subordinates' self-confidence and skills.

LEADER LEGITIMACY

In addition to types of leadership behavior, our analyses examine the role of the influencer in the organization. Legitimate leaders are those who occupy formal leadership positions in an organization, volunteer community or other social system. The legitimacy stems from the selection process, whether they are appointed by supervisors, elected by the membership or appointed because they fulfilled more or less explicit criteria [30]. The specific procedures for selecting the leader are often based on tradition and the provisions of the organizations. Take two online communities as examples. In Newgrounds, a collective movie creation community, the role of project leader is typically assumed by the person who conceives the idea for the project [20], while in Wikipedia, administrators are appointed through a peer review and election procedure [26]. Deviations from the selection process that members consider legitimate will weaken the leader's legitimate power [30].

Their legitimacy gives these leaders the right to make requests within their leadership domain and increase the potency of these requests [30]. Legitimate leaders often have defined privileges, obligations and responsibilities. For example, administrators in Wikipedia have access to restricted technical features, such as protecting, restoring and moving pages [26].

HOW PEER LEADERS MOTIVATE OTHERS

"*To be successful, online communities need the people who participate in them to contribute the resources on which the group's existence is built…. In almost every online community, there are important contributions not being made.*" [19]. For example, although Wikipedia is among the ten most visited sites in the world, two-thirds of assessed articles are categorized as low quality [19]. Furthermore, researchers have found a gradual decline in the growth of Wikipedia's active editor base in recent years [22]. The above suggest that promoting members' motivation to contribute is a crucial goal for online communities.

We use path-goal theory [11, 16, 30] to explain how peer leaders influence the motivation of other members in online communities. According to House [16, p. 324], "*the motivational function of the leader consists of increasing personal payoffs to subordinates for work-goal attainment and making the path to these payoffs easier to travel by clarifying it, reducing roadblocks and pitfalls, and increasing the opportunities for personal satisfaction en route.*", Path-goal theory borrows concepts particularly from expectancy value theory [14, 25] to explain how leadership behaviors affect targets' efforts. There are different versions of expectancy theories, but they all explain motivation as a rational choice process in which individuals decide how much effort to devote. According to the theory, in determining their level of effort, individuals consider the likelihood a given level of effort will lead to desirable outcomes (e.g. high pay, promotion, acceptance by a community, sense of achievement) and avoidance of undesirable outcomes (e.g. termination, rejection from the community, reprimands). The perceived probability of an outcome is its expectancy, and the perceived desirability of an outcome is its valence. If people believe that the outcomes are attractive (high valence) and the effort will result in the outcome (high expectancy), they will make the effort. Leadership behaviors have their effects primarily by modifying these perceptions and beliefs [30].

Task-focused leadership (transactional leadership, aversive leadership and directive leadership) have positive motivational effects to the extent that they reduce ambiguity and increase the expectation of successfully accomplishing tasks and goals.

Task-focused leadership also changes the valence of the outcomes. More specifically, transactional and aversive leadership behaviors change targets' motivation by changing their perception of the valence associated with their efforts— gaining rewards and avoiding punishments, respectively [16]. Transactional leadership is based on reward power, increasing the size of incentives for task success, whereas aversive leadership is based on coercive power [13]. Transactional and aversive leadership behaviors will influence targets' behavior only to the extent that the influencers control rewards that have positive valence for the target or punishments that have negative valence for the target, respectively. That is, to influence effort the influencer must control rewards that are contingent on hard work and punishments contingent on insufficient work, and the contingency must be clearly perceived by the targets.

In addition to these extrinsic rewards and punishments, aversive leadership behaviors may modify valence by decreasing the intrinsic attractiveness of the work. Previous research shows that aversive leadership tends to negatively impact subordinates' satisfaction in conventional organizations [9]. These negative effects on intrinsic motivation may be especially problematic in volunteer organizations where extrinsic motivators (e.g., money) are less powerful.

We predict that person-focused leadership should have a positive influence on members' effort, for a number of reasons. First, person-focused leadership behavior helps members build self-confidence and develop skills, which in turn influences their motivation by increasing their effort-performance expectation. In addition, person-focused leaders focus on interpersonal relationships, which will result in social support, friendliness among group members, increased cohesiveness, and team effort. These social outcomes have positive valence for most members [16]. Person-focused leadership should be especially effective in online communities because many peer production tasks (e.g., editing Wikipedia articles) often require cooperation and team spirit despite weak bonds between members, due to the distributed nature of the communities.

