

## **Retrospective versus Prospective Evidence for Promotion: The Case of Wikipedia**

Ben Collier  
Moirra Burke  
Aniket Kittur  
Robert Kraut

Carnegie Mellon University

### Abstract

Wikipedia is a large, online encyclopedia written and edited by volunteers. With a distributed work force of thousands of editors autonomously making changes, coordination in Wikipedia is a challenge. This paper examines how the process used to promote some editors to become administrators selects those who have the skills and temperament to act like managers in an organization without formal management roles. Studies of promotions suggest they have two complementary functions: (1) providing workers with incentives and rewards to deliver the effort and performance valued by the organization and (2) selecting workers for roles where they can contribute effectively to the organization. The incentive function leads organizations to promote based on retrospective indicators of past achievement, while the selection function leads them to promote based on prospective predictors of future job performance. We examined 1502 promotion decisions in Wikipedia to determine the influence of retrospective and prospective evidence for promotion. Two studies – a qualitative analysis of reasons cited in promotion decisions and a quantitative, policy-capture analysis relating users’ historical editing patterns to promotion success – show that Wikipedia bases its promotion decisions on both prospective and retrospective evidence. These findings are consistent with the need in Wikipedia to recruit and select individuals without formal authority or explicit hierarchy to do the difficult job of coordinating the work of others. A comparison of the two studies, however, suggests that while Wikipedians talk about promoting editors based prospective evidence, particularly their knowledge of policy and ability to help with common chores, data from the policy capture study indicates they do not, in fact, use these criteria. We discuss the implications of these findings for theories of promotion, policy capture, and online production communities.

## Introduction

One of the largest concerns in the organizational behavior literature is how social systems organize, coordinate, and achieve collective outcomes. While bureaucratic forms of organizing have received dominant attention in the literature (Blau, 1955; March & Simon, 1958), community forms of organizing have received much less attention (Stern & Barley 1996).

Community forms of organizing have a long history from kibbutzim and cooperatives (Ingram & Simons, 2000; Rothschild & Russell, 1986; Simons & Ingram, 1997) but have more recently have been given visibility as production communities have organized in a distributive setting to produce goods such as scientific knowledge<sup>1</sup>, art<sup>2</sup>, academic resources<sup>3</sup>, and software<sup>4</sup>. As production communities become more prominent and evolve as organizations, it becomes more important to not only understand how they motivate contributors and coordinate activities, but also how they govern and promote leaders within their community. By diversifying our knowledge of organizational processes by understanding a variety of organizational forms, we increase our knowledge of solutions society can leverage for organizing and tools available for success (Romanelli, 1991).

This research takes a step forward by filling in gap in research on governance and promotion in production communities (O'Mahony & Ferraro 2007). While production communities often consist of thousands of contributors, the community must enact a decision model for selecting contributors to move to positions that focus on macro level coordination, policy making, developing community norms, and achieving the collective goals. Production communities have been historically characterized as a meritocracy in which members are given status based on their

---

<sup>1</sup> <http://www.sciencecommons.com>

<sup>2</sup> <http://www.creativecommons.com>

<sup>3</sup> <http://www.opentextbooks.org>

technical contributions to the project; however, technical expertise may not be the most suitable promotion criteria for organizational building (Kogut & Metiu, 2001; Lee & Cole, 2003; Raymond, 1999). O'Mahony & Ferraro (2007) found that early on in an open source software organization the community selected leaders based on their retrospective technical contributions as developers and technical goals for the project, but later as it transformed into a more solidified organization it required leaders that had more prospective visionary goals and organizational building capabilities. In the first part of this paper examines how community members of Wikipedia, an open online encyclopedia, publicly discuss and vote for candidates for promotion using qualitative coding of the promotion discussions. We find that while the community does give significant attention to retrospective attributes for the candidate, they primarily focus their discussion around the prospective attributes of each candidate. In the second part of the paper we examine the empirical behavioral history of each candidate to determine which factors lead to promotion. We find that retrospective indicators are highly significant, and prospective indicators carry much less predictive power than the community may realize.

### **Governance in Production Communities**

Community forms of production are characterized in contrast to markets and hierarchies (Adler, 2001) as achieving a common goal while minimizing the use of bureaucratic elements and authoritarian division of labor (Roshchild-Whitt, 1979). In an offline setting community production has a history in kibbutz (Ingram & Simons, 2000) and cooperatives (Rothschild-Whitt, 1979) and has sociological roots dating back to medieval craft guilds (Kieser, 1989). As summarized by O'Mahony & Ferraro (2007), modern production communities differ from

---

<sup>4</sup> <http://www.sourceforge.net>

community forms previously research in three ways: (1) modern forms are not associated with a single employer or workplace, (2) unlike online communities, production communities must integrate the contributions of its members to be aligned with a common product or goal, which can lead to heightened interdependencies and coordination tasks, and (3) production communities own the output of their efforts collectively and work towards that goal outside of their own employment.

Production communities often find tensions in the contrast between their ideology of collective goods and openness, and the economic markets in which they operate (Simons & Ingram, 1997), and organizational tensions between the status and promotion of their members based on production merit rather than social ties or authority (Ingram & Simons, 2000; Kollock, 1998; O'Mahony & Ferraro, 2007; Stewart, 2005). In hierarchical organizations the goals and processes for the organization are typically developed from the top down based the organizations interests, however, in production communities democratic decision making is often a community ideal and thus the decision making is a very public and inclusive process (Adler, 2001). Thus, two social dilemmas exist in production organizations, (1) how does the community determine the structure of the decision making processes in the organization and (2) how do the individual community members carry out the democratic ideals established?

Research in online production communities have examined how the organization motives contributors (Roberts et al., 2006), manage members (Shah, 2006), establish joining processes (von Krogh et al., 2003), and develop status differences between members (Stewart, 2005). Much less attention has been given to how production communities govern and select leadership (O'Mahony & Ferraro, 2007). In this paper we will give a background on the administrators

within the production community of Wikipedia, and address this gap in research through both qualitative and quantitative analysis of the selection processes and democratic decision making of its members.

### **Production Communities: The Case of Wikipedia**

Wikipedia is a large, open source, online encyclopedia written and edited by volunteers. As of December 2007 over 75,000 editors had contributed over 2.1 million articles to the English version alone. Smaller but active Wikipedia versions exist in German, French, Japanese, Polish and other languages. Although Wikipedia is written and edited by volunteers and is not supervised by a professional staff, evidence suggests that the quality in Wikipedia is comparable to that of the Encyclopedia Britannica (Giles, 2005). Although organizational scholars have asked why volunteers contribute to Wikipedia and similar environments where unpaid volunteers contribute content, they have written less about how these highly decentralized environments coordinate work to produce high quality products (Bryant, Forte, & Bruckman, 2005; Roberts, Hann & Slaughter, 2006). Wikipedia as an organization is substantially less hierarchical than a conventional production company. By contrast, Encyclopedia Britannica has a CEO and other corporate officers, an editor-in-chief, a scholarly advisory board, and a staff of 19 full-time editors who vet and supervise the work of over 4000 contributors.

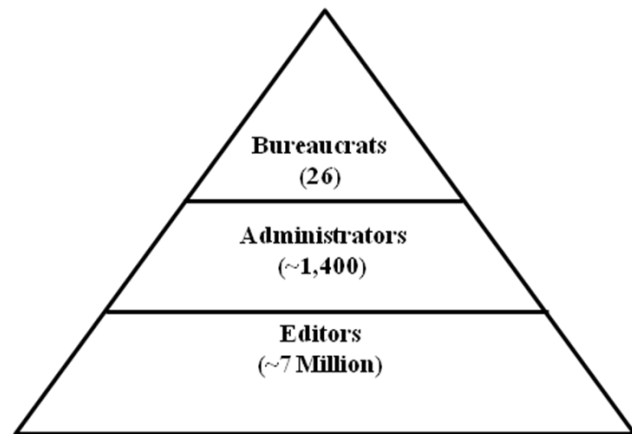
While any one may alter an article page within Wikipedia, a vast majority of changes made to Wikipedia is done by registered users. Within the community of registered users of Wikipedia, anyone who makes changes to articles may be considered an editor. Editors makes up the largest proportion of Wikipedia users, recently approaching 7 million members. Administrators are promoted from the group of editors after going a stringent week long democratic process in

which users discuss and vote on their fit for the Administrator position. From within the Administrator position, a very select group of users are then promoted to Bureaucrats, and have a very public role as the policy developers, mediators, and community leaders. The focus of this paper will be on the promotion of Editors to Administrators.

