Human Factors: The Journal of the Human Factors and Ergonomics Society
http://hfs.sagepub.com/

Effectiveness of Shared Leadership in Wikipedia

Haiyi Zhu, Robert E. Kraut and Aniket Kittur

Human Factors: The Journal of the Human Factors and Ergonomics Society 2013 55: 1021
DOI: 10.1177/0018720813515704

The online version of this article can be found at:
http://hfs.sagepub.com/content/55/6/1021

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
HF ES
Human Factors and Ergonomics Society

Additional services and information for *Human Factors: The Journal of the Human Factors and Ergonomics Society* can be found at:

Email Alerts: http://hfs.sagepub.com/cgi/alerts

Subscriptions: http://hfs.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Dec 19, 2013

What is This?

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

SPECIAL SECTION: 2013 **Human Factors** *Prize Winner*

Effectiveness of Shared Leadership in Wikipedia

Haiyi Zhu, **Robert E. Kraut**, and **Aniket Kittur**, Carnegie Mellon University, Pittsburgh, PA, USA

Objective: The objective of the paper is to understand leadership in an online community, specifically, Wikipedia.

Background: Wikipedia successfully aggregates millions of volunteers' efforts to create the largest encyclopedia in human history. Without formal employment contracts and monetary incentives, one significant question for Wikipedia is how it organizes individual members with differing goals, experience, and commitment to achieve a collective outcome. Rather than focusing on the role of the small set of people occupying a core leadership position, we propose a shared leadership model to explain the leadership in Wikipedia. Members mutually influence one another by exercising leadership behaviors, including rewarding, regulating, directing, and socializing one another.

Method: We conducted a two-phase study to investigate how distinct types of leadership behaviors (transactional, aversive, directive, and person-focused), the legitimacy of the people who deliver the leadership, and the experience of the people who receive the leadership influence the effectiveness of shared leadership in Wikipedia.

Results: Our results highlight the importance of shared leadership in Wikipedia and identify trade-offs in the effectiveness of different types of leadership behaviors. Aversive and directive leadership increased contribution to the focal task, whereas transactional and person-focused leadership increased general motivation. We also found important differences in how newcomers and experienced members responded to leadership behaviors from peers.

Application: These findings extend shared leadership theories, contribute new insight into the important underlying mechanisms in Wikipedia, and have implications for practitioners who wish to design more effective and successful online communities.

Keywords: shared leadership, online communities, Wikipedia, feedback, transactional leadership, aversive leadership, directive leadership, person-based leadership

Address correspondence to Haiyi Zhu, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15217, USA; e-mail: haiyiz@cs.cmu.edu.

HUMAN FACTORS
Vol. 55, No. 6, December 2013, pp. 1021–1043
DOI: 10.1177/0018720813515704
Copyright © 2013, Human Factors and Ergonomics Society.

INTRODUCTION

A variety of social media technologies provide virtual spaces where people all over the world can interact around a shared purpose and open up exciting opportunities to carry out projects of unparalleled scope and scale. One example is Wikipedia, which provides platforms for people to collaboratively edit encyclopedia articles. As of September 2013, Wikipedia had more than 19 million editors, contained 30 million articles in 287 languages (Wikipedia, 2013a), ranked as the sixth most frequently visited Web site in the world (Alexa Internet, 2013), and had an estimated 365 million readers worldwide (West, 2010). One significant question regarding Wikipedia is how it organizes the actions of millions of individuals with differing goals, experiences, and commitment to achieve collective outcomes, given the large number of members, high turnover, lack of employment contracts, lack of external incentives, and weak interpersonal bonds (Choi, Alexander, Kraut, & Levine, 2010; Kraut & Resnick, 2012; Zhu, Kraut, & Kittur, 2012b).

Empirical research from conventional organizations demonstrates the importance of leadership in helping group members effectively achieve their goals. Burke et al.'s meta-analysis (2006) showed that both task-oriented leadership, which focuses on the group's work, and relational leadership, which focuses on interpersonal relationships within the group, are valuable in improving productivity, developing teamwork, and developing increased capabilities. In a conventional organization, formal leadership roles are easily identified. For example, the chief executive officer, department heads, and supervisors serve as leaders of a firm. However, identifying the leaders in Wikipedia and understanding how they lead is a more difficult task.

Who are the leaders in Wikipedia? Previous researchers who investigated leadership in online communities have tended to use traditional vertical leadership models (Bass, 1990; Hogan,

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

Curphy, & Hogan, 1994). They have suggested that the leadership role is a specialized one. People who are appointed or elected to perform this role are designated as "leaders" (Cassell, Huffaker, & Tversky, 2006; Luther & Bruckman, 2008; Luther, Caine, Ziegler, & Bruckman, 2010; Misiolek & Heckman, 2005; Yoo & Alavi, 2004). 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 Wikipedia. The shared leadership framework was originally proposed by researchers investigating offline, leaderless groups such as self-managing teams, volunteer organizations, and employee-managed companies (Pearce & Conger, 2002; Pearce & Sims, 2002; Yukl, 1998). They argued that leadership—which involves 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 one another about what tasks are to be done, how tasks should be done, and the ways those tasks relate to each other. Leadership is viewed as a shared influence process rather than being seen as invested in specialized roles.

Although in the shared leadership model each member can display some level of leadership behavior, the model does not assume that all individuals' leadership behaviors are effective to the same extent. This paper investigates how distinct types of leadership behaviors, the legitimacy of the people who deliver the leadership, and the experience of the people who receive the leadership influence the effectiveness of leadership behaviors. Here, we operationalize the effectiveness of leadership in terms of the extent to which those exhibiting leadership behaviors can influence others to invest effort and contribute to the community, specifically efforts directed to a given focal task as well as general motivations to contribute.

In this article, we use Feedback Intervention Theory (FIT; Kluger & DeNisi, 1996) to explain the process by which leaders influence others' efforts on focal tasks and general motivation to work. We offer several hypotheses regarding the effectiveness of different types of leadership behavior, moderated by different types of leaders and different types of receivers. Subsequently, we describe our two studies, present our results, and discuss theoretical and practical implications.

THEORY AND HYPOTHESES

In this section we briefly summarize the shared leadership framework and the four types of leadership behaviors. We then use FIT to predict how leadership behaviors affect people's performance on specific tasks and their general motivation to work.

Shared Leadership

The traditional leadership literature has typically focused on the attributes and behaviors of the appointed or elected leader of some group or organization (cf. Bass, 1990). In contrast to this traditional, "heroic" view of leadership, shared leadership conceptualizes leadership as a collective social process emerging through the interaction among multiple actors (Pearce & Conger, 2002; Pearce & Sims, 2002; Yukl, 1998). The concept of shared leadership was developed in the mid-1990s in response to the increasing use of self-managed teams in conventional organizations, along with the rising speed of delivery, the increasing richness of information, and greater job complexity (Pearce & Conger, 2002). Unlike vertical leadership in a hierarchical managerial system, shared leadership was defined by Pearce and Conger as "a dynamic, interactive influence process among individuals in groups for which the objective is to lead one another to the achievement of group or organizational goals." These authors summarized three main characteristics of shared leadership as distributed and interdependent among people at all levels, a social process embedded in the social context in which it occurs, and focused on the particular social interactions that lead to mutual learning, greater shared understanding, and, eventually, positive actions.

Researchers investigating traditional vertical leadership have identified a range of effective leadership behaviors (Bass, 1990; Burke et al.,

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

2006; Yukl, 1998). In a shared leadership context, these strategies continue to be relevant (Pearce & Sims, 2002). Rather than prejudge results with terms like *leader* and *follower*, in the following sections we use the term *influencer* to indicate those who exercise influence and *targets* to indicate those whom the influencers 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 University's program on leadership research (Fleishman, 1953), task-oriented versus relationship-oriented in the University of Michigan research program (Katz, Maccoby, & Morse, 1950), and task-focused versus person-focused behaviors in Burke et al.'s 2006 review of leadership behaviors. We follow this general classification and then specifically differentiate three subcategories of task-based leadership behaviors, using Pearce and Sims's 2002 classification.