Legitimate leaders, who occupy formal leadership positions, are in general more powerful in influencing and motivating others' activities compared to peers who perform comparable leadership behaviors [8]. People in leadership roles are perceived to have the legitimate right to issue directions and distribute rewards and punishment. Because of past socialization experiences (e.g., with parents, teachers, religion), complying with legitimate requests from authorities is often intrinsically satisfying [13]. Furthermore, formal leaders are often perceived as central members of the social system and may induce a sense of connection and identification with the community, which in turns increases the positive valence of contributing to the community.

The following hypotheses summarize this reasoning.

Hypothesis 1. Shared leadership behaviors performed by members at all levels influence targets' motivation to contribute in online communities.

Hypothesis 2. Among task-focused leadership behaviors (transactional, aversive and directive), transactional leadership behaviors are most effective in promoting motivation; aversive leadership behaviors are least effective.

Hypothesis 3. Person-focused leadership behaviors will motivate members' contribution.

Hypothesis 4. Legitimate leaders are more powerful in influencing members' contribution than regular members.

STUDY PLATFORM

Wikipedia is the site of our empirical investigation. Wikipedia, formally launched in January 2001, is a free, web-based, collaborative, encyclopedia project and is the largest encyclopedia in the world. We used a complete download provided by the Wikimedia Foundation from Wikipedia's inception to January 2008 (approximately 182 million revisions). To handle this data volume, we used the Yahoo! M45 computing cluster running Hadoop and Pig.

MEASUREMENT

Measurement of leadership behaviors

We measured leadership behaviors as exchanged in the communication among Wikipedia editors by examining the messages editors left on each others' personal profile pages. We used machine learning techniques to automatically classify the messages into four categories of leadership behaviors: positive feedback, negative feedback, directive message, and social messages, which correspond respectively to transactional leadership, aversive leadership, directive leadership and person-based leadership. Messages could be assigned to multiple categories. Table 1 shows the sample messages sent by Wikipedia editors and their corresponding categorizations automatically generated by the machine learning tools.

We used machine learning models created by Zhu et al. [31]. They had two trained judges hand-tag 500 messages randomly selected from the Wikipedia corpus for the presence or absence of each of the four types of peer leadership behavior. Then they used support vector machine algorithms on 21 language pattern features to predict the human judgments. The machine learning models classified the four categories with high accuracy (89% on average). The agreement between the machine classifications and human judgments was moderate for negative feedback (Kappa = 0.48), but is very substantial or excellent for the other three categories (Kappa = 0.75, 0.71, 0.80) [31].

The most important features predicting each of the four types of leadership behavior are listed below [31].

- Transactional leadership: the frequency of the word "barnstar" (a barnstar is a type of virtual reward in Wikipedia) and the frequency of the phrases "thanks for" and "thank you for"
- Aversive leadership: the frequency of a set of strong negative words, including "block," "revert" and "remove";
- Directive leadership: the frequency of "you" followed by modal words such as "should," "could," "might," and the frequency of the word "please" followed by a verb;
- Person-based leadership: the frequency of greeting words and smiley emoticons.

Applying these classifications to 4 million messages sent between editors, Zhu et al. [31] found that a large proportion of leadership behaviors were performed by editors without formal leadership roles in Wikipedia. For example, non-administrators contributed 64% of directive leadership behaviors. In this paper we go beyond characterizing the types of leadership behaviors evinced in Wikipedia to examining the effects they have on their targets.

Measurement of leader legitimacy

We measure leader legitimacy by examining whether the leadership exerciser was an administrator or not. As discussed previously, administrators are formal leader-like roles in Wikipedia, selected by a rigorous peer review process and given special privileges, including protecting, deleting and restoring pages, moving pages over redirects, hiding and deleting page revisions and blocking malicious editors [26].

Measurement of the effectiveness of leader behavior

We measured the effectiveness of leadership behaviors by examining the extent to which they affected others' efforts. Particularly, we looked at changes in targets' editing before and after receiving leadership messages, depending on the types of leadership messages they received, and whether the leadership messages came from a legitimate leader or not.

Machine learning (ML) categories & Leadership type	Sample messages
ML category: Positive feedback **Leadership type:** Transactional leadership (Task-focused)	*"I award this barnstar[1] to XXX for your help and assistance in getting the WikiProject user warnings to the review phase, and to let you know your work has been appreciated."*
	"Thanks for all your work on the Survivor articles"
ML category: Negative feedback **Leadership type:** Aversive leadership (Task-focused)	*"If you continue in this manner you will be blocked from editing without further warning."*
	"…there is a concern that the rationale you have provided for using this image under "fair use" may be invalid. ... If it is determined that the image does not qualify under fair use, it will be deleted within a couple of days according to our criteria for speedy deletion."
ML category: Directive message **Leadership type:** Directive leadership (Task-focused)	*"Please read the instructions at… Using one of the templates at…, but remember that you must complete the template…"*
	"… one of these days do you think you could take some pictures at Mission Mill? I'd like to spruce up the article but it really needs some photos…"
ML category: Social message **Leadership type:** Person-focused leadership	*"Hi XX. Welcome to WikiProject XXX! I saw your name posted on the members list and wanted to welcome you... Anyway we are glad to have you. If I can help at all let me know :) ..."*
	"[[Image:Smiley.svg]] has smiled at you Smiles promote WikiLove and hopefully this one has made your day better… Happy editing"