Editors in Wikipedia may add content, make changes, delete content, and change an article back to a previous version if unwanted changes are made by another editor (commonly referred to as vandalism<sup>5</sup>). Editors often maintain “watch lists” of articles they have an interest in developing so that if any changes are made to those articles by other editors they can review the changes and decide if they are of sufficient quality and have a neutral point of view as required by

encyclopedia standards. An editor’s role in Wikipedia is characterized by developing the content, and an editor’s merit is often judged by bringing an article up to peer-reviewed quality standards in Wikipedia such as that of a “Good Article”, or the gold standard in Wikipedia a “Featured Article” which will at some point be featured on the front page of Wikipedia.

**Figure 1: Organizational Structure**



To help coordinate work in Wikipedia in the face of a large and largely unsupervised workforce, Wikipedia has developed an elaborate system of policies that lay out Wikipedian values such as civility and neutral point of view and work procedures such as those for deleting pages, resolving disputes and electing administrators (Butler, Joyce, & Pike, 2008). It also empowered a class of

administrators with special tools to enforce some of the policies. For example, these administrators can ban other editors from working in the encyclopedia, protect pages from changes, and often serve as formal and informal mediators in resolving disputes. Administrators in Wikipedia differ from regular editors in two main ways. First, they have additional tools for page deletion and user maintenance, and regular editors request assistance from administrators at a public notice board for such issues as vandalism, page protection, inappropriate usernames, conflict between users, and page deletion. Second, administrators have elevated social status, as trusted editors who understand policy and its application. Though their position is ostensibly a lowly, janitorial one (as symbolized by their logo of a bucket and a mop), administrators in practice are well regarded as acting in the community's best interest and having enough experience to use their tools wisely. Approximately 2700 editors have been nominated for administrator status since 2001 with an overall promotion success rate of 53%. However, the process has gradually grown more rigorous, dropping from a 75.5% success rate through 2005 to 42% in 2006 and 2007, and some early administrators have expressed doubt that they would pass muster if their promotion debate were held today (Forte & Bruckman, 2008).

### **Promotion in Production Communities**

According to Milgrom and Roberts (1992), promotions serve two roles in an organization. First, they are part of a reward and incentive system to induce employees at one level of an organization to work hard in the organization's interest so that they can be promoted to positions of more power and responsibility. Second, they are part of a selection process in which people are vetted and assigned to the roles where they can best contribute to the organization's

---

<sup>5</sup> <http://en.wikipedia.org/wiki/Wikipedia:Vandalism>

performance. This paper examines the criteria that Wikipedia uses to promote regular editors to administrator status and the way in which the relatively flat organizational structure and large, decentralized workforce in Wikipedia influence these criteria.

Practitioners, organizational scholars and economists alike agree that rational contest models of promotion, which focus on the quality of candidates' prior job performance, are a partial explanation for who will be promoted (Turner, 1960). In a contest model, candidates at one level of a corporate hierarchy compete against each other for promotion to the next level based on their job performance in the lower position. "What makes the greatest difference in getting ahead . . . is performance on the job and adding value . . . People compete with each other in an open and fair contest for advancement, and victory comes to those who demonstrate the greatest accomplishments" (Ng, Eby, Sorenson, & Feldman, p. 369). Contest models are consistent with common organizational practices of counting sales commissions, journal publications or 3-point shots as criteria for promotion among sales organizations, professors and National Basketball Association players, respectively. Scholars often use these rational, contest models as the foil when discussing models of promotion that appear less rational and that emphasize factors less relevant to job performance, including height (Melamed & Bozionelos, 1992), good looks (Hamermesh & Biddle, 1994), gender (Maume, 1999) or social ties (Burt 2001) or other types of sponsorship (e.g., Hargen & Hagstrom, 1967).

Whether promotion functions as an incentive or a selection mechanism is likely to determine the types of evaluation criteria managers use in making promotion decisions. To the extent that promotions serve as incentives, managers should look for evidence that candidates for promotion have performed well in the past, independent of the degree to which prior job performance

predicts performance in a subsequent one. Because the promotions are being used to reward past behavior and encourage effort and work quality that the organization judges as valuable, managers should base promotions on the type of behavior the organization wants to encourage, whether or not this behavior would serve the organization or candidate when the candidate ascends to a new position. For example, in a production environment like a factory, managers should promote the most productive workers to supervisor positions. In this sense, the criteria that managers use to make promotion decisions should be retrospective or backward-looking; they should be looking for evidence of good performance in the prior job.

This retrospective use of evidence is a partial explanation for the "Peter Principle," in which employees rise to their level of incompetence (Peter & Hull, 1969). People who are good at one job may not necessarily be good in the job into which they are promoted. If promotion candidates are being judged on a set of behavior, such as their individual productivity, for a position that requires a different set of knowledge, abilities, and skills then they will not necessarily be well qualified for the new position. This might be the case, for example, when professors are promoted to be department heads based on the quality of their research or basketball players promoted to be coaches based on their athletic talents; the skills needed for the more managerial positions of department head or coach may have had little to do with the candidate's successes as academic or athletic superstars.

In contrast, to the extent that promotions serve as a selection device to identify appropriate candidates for higher-level, more difficult or more valuable organizational roles, then managers should use prior job performance primarily as evidence of future performance in these new roles. In this view, the promotion view is a prediction task, and the promotion criteria that managers

use should be prospective or forward-looking, rather than retrospective, as they try to use prior job performance to predict performance in the new job.

Retrospective and prospective criteria for promotion are not necessarily mutually exclusive; in many cases some behavioral criterion for promotion may serve both as an incentive to encourage incumbents in the current job and to predict performance in a future one. Thus behaviors that indicate conscientiousness or willingness to work hard may generalize well, and using them as promotion criteria is valuable both as an incentive and as prediction. Many human-capital factors commonly used as criteria for promotion—such as work ethic or educational level (e.g., Ng et al, 2005)—have both functions. However, other potential promotion criteria, such as those indicating individual productivity in a production task (e.g., papers published in academia or points scored in basketball) may not be relevant to promotion to a more managerial position.

In this paper we use data from over 1500 promotion decisions elevating regular editors in the online encyclopedia Wikipedia to administrator status. Wikipedia is an interesting and important organizational phenomenon in its own right. Moreover, because all the work done on the encyclopedia and all the communication among participants is archived, Wikipedia makes visible organizational processes that are difficult to observe in conventional work organizations. Here we examine the process in Wikipedia in which ordinary editors become administrators in order to distinguish the use of retrospective and prospective evidence in promotion.

Our primary research goal is to understand the type of evidence judges use to promote editors to administrator status. The workforce in Wikipedia is largely autonomous and distributed.

Individual editors decide which articles to work on and how much to work. There is no employment contract binding the volunteers to the organization and no supervisors assigning

them tasks. Under these conditions one might expect that Wikipedia needs to use promotion as part of a reward and incentive system to motivate editors to do the needed production work. If this is the case, then Wikipedia should promote people who are highly productive editors -- contributing to many articles and projects and having discussions with other Wikipedians about these editing tasks. That is, if Wikipedia is using promotions primarily for purposes of rewarding and encouraging desired behavior on the part of editors, then we'd expect that the criteria for promotion in Wikipedia should primarily be retrospective and production oriented. However, given that Wikipedia has a very large population of volunteers who already do production work without concrete rewards, promotion to reward production work might not be especially valuable to the organization.