Task-Based Leadership

Transactional leadership is generally similar to the components of the transactional-transformational paradigm of leadership (Bass, 1990). 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. The basis of transactional leadership is that people engage in behaviors that will maximize their expected return from performance. Representative transactional leadership behaviors include (a) providing personal rewards, (b) providing material rewards, (c) managing by exception (active), and (d) managing by exception (passive; see Pearce & Sims, 2002). Sample questionnaire items measuring this type of leadership include "X will recommend that I am compensated well if I perform well"; "X gives me positive feedback or special recognition when I perform well"; "X tracks mistakes"; and "X delays taking action until problems become serious."

Aversive leadership. In contrast to transactional leadership, aversive leadership relies on coercive power (French & Raven, 1959). According to Pearce and Sims (2002), aversive leadership uses intimidation and reprimands to decrease undesired behaviors from targets. Sample questionnaire items measuring aversive leadership include "X tries to influence me through threat and intimidation" and "X lets me know about it when I perform poorly."

Directive leadership emphasizes the need to provide direction to targets and specify their roles and responsibilities. Directive behaviors include issuing instructions and commands and assignment goals (Pearce & Sims, 2002). Sample questionnaire items measuring directive leadership include "When it comes to work, X gives me instructions on how to carry it out" and "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 (Burke et al., 2006).

Consideration was first proposed as a type of person-based leadership behavior in the 1950s in the Ohio State University leadership research program (Fleishman, 1953). 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 a close social relationship with them (Yukl, 1998). In general, dyadic relationships characterized by consideration reflect two-way open communication, mutual respect and trust, and an emphasis on satisfying employee needs. The concept of relationship-oriented leadership in the University of Michigan research on leadership (Katz et al., 1950) is similar.

Later, researchers developed and elaborated on the concept of person-based leadership by proposing ideas such as *transformational leadership* (i.e., a component of the transactional-transformational leadership paradigm), which highlights encouragement, inspiration, and intellectual stimulation (Bass, 1990; Burns, 1978; Pearce & Sims, 2002), and *empowering leadership*, which focuses

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

on self-management skills and teamwork (Pearce & Sims, 2002; Thorenson & Mahoney, 1974).

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

Effects of Shared Leadership

Leadership behaviors as feedback. We assume here that without power derived from formal leadership positions, many attempts by peers to influence others will be interpreted by recipients as feedback about their prior behavior. Although feedback interventions (defined as intentional feedback given by an external agent) are not identical to leadership behavior, the concepts overlap substantially. (See Table 1 for the relationships between leadership behavior and feedback.) As such, Feedback Intervention Theory (FIT) can help to improve understanding of the effects of leadership behaviors on general motivation and performance on specific tasks (Kluger & DeNisi, 1996).

The first key assumption of FIT is that feedback is processed hierarchically. To simplify the presentation, the hierarchy can be divided into two levels: meta-task processes involving the self (e.g., self-goals and self-beliefs) and task processes involving the focal task and the detail of the task. Processes at the higher level (i.e., the meta-task processes) can supervise performance in the lower level (task-level processes). The processes in the lower level may also divert attention up the hierarchy and influence higher-level process.

The second key assumption of FIT is that people use feedback to evaluate their performance relative to their standards, which is often referred to as *feedback-standard comparison.* When they note a discrepancy between performance and the standard, people are motivated to reduce it. Typically they choose to eliminate the discrepancy by attempting to attain the standard.

Based on these two assumptions, we can predict people's reaction toward four types of leadership behaviors: transactional leadership (i.e., providing positive feedback and rewards), aversive leadership (providing negative feedback and punishment), directive leadership (giving directions and instructions), and person-based leadership (socializing and building personability and interpersonal relationships).

First, transactional leadership behaviors, aversive leadership behaviors, and directive leadership behaviors are all task-oriented and focus on details and progress toward a focal task. Aversive leadership, which provides negative feedback, signals that performance falls short of a standard and will lead people to increase effort toward the focal task. Directive leadership, which provides instructions to either achieve or raise standards, will also lead people to invest more effort in the focal task and improve performance. In contrast, providing positive feedback and rewards signals that performance exceeds the standard. Therefore, when people are subject to transactional leadership behavior, they typically maintain their effort or even reduce it (Kluger & DeNisi, 1996). In contrast, person-based leadership focuses on the person level rather than the task level and therefore should have little effect on people's performance on a specific task. This leads to our first hypothesis:

> *Hypothesis 1:* Aversive leadership and directive leadership can increase people's effort on focal tasks and improve task performance, whereas transactional leadership and person-based leadership should have less effect on focal task performance.

Although transactional leadership tends to have little effect on performance of specific tasks, it has effects at the meta-task level, influencing people's view of themselves. Positive feedback and rewards might increase people's self-efficacy and self-esteem and thus increase their general motivation to work. This increased motivation might spill over to nonfocal tasks (Kluger & DeNisi, 1996), lead to persistence in an activity, and increase self-report interest in the activity (Deci, Koestner, & Ryan, 1999). Similarly, although person-based leadership behavior does not affect specific task performance, it can help to develop people's self-confidence, build commitment toward the community, and thus increase general motivation. In contrast, aversive leadership might

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 1: Four Types of Leadership Behaviors, the Corresponding Feedback Types, Example Messages, and Hypotheses

Leadership Type	Corresponding Feedback Type	Hypotheses: Effects on the focal task (H1)	Hypotheses: Effects on general motivation (H2)
Transactional leadership (Task-focused)	Positive feedback	**No effects** Positive feedback signals that performance already exceeds the standard, so people do not invest extra efforts on the specific tasks receiving feedback.	**Increase** Positive feedback and rewards increase people's self-efficacy and self-esteem, and thus increases general motivation.
Definition: Behaviors intended to energize people through acknowledging work and provides rewards. **Example 1:** *"I award this barnstar* 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."* **Example 2:** *"Thanks for all your work on the Survivor articles."*			
Aversive leadership (Task-focused)	Negative feedback	**Increase** Negative feedback signals that performance falls short of a standard, so people invest more effort on the specific task to reach the standards.	**Decrease** Negative feedback decreases people's self-efficacy and self-esteem and thus decreases general motivation.
Definition: Behaviors intended to regulate people through negative messages, warnings, and reprimands. **Example 1:** *"If you continue in this manner you will be blocked from editing without further warning."* **Example 2:** *"…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."*			
Directive leadership (Task-focused)	Directive feedback	**Increase** Directive behavior provides instructions to either achieve standards or raise standards; it will also lead people to invest more effort in the specific task.	**No effects** Has no effects on people's general motivation.
Definition: Behaviors intended to direct people through issuing instructions and commands, assigning tasks, setting goals. **Example 1:** *"Please read the instructions at… Using one of the templates at…, but remember that you must complete the template…"* **Example 2:** *"… 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…"*			
Person-focused leadership	Social feedback	**No effects** Person-based leadership behavior (social feedback) is not directly related to any specific task.	**Increase** Develops people's self-confidence, builds commitment toward the community, and thus increases general motivation.
Definition: Behaviors intended to maintain close social relationships, support group cohesion, and develop subordinates' self-confidence and skills. **Example 1**: *"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 :) …"* **Example 2**: *"[[Image:Smiley.svg]] has smiled at you Smiles promote WikiLove and hopefully this one has made your day better… Happy editing"*			

*Barnstar is a type of virtual award in Wikipedia.

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

be perceived as a threat to self-esteem and could decrease motivation. Directive leadership behaviors do not draw attention to the self-level and should not influence motivation. This leads to our second hypothesis:

> *Hypothesis 2:* Transactional leadership and person-based leadership can increase people's general motivation to work; directive leadership has limited effects on general motivation, whereas aversive leadership might decrease people's motivation.