Table 1. Sample messages sent by Wikipedia editors and their corresponding machine learning categories and leadership types.

ANALYSIS

The goal of this paper is to identify the effects of receiving different types of leadership messages from other Wikipedia editors on changes in recipients' editing behavior. In an analogy to a true experiment, we will compare the changes in editing behavior of those who received leadership messages (i.e., treated group) to those who do not receive messages (i.e., control group).

The sine qua non of a true experiment is random assignment of treatment to experimental units. Random assignment is the best way to assure that potential confounding variables, both measured and unmeasured, are on average equal among the treatment group and the control group [5].

Unfortunately, although Wikipedia has an enormous amount of archival data, these data are observational, and the receipt of a leadership message is not a true experimental treatment. The treatment here, as with most events in the real world, is endogenous in the sense that it is caused by other factors inside the system. In our data, the messages a recipient gets are partially a response to the recipient's previous behaviors. For example, the number of edits one person made in a previous week may cause others to send them messages in the following week. Similarly, editors who produce good edits may cause others to send them transactional leadership messages, while those who produce poor edits may cause others to send them aversive leadership messages in a subsequent week. Not controlling for confounding factors that influence both the treatment and the outcome can lead to biased estimates of the treatment effects.

To ameliorate the endogeneity problem, we use propensity score matching (PSM) to approximate randomization. PSM builds experimental and control groups by balancing the groups on potential confounding factors. PSM can effectively reduce the bias caused by these conditioning factors [10, 21]. However, because PSM balances only on measured variables, it cannot adequately control for all variables relevant to treatment. However, the results approximate a true experiment to the extent that one can reasonably presume that omitted variables are uncorrelated with the outcome.

Data preparation

We restricted the analysis to registered Wikipedia editors who had edited any Wikiproject page at least once, since this provided a basic filter against vandals and guaranteed that the editors had some experience in Wikipedia. The data were longitudinal, following the same editors across different weeks. For the analysis, we first defined whether an editor was active in a given week (the focal week) in terms of whether the editor made any edits during a five-week period (including the focal week, two weeks before, and two weeks after the focal week). Then we did an editor-week level analysis, restricted to the weeks in which the editor was active. The data comprised 31,676 unique editors, 2,053,405 editor-week observations and 1.6 million messages.

Dependent Variable

- *Contribution_change*. We measured editors' contributions towards Wikipedia articles through their revision count (i.e., number of edits). Edits are a direct measure of editors' effort, indicating the number of changes they made to articles during a period of time. Each edit indicates a set of editing actions, for example adding, changing, deleting or reverting text, references or illustrations, or communicating with other editors. The dependent measure was the log transformed edits in the week after the focal week minus the log transformed edits in the week prior to the focal week. Because the logarithm of zero is undefined, we added one before computing the logarithm.

$$contribution\_change = \ln(edits_{t+1} + 1) - \ln(edits_{t-1} + 1)$$

Independent Variable

- *Receive_msg.* This dummy variable[1] indicates whether the editor received any messages during the focal week. One indicates that the editor received at least one message, while zero indicates that the editor received no messages.
- *Transactional.* This dummy variable indicates whether in the focal week the editor received any message categorized as transactional (i.e., providing positive feedback). One indicates that the editor received at least one message with positive feedback, and zero indicates that the editor received no positive feedback. The following three variables are similar.
- *Aversive.* This dummy variable indicates whether the editor received any message categorized as negative feedback during the focal week.
- *Directive.* This dummy variable indicates whether the editor received any message categorized as directive during the focal week.
- *Person.* This dummy variable indicates whether the editor received any message categorized as social exchange during the focal week.
- *Admin.* This dummy variable indicates whether the editor received any messages from any administrator during the current week. One indicates that the editor received at least one message from an administrator, while zero indicates that the editor received no messages from any administrator.

We also examine the interaction between message type and legitimacy of leadership.

[1] For ease of interpretation, we use dummy variables to represent the independent variables. The results are substantively the same when using log transformed number of different types of messages as the independent variables.