On the other hand, there is much managerial work such as coordination and conflict resolution that needs to be accomplished in Wikipedia, with relatively few formal mechanisms to do so. If Wikipedia is using the promotion process to recruit and select individuals to do the difficult job of coordinating the work of others without formal authority or explicit hierarchy, then it should use more managerial-like criteria for evaluating candidates for promotion. The goals should be to promote people whose behavior as editors signals that they can become good administrators. Thus, the decision-makers should be looking for evidence that the candidate has the breadth of experience to handle the unanticipated demands made of administrators, the interpersonal skills needed to be civil while still handling the conflict frequently encountered in Wikipedia, and the temperament and interests to deal with Wikipedia policy.

We base our observations on two types of data. First, we examine formal Wikipedia policy and the standards members of the community explicitly invoke in their debates with each other about

whom to promote to administrator status. These linguistic data provide evidence about the criteria the community claims to use. Second, because people often possess poor insight into their own evaluation policies, we use a policy-capture technique (Hobson & Gibson, 1983) to examine the details of Wikipedians' behavior that predicts whether they will be selected to be administrators or not.

## **Part I: Inductive Approach**

To become an administrator, an editor must undergo a week of scrutiny known as the Request for Adminship (RfA), during which the community builds consensus about the candidate's experience and trustworthiness. Administrator tools are not granted lightly; an inexperienced, biased, or ill-intentioned editor could cause significant damage, reducing the encyclopedia's credibility or demotivating other editors. Thus, RfA contributors—those who voice support or opposition to the candidate's adminship—examine a broad range of evidence including a strong edit history, diverse experience, polite interaction with other users, policy understanding, conflict resolution, and high quality work.

The process consists of three parts: an introductory nomination statement, the nominee's answers to questions about past and future behavior, and statements of support, opposition, or neutrality by community members. Any registered Wikipedia member can voice an opinion, but the RfA is not a strict vote: at the end of a week, a bureaucrat—one of approximately 25 editors with privileges greater than administrators—reads through the opinions and decides whether consensus was reached. Candidates with more than 75% support are generally successful, though bureaucrats weigh voters' reasons, not just their votes. Votes by suspected sockpuppets (multiple identities held by the same person) or meatpuppets (new users recruited by a voter to back up the

voter's opinion) are discounted. Nominees may withdraw at any time, and the “snowball clause<sup>6</sup>”—a derivative of Wikipedia's famous “ignore all rules<sup>7</sup>” rule—allows any editor to close a nomination early to avoid wasting the community's time if the nominee is so inexperienced as to not stand a “snowball's chance in hell” of passing.

Community members within Wikipedia are at full discretion to vote for candidates based on their own criteria and often approach the decision with a personal policy for what they consider adequate. Decision policies for some members are even publicly listed as to which criteria that particular member finds most important and those which they find less relevant<sup>8</sup>. However, the community has provided a public guide for candidates for promotion as well as voters to assist in establishing a unified community norm on what behaviors and personal characteristics should be considered when making a promotion decision. Exhibit A displays the guidelines of what attributes are desired in a candidate and those attributes that are desired to not present in a candidate. Through the qualitative coding study in this paper further insights will be gathered by comparing the community standards with the comments voters make to determine (1) which criteria they are using and (2) to what extent each criteria is discussed when making decisions.

---

<sup>6</sup> [http://en.wikipedia.org/wiki/Wikipedia:Snowball\\_clause](http://en.wikipedia.org/wiki/Wikipedia:Snowball_clause)

<sup>7</sup> [http://en.wikipedia.org/wiki/Wikipedia:Ignore\\_all\\_rules](http://en.wikipedia.org/wiki/Wikipedia:Ignore_all_rules)

## Exhibit A: Community Guidelines for the Request for Adminship (RfA) Process<sup>9</sup>

### What [decision makers] look for and hope to see

**Strong edit history** with plenty of material contributions to Wikipedia articles.

**Varied experience.** RfAs where an editor has mainly contributed in one way (little editing of articles, or little or no participation in AfDs, or little or no participation in discussions about Wikipedia policies and processes, for example) have tended to be more controversial than those where the editor's contributions have been wider.

**User interaction.** Evidence of you talking to other users, on article talk or user talk pages. These interactions need to be helpful and polite.

**Trustworthiness.** General reliability as evidence that you would use administrator rights carefully to avoid irreversible damage, especially in the stressful situations that can arise more frequently for administrators.

**Helping with chores.** Evidence that you are already engaging in administrator-like work and debates such as RC Patrol and articles for deletion.

**High quality of articles.** A good way to demonstrate this is contributing to getting articles featured, although good articles are also well-regarded.

**Observing consensus.** A track record of working within policy, showing an understanding of consensus.

**Edit summaries.** Constructive and frequent use of edit summaries is a quality some RfA contributors want to see. Some expect use of edit summaries to approach 100% of the time. See Wikipedia:Edit summary.

A **clean block log** as evidence of good editing behavior.

### What [decision makers] for and hope not to see

**Vandalism:** A persistent and unreformed vandal will never be made an administrator; one of the primary tasks of administrators is fighting vandalism (and a truly bad administrator could cause serious damage to the site). Even a relatively minor disruption, like making a joking edit to an article, can cause problems.

**Incivility:** If a nominee has responded to unpleasant or irritating users by leaving insulting messages which violate the spirit of civility.

**Intransigence:** If a nominee has ever refused to be involved in good faith efforts to reach consensus on talk pages, and instead engaged in edit wars.

**Controversial activity on AfD:** Voting according to criteria not relevant to the purpose of AfD, persistently starting AfDs on articles on the kinds of subject generally (let alone explicitly) recognized as worth an article.

**Abuse of process:** If a nominee has ever started an inappropriate RfC, or made seemingly frivolous complaints via official channels.

**Edit wars:** If a candidate is prone to repeating a single edit after it becomes obvious that there is a disagreement with it. To most RfA contributors, it does not matter who is right, it matters how a candidate handles themselves during a debate.

**"Advertising" your RfA:** Some editors do not like to see an RfA "advertised" by the nominee on other people's talk pages or on IRC. RfA is not a political campaign. The intent is to develop consensus. Impartial evaluation of a candidate, not how popular they are, is the goal. Canvassing is often looked down upon.

**Blocks:** If your block log has activity and shows you've been blocked in the past several months.

**Elaborate signatures:** Some have a low opinion of editors who create fancy signatures, especially ones with special characters and images.

**Long gaps in editing:** Unless you have a good reason and you state on your page, a steady edit history is preferred.

**Use of sockpuppet accounts** to avoid scrutiny, or to mislead the community about your past editing history.

**Young age:** If you are a teenager or younger, many people will oppose based simply on your age.

<sup>8</sup> [http://en.wikipedia.org/wiki/Wikipedia:Requests\\_for\\_adminship/Standards/A-Z](http://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship/Standards/A-Z)

<sup>9</sup> [http://en.wikipedia.org/wiki/Wikipedia:Guide\\_to\\_requests\\_for\\_adminship](http://en.wikipedia.org/wiki/Wikipedia:Guide_to_requests_for_adminship)

RfA candidates answer three standard questions as part of their nomination: (1) What chores do you intend to help with? (2) What are your best contributions, and why? and (3) Have you been involved in conflicts over editing or have other users caused you stress? Nominees answer these and any additional questions posed by other community members, citing records of past events to demonstrate their competence and handling of controversy.

While RfA contributors have differing standards, many look for an answer to the first question that demonstrates the candidate is already behaving like an administrator, helping with chores such as monitoring recent changes, welcoming newcomers, or participating in debates at Requests for Comment (RfC) or Articles for Deletion (AfD). Editors are encouraged to first help the wiki using their existing privileges, such as reverting vandalism, organizing collections of articles in a subject area (WikiProjects), and reducing the non-administrator backlog, before considering promotion to administrator status. Only when new privileged abilities are needed are users encouraged to seek adminship, such as if a diligent vandal fighter is hampered in fighting vandalism by frequently waiting for administrators to block users she identifies.

The second RfA question demonstrates the significance of a candidate's contribution, and many successful candidates discuss their work on articles that reached high levels of quality, such as Featured Article status. Answers to the third question highlight incidents in which the nominee has dealt with interpersonal conflict, a common occurrence in an encyclopedia edited by thousands of people with differing viewpoints yet striving for neutrality. Previous work has shown conflict on even seemingly neutral topics like chocolate, and that administrators often

serve as unofficial mediators in “edit wars<sup>10</sup>” on more controversial topics, such as euthanasia and evolution. In answering the third question, the nominee demonstrates how he has dealt with controversy or uncivil comments from other editors, linking to evidence on talk pages.