Moderating effects of leader legitimacy. Although shared leadership behaviors do not necessarily require formal leadership positions to be effective, such behaviors exercised by legitimate leaders might be more powerful than those of ordinary members. Legitimate leaders are those who occupy formal leadership positions in an organization, volunteer community, or other social system. Their legitimacy stems from the selection process, whether appointed by supervisors, elected by the membership, or appointed because they fulfilled more or less explicit criteria (Yukl, 1998). The specific procedures for selecting the leader are often based on tradition and the provisions of the organizations. Deviations from the selection process that members consider legitimate will weaken the leader's legitimate power (Yukl, 1998).

In Wikipedia, legitimate leaders are the administrators who are appointed through a peer review and election procedure. This legitimacy gives these leaders the right to make requests within their leadership domain and requires the targets of their requests to obey (Yukl, 1998). 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 (Wikipedia editors).

Legitimate leaders, who occupy formal leadership positions, are in general more powerful in influencing and motivating others' activities compared with peers who perform comparable leadership behaviors (Cialdini & Goldstein, 2004). 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 (French & Raven, 1959). 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 turn increases the positive valence of contributing to the community. This leads to our third hypothesis:

> *Hypothesis 3:* Legitimate leaders are more powerful in influencing members' behaviors than are regular members.

Moderating effects of prior experience. The effect of leadership behavior is substantially influenced by the willingness of the recipients to respond to the leadership influence (Ilgen, Fisher, & Taylor, 1979). Specifically, prior experience is an important variable moderating reactions to the leadership behavior. People with little experience in a task are less certain about standards and their abilities. In conventional organizations, newcomers, in contrast to more established members, have greater uncertainty regarding role requirements. As a result, they are especially eager to try to learn the beliefs, values, orientations, behaviors, and skills necessary to fulfill their new roles and function effectively within an organization (Ashforth & Saks, 1996). Therefore, one expects that newcomers will be particularly susceptible to influence compared with experienced members (1996). This leads to our fourth hypothesis:

> *Hypothesis 4:* Leadership is more influential on newcomers than on experienced users.

STUDY 1

Wikipedia is the site of our empirical investigation. Wikipedia, formally launched in January 2001, is a free, Web-based, collaborative encyclopedia project and has become 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 analyze Wikipedia editors' behavior. To handle this data

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

volume, we used the Yahoo! M45 computing cluster running Hadoop and Pig.

Measurement of Shared Leadership Behavior

In online communities like Wikipedia, people communicate and interact with one another predominantly through written text that is visible to all other community members. People tend to exert influence on one another through text-based communication. Therefore, we measured leadership behaviors by examining the messages exchanged between Wikipedia editors, specifically those messages they left on one another's personal profile pages.

Without automated coding of behavior, research on leadership is restricted to relatively small samples. For example, meta-review (Burke et al., 2006) shows that the average sample size is in the range of several hundred. In this paper, we demonstrate the possibilities of going beyond these small samples by using automated coding of leadership behaviors. We propose that we can use machine-learning techniques to automatically classify the messages into different leadership categories. The four categories and sample messages for each category are shown in Table 1.

A machine-learning approach has three main components: training sets (hand-coded data), representation of messages for machine learners (feature sets), and training algorithms. To begin the process, we train statistical models on a small set of human-coded data and evaluate them using a separate set of human-coded data. If the evaluation shows that the model is accurate, we can apply the model to a larger data set that had not been human coded. Details of the machine-learning approach are shown in Table 2, where we report results of tenfold cross-validation of the trained model. The accuracy of four categories is quite high (0.91, 0.87, 0.86, and 0.92). Kappa, which represents agreement between machine learners and human judges (Stemler, 2001), is moderate for aversive leadership (0.48) but is substantial or excellent for the other three categories (0.75, 0.71, and 0.80).

In applying these classifications to 4 million messages between editors, we found that a large proportion of leadership behaviors were performed by editors without formal leadership roles in Wikipedia (Table 3). For example, nonadministrators contributed 64% of directive leadership behaviors. In this study, we go beyond characterizing the types of leadership behaviors demonstrated in Wikipedia to examine the effects they have on their targets.

Analysis Strategy

We can measure the effects of different messages on people's general motivation by looking at the total number of revisions they make on any Wikipedia articles before and after receiving leadership messages. However, it is impossible to hand-code the millions of messages to identify which specific tasks these messages target, such as whether the message is about adding a photo to Article A or about changing the reference for Article B. Because there are too many potential categories, it is also not feasible to build machine learning to automatically categorize the messages. Therefore, Study 1 can test only Hypothesis 2 (effects on general motivation) but not Hypothesis 1 (effects on specific tasks).

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

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 that person messages in the following week. Similarly, experienced editors who produce good edits may cause others to send them transactional leadership messages, whereas those newcomers who produce

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 2: Creating Automatic Measurement for Leadership Behaviors Using Machine Learning

Training Sets	We hand-coded 500 messages into each of the four leadership behaviors to provide training data for the model. Messages could be assigned to multiple categories if they exhibited more than one leadership behavior. To assess the reliability of the coding, two human judges annotated 100 messages. The Cohen's Kappa measure of interjudge agreement averaged across the four categories was 0.82 (positive 0.81, negative 0.80, directive 0.79, social 0.88), which is very high (Stemler, 2001).
Representation of Messages (Feature set)	We used features based on domain knowledge, realizing that message senders tend to frequently use certain words and phrase patterns to express different intents. We identified 21 domain knowledge features: Strong/weak, positive/negative polarity words. Four features based on the combination of strength and polarity derived from the subjectivity lexicon of OpinionFinder (Wilson, Wiebe, & Hoffmann, 2009). • Strong positive adjectives. Seventeen strong positive adjectives used in praise, such as "excellent," "great," and "impressive." • Negation. Seventeen negation words and phrases (e.g., "not," "shouldn't," "doesn't"). • Negative jargon. Nineteen Wikipedia-specific negative words such as "vandalism" and "blocked." Causative/subjunctive verbs. Twenty-seven causative or subjunctive verbs including "make," "suggest," "recommend," "wish," and "need." • <You+modal>. Sentences starting with a pronoun "you" immediately followed by a modal word (e.g., "should,, "might," "must") or vice versa. • Acknowledgements. Phrase patterns of "thank you/thanks for." • Smiley. Textual expressions such as :),;). • Greetings. Greeting words/phrases, such as "hello," "congratulations," and "happy birthday." • He/she. Number of "he, him, his, she, her." • Length. Number of word tokens in a message. • Variants of the following words/phrases were included as a separate feature: "if you," "newsletter," "Wikiproject," "congrats," "welcome," and "please"+ verb.
Learning Algorithm	Support Vector Machine (Sebastiani, 2002)
Validation of the Measurement	• Accuracy: Transactional (0.91) Aversive (0.87) Directive (0.86) Person-based (0.92) • Kappa agreement between machine learning results and human coders: Transactional (0.75) Aversive (0.48) Directive (0.71) Person-based (0.80)

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 a biased estimation of the treatment effects.

To ameliorate the endogeneity problem, we use propensity score matching (PSM) to approximate randomization. With PSM it is possible to build experimental and control groups by balancing the groups on potential confounding factors. These confounding factors include the number of edits the editors made before, the number of messages they received or sent before, and their tenure in Wikipedia. PSM can effectively reduce the bias caused by these conditioning factors (Angrist & Krueger, 1999; Rosenbaum & Rubin, 1983).