- *Admin*Transactional.* This dummy variable indicates whether the editor received any messages categorized as positive feedback from any administrator during the focal week. One indicates that the editor received at least one positive message from an administrator, while zero indicates that the editor received none. The other three interactions were constructed similarly.
- *Admin*Aversive.* This dummy variable indicates whether the editor received any messages categorized as negative feedback from an administrator during the focal week.
- *Admin*Directive.* This dummy variable indicates whether the editor received any messages categorized as directive messages from an administrator during the focal week.
- *Admin*Person.* This dummy variable indicates whether the editor received any messages categorized as social messages from an administrator during the focal week.

Propensity Score Matching

For each editor who received messages, we selected a comparison editor who was most similar on confounding variables but did not receive messages. We used Propensity score matching (PSM) to pair editors. PSM involved three steps. First was to estimate the propensity score (i.e., the probability of receiving messages from others) from a set of conditioning variables. The variables used to predict receiving a message were the editors' prior activities (e.g., number of edits in previous week, number of messages received in previous week, tenure in Wikipedia). The rationale was that these factors might both cause other editors to communicate with them and also be correlated with subsequent changes in effort. Therefore, we chose six of the editors' previous activities listed below as conditioning variables.

In the second step, we matched each editor who received leadership messages in a particular week with another editor who did not receive a message, but who had the most similar propensity score based on six behavioral indicators. Propensity scores allow researchers to control for many variables simultaneously by matching on a single scalar variable. The variables with higher correlation with the treatment (also having higher risk to introduce bias) will be balanced better than the variables with lower correlation with the treatment. At the end of the second step, we tested whether the treatment group and control group were well matched in terms of the conditioning variables.

In the third step, we ran fixed effects regression analyses to estimate the effect of receiving messages, especially different types of leadership messages, on the treated groups and matched controls.

***Step 1**: Estimate propensity score*

We first used logistic regression to estimate the probability of an editor receiving messages in a given week (t) based on the editor's previous activities. The estimated probability is the propensity score. The six predictors are listed below, and the results are shown in Table 2.

- $Edits_{t-1}$. The log of the number of edits done by the editor in the week before the focal week.
- $MsgReceived_{t-1}$. The log of the number of messages the editor received in the week before the focal week.
- $MsgSent_{t-1}$. The log of the number of messages the editor sent in the week before the focal week.
- $MsgReceived_{<t-1}$ The log of the total number of messages the editor received any time prior to the week before the focal week.
- $MsgSent_{<t-1}$ The log of the aggregate number of messages the editor sent any time prior to the week before the focal week.
- *Tenure.* The number of weeks between the editor's first edit and the focal week.

Receive Msg	Coef.	Std. Err.
Intercept	-2.8803**	.0046
$Edits_{t-1}$	.2906**	.0014
$MsgReceived_{t-1}$	.8926**	.0044
$MsgSent_{t-1}$	.1682**	.0039
$MsgReceived_{<t-1}$	.4730**	.0024
$MsgSent_{<t-1}$	-.0147**	.0018
Tenure	-.0062**	<.0001
Log likelihood	-828366.63	
Pseudo R2	0.2756	
Number of obs	2,053,405	

Table 2. Estimate the probability of receiving messages (propensity score) with logistic regression. ** indicates that p value is less than 0.0001

Number of obs	Full Matched	Treat 503,259 Treat 503,259	Ctrl 1,550,146 Ctrl 503,259		
Variable	Sample	Treat Mean	Ctrl Mean	% bias	Reduced bias
$Edits_{t-1}$	Full	3.33	1.44	109.8	
	Matched	3.33	3.36	-1.4	98.7
$MsgReceived_{t-1}$	Full	0.78	0.12	100.8	
	Matched	0.78	0.78	-0.7	99.3
$MsgSent_{t-1}$	Full	0.76	0.11	78.8	
	Matched	0.76	0.81	-5.8	92.6
$MsgReceived_{<t-1}$	Full	3.34	1.77	97.7	
	Matched	3.34	3.24	6.3	93.6
$MsgSent_{<t-1}$	Full	3.12	1.43	84.2	
	Matched	3.12	3.06	2.8	96.7
Tenure	Full	68.1	61.4	13.0	
	Matched	68.1	63.6	8.7	33.1
Dependent Var Contri_change	Full Matched	-0.055 -0.055	-0.011 -0.606		

Table 3. Comparison between treatment editors who received messages in the focal week (treat) and control editors (ctrl) before and after propensity score matching (full versus matched).