Candidates involved in heated edit wars are unlikely to be well received by the community, though candidates who make full disclosures of previous mistakes and have recent histories of good behavior are more likely to succeed.

## **Study 1: Explicit Mentions of Criteria for Promotion**

### **Rationale**

In this study we used Grounded Theory methods to examine the criteria cited by voters in promotion decisions (Glaser & Strauss, 1967; Corbin & Strauss 2007). During the RfA process voters are strongly encouraged to leave a comment as to why they voted the way they did. By characterizing these comments we aim to understand whether administrator promotion is prospective or retrospective in nature. A major advantage of examining this question in the context of Wikipedia is that it provides us with complete records of an actual promotion decision context, rather than having to rely on laboratory or other artificial settings which may bias results (Hobson & Gibson, 1983).

### **Method**

We randomly selected 25 successful and 25 unsuccessful administrator candidates from the period of November 2005 to May 2007 drawing uniformly over the time frame. Evaluators were blind to the criteria in the guide to the RfA and those outlined by current administrators. A

---

<sup>10</sup> [http://en.wikipedia.org/wiki/Wikipedia:Edit\\_war](http://en.wikipedia.org/wiki/Wikipedia:Edit_war)

preliminary analysis of each candidate revealed several salient criteria mentioned by voters for both successful and unsuccessful candidates. In some cases, the criteria was mentioned as a reason the candidate should not be promoted (Negative Comment), in others it was mentioned as a reason the candidate should receive the promotion (Positive Comment). For example, if a voter comments, “Oppose – [candidate] shows a lack of knowledge of policy has a poor history of dealing with conflict” that comments would be coded once for “Negative – Knowledge of Policy” and once for “Negative – Conflict Management”. In the same comment, voters will occasionally make a both negative and positive descriptions of the candidate, such as, “Oppose - while [candidate] has a strong edit count and has a history of fighting vandalism, they have only been with Wikipedia for 3 months which is far too short for an Admin”. In this case, this comment would be coded for, “Positive – Edit Count”, “Positive – Fighting Vandalism”, and “Negative – Time in Wikipedia”. These criteria were categorized into either Retrospective criteria that demonstrated the candidate’s abilities and experience as an editor or Prospective criteria more suggestive of skills and personal characteristics required of an Administrator.

After two passes through the promotion decision comments were made 15 distinct categories for voter comments on candidate attributes emerged (described further below). Using Atlas TI, two independent coders tagged voter comments of the 50 discussions into the 15 categories using a standardized codebook. While Cohen’s Kappa is typically used in determining inter-rater reliability for qualitative work, the coding in our study was not mutually exclusive, that is, each candidate had an entire matrix of categories for the comments voters made (Cohen, 1960). Thus, rather than report a kappa, the inter-rater reliability is reported as a correlation (Neuendorf, 2002). The overall coder correlation resulting from the analysis is very strong at 0.976 and the weakest category in terms of inter-rater correlation was 0.872 (see Table 2 for complete inter-

rater correlations). Counts were then extracted of the number of distinct criteria used for each candidate (e.g., if a candidate had 8 voters each say he had a poor knowledge of Wikipedia policy it would only be counted once in the results). Using distinct criteria per candidate rather than raw counts results in a data set that is more interpretable and is not highly skewed for a high volume of redundant comments about a single aspect of any one candidate.

### **Retrospective Criteria**

*Number of Edits* is perhaps the simplest and most salient characteristic voters have available to them when evaluating an RfA. The community often discusses edits in terms of a necessary, but not sufficient condition to becoming an administrator. Voters suggest that edits should be "sufficient" (P37, 1) or "above the threshold" (P6, 52) before considering other factors.

*Time in Wikipedia* relates to how long the candidate has been a registered editor of Wikipedia. Candidates who have been a part of Wikipedia for extended periods of time are often noteworthy as it shows both a dedication to continue working and a sense that this candidate is ready for promotion. One voter comments, "[the candidate has] served his time, has stayed around to help the project, and does good work" (P26, 82).

*Article Quality* describes mentions of featured articles the candidate may have worked on, or other mentions of producing high quality article work. Often voters cite articles the candidate has worked on that have reached a community standard of quality such as "Good Articles<sup>11</sup>" or "Featured Articles<sup>12</sup>".

*Edit Summary Usage* indicates how often and to what quality candidates use an "edit

---

<sup>11</sup> [http://en.wikipedia.org/wiki/Wikipedia:Good\\_Articles](http://en.wikipedia.org/wiki/Wikipedia:Good_Articles)

summary” which is a short description of the change they made to an article and why they made the change. According to guidelines, edit summaries should be used for every change to an article and should approach 100%.

### **Prospective Criteria**

*Knowledge of policy* describes whether the voter perceives the candidate understands the policies, guidelines, and processes that govern the Wikipedia community. While Wikipedia may on the surface appear to be ad hoc, thousands of policies are in place to govern the production of articles and the community (Butler, et al., 2008). Administrators are one of the key enforcers of policy and must know how to intervene in disputes; in addition, administrators play a key role in discussing and creating new policies to confront new challenges. A voter who provides a positive comment regarding a candidate’s knowledge of policy states, “In my experience with [candidate], I found him to be reasonable in confronting POV-pushing editors. I believe [candidate] has demonstrated sufficient knowledge of Wikipedia policy to wield the tools” (P25:81). Conversely, those who are clearly not up to snuff on Wikipedia policy may receive an oppose vote that comments “[candidate] is an excellent editor but I don’t think you are hot enough on policy yet” (P10, 38).

*Interaction with others* encompasses how voters comment on their communications with the candidate in Wikipedia. Whether they were working on an article together, discussing a project, or giving pointers on how to improve their style users often comment on whether the person was polite, civil, friendly, or helpful. Examples of a positive comment on a candidate's interpersonal skills include, "I have had frequent interaction with [candidate], and he's extremely polite and

---

<sup>12</sup> <http://en.wikipedia.org/wiki/Wikipedia:FA>

civil" (P9, 126). If a candidate has a history of uncivil interactions with users since Wikipedia is archival in nature it is available for all to see and will often not go unnoticed. "I get a bad feeling about this candidate's temperament and skills at dealing with other editors. Seems a little to prone to snap at others and be a little blunt" (P20, 086).

*Trustworthy* candidates have somehow demonstrated a history in Wikipedia that voters deem that unlikely to abuse the administrative tools or status. After an examination of possible faults in the candidate such as uncivil behavior, edit waring, or vandalisms, voters often comment they have found no cause for concern in this candidate. One voter's only comment on a supportive vote states: "Everything checks out, trustworthy" (P14, 62).

*Neutral point of view* is one of the key policies in Wikipedia and a necessary quality for anyone mediating conflict. Voters to describe someone as "fair" (P4, 63), "level-headed" (P9,12), and "calm and neutral" (P16, 121) when a candidate has demonstrated they do not take sides on a topic but rather works to build consensus.

*Conflict management* relates strongly to a candidates demeanor and interpersonal skills in negotiating through difficulty. Often two parties have differing opinions on what an article should say or how something should be done and administrators need to have a calm, reasoned, and friendly mediation to resolve conflict. Often the voters who comment on a candidates behavior have personally interacted with the candidate and had negative or positive experiences with the candidate's conflict management skills. Comments such as "helpful in our conflict" (P4, 25) and "very good mediator" (11, 152) are interpreted as positive comments, while the comments such as "I am concerned about [the candidate] being able to keep a cool head in confrontational situations" (p24, 221) suggest the candidate has poor conflict management

skills.

*Community involvement* criteria that voters consider when determining if the candidate is not only editing articles but becoming involved in the Wikipedia community at large. Often voters will acknowledge the candidate has sufficient editing history, but note that "admins ought to be active within the community" (P7, 150). An ideal candidate for Adminship according to some voters should be "a well rounded user with heavy community involvement" (P53, 16).