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 3: Distributions of the Leadership Messages Among Administrators and Nonadministrators

	Administrators		Nonadministrators	
Number of People	1,723		131,848	
	Per person	Aggregate	Per person	Aggregate
Transactional leadership	154.7	267K	4.3	569K
Aversive leadership	155.9	269K	3.9	509K
Directive leadership	483.8	834K	11.4	1503K
Person-based leadership	244.0	386K	4.6	602K
Overall Wikipedia activities per person	16977.7		573.7	

However, because PSM balances only on measured variables, it cannot adequately control for all variables relevant to treatment.

Given that editors' prior experience is one important confounding factor for examining the effects of receiving different types of leadership messages, PSM will balance experimental and control groups on their prior experience. In other words, editors with similar experience in Wikipedia are compared. Therefore, Hypothesis 4 was not examined in Study 1.

In sum, we tested Hypothesis 2 and Hypothesis 3 in Study 1, examining the effects of receiving different types of leadership messages on recipients' total amount of contribution (i.e., a proxy of general motivation) and the moderating effects of the roles of message senders. We used propensity score matching to ameliorate the endogeneity problem. The results of Study 1 have been reported previously (Zhu, Kraut, & Kittur, 2012a).

Data Preparation

We restricted the analysis to registered Wikipedia editors who had edited any Wikiproject page at least once, as 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). 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. All the variables are described in Table 4.

Propensity Score Matching

PSM was involved in three steps, first to estimate the propensity score (i.e., the probability of receiving messages from others) from a set of conditioning variables. The variables we 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. Table 5 shows the results of estimating the probability of receiving messages (propensity score) with logistic regression, with six of the editors' previous activities as conditioning variables.

In the second step, we matched each editor who received leadership messages in a particular week (treatment group) with an editor who did not receive a message (control group) but who had the most similar propensity score based on the six behavioral indicators. Propensity scores allow researchers to control for many variables simultaneously by matching on a single scalar variable. At the end of the second step, we checked whether the treatment group and control group were well matched in terms of the conditioning variables in which we were interested. From Table 6, it can be seen that the bias was reduced over 90% for five of the six conditioning variables, indicating that the treatment group and the control group were well balanced.

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 4: Variables of Study 1

Variable Name	Definition
Dependent Variable	
General motivation (change)	We measured editors' general motivation by calculating 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. To alleviate the endogeneity caused by individual differences, we measure the contribution change after receiving the message. 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. Therefore, this variable is defined as $\ln(edits_{t+1}+1)-\ln(edits_{t-1}+1)$
Independent Variables	
Receive_msg	This dummy variable indicates whether the editor received any messages during the focal week. One indicates that the editor received at least one message, and 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 transactional leadership message, and zero indicates that the editor received no transactional leadership message. The following three variables are similar.
Aversive	This dummy variable indicates whether the editor received any message categorized as an aversive leadership message during the focal week.
Directive	This dummy variable indicates whether the editor received any message categorized as a directive leadership message during the focal week.
Person	This dummy variable indicates whether the editor received any message categorized as a person-based leadership message 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, and zero indicates that the editor received no messages from any administrator.
Admin X Transactional	This dummy variable indicates whether the editor received any messages categorized as a transactional leadership message from any administrator during the focal week. One indicates that the editor received at least one transactional leadership message from an administrator, and zero indicates that the editor received none. The other three interactions were constructed similarly.
Admin X Aversive	This dummy variable indicates whether the editor received any messages categorized as an aversive leadership message from an administrator during the focal week.
Admin X Directive	This dummy variable indicates whether the editor received any messages categorized as a directive leadership message from an administrator during the focal week.
Admin X Person	This dummy variable indicates whether the editor received any messages categorized as a person-based leadership message from an administrator during the focal week.

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 5: Estimating the Probability of Receiving Messages (Propensity Score) With Logistic Regression

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	

***$p < 0.0001$.

TABLE 6: Comparison Between Treatment Editors Who Received Messages in the Focal Week (treat) and Control Editors (ctrl) Before and After Propensity Score Matching (Full vs. Matched)

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

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. Results are shown in Table 7.

Results

To understand Table 7, one must first understand how to interpret the dependent variable. 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 the 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 prior week constant,

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 7: Regression Predicting the Effects of Leadership Behaviors on Subsequent Change in Editors

Dependent Variable *General motivation (change)*	Descriptive Statistics Mean	Std. Dev.	Model 1 Coef.	Std. Err.	Model 2 Coef.	Std. Err.	Model 3 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	.2511	.4336			.2290***	.0061	.1698***	.0071
Admin	.2264	.4185					.1584***	.0086
Admin X Transactional	.0657	.2478					.0278 **	.0118
Admin X Aversive	.0174	.1306					–.0945***	.0205
Admin X Directive	.1103	.3133					.0174	.0102
Admin X Person	.1090	.3117					.0579***	.0103
Number of observations				1,006,518				
Number of groups				503,259				

**$p < 0.05$.
***$p < 0.0001$.

the edits in the subsequent week increased approximately x%.

First, Model 1 in Table 7 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. However, receiving messages slows this decline.

Second, Model 2 demonstrates that different types of leadership behaviors differentially influenced subsequent motivation (Hypothesis 2). The effects of messages that were not one of the four leadership types resulted in a 33% increase in edits in the subsequent week. That 33% increase in people's motivation and contributions can be explained by considering 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. 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 of 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%. Therefore, Hypothesis 2 is confirmed.

Third, Model 3 demonstrates that messages sent by administrators were more influential than those sent by peers (Hypothesis 3). Receiving a nonleadership message from an administrator increased edits by 15% compared with receiving messages from nonadministrators. Transactional messages sent by administrators increased editing an additional 2.8% compared with those sent by peers. Conversely, aversive messages sent by administrators decreased editing an additional 9.5% compared with aversive messages sent by peers. Finally, person-based messages sent by administrators increased editing by 5.8% compared with those sent by peers. Hypothesis 3 is also confirmed.

Limitation of Study 1 and Motivation for Study 2

The first study has three limitations. First and most important, as previously indicated, one cannot conclude from correlational research that leadership behavior actually changes the behavior

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

of those who receive it. Although we have used sophisticated propensity score matching to try to equate pre-existing characteristics, some unmeasured variables—such as politeness or extraversion, which can potentially predict both the type of messages people receive and their subsequent behavior—can still undermine causal inferences. In Study 2, we randomly assigned someone either to receive or not receive a particular type of leadership message. By doing so, we ensured that, within the limits of chance, those two groups were equivalent on both measured and unmeasured variables before the intervention.

Second, Study 1 investigated only how leadership behaviors affected receivers' general motivation to work (e.g., total number of edits). It failed to examine how leadership affects people's performance on the specific tasks that the leadership behaviors explicitly target. As Hypotheses 1 and 2 suggest, different types of messages have different effects on people's efforts on focal tasks and general motivations. Study 2 examined how different types of leadership messages influence both general work motivation and specific task performance (Hypotheses 1 and 2).

Finally, Study 1 suggested that leadership messages have stronger effects when delivered by formal leaders (Hypothesis 3), but it failed to examine how effectiveness varies among people who receive them (Hypothesis 4). We tested this distinction more definitively in Study 2.

In sum, Study 2 aimed to resolve the limitation of Study 1 and examine Hypotheses 1, 2, and 4.

STUDY 2

We conducted a field experiment in Wikipedia. In the experiment, we randomly sent different types of leadership messages or no message at all to editors who had recently created new articles.

Participants

Research participants were the original authors of newly created Wikipedia articles. They were randomly selected without replacement via a computer script from Wikipedia's new article list.

Each new article was evaluated on several dimensions to ensure that potential leadership messages were relevant to the content. If the article was not relevant to at least one template, the author was excluded. For example, authors of new articles with nothing explicitly incorrect were excluded, because that editor could not be randomly assigned to receive or not receive aversive leadership. Similarly, editors of an article that contained nothing praiseworthy were dropped because the article could not randomly receive transactional message.

We included 703 editors in the experiment, which lasted from August to November 2011.