Bias is defined as $100(\bar{x}_t - \bar{x}_c)/\sqrt{\{(S_t^2 + S_c^2)/2\}}$, where $\bar{x}_t$ and $\bar{x}_c$ are the sample means, S_t^2 and S_c^2 are the sample variances

Step 2: *Matching based on propensity score*

In this step, we matched editors who received messages during the focal week with non-recipients, based on the logit of the estimated propensity score. To do this we ordered the treated and control editors on their propensity scores and then for each treated subject selected a control editor with the closest propensity score within a maximum distance, called a caliper. We defined the caliper, the maximum distance, as 0.1. Following Rosenbaum and Rubin [21], we used the logit of the estimated propensity score to match because its distribution is generally approximately normal.

Table 3 reports the descriptive statistics (means, percentage of bias and the percentage reduction in the bias after matching), showing that the treatment group and the control group are balanced. The percentage of bias is the mean difference as a percentage of the average standard deviation: $100(\bar{x}_t - \bar{x}_c)/\sqrt{\{(S_t^2 + S_c^2)/2\}}$, where for each covariate $\bar{x}_t$ and $\bar{x}_c$ are the sample means in the treatment groups (editors who received messages in the given week) and control groups (editors who did not receive messages in the given week), respectively, and S_t^2 and S_c^2 are the corresponding sample variances [21].

Here are several points to note about Table 3. First, while editors who received or failed to receive messages differed substantially on all the conditioning variables before matching, they had similar means after matching. The reduction in bias was over 90% for five of the six conditioning variables. Secondly, we looked at the dependent variable—change in contributions—before and after matching. Prior research shows that editors decrease the amount they edit over time [6]. Without matching, one would be left with the erroneous conclusion that editors who received messages in a focal week were more likely to decrease their activities (-.055) than editors who did not receive messages (-.011). However, after controlling for previous activities with propensity score matching, one can see that receiving messages in the focal week slowed the general decline in contributions (-.055 versus -.606).

Step 3: *Run the analysis*

Using the matched sample, we then examined the effects of receiving messages on changes in contribution and how the effects interact with the types of leadership behaviors and the source. We used fixed effects linear regression to predict change in editors' contribution, with each treated-control pair as a group. Fixed effects allowed each pair a different intercept (pre-existing difference among pairs were embodied in the intercept); independent variables—the event of receiving messages and types of messages—determined the slope, which we assume was the same for all pairs (representing an average effect of receiving a certain type of message). Table 4 shows the descriptive statistics of the independent variables and the analysis results.

RESULTS

To understand Table 4, we first need to understand how to interpret the dependent variable *Contribution_change*. The dependent variable is the log transformed edits in the week after the focal week minus the log transformed edits in the week prior to the focal week. Therefore, the sign of dependent variable indicates whether the editor's editing increased (positive sign) or decreased (negative sign) surrounding the focal week. Furthermore, an increase of x in the dependent variable indicates that, holding the edits in the prior week constant by PSM, the edits in the subsequent week increased approximately x%.

First, model 1 in Table 4 shows that editors who received messages in a focal week subsequently edited more than those who did not. Note that the intercept is significantly negative, indicating that those who received no messages reduced their editing surrounding a focal week by 61%. However, receiving messages slowed this to a 6% decline

Dependent Variable	Descriptive Statistics		Model 1		Model 2		Model 3	
Contribution_change	Mean	Std.Dev.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept			-.6059**	.0021	-.6059**	.0021	-.6059**	.0021
Receive_ msg	.5000	.5000	0.5507**	.0030	.3326**	.0054	.2956**	.0060
Transactional	.1872	.3901			.1927**	.0067	.1615**	.0079
Aversive	.0646	.2458			-.1442**	.0098	-.1003**	.0115
Directive	.2884	.4530			.0859**	.0064	.0585**	.0072
Person-based	.2511	.4336			.2290**	.0061	.1698**	.0071
Admin	.2264	.4185					.1584**	.0086
AdminXTransactional	.0657	.2478					.0278 *	.0118
AdminXAversive	.0174	.1306					-.0945**	.0205
AdminXDirective	.1103	.3133					.0174	.0102
AdminXPerson	.1090	.3117					.0579**	.0103
Number of obs	1,006,518							
Number of groups	503,259							

Table 4. Regression predicting the effects of leadership behaviors on subsequent change in editors.
**** indicates that p value is less than 0.0001; * indicates that p value is less than 0.05**

(-61%+55%).