*Commitment to Wikipedia* is a rather abstract criterion that voters perceive based on a number of candidate features such as tenure in Wikipedia, steady involvement, and consistency in their interactions. Voters are looking for candidates who are "clearly committed to Wikipedia in a long-term way" (P13, 21).

*Need for admin tools* describes criteria relating to whether the user has demonstrated they have previously worked on tasks that would benefit from having the Administrator tools. If a user has submitted requests for page moves or has demonstrated work in fighting vandalism, voters comment that the candidate "could do even better things to help with admin tools" (P37, 53).

*Vandalism to Wikipedia pages* includes deletion of pages by users and other defamation of Wikipedia content. In addition to editing, some users have taken it upon themselves to watch a set of pages to ensure users with ill intent do not destroy valuable content, and thus become "experienced vandal fighters" (P11, 80).

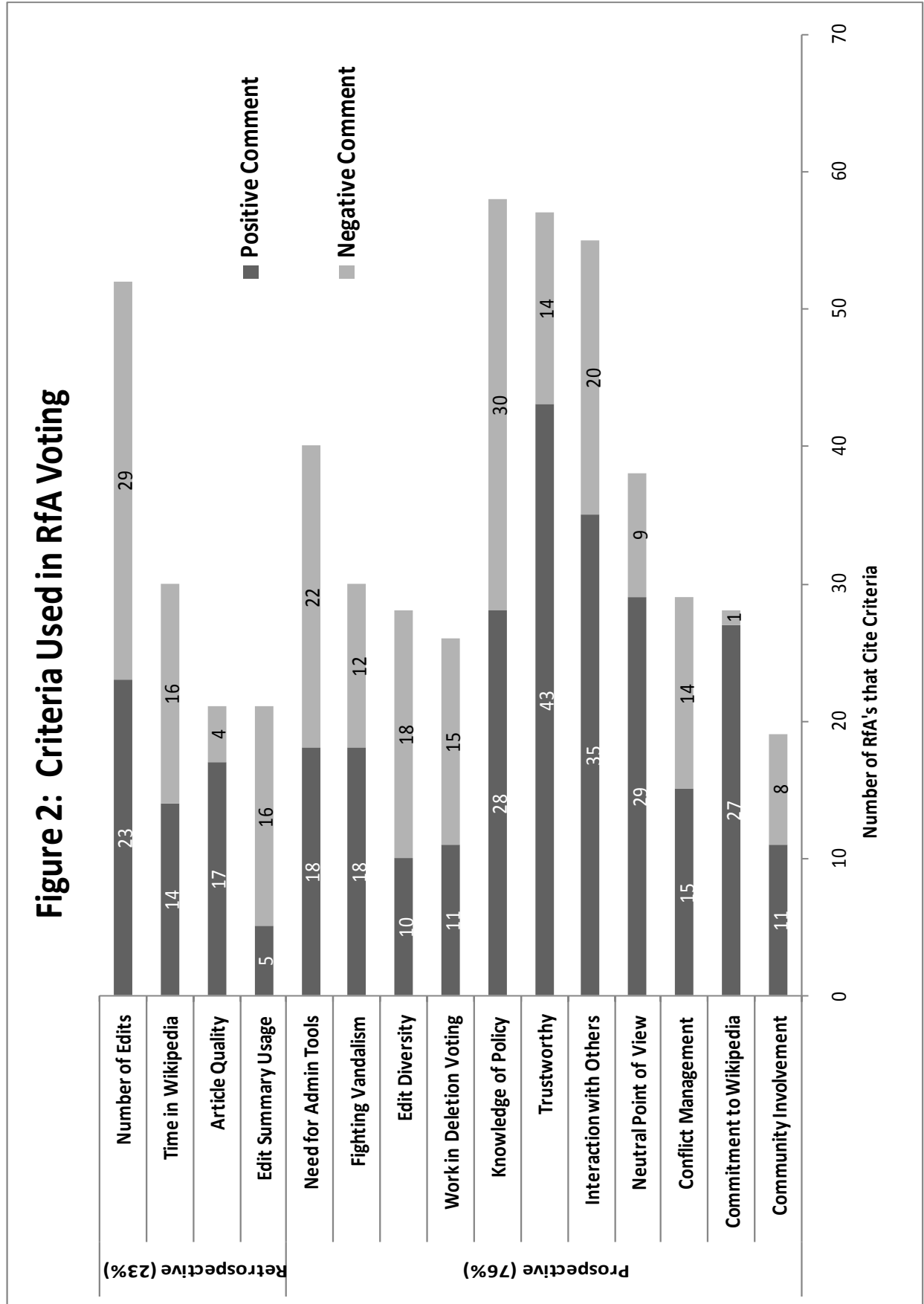
*Work in deletion voting* shows that candidates are working in a forum for unneeded articles, categories, and templates that are voted on to be deleted. Work in this area is characteristic of work an administrator will do once they have the position, but voting and discussion surrounding

content to be deleted reflects positively as it demonstrates familiarity with administrative tasks.

*Edit diversity* reflects the extent to which a candidate is working on various sections of Wikipedia rather than only contributing to articles. Voters prefer candidates who have experience working on projects, discussing policy, and interacting with other users. Candidates who have "made excellent contributions across several namespaces" (P47, 38) are likely to have a variety of experience to draw from when they become administrators.

### ***Results***

To empirically summarize the results this study used the number of unique counts of a comment category per candidate, for example, for a given RfA if a handful of voters all mention that this candidate has a strong history of fighting vandalism the results show only one unique comment on positive mention of vandalism work. By counting unique comments per candidate we eliminate a skew in the data that results from having one highly qualified or highly unqualified candidate have dozens of comments for a given criteria such as when a user clearly has a low number of edits and dozens of voters point out that they are opposing due to a lack of edits. Figure 2 shows the aggregate counts of how many RfA's cite each criteria. While only 50 candidates were evaluated, the counts for a single category, such as Number of Edits may have more than 50 unique comments since they are counted once for positive comments, and once for negative comments for a total of 52 number of times voters cite that criteria in their evaluations. Separate graphs are given in the Appendix B to show the distribution of criteria used for candidates were promoted to Administrator (Figure 3) and those who were not promoted (Figure 4). By comparing the graphs of the successful and unsuccessful candidates we can see that regardless of whether candidates are promoted or not the distribution of their criteria by category



appear to be nearly identical. Whether candidates are evaluated positively or negatively on each category seems to be what makes the difference for the candidate as to whether they are promoted.

*Prospective criteria.* Candidates who had not sufficiently shown a need for the Administrative tools were heavily noted by voters and punished for their lack of participation tasks characteristic of Administrators. This suggests evaluators have a forward-looking view of candidates, and want to know candidates for promotion are already beginning to work on tasks that they will undergo once promoted. Candidates with a strong history of fighting vandalism in Wikipedia received notable attention for their efforts, but those who did not were rarely punished. The diversity of candidates work and their participation in deletion discussion were mentioned in roughly a third of the candidates primarily in a negative light. Those who lacked diversity or participation were noticed and called out for their shallow experience.

Candidates' knowledge of policy was the most widely discussed feature during their evaluation, with both high discussion of positive and negative aspects of their knowledge. A majority of the negative comments came from evaluation of candidates were not promoted to be an Administrator. This suggests that regardless of other criteria, if the candidate does not have the cognitive knowledge and has not demonstrated their familiarity with policy they do not have the personal qualities necessary to be promoted. Throughout a candidate's history voters evaluate the candidate's interpersonal skills and trustworthiness as next two most cited criteria. Voters appear to be keenly aware that upon promotion editors will be in the spotlight to other users and have more frequent interactions with other administrators; promoting someone who is a strong worker but difficult to get along with or untrustworthy does not benefit the community. As

expected, a candidates apparent neutral point of view, conflict management abilities, and commitment to Wikipedia are all widely cited as criteria for promotion. Additionally it is interesting to note that when considering policy knowledge, conflict management, and community involvement voters cite a fairly even split between positive and negative comments perhaps due to the subjectivity of the evaluation process and variation in the personal experiences with that particular candidate.