Experiment Design

We randomly assigned 80% of selected Wikipedia editors to receive a message, and the remaining 20% who did not receive a message served as a control group. All messages contained some common content (the base). The additional components—positive feedback, negative feedback, directive message, and a social message (including a social greeting and a social closing)—each had a 50% chance of inclusion. Positive feedback corresponds to transactional leadership; negative feedback corresponds to aversive leadership; directive message corresponds to directive leadership; and social message corresponds to person-based leadership. We used a 2 (positive feedback vs. not) × 2 (negative feedback vs. not) × 2 (direction message vs. not) × 2 (social message vs. not) between-subjects factorial design for the 80% who received a message. To understand the effects of different types of messages, we measured the users' contribution to the particular article on which we gave feedback (efforts on focal task) as well as their contributions to any Wikipedia articles (general motivation) over the following month.

Materials

All messages contained some or all of the following components.

> "[Social Greeting] + [Base Message] + [Positive Feedback] + [Negative feedback] + [Directive Message] + [Social Closing] + [Signature]."

Figure 1 is an example that contains all the components. All messages contained a base and

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

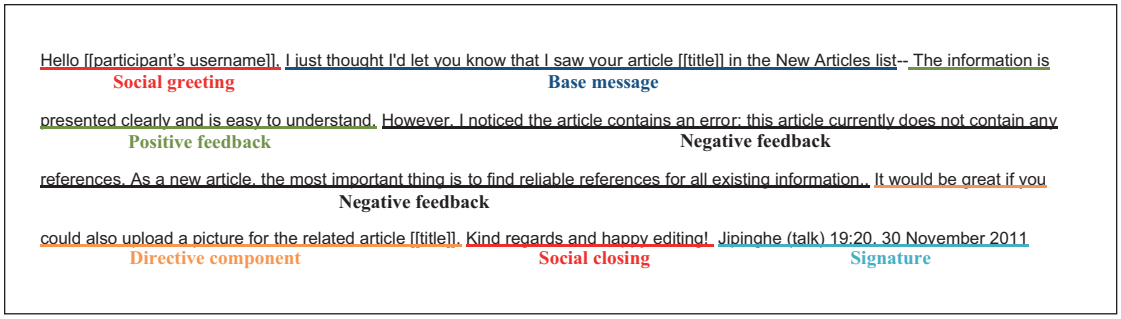

Figure 1. An example message containing all the elements.

signature. To provide experimental control, a computer script randomly decided whether to include the additional components: positive feedback, negative feedback, directive message, or a social message (social greeting plus social closing).

We created 12 templates for positive feedback, 10 templates for negative feedback, 9 templates for directive messages, 4 templates for social greeting, and 8 templates for social closing. Table 8 shows two examples of each message component, and Figure 1 shows an example of a message assembled from the components.

To generate different components, we used a script to run through the various templates in random order, asking the researcher if a specific positive or negative template applied to the article. This ensured that the aspect was both appropriate and randomly chosen. Note that the negative feedback only politely critiqued the editor's work by pointing out an error but was not directive, such as requesting that the editor make a particular change. In contrast, directive messages asked for the editor's help with improving a related article without being positive or negative about the new article that the user created. We used Suggestbot (Cosley, Frankowski, Terveen, & Riedl, 2007) to help find related articles that needed work.

Research Ethics

We designed this experiment with the twin goals of observing how different types of leadership messages naturally affect Wikipedia editors while at the same time minimizing potential risk to Wikipedia editor participants and the Wikipedia community as a whole.

First, we made sure that the leadership messages sent to Wikipedia editors who had created a new page were natural and appropriate. The researchers posting the messages were members of the New Page Patrol, a collection of Wikipedia editors who evaluate and comment on new articles. They both had experience editing in Wikipedia. Furthermore, all the component templates sent to editors were based on observations of messages on Wikipedia, suggestions by senior Wikipedia editors, and the Wikipedia civility guidelines. Thus, these messages are very similar to those that Wikipedia users might encounter in their everyday interactions on the Web site, although perhaps more polite.

In particular, negative feedback components in the experiment are milder than the messages categorized as aversive leadership sent between editors. In the "wild," some editors use intimidation, threat, and harsh language to decrease undesired behaviors from targets. Here are two examples: "If you continue in this manner you will be blocked from editing without further warning" and "Blech. This really needs [[WP:TNT]]," which is Wikipedia's jargon for "Blow it up and start over." In our experimental design, negative feedback consisted only of constructive criticism.

The experiment was approved by the Carnegie Mellon University Institutional Review Board, as well as the Wikipedia research committee. Information about the experiment was posted on public Wikipedia pages and received unanimous agreement of active discussants from the Wikipedia community (Wikipedia, 2013b).

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 8: Example Templates for Message Components

Component Type	Leadership Type	Template 1	Template 2
Social Opening	Person-based Leadership	Hi XX,	Hey XX,
Base Message		I'm posting this message on your talk page because you've recently created the new article XX --	I saw your article XX in the new articles list --
Positive Feedback	Transactional Leadership	The content seems well-organized.	There is a good number of citations and references.
Negative Feedback	Aversive Leadership	However, I noticed the article contains an error: this article currently does not contain any references. As a new article, the most important thing is to find reliable references for all existing information.	However, I noticed the article contains an error: the article does not contain any Wikilinks, and so doesn't follow Wikipedia style guidelines.
Directive Component	Directive Leadership	It would be great if you could also improve the related article XX.	It would be great if you could also clean up the related article XX.
Social Closing	Person-based Leadership	Happy editing! Hope your day is going well and you are having fun.	It's always nice to see users contributing to make Wikipedia better!

Analysis

The goal of the analysis was to measure the effects of leadership messages on participants' efforts on focal task and general motivation, moderated by the experience of receivers. Variables are described in Table 9.

Analysis strategy. Because the dependent variables (the number of edits editors made on particular target articles and other Wikipedia articles) are count data, and because editors might not log in to Wikipedia and have a chance to see the messages during the time window (one month after receiving the message), we analyzed the data using a zero-inflated negative binomial regression.

Zero-inflated negative binomial regression (Hall, 2004) is often used when the dependent variable is an upper-bounded count value and is overdispersed, with more zeros than predicted by a regular binominal distribution. The basic idea is that the excess zeros can be generated by a separate process that can be modeled independently. In our case, the goal was to predict whether reading the leadership messages changed participants' behavior. Some recipients might not have been influenced by the message because they were not persuaded by its content. However, others might have failed to log in recently and hadn't actually seen the leadership message meant for them. To model these two separate processes, the zero-inflated negative binominal analysis has two stages. In the first stage, we used a logit regression to predict the excess zero (i.e., the likelihood of not seeing the message). In the second stage, given the likelihood of being exposed to the message, we predicted the effects of leadership messages on the number of edits. Specifically, we used the following two

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 9: Variables of Study 2

Variable Name	Definition
Dependent Variables	
Performance on focal task.	To measure participants' performance on their focal task (which the leadership message specifically targets), we calculated the number of edits they made in the month after receiving a leadership message on the article that was the target of the message. Note that for participants who received a directive message asking them to improve a related article, efforts on focal task also included edits on that related article.
General motivation	To measure the effects of leadership messages on participants' general motivation to work, we calculated the number of edits on any Wikipedia articles excluding the focal article(s) that the leadership messages target.
Independent Variables	
Base message	This dummy variable indicates whether or not the participant received a base message. One indicates that the editor was randomly assigned to receive a base message, and zero indicates that the editor did not receive one from us.
Transactional	This dummy variable indicates whether the participant received a message with the positive feedback component (1) or without this component (0).
Aversive	This dummy variable indicates whether the participant received a message with the negative feedback component (1) or without this component (0).
Directive	This dummy variable indicates whether the participant received a message with the directive component (1) or without this component (0).
Person	This dummy variable indicates whether the participant received a message with the social component (1) or without this component (0).
Receiver is a newcomer	This dummy variable indicates whether the receiver is a newcomer (1) or not (0). We define newcomers as editors with less than six months' experience in Wikipedia and received fewer than four messages before receiving our message.
Newcomer X Base message	This variable indicates the interaction effects of receiver experience and message type. This variable is one when newcomers receive base message; otherwise, it is zero.
Newcomer X Transactional	This variable indicates the interaction effects of receiver experience and message type. This variable is one when newcomers receive a message with positive feedback element; otherwise, it is zero.
Newcomer X Aversive	This variable indicates the interaction effects of receiver experience and message type. This variable is one when newcomers receive a message with negative feedback element; otherwise, it is zero.
Newcomer X Directive	This variable indicates the interaction effects of receiver experience and message type. This variable is one when newcomers receive a message with directive feedback element; otherwise, it is zero.
Newcomer X Person	This variable indicates the interaction effects of receiver experience and message type. This variable is one when newcomers receive a message with social elements; otherwise, it is zero.