Secondly, model 2 demonstrates that different types of leadership behaviors differentially influenced subsequent motivation. (Hypotheses 1 to 3). The effects of messages that were not one of the four leadership types resulted in a 33% increase in edits in the subsequent week[2]. Among task-focused leadership behaviors, receiving transactional leadership messages (i.e., positive feedback) led to an additional 19% increase in subsequent edits. Directive behavior messages led to an addition 8.6% in subsequent edits. In contrast, aversive leadership messages (negative feedback) decreased members' contribution by 14%. We can also see that the influence of person-based leadership was substantial, increasing edits by 23%.

Thirdly, Model 3 demonstrates that messages sent by administrators were more influential than those sent by peers. Receiving a non-leadership message from an administrator increased edits by 15% compared to messages from non-administrators. Transactional messages sent by administrators increased editing an additional 2.8% compared to those sent by peers. Conversely, aversive messages sent by administrators decreased editing an additionally 9.5% compared to aversive messages sent by peers. Finally, person-based messages sent by administrators increased editing by 5.8% compared to those sent by peers.

DISCUSSION

Endogeneity

We used sophisticated matching procedures in the form of propensity score matching to control for endogeniety, in which something about the targets and their behavior induced others to send them leadership messages and was also associated with changes in their contribution from before to after receiving the messages. We built balanced experimental and control groups based on prior levels of activity (see Table 3). However, the validity threat associated with the propensity score matching procedures is omitted variable biases. It is possible that our experimental group and control group were unbalanced in some unmeasured variable. For example, we did not control for the quality of the work. It is likely that the editors who produced higher quality work than average were more likely to fall into the treatment group because their work attracted attention, and thus they were more likely to receive transactional leadership messages from others. (Conversely, editors who produced work of lower quality than average might have been more likely to receive aversive messages.) According to econometric theory, omitted variables are problematic only if they are correlated with the outcome of interest; otherwise the estimated effects of treatment will be unbiased [28]. Note that the outcome of our analysis is the *change* of the contribution from before to after receiving the messages, rather than the pure quantity of the contributions. Even though it is easy to speculate about potential omitted variables that can compromise matching, we were unable to identify pathways by which omitted variables—such as the quality of the work—would be responsible for the dramatic changes in editors' contributions in the week surrounding the receipt of leadership messages. There was no good reason for editors to suddenly increase or decrease their efforts solely because they were producing high or low quality wok. As long as we do not find omitted variables that substantially influenced both the receipt of leadership messages and the behavior change in the weeks surrounding the receipt, we believe that the analysis results shown in Table 4 capture the unbiased causal effect of leadership messages on the change in members' effort.

Natural experiment

Another way to deal with threats to validity associated with endogeneity is to identify external events that change the probability of receiving a treatment, but that cannot plausibly be attributed to the behavior of those receiving the treatment. For example, in the past researchers used natural experiments, such as arbitrary changes in laws, to examine how the threat of punishment influences the likelihood of obeying social norms, like obeying the speed limit or paying taxes [5]. We identified a similar natural experiment in Wikipedia, which changed the frequency with which aversive leadership messages were sent independent of the recipients' behavior.

Results described previously indicate that editors receiving aversive leadership messages decreased their editing. More than half of the messages categorized as aversive leadership were actually sent by a single editor who created a program

Disputed fair use rationale for Image: XXX.png

Thanks for uploading XXX.png. However, there is a concern that the rationale you have provided for using this image under "fair use" may be invalid. Please read the instructions at Non-free content carefully, then go to the image description page and clarify why you think the image qualifies for fair use. Using one of the templates at Fair use rationale guideline is an easy way to ensure that your image is in compliance with Wikipedia policy, but remember that you must complete the template. Do not simply insert a blank template on an image page. If it is determined that the image does not qualify under fair use, it will be deleted within a couple of days according to our criteria for speedy deletion. If you have any questions please ask them at the media copyright questions page. Thank you. – BetacommandBot (talk).

Figure 1. Template message about copyright issue sent by the editor Betacommand.

[2] Non-leadership messages include newsletters, notice of a Wikiproject-wide or Wikipedia-wide event, reports of the status of certain work, and discussions and arguments about certain issues, etc. The 33% increase in people's motivation and contributions can be explained by the fact that receiving messages from other members or from the whole community, even without any specific directions, criticisms or praise, can elicit a sense of belonging to and identification with the community [32].

he labeled Betacommandbot to automatically send messages warning other editors who uploaded pictures without proper fair use rationale. The sample template is shown in Figure 1. The editor used an aversive leadership style and tried to influence other contributors by threatening to remove their contributions[3]. From the time this automatic program was approved on May 2007 [27] until it was removed in March 2008, this editor sent out more than 150,000 warning messages affecting more than 40,000 distinct editors. Importantly, Betacommandbot's messages were sent to all past violators (even ones whose edits had been far in the past), thus controlling for issues such as the likelihood of receiving a message. Thus the Betacommandbot warning was a natural experiment, like a change in speeding laws, that was not induced by recipients' behavior.