*Retrospective criteria.* Voter frequently discuss the raw number of edits a user has, perhaps because that information is so readily available and is thought to be a general proxy for experience and worth ethic. Time within Wikipedia, article quality, and edit summary usage are of relatively little consideration when compared to other criteria.

The results summarized in Figure 2 show that a vast majority of voting behavior (76% of unique category votes) reflects a Prospective look at Administrator promotion. Having a strong history of editorial behavior is still highly considered, but these results would suggest that they are a necessary but not sufficient condition for promotion. The three most frequently cited criteria are all within the prospective view of promotion. The Wikipedia voting community appears to regard the promotion to Administrator as a matching of finding the most qualified person for the position in terms of personal factors and skills rather than as a reward for hard work or longevity within the community.

## **Part II: Empirical Comparison Using Behavioral Data**

### ***Policy Capture***

Policy capture research focuses on the criteria and processes decision makers use to make

promotion and selection decisions of personnel. The methodology generally examines how decisions makers are evaluating candidates by giving them a set of hypothetical candidates to evaluate who vary on a set of attributes such as potential, current position, gender, and recommendations from others and asking the decision makers to rate the candidates on whether they would hire them for a position. By clustering candidates by similar recommendation patterns researchers find patterns in how decision makers evaluate (Stumpf & London, 1981). Additionally, researchers can examine how decision makers think they are evaluating and compare it to how they actually behave in the experiment (Hobson et al., 1981). Two key findings relevant to this study that Policy Capture research has found are that in general (1) decision makers often possess poor insights into their own rating policies and specifically (2) decision makers tend to perceive a more equal weighting of performance metrics than they actually use to make decisions (Taylor & Wilsted, 1974; Zedeck & Kafry, 1977). Specifically, decision makers tend to underestimate the importance of major dimensions and underestimate the importance of minor ones. In Study 1 we developed the criteria that decision makers acknowledge they are using but to shed light on whether their decisions reflect those policies in Study 2 we will examine the behavioral data of the candidates to compare.

## **Study 2: Modeling Successful RfA Candidates**

### ***Rationale***

The analysis of reasons cited by voters for promotion decisions in Study 1 suggests a focus on prospective rather than retrospective factors. Administrator promotion in Wikipedia has a number of characteristics ideal for policy capture research: the role of the Wikipedia administrator is well-defined; there are many judges involved in the promotion decision; and

there is high transparency in the records of past action since all edits are recorded. In Study 2 we examine the link between quantitative prospective and retrospective factors of candidate performance and promotion success. We also compare the factors cited in Study 1 as important to promotion success with the factors actually used in promotion decisions.

## **Methods**

The data include all 1551 Requests for Adminship from January 2006 through October 2007, with 49 RfAs removed for being multiple attempts by the same candidate in one month (all of which failed), bots, sockpuppets (multiple identities held by the same person), or because the nominee's edit history prior to the RfA was not available. For each RfA, data from the user's contribution history page up to the month before the RfA was parsed and counted, and grouped into the categories introduced in Study 1. Three categories from Study 1—Trustworthiness, Commitment to Wikipedia, and Article Quality—were not included in this analysis because they could not be captured with empirical metrics. These categories are discussed further in the discussion section. Table 1 provides summary statistics for the features.

### **Retrospective criteria**

*Number of edits* includes the number of edits to articles and to WikiProjects.

*Time in Wikipedia* is the number of months between a user's first edit and his or her RfA.

*Edit Summary Usage* is measured on two dimensions, *Summarized %* reflects the percent of the time the candidate write a summary for an article edit and *Average Summary Length* is a measure of the length of the edit summary.

## **Prospective-criteria**

*Edit diversity* includes a count of the number of different areas in which the user has participated, from the set {article, article talk, Wikipedia, Wikipedia talk, user, user talk, articles/categories/templates for deletion (XfD), (un)deletion review, other RfAs, village pump, administrator intervention against vandalism (AIV), requests for protection (RfP), administrator noticeboard, arbitration committee, mediation committee, and wikiprojects}. So, a user who has edited articles, her own user page, and posted once at the Village Pump would have a diversity score of 3. Actual number of edits in each of these sections is accounted for in the following categories to determine their relative importance.

*Need for admin tools* includes instances in which the editor requested help from administrators to perform a chore not available to normal editors. It is a count of posts to the administrators' noticeboard (e.g. for reporting inappropriate usernames or edit wars).

*Vandalism* includes reversion of vandalism (noted by “revert” or “rv” in the comment text accompanying an edit), requesting administrator intervention for specific vandals (AIV), and requesting protection for particular pages from malicious editors.

*Work in deletion voting* includes participating in deletion discussions, including articles/categories/templates for deletion (XfD), (un)deletion reviews, and votes for deletion.

*Knowledge of policy* includes edits to the Wikipedia namespace, a section of the encyclopedia devoted to policy and guideline discussions. Edits to WikiProjects, which are also part of the Wikipedia namespace, have been separated and counted below.

*Interaction with others* includes edits to "talk" pages. Every piece of content on Wikipedia (including articles, user profiles, and policies) has an associated "talk" page, on which changes to the content are discussed. Simple politeness markers, including any variant of "please" (including "pls" and "plz") or "thanks" (including "thx") in comment text are also counted.

*Neutral point of view* includes the number of edits in which neutrality fixes were made, noted by "(n)pov" for "neutral point of view" in the edit summary text.

*Conflict management* includes the number of posts to the arbitration or mediation committee pages, or "wikiquette" alerts (an early stage in dispute resolution).

*Community involvement* includes participation in other editors' Requests for Adminship, posts to the community discussion board known as the Village Pump, and welcoming posts to newcomers' talk pages.

## **Results**

To examine the impact of these behavioral factors on the likelihood of a candidate's promotion, we performed a probit regression on the binary dependent variable, RfA success. All variables were standardized, so the coefficients in Table 1 represent the change in probability of success when a continuous variable is increased by one standard deviation. For example, every additional 39 instances of saying "thanks" increases a candidate's likelihood of success by 9.8%. All variance inflation factors (VIFs) were well below 10, indicating low multicollinearity between factors. Multiple attempts by the same candidate in a single month were excluded, leaving only one attempt per month, and the candidate's number of previous RfA attempts (in other months) is included as a control variable; each subsequent attempt has a 14.8% lower

<b>Table 1. Behavioral Predictors of Promotion</b>					
	<b>Criteria</b>	<b>Mean</b>	<b>Std. Dev</b>	<b>Δ Prob.</b>	
<b>Prospective Criteria</b>					
Edit diversity					
	Number of sections edited	11	3.7	11.3%	***
Need for admin tools					
	Admin attention/noticeboard edits	18.7	57.9	-3.1%	†
Vandalism					
	"Revert" in comments	257.6	563.2	4.3%	†
	Vandal-fighting (AIV) edits	26.5	108.7	-3.3%	ns
	Requests for protection	3.7	12.2	-3.0%	†
Work in deletion voting					
	X for deletion/review edits	504.5	1027.3	2.5%	ns
	Votes for deletion	3.4	15.1	1.0%	ns
Knowledge of policy					
	Wikipedia policy edits	433.8	913.8	6.0%	ns
Interaction with others					
	Article talk edits	415.2	775.4	3.6%	†
	User talk edits	786.6	1169.9	1.3%	ns
	Wikipedia talk edits	87.6	179.5	1.6%	ns
	"Please" in comments	31.7	83.8	1.3%	ns
	"Thanks" in comments	21.8	39.3	9.8%	***
Neutral point of view					
	"POV" in comments	26.7	46.9	3.4%	†
Conflict management					
	Arb/mediation/wikiquote edits	9.8	47.1	-6.8%	***
Community involvement					
	Other RfAs	27.4	58.9	-3.2%	†
	Village pump	9.5	34.6	-0.1%	ns
	Newcomer welcomes	76.9	321.1	-2.2%	ns
	WikiProject edits	144	569.1	0.6%	ns
<b>Retrospective Criteria</b>					
Number of edits	Article edits	2611.1	3804.3	7.4%	***
Time in Wikipedia	Months since first edit	9.6	8.0	2.9%	*
Edit Summary Usage					
	Summarized (%)	80.0	20.0	25.3%	***
	Avg. summary length (log2chars)	5.0	0.8	0.9%	ns
N=1502    *** p < .001    ** p < .01    * p < .05    † p < 0.1					

chance of success than the previous one. The model is 74.8% accurate in classifying RfA

attempts as successful or not.