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 10: Descriptive Statistics of Participants

	Newcomers	Experienced editors
Number of people	132	473
Efforts on focal task Unit: # of edits	*M* = 2.1; *SD* = 7.6	*M* = 1.3; *SD* = 3.7
General motivation Unit: # of edits	*M* = 128; *SD* = 25	*M* = 403; *SD* = 959
# of people receiving messages	106	362
# of people receiving positive feedback	45	183
# of people receiving negative feedback	48	164
# of people receiving directive feedback	47	126
# of people receiving social feedback	61	194

estimates of editors' recent activity to predict the likelihood of their seeing the message.

- *Number of edits one day before receiving our message.* The more edits the participant did in the 24 hours before we sent them messages, the more active they were and the more likely they were to have seen our message.
- *Number of days between last edit and receiving our message.* Similarly, we included the number of days between the last edit the participant made and the time we sent our message.

Results

The descriptive statistics of participants in different condition are shown in Table 10. The results of zero-inflated negative binominal regression are shown numerically in Table 11 and graphically in Figures 2a through 2d. The error bars in those figures indicate a 95% confidence internal. We report the main effects of receiving a particular type of leadership component. For example, in the figures, the condition "with transactional components" includes "transactional" and "transactional + aversive" and "transactional + directive," etc.; the condition "without transactional components" includes "base," "aversive," and "directive," etc. We did not find significant interaction effects between different types of leadership components.

The bottom panel of Table 11 indicates that the likelihood ratio test of alpha = 0 is significantly different from zero. This suggests that our data are overdispersed and that a zero-inflated negative binomial model is more appropriate than a zero-inflated Poisson model. The Vuong test suggests that the zero-inflated negative binomial model is a significant improvement over a standard negative binomial model. These results suggested that we used the right statistical model.

The top panel of Table 11 shows analyses testing Hypotheses 1, 2, and 4. Model 1 tested whether receiving a leadership message led editors to edit more on the article that the leadership message targeted (focal task). Model 2 tested whether receiving a leadership message increased editors' activities in general. Each coefficient represents the change in the log of the expected number of edits the editor would produce when increasing the independent variable by one unit, when other variables in the model were held constant at zero. For ease of interpretation, we also included the change in edit counts in the original units. Thus, the intercept indicates that "old-timers" who received no messages (baseline) could be expected to make 1.27 (e^.24) edits to the focal article. Newcomers made edits 2.44 ((e^0.89)) times compared with experienced editors because the coefficient of the variable of *Receiver is newcomer* is 0.89. Therefore, newcomers who received no messages made 3.10 edits (1.27*2.44) to the focal article.

For experienced editors, receiving any type of leadership message had no significant impact on their subsequent editing behavior, either for the specific articles on which we

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 11: Effects of Leadership Messages on Focal Task (Particular Articles That Our Leadership Message Targets) and General Motivation (Contributions on any Wikipedia Article Excluding the Particular Articles That the Leadership Message Explicitly Targets)

	Dependent Variable: Focal Task		Dependent Variable: General Motivation	
	Model 1		Model 2	
Predictors	Coef	S.E. Change in # of edits	Coef	S.E. Change in # of edits
Intercept	.24	(.26) N/A	6.3**	(.17) N/A
Base message	.29	(.34) 1.34	-.090	(.27) 0.91
Transactional	.10	(.25) 1.11	-.051	(.19) 0.95
Aversive	.04	(.25) 1.04	-.16	(.20) 0.85
Directive	-.10	(.26) 0.90	-.038	(.20) 0.96
Person	.06	(.25) 1.06	-.13	(.19) 0.88
Receiver is newcomer	.89	(.65) 2.44	-3.8**	(.46) 0.02
Newcomer X Base message	-2.1**	(.94) 0.12	-.67	(.69) 0.51
Newcomer X Transactional	-.47	(.73) 0.63	1.3**	(.54) 3.67
Newcomer X Aversive	1.4**	(.67) 4.06	-.25	(.54) 0.78
Newcomer X Directive	2.2**	(.68) 9.03	.58	(.51) 1.79
Newcomer X Person	.23	(.71) 1.26	2.2**	(.50) 9.03
Inflate				
Number of edits during one day before receiving our message	-..03	(.09)	-20	(14580)
Number of days between last edit before receiving our message and the time they receive the message	.48**	(.14)	.36**	(.06)
Alpha	3.70		2.73	
Likelihood-ratio test of alpha = 0	chibar2(01) = 624; Pr>=chibar2 = 0.0000		chibar2(01) = 3.9e+5; Pr>=chibar2 = 0.0000	
Vuong test of zinb vs. standard negative binomial	z = 3.60 Pr>z = 0.0002		z = 1.5 Pr>z = 0.07	

*$p < 0.1$.
**$p < 0.05$.

gave feedback (focal task) or for any other articles (general motivation). For newcomers, the effects are significant. Therefore, Hypothesis 4 is supported.

Model 1 shows that leadership messages had significant effects on newcomers' subsequent editing of the target, as our hypotheses predicted. Whereas receiving a base message reduced the amount that newcomers changed the target article compared with receiving no messages, receiving aversive and directive leadership messages increased their editing of the target article. The coefficient of Newcomer X Aversive was 1.4, indicating that newcomers who received aversive leadership messages were estimated to make edits on focal articles approximately four times compared with newcomers who did not receive aversive leadership messages. The coefficient of Newcomer X Directive component was 2.2, indicating that newcomers who received directive messages were estimated to make edits on focal articles approximately nine times