According to Table 4, the editors receiving aversive messages decreased their contributions in the following week by 14%, on average. We conducted analyses to isolate the effects of receiving the Betacommandbot warning. We used two matching methods: 1) within-editor matching, in which we matched the work of each editor who received a Betacommandbot warning with the same editor's work before the Betacommandbot was created; 2) delay matching, in which we matched the work of editors who received the Betacommandbot warning during that week with others who had also uploaded images without copyright rationale but who did not receive the Betacommandbot warning during that week (although they would receive it in the future). Across these variations, editors who received the Betacommandbot warning decreased their editing more than 10%. The results provide additional evidence to support our main findings.

Elite and peer

The shared leadership model claims that leadership emanates from members from all levels rather than a few select individuals. Although each member can enact some level of leadership behaviors, it does not assume an equal distribution and the same effectiveness of leadership behaviors across individuals. Actually, administrators issue more leadership messages than do non-administrators on a per capita basis[31], and the messages they send have more impact on the targets' editing behavior in the week following the message compared to messages sent by non-administrators. The administrators, representing only 1.3% of those in the current sample, sent approximately forty percentage of the leadership messages. Moreover, both their task-oriented and person-oriented messages were more powerful than those of the non-administrators in changing others' rates of editing.

[3] The warning message could also be considered (and categorized by machine learning models) as directive leadership, since it clarified instructions, though not person-based, since the messages used standardized templates.

However, in the aggregate, leadership exercised by non-vertical leaders (i.e., regular editors) was very important. For example, the large numbers of non-administrators sent more than half of leadership messages [31]. Moreover, these leadership messages had substantial effects on others' editing. For example, their transactional leadership messages induced a 16% increase in others' editing in the week following their message, while their person-based behavior increases other members' efforts by 17%.

Implications

There are several important theoretical and practical implications from this research. First, shared leadership, which has heretofore been a relatively neglected area of research on online communities, seems to be an important force influencing the success of peer production communities. Even though members (especially those without formal leadership roles) might not perceive themselves as "leaders," they actually perform leadership actions which can substantially influence others' behaviors and thus significantly affect the continuous functioning of the communities.

Secondly, practitioners should consider how to encourage effective leadership behaviors and prevent ineffective leadership behaviors from members. Our research shows that the four leadership types—transactional, aversive, directive and person-based—are not equally beneficial. In particular, transactional and person-based leadership had the strongest effects, suggesting that interfaces and mechanisms that make it easier for editors to connect with, reward, and express their appreciation for each other may have the greatest benefits.

Aversive leadership is the least effective leadership style in terms of motivating members to work. People may argue that reducing the activity of harmful editors is a positive impact of aversive leadership. However, considering the fact that there is much work to be accomplished in Wikipedia and the recent downward trend of active editors, pure aversive leadership should be avoided. It is important to find ways to reduce low quality work while maintaining editors' motivation. Choi et al [7] showed that constructive criticism (i.e., gentle and sociable corrective feedback) exchanged between in-group members (Wikipedia editors in the same subgroups) increased, not decreased recipients' motivation. In future work, we plan to explore the effects of these nuanced variations of aversive leadership behaviors in terms of encouraging high-quality work and maintaining motivation.

CONCLUSION

This research introduced a framework of shared leadership to study the leadership process in online communities. We examined in particular how leadership types and leader legitimacy facilitate the effectiveness of shared leadership. We found that 1) leadership performed by members at all levels significantly influenced other members' motivation; 2) transactional leaders and person-focused leaders were effective in motivating others, whereas aversive leaders

decreased other contributors' motivations; 3) legitimate leaders were in general more influential than regular members.

ACKNOWLEDGEMENT

We thank the members of the Social Computing Group at Carnegie Mellon University for helpful feedback. We also thank Aya Betensky, Yarun Luon and Raechel Lee for editing help. This research was supported by NSF grants OCI-09-43148, IIS-0968484, IIS-1111124 and a grant from Carnegie Mellon's Center for the Future of the Work.