*Prospective criteria.* The editor's diversity of experience in Wikipedia had a strong impact on the likelihood of promotion. Branching out to additional regions demonstrates an understanding of the encyclopedia and its community as a whole, and editors with experience in additional sections were more likely to become administrators. Editing approximately 4 new regions was associated with an 11.3% increased likelihood of promotion. Reverting vandalism tended to be associated with an increased likelihood of promotion, as well, however, it was only significant at a  $p < 0.10$  level. However, demonstrating a need for the tools by elevating issues to administrators' attention tended to be associated with a decreased likelihood of promotion: posts to the administrator noticeboard, requesting administrator intervention against vandalism, or requesting page protection were non-significant but in the negative direction. Finally, participation in page deletion discussions had no impact on promotion.

Contrary to the results of Study 1, in which RfA evaluators frequently stated candidates needed policymaking experience, participation in policymaking discussions was not a predictor of RfA success. Neither were edits to talk pages in the policymaking sections of the encyclopedia. Edits to article talk pages modestly improve the likelihood of promotion, consistent with the norm in Wikipedia of coordinating article content there. However, when disagreements over content covers multiple articles or a user's overall behavior, the discussion often moves to user talk pages. Thus, user talk edits are likely to be mixed, and may have higher interpersonal conflict. As a result, user talk edits do not predict promotion. In fact, interpersonal conflict may also explain why posts to the Arbitration or Mediation Committee pages, or to Wikiquote notices, all of which are venues for dispute resolution, decrease the likelihood of success, as well. Thus, successfully promoted Wikipedia administrators appear to avoid conflict-related interpersonal

behavior. Politeness helps modestly; though it was rare, every 39 instances of saying “thanks” in comments increased the likelihood of success by 9.8% . Saying “please” had no effect.

Upholding Wikipedia's much valued neutrality by fixing point-of-view issues also tended to be associated with a greater likelihood of promotion. Somewhat surprisingly, involvement in the community through welcoming newcomers, participating in other RfAs, or chatting at the Village Pump also had no effect.

*Retrospective criteria.* Successful editors had to demonstrate extensive article edits and tenure in the community. Every 3800 article edits increased likelihood of success by 7.4% Length of experience helps slightly; every 8 additional months between a user's first edit and her RfA tended to increase the likelihood of success by 2.9%. Observing a consistent use of edit summaries shows a regard for documenting your work as a strong editor, and has a highly significant effect and an increased likelihood of promotion: candidates with near 100% usage were 25.3% more likely to be promoted.

## Discussion

The results from this study are consistent with research in policy capture suggesting that decision makers often perceive a more equal weighting of performance metrics than they actually use in making decisions (Taylor & Wilsed, 1974; Zedeck & Kafry, 1977). In the production community setting, while the community has collectively established guidelines and norms on criteria for promotion decision the task of implementing the decisions rests on the shoulder of its members. The comparison between the voters’ comments in Study 1 and the behavioral data from Study 2 demonstrate that voters appear to believe they are voting based on the community standards and a balance of Prospective and Retrospective criteria while the decision outcomes

reflect a strong reliance on easily available public metrics. Through qualitative examination of voter discussions three explanations emerge to explain the differences between the discussion and the voting behavior.

*Bounded Rationality.* Due to the large number of promotion decisions the community must make and the abundance of available data on the candidates must make a boundedly rational decision without reviewing all available information (Simon, 1991). While the entire history of each candidate's talk with other users, edits to article pages, and contribution to administrative tasks are publicly available, the community does not have a public means to display or evaluate all this data into interpretable information for its users. Thus, at the expense of spending hours reviewing their politeness with other users and their administrative contributions voters may rather limit their decision criteria to a few key metrics that can ostensibly proxy for the community criteria.

*Empirical Democracy.* Since retrospective criteria are largely tangible, historical, and verifiable and prospective criteria are predominantly interpersonal, subjective, and knowledge related community members may have made a conscious decision to rely largely on retrospective criteria. As a community Wikipedia places a high value on a neutral point of view and verifiable information in producing content and interacting with other users. For some members these values extend to promotion decisions as demonstrated by their voting based on highly visible empirical evidence rather than subjective discussions of trustworthiness and knowledge of policy. The community has recognized this tendency of voters to rely on raw edit counts, and termed it "editcountitis". A Wikipedia page on the term states that some members possess the belief that "a Wikipedian's overall contribution level can be measured solely by their edit count.

This is a phenomenon which some think may be harmful to processes such as requests for adminship, as well as to the Wikipedia community in itself.<sup>13</sup>”

*Collective Inconsistency.* Community members recognize the guideline criteria and attempt to use all available information but are unable to apply their decision criteria in a consistent manner. While it may be that community members agree on the criteria to evaluate candidates the variance in their beliefs as to the importance of the criteria may be distributed in such a way that the predictive value of the criteria is full of noise and not predictive of success for the candidates. In candidate voting discussions one voter downplays the need for candidate to be involved in administrative work, “You can't really expect someone to already be doing administrative type things before becoming an administrator because ... well ... they aren't an administrator” (P27, 3) while another voter expresses his concern that the candidate “can't support a candidate whose shown such minimal interest in administrative processes” (P41, 9). Thus, as frequently shown in policy capture literature individual decision makers more heavily weight salient criteria and downplay perceptively less relevant criteria. When aggregate across thousands of decisions and thousands more decision makers the criteria are applied inconsistently and idiosyncratically and produce a noisy voting process that is not reflective of candidates behaviors.

## **Limitations**

While Wikipedia is one of the more public and successful of the online production communities, this study's findings are limited in that results generated may not be generalizable into the forms and governance structures of other production communities. Future research on promotion and

---

<sup>13</sup> <http://en.wikipedia.org/wiki/Wikipedia:Editcountitis>

governance in other production communities with alternative processes will be beneficial to our further understanding of organizations.

Although this research has shown that judges pay attention to candidates' job-relevant behavior it is silent about whether other factors identified in the organizational literature such as social networks, irrelevant attributes, or strategic self-presentation (Ng et al. 2005). Indeed, recent evidence that Wikipedia administrators use a secret mailing list to coordinate their actions toward others suggest that sponsorship may also play a role in promotion (Metz, 2007). Future research in Wikipedia using techniques like those in the current paper can be used to test theories in organizational behavior about criteria for promotion.

Another important limitation of the current model is that it does not take the quality of contribution into account. We plan to improve the model by examining measures of length, persistence, and page views of edits, which are already being used in more processor intensive models of existing administrator behavior and impact of edits.

Additional research is also needed to determine whether the prospective criteria used by RfA voters to select administrators actually leads to better administrator performance. Our data currently only describe candidates who become administrators, not necessarily those who become good administrators. An important next step is to determine if administrators change their behavior after the RfA and to measure the quality of that administrator's future behavior.

## **Contribution**

In a qualitative study of promotion in a large production community, people weighted prospective factors (predicted performance in a new role) more heavily than retrospective factors (previous role performance). However, a quantitative study of promotion success suggests that

retrospective factors are far predictive of actual promotion. This suggests that those making promotion decisions in Wikipedia believe they are assigning people to roles in which they can maximally benefit the organization, rather than as an incentive or reward for past achievement. However, there are disconnects between the criteria they cite and those that actually predict promotion success. Further research is needed to determine the role of other factors in administrator promotion and to link promotion decision criteria to post-promotion performance.