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

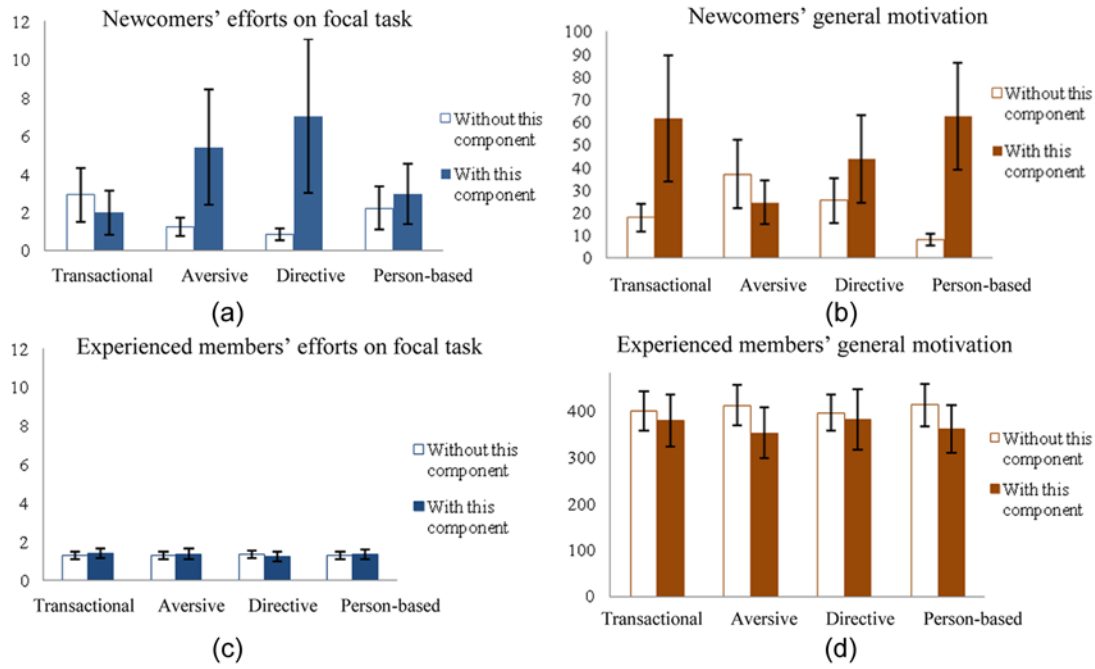

Figure 2. (a) The effects of receiving messages on newcomers' efforts on focal task. (b) The effects of receiving messages on newcomers' general motivation. (c) The effects of receiving messages on experienced members' efforts on focal task. (d) The effects of receiving messages on experienced members' general motivation.

compared with newcomers who did not receive directive messages. Transactional and person-based leadership message did not have effects on local tasks. The results are shown graphically in Figure 2a. Hypothesis 1 is confirmed.

Results of Model 2 confirm our Hypothesis 2 about the effects of leadership messages on editors' general motivation. In contrast to Model 1, aversive and directive leadership messages do not have effects on general motivation. Instead, transactional and person-based leadership substantially increase newcomers' general work motivation. The coefficient of Newcomer X Transactional is 1.3, indicating that positive feedback causes 3.67 times change in number of edits for newcomers. The coefficient of Newcomer X Person-Based is 2.2, indicating that messages with a social component cause 9.03 times change in number of edits for newcomers. The results are also graphically shown in Figure 2b and are consistent with Hypothesis 2, except that aversive leadership does not have significant negative effects. However, in Study 1 we found that aversive leadership reduced motivation. Remember that the aversive leadership messages in Study 2 were intentionally designed to be milder than aversive leadership messages actually sent between Wikipedia editors, as in Study 1.

DISCUSSION

The results of two studies basically confirmed our hypotheses:

1. Aversive leadership and directive leadership increase recipients' efforts on specific tasks the leadership targets, whereas transactional leadership and person-based leadership have no effects on performance on a specific task.
2. Transactional leadership and person-based leadership increase people's general motivation to work, whereas aversive leadership and directive leadership do not.
3. The effects are stronger when senders are formal leaders.
4. The effects are stronger when receivers are newcomers.

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 12: The Effects of Leadership Messages on the Number of Words Added on the Focal Article

	Dependent Variable: The Number of Words Added to the Focal Articles	
Predictors	Coef.	S.E. Change
Intercept	5.4**	(.42) 221
Base message	-.95	(.51) 0.39
Transactional	-.07	(.35) 0.93
Aversive	.51	(.37) 1.67
Directive	-1.0**	(.38) 0.37
Person	-.25	(.39) 0.78
Receiver is newcomer	-.94	(.91) 0.39
Newcomer X Base message	-.07	(1.2) 0.93
Newcomer X Transactional	-.07	(1.2) 0.93
Newcomer X Aversive	-.72	(.98) 0.49
Newcomer X Directive	2.4**	(.93) 11.0
Newcomer X Person	1.3	(1.2) 3.67

*$p < 0.1$.
**$p < 0.05$.

Experienced Members' Reaction

Although we predicted that the effects should be stronger for newcomers because they are particularly susceptible to influence, we were still surprised to see that in Study 2, the messages had no significant effects at all on experienced members. When we dug deeper about the participants' editing behaviors on focal articles in addition to calculating the raw counts of edits, we even found evidence that experienced members went in the opposite direction that our leadership messages wanted them to, as if being influenced by a counterforce.

First, we examined the total number of words added to the focal articles (see Table 12). Similarly, we used zero-inflated negative binominal regression to measure the effects of different types of leadership messages. Experienced editors who received a directive message even added fewer words compared with the condition in which they did not receive a directive message: The expected number of words added to focal articles decreased by 63% (Coef. = –1, Change = 0.37) when they received a directive message. In contrast, the newcomers added 10 times more words when they received a directive message.

Second, we examined the likelihood of participants' revisions being "self-removed." Removing one's own work indicates that the person accepts the external suggestions and is willing to revise and refine the previous work. To quantify the effects, we conducted a revision-level survival analysis. We defined the "death" of a particular revision as being the case when more than 50% of the words were removed by the same editor. The random-effect model was applied to control the intrapersonal similarity when the same person did multiple revisions. The results are represented as a hazard ratio in Table 13, which can be interpreted as the ratio change of the likelihood of being self-removed. The results show that aversive leadership reduced the likelihood of experienced users removing their previous edits by 61%, whereas newcomers were 550% more likely to remove and refine their own edits after receiving aversive leadership.

We also found some qualitative evidence from the messages the participants sent back to the researchers' user pages. For example, some participants wrote to us and said the following:

> "Well, er, yes, I am not new here and the stub tag was intended as a cheerful

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

TABLE 13: The Effects of Leadership Message on the Likelihood of Being Self-Removed

	Dependent Variable: The Likelihood of Being "Self-Removed" for the Revisions on the Focal Articles	
Predictors	Haz . Ratio	S.E.
Intercept	.02**	(.005)
Base message	1.5	(.95)
Transactional	.90	(.42)
Aversive	.39*	(.21)
Directive	.77	(.46)
Person	1.5	(.68)
Receiver is newcomer	2.6	(2.2)
Newcomer X Base message	.70	(.90)
Newcomer X Transactional	.80	(.63)
Newcomer X Aversive	6.5**	(6.0)
Newcomer X Directive	.44	(.36)
Newcomer X Person	.90	(.68)

*$p < 0.1$
**$p < 0.05$.

acknowledgement of the effort's insufficiency." – P1.

"There are plenty of external references on that page for John Hess (journalist) for the information given. I can show you plenty of pages that do not have any external references - worry about those first…" – P2.

"You're still wet behind the ears and have too little experience to have perspective." – P3.

We believe that experienced members might have psychological reactance to our messages. *Psychological reactance* was originally proposed by Brehm, in which a person has a negative emotional response in reaction to being persuaded and thus chooses the option that is being advocated against (Brehm, 1966). Experienced members might perceive aversive leadership and directive leadership as a challenge to their knowledge and expertise (P1 and P2), especially when noticing that the message senders have less experience than themselves (P1 and P3). Previous research shows that when people perceived feedback as self-threatening, they might avoid exposure to the feedback or even abandon the entire task (Kluger & DeNisi, 1996). It is possible that experienced editors chose not to follow what their "newbie" colleagues suggested, so as to preserve positive self-belief about their expertise. The results suggest that although any member can try to exhibit leadership behavior with others in Wikipedia, the relative status of the sender still might matter. Therefore, to ensure the effectiveness of shared leadership on senior community members, it is probably better to have other senior community members deliver the leadership messages.

Theoretical Contribution

Our paper investigates the shared leadership model in an online community setting, a condition that prior work has not studied. Our results confirm prior theory in this new condition by demonstrating the prevalence and effectiveness of shared leadership in Wikipedia. Our results suggest that the shared leadership model can not only effectively manage dozens to hundreds of employees in offline organizations but also can scale to managing millions of volunteers who do not know each other face-to-face, have differing experience and commitment, and interact only in an online community.