REFERENCES

1. Angrist, J.D., Krueger, A.V. (1999) Empirical strategies in labor economics. In Handbook of Labor Economics. Orley Ashenfelter and David Card, eds. Amsterdam: North Holland.
2. Bass, B. M. (1990) Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications (3rd ed.). New York, NY, US: Free Press.
3. Burke, C. S, Stagl, K. C, Klein, C, Goodwin, G. F, Salas, E. (2006) What type of leadership behaviors are functional in teams? A meta-analysis The Leadership Quarterly,
4. Burns, J. M. (1978) Leadership. New York: Harper & Row.
5. Campbell, D.T., Stanley, J.C. (1963) Experimental and Quasi-experimental Designs for Research, Houghton Mifflin Company, Boston.
6. Cassell, J., Huffaker, D., Tversky, D. (2006) The Language of Online Leadership: Gender and Youth Engagement on the Internet. Developmental Psychology. Vol 42(3) May, 436-449.
7. Choi, B. R., Alexander, K., Kraut, R. E., & Levine, J. M. (2010). Socialization tactics in Wikipedia and their effects. In Proceedings of the 2010 ACM Conference on Computer-Supported Cooperative Work. New York: ACM Press.(CSCW'10)
8. Cialdini, R. B., and Goldstein, N. J. Social influence: Compliance and Conformity. Annual Review of Psychology, 55, 1 2004), 591-621.
9. Cox, J. F. (1994). The effects of superleadership training on leader behavior, subordinate self-leadership behavior and subordinate citizenship. Unpublished doctoral dissertation, University of Maryland, College Park.
10. D'Agostino, R.B., Jr. (1998). Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Statistics in Medicine, 17, 2265-2281.
11. Evans, M.G.(1970) The effects of supervisory behavior on the path-goal relationship, Organizational Behavior and Human Performance, Volume 5, Issue 3, May 1970, Pages 277-298
12. Fleishman, E. A. (1953) The description of supervisory behavior. Journal of applied psychology
13. French, J., & Raven, B.H. (1959). The bases of social power. In D. Cartwright(Ed.), Studies of social power. Ann Arbor, MI: Institute for Social Research, pp. 150-167
14. Georgopoulos, Basil S.; Mahoney, Gerald M.; Jones, Nyle W., Jr. (1957) A path-goal approach to productivity. Journal of Applied Psychology, Vol 41(6), Dec 1957, 345-353.
15. Hogan, R., Curphy, G. J., & Hogan, J. (1994). What We Know About Leadership. American Psychologist, 49(6), 493-504.
16. House, R. J. (1971) A Path Goal Theory of Leader Effectiveness Administrative Science Quarterly Vol. 16, No. 3.
17. Katz, Daniel; Maccoby, Nathan; Morse, Nancy C. (1950). Productivity, supervision, and morale in an office situation. Part I. Oxford, England: Institute for Social Research.
18. Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. In CSCW'08: Proceedings of the ACM conference on computer-supported cooperative work. New York: ACM Press.
19. Kraut, R. E. & Resnick, P. (In press). Evidence-based social design: Mining the social sciences to build online communities. Cambridge, MA: MIT Press.
20. Luther, K and Bruckman, A. (2008). Leadership in online creative collaboration. In Proceedings of the 2008 ACM conference on Computer-Supported Cooperative Work (CSCW '08).
21. Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41-55.
22. Suh, B., Convertino, G., Chi, E. H.,and Pirolli, P. (2009). The singularity is not near: slowing growth of wikipedia. In WikiSym '09: Proceedings of the 5th International Symposium on Wikis and Open Collaboration, pages 1–10, New York, NY, USA. ACM
23. Pearce, C.L., and Sims H.P. (2002) Vertical Versus Shared Leadership as Predictors of the Effectiveness of Change Management Teams: An Examination of Aversive, Directive, Transactional, Transformational, and Empowering Leader Behaviors, Group Dynamics, 6(2).
24. Pearce, C.L., and Conger,J. A. (2003) Shared leadership: Reframing the hows and whys of leadership, Sage, Thousand Oaks.
25. Vroom, V.H. (1964) Work and motivation. Oxford, England: Wiley.
26. Wikipedia editors. http://en.wikipedia.org/wiki/Wikipedia:Administrators
27. Wikipedia editors http://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/BetacommandBot_Task_5
28. Wooldridge, J. M. (2003) Introductory econometrics, 2nd Ed. Mason, OH: South-Western.
29. Yoo, Y., & Alavi, M. (2004). Emergent leadership in virtual teams: what do emergent leaders do? Information and Organization, 14(1), 27-58.
30. Yukl, G. (1998) Leadership in Organizations
31. Zhu, H., Kraut, R.E., Wang, Y.C. and Kittur, A. (2011) Identifying shared leadership in Wikipedia. In Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11).
32. Zhu, H., Kraut, R.E., & Kittur, A. (in press) Organizing without formal organization: Group Identification, Goal Setting and Social Modeling in Directing Online Production. In Proceedings of the 2012 ACM conference on Computer-Supported Cooperative Work (CSCW '12).