Our data is consistent with prior literature on contest theories of promotion which emphasize prior job performance as an explanatory of promotion decisions (Turner, 1960). In our study, several metrics of a user's prior experience were significantly related to promotion success. Importantly, we also introduce a distinction between prospective factors of evaluation which focus on a person's likely success in their new role, and retrospective factors which emphasize past achievement. This distinction mirrors the dual role of promotion in assigning people to roles they are best able to contribute and as a source of incentives and rewards (Milgrom & Roberts, 1992). The greater importance of prospective factors in the qualitative analysis suggests that Wikipedians believe they use promotion as an assignment and selection mechanism for increased user privileges, although the final outcomes suggest that promotion may largely serve as a reward mechanism for past contribution.

These findings also contribute to the literature on policy capture, in which several studies have found moderate differences between subjective estimates of policy importance and objective captured policy (Zedeck & Kafry, 1977; Hobson, Mendel, & Gibson, 1981; Stumpf & London, 1981). However, most extant studies of policy capture have employed artificial rating situations which some researchers argue are very different from actual settings and make generalization to

real-world situations difficult (Hobson & Gibson, 1983). Here we demonstrate the benefits that archival production communities such as Wikipedia present as a way to study policy capture in actual promotion settings. Furthermore, we extend the study of policy capture to production communities, which are growing in real-world importance in the knowledge economy.

## References

- Armstrong, A., & Hagell, J. (1996). The Real Value of Online Communities. *Harvard Business Review*, 74(3), 134-141.
- Blau, P. M. (1955). *The Dynamics of Bureaucracy*. Chicago: University of Chicago Press.
- Bryant, S. L., Forte, A., & Bruckman, A. (2005). *Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia*. Paper presented at the Group '05, Sanibel Island, Florida.
- Burt, R. S. (1982). Positions in Networks. *Social Forces*, 51, 93-122.
- Butler, B., Joyce, L., & Pike, J. (2008). *Don't Look Now, But We've Created a Bureaucracy*. Paper presented at the CHI 2008: Proceedings of the ACM Conference on Human-factors in Computing Systems, Florence.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 76(5), 37-46.
- Corbin, J., & Strauss, A. C. (2007). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA: Sage Publications.
- Forte, A., & Bruckman, A. (2008). *Scaling Consensus: Increasing Decentralization in Wikipedia Governance*. Paper presented at the HICSS 2008.
- Giles, J. (2005). Internet encyclopedias go head to head. *Nature*, 438, 900-901.
- Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory*. Chicago, IL: Aldine Transaction.
- Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the Labor Market. *The American Economic Review*, 84(5), 1174-1194.
- Hargens, L. L., & Hagstrom, W. O. (1967). Sponsored and Contest Mobility of American

- Academic Scientists. *Sociology of Education*, 40(1), 24-38.
- Hobson, C. J., & Gibson, F. W. (1983). Policy Capture as an Approach to Understanding and Improving Performance Appraisal: A Review of the Literature. *The Academy of Management Review*, 8(4), 640-649.
- Hobson, C. J., Mendel, R. M., & Gibson, F. W. (1981). Clarifying Performance Appraisal Criteria. *Organizational Behavior and Human Performance*, 28, 164-188.
- Kieser, A. (1989). Organizational, Institutional, and Societal Evolution: Medieval Craft Guilds and the Genesis of Formal Organizations. *Administrative Science Quarterly*, 34(4), 540-564.
- Kittur, A., Chi, E., Pendleton, B. A., Suh, B., & Mytkowicz, T. (2007). *Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie*. Paper presented at the CHI 2007: Proceedings of the ACM Conference on Human-factors in Computing Systems, San Jose.
- Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H. (2007). *He Says, She Says: Conflict and Coordination in Wikipedia*. Paper presented at the Chi, San Jose, CA.
- Kogut, B., & Metiu, A. (2001). Open-Source Software Development and Distributed Innovation. *Oxford Review of Economic Policy*, 17, 248-264.
- Lazear, E. P. (2004). The Peter Principle: A Theory of Decline. *Journal of Political Economy*, 112(1), 141-163.
- March, J. G., & Simon, H. A. (1958). *Organizations*. New York: Wiley.
- Maume, D. J., Jr. (1999). Glass Ceilings and Glass Escalators: Occupational Segregation and Race and Sex Differences in Managerial Promotions. *Work and Occupations*, 26(4), 483-509.

- Melamed, T., & Bozionelos, N. (1992). Managerial Promotion and Height. *Psychological Reports, 71*(2), 587-593.
- Metz, C. (2007). Secret mailing list rocks Wikipedia. from [http://www.theregister.co.uk/2007/12/04/wikipedia\\_secret\\_mailing](http://www.theregister.co.uk/2007/12/04/wikipedia_secret_mailing)
- Milgrom, P., & Roberts, J. (1992). *Economics, Organization and Management*. Englewood Cliffs, N.J: Prentice Hall.
- Neuendorf, K. A. (2002). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage Publications.
- Ng, T. W. H., Eby, L. T., Sorenson, K. L., & Feldman, D. (2005). Predictors of Objective and Subjective Career Success: A Meta-Analysis. *Personnel Psychology, 58*(2), 367-408.
- Peter, L. J., & Hull, R. (1969). *The Peter Principle*. New York: W. Morrow.
- Raymond, E. S. (1999). *Cathedral and the Bazaar*. Sebastopol, CA: O'Reilly & Associates.
- Roberts, J., Hann, I.-H., & Slaughter, S. (2006). Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Project. *Management Science, 52*(7), 984-999.
- Romanelli, E. (1991). The Evolution of New Organizational Forms. In W. R. Scott & J. Blake (Eds.), *Annual Review of Sociology* (Vol. 17, pp. 79-103). Palo Alto, CA: Annual Reviews.
- Rosenbaum, J. E. (1979). Tournament Mobility: Career Patterns in a Corporation. *Administrative Science Quarterly, 24*(2), 220-241.
- Sawhney, M., & Prandelli, E. (2000). Communities of Creation: Managing Distributed Innovation in Turbulent Markets. *California Management Review, 42*(4), 24-54.
- Shah, S. K. (2006). Motivation, Governance, and the Viability of Hybrid Forms in Open Source

- Software Development. *Management Science*, 52(7), 1000-1014.
- Stern, R. N., & Barley, S. R. (1996). Organizations and Social Systems: Organization Theory's Neglected Mandate. *Administrative Science Quarterly*, 41(1), 146-162.
- Stumpf, S. A., & London, M. (1981). Capturing Rater Policies in Evaluating Candidates for Promotion. *The Academy of Management Journal*, 24(4), 752-766.
- Taylor, R. L., & Wilsted, W. D. (1974). Capturing Judgement Policies: A Field Study in Performance Appraisal. *The Academy of Management Journal*, 17, 440-449.
- Tharenou, P., Latimer, S., & Conroy, D. (1994). How Do You Make it to the Top? An Examination of Influences on Women's and Men's Managerial Advancement. *The Academy of Management Journal*, 37(4), 899-931.
- Turner, R. H. (1960). Sponsored and Contest Mobility and the School System. *American Sociological Review*, 25(6), 855-867.
- von Krogh, G., Spaeth, S., & Lakhani, K. R. (2003). Community, Joining, and Specialization in Open Source Software Innovation: A Case Study. *Research Policy*, 32(7), 1217-1241.
- Zedeck, S., & Kafry, D. (1977). Capturing Rater Policies for Processing Evaluation Data. *Organizational Behavior and Human Performance*, 18(2), 269-294.

## Appendix A: Inter-Rater Reliability in Qualitative Coding

**Table 2: Inter-Rater Correlation for Category Coding**

Category	Positive	Negative	Category	Positive	Negative
Article Quality	0.955	1.000	Knowledge of Policy	0.959	0.919
Commitment	1.000	1.000	Need for Admin Tools	1.000	1.000
Community Involvement	0.941	0.924	Neutral Point of View	1.000	0.937
Conflict Management	0.872	0.950	Number of Edits	0.878	0.958
Deletion Experience	1.000	1.000	Tenure in Wikipedia	1.000	1.000
Edit Diversity	1.000	1.000	Trustworthiness	1.000	0.952
Edit Summary Usage	1.000	1.000	Interaction with Others	0.900	1.000
Fighting Vandalism	0.958	1.000			
			<b>Overall Correlation</b>		<b>0.976</b>

### Appendix B

