Revisiting "The Rise and Decline"
in a Population of Peer Production Projects

Nathan TeBlunthuis
University of Washington
Seattle, Washington, USA
nathante@uw.edu

Aaron Shaw
Northwestern University
Evanston, IL, USA
aaronshaw@northwestern.edu

Benjamin Mako Hill
University of Washington
Seattle, Washington, USA
makohill@uw.edu

ABSTRACT

Do patterns of growth and stabilization found in large peer production systems such as Wikipedia occur in other communities? This study assesses the generalizability of Halfaker et al.'s influential 2013 paper on "The Rise and Decline of an Open Collaboration System." We replicate its tests of several theories related to newcomer retention and norm entrenchment using a dataset of hundreds of active peer production wikis from Wikia. We reproduce the subset of the findings from Halfaker and colleagues that we are able to test, comparing both the estimated signs and magnitudes of our models. Our results support the external validity of Halfaker et al.'s claims that quality control systems may limit the growth of peer production communities by deterring new contributors and that norms tend to become entrenched over time.

ACM Classification Keywords

H.5.3. Information Interfaces and Presentat ion (e.g. HCI): Group and Organization Interfaces – Computer-supported cooperative work

Author Keywords

governance; peer production; online communities; quality control; retention; replication; Wikipedia; wikis

INTRODUCTION

"Peer production" describes a way of organizing collaborative information production in online commons [2]. Over the last decade, peer production has become a central object of HCI research. However, the vast majority of peer production research has studied a small number of the largest communities [3, 6]. An enormous portion of empirical studies of peer production in HCI are of the English-language version of Wikipedia. Unfortunately, HCI's historical focus on novelty has meant that tests of the applicability of findings shown in one setting to other contexts rarely happens [27]. As a result, we know little about the degree to which theory and design claims from studies of Wikipedia apply more broadly.
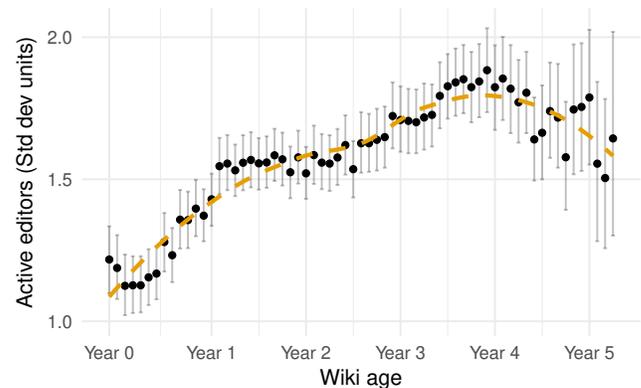
Figure 1. Mean of the number of editors with at least 5 edits per month in standard deviation units for wikis in our sample. The dashed lines represent the results of a LOESS regression. The error bars represent bootstrap 95% confidence intervals. This replicates Figure 2 in RAD.

This paper replicates analysis from Halfaker et al.'s "The Rise and Decline of an Open Collaboration System" [11] (which we abbreviate "RAD") in a sample of 740 active wikis hosted on Wikia.[1] RAD makes one of the most influential and highly cited claims about peer production dynamics, attributing English Wikipedia's decline in contributors since 2007 to entrenchment (RAD uses the term "calcification") within the community as norms and policies become difficult to change, especially for newer users. Our results reproduce most of RAD's findings. Like RAD, we find that the average community in our dataset experiences a "rise and decline," that newcomers are less likely to survive over time, that rejected newcomers are less likely to survive, that editors with longer tenure have more influence over norms, and that norms become entrenched as wikis age. In addition to providing an external validation of RAD's findings, we rule out alternative explanations of RAD's results that emphasize unique attributes of Wikipedia or the timing of the editor decline in that community.

BACKGROUND

Entrenchment in Wikipedia and Peer Production

Active peer production communities often experience a period of rapid growth followed by stabilization [20, 23]. Following

---

[1]Wikia is a wiki hosting platform where anyone can start a wiki. In 2016, Wikia partially rebranded as "Fandom" to emphasize support for fan communities. See: **https://www.wikia.com/** (**https://perma.cc/TL79-VB57**).

this pattern, the number of contributors to English Wikipedia grew exponentially until March 2007, when it began to decline [26]. Although early accounts of peer production argued that projects such as Wikipedia and the Linux kernel mobilized massive collaboration without the sorts of formal hierarchies or bureaucracies used in formal organizations [2, 17], organizational research has argued that the formalization of rules, norms, and routines accompany this trajectory in many types of organizations [13, 18, 24]. Drawing from this work, early explanations for Wikipedia's decline included bureaucratic overhead and increasing resistance to contributions from less active editors [26, 