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

Practical Implications

Our results provide practical implications to better manage Wikipedia and other Wikipedia-like online communities. Our results demonstrate the trade-off of different types of leadership behavior on recipients' focal task performance and general work motivation. Aversive leadership and directive leadership benefit focal task performance but do not affect general work motivation, whereas transactional leadership and person-based leadership can positively influence general work motivation but do not have effects on focal tasks. Practitioners can consider their primary goal (e.g., accomplishing the current task or encouraging long-term motivation) when designing interfaces and mechanisms to encourage certain types of shared leadership behaviors. For example, to encourage general motivation, interfaces and mechanisms should be designed to make it easier for members to connect with, reward, and express their appreciation for one another. Our findings also reveal opportunities to design computer-supported shared leadership systems, and suggest that automatically generated leadership messages might be particularly effective for influencing the behaviors of newcomers in the communities.

Generalization

In this study, we examined the leadership behaviors in Wikipedia. Note that Wikipedia is different from many other online communities. For example, Wikipedia is a peer production community, and building the main products—encyclopedia articles—does not require in-depth domain knowledge. It remains unproven whether our results about shared leadership can be applied to other non-production-focused communities (e.g., those that focus on social bonding, such as Facebook) or production communities, which rely heavily on domain experts (e.g., open-source projects). We expect that further comparative studies can confirm the extent to which these findings are generalizable.

Contribution to Social Media Research

Our paper provides a unique contribution to social media research by combining social science theories (i.e., shared leadership model) and various quantitative methods (i.e., machine-learning coding, statistical analysis of large-scale observational data, and field experiment) to understand the underlying mechanisms in social media systems.

CONCLUSION

We conducted two studies in Wikipedia to examine how different types of leadership behavior affect receivers' focal task performance and general work motivation, moderated by receivers' prior experience and senders' role. Our findings extend shared leadership theories, contribute new insight into important management mechanisms in Wikipedia, and provide implications for practitioners to design more effective and successful online communities.

KEY POINTS

- Aversive leadership and directive leadership increase recipients' efforts on specific tasks that the leadership targets.
- Transactional leadership and person-based leadership increase people's general motivation to work.
- The effects are stronger when senders are formal leaders.
- The effects are stronger when receivers are newcomers. Experienced members might have psychological reactance to leadership messages sent from newcomers.

REFERENCES

Alexa Internet. (2013). *Five-year traffic statistics for Wikipedia.org*. Retrieved September 27, 2013, from http://www.alexa.com/siteinfo/wikipedia.org?range=5y&size=large&y=t

Angrist, J. D., & Krueger, A. V. (1999). Empirical strategies in labor economics. *Handbook of Labor Economics*, *3*, 1277−1366.

Ashforth, B. E., & Saks, A. M. (1996). Socialization tactics: Longitudinal effects on newcomer adjustment. *Academy of Management Journal*, *39*, 149−178.

Bass, B. M. (1990). *Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications* (3rd ed.). New York, NY: Free Press.

Brehm, J. W. (1966). *A theory of psychological reactance*. Oxford, UK: Academic Press.

Burke, C., Stagl, K., Klein, C., Goodwin, G., Salas, E., & Halpin, S. (2006). What type of leadership behaviors are functional in teams? A meta-analysis. *Leadership Quarterly*, *17*, 288−307.

Burns, J. M. (1978). *Leadership*. New York, NY: Harper & Row.

Cassell, J., Huffaker, D., & Tversky, D. (2006). The language of online leadership: Gender and youth engagement on the Internet. *Developmental Psychology*, *42*, 436–449.

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014

Choi, B. R., Alexander, K., Kraut, R. E., & Levine, J. M. (2010). Socialization tactics in Wikipedia and their effects. In *CSCW'10: Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (pp. 107−116). New York, NY: ACM Press.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591−621.

Cosley, D., Frankowski, D., Terveen, L., & Riedl, J. (2007). SuggestBot: Using intelligent task routing to help people find work in Wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces* (IUI '07; pp. 32−41). New York, NY: ACM Press.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*, 627−668.

Fleishman, E. A. (1953). The description of supervisory behavior. *Journal of Applied Psychology*, *37*(1), 1–6.

French, J., & Raven, B. H. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies of social power* (pp. 150−167). Ann Arbor, MI: Institute for Social Research.

Hall, D. B. (2004). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, *56*, 1030−1039.

Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership. *American Psychologist*, *49*, 493−504.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*, 349−371.

Katz, D., Maccoby, N., & Morse, N. C. (1950). *Productivity, supervision, and morale in an office situation, Part I*. Oxford, UK: Institute for Social Research.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254−284.

Kraut, R. E., & Resnick, P. (2012). *Evidence-based social design: Mining the social sciences to build online communities*. Cambridge, MA: MIT Press.

Luther, K., & Bruckman, A. (2008). Leadership in online creative collaboration. In *CSCW'08: Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (pp. 343−352). New York, NY: ACM Press.

Luther, K., Caine, K., Ziegler, K., & Bruckman, A. (2010). Why it works (when it works): Success factors in online creative collaboration. In *Proceedings of the 16th ACM International Conference on Supporting Group Work* (GROUP '10, pp. 1−10). New York, NY: ACM Press.

Misiolek, N. L., & Heckman, R. (2005). Patterns of emergent leadership in virtual teams. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (pp. 49–58). Washington, DC: IEEE Computer Society.

Pearce, C. L., & Conger, J. A. (2002). *Shared leadership: Reframing the hows and whys of leadership*. Thousand Oaks, CA: SAGE.

Pearce, C. L., & 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), 172–197.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41−55.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, *34*, 1−47.

Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation*, *7*(17), 137−146.

Thoresen, C. E., & Mahoney, M. J. (1974). *Behavioral self-control*. New York: Holt, Rinehart and Winston.

Web Technology Surveys. (2013). Usage of web servers for websites. Retrieved July 11, 2013, from http://w3techs.com/technologies/overview/web_server/all

West, S. (2010). Wikipedia's evolving impact. *Slideshow presentation at TED2010*. Retrieved November 7, 2013, from http://upload.wikimedia.org/wikipedia/meta/3/3a/TED2010%2C_Stuart_West_full_presentation_updated_with_January_data.pdf

Wikipedia. (2013a). Wikipedia statistics. Retrieved September 27, 2013, from http://en.wikipedia.org/wiki/Wikipedia:Statistics

Wikipedia. (2013b). Research talk: Effects of feedback on participation in Wikipedia. Retrieved July 11, 2013, from http://meta.wikimedia.org/wiki/Research_talk:Effects_of_Feedback_on_Participation_in_Wikipedia

Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, *35*, 399−433.

Yoo, Y., & Alavi, M. (2004). Emergent leadership in virtual teams: What do emergent leaders do? *Information and Organization*, *14*, 27−58.

Yukl, G. (1998). *Leadership in organizations*. Upper Saddle River, NJ: Prentice Hall.

Zhu, H., Kraut, R. E., & Kittur, A. (2012a). Effectiveness of shared leadership in online communities. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 407−416). New York, NY: ACM Press.

Zhu, H., Kraut, R. E., & Kittur, A. (2012b). Organizing without formal organization: Group identification, goal setting and social modeling in directing online production. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 935−944). New York, NY: ACM Press.

Haiyi Zhu is a PhD student in the Human-Computer Interaction Institute at Carnegie Mellon University, where she obtained a master's in human-computer interaction in 2012.

Robert E. Kraut is the Herbert A. Simon Professor of Human-Computer Interaction in the Human-Computer Interaction Institute, School of Computer Science, Tepper School of Business, and Center for the Future of Work, Heinz College at Carnegie Mellon University. He obtained a PhD in social psychology from Yale University in 1973.

Aniket Kittur is an assistant professor in the Human-Computer Interaction Institute at Carnegie Mellon University. Kittur obtained a PhD in cognitive psychology/quantitative from the University of California, Los Angeles, in 2007.

Date received: July 15, 2013
Date accepted: November 13, 2013

Downloaded from hfs.sagepub.com at CARNEGIE MELLON UNIV LIBRARY on August 18, 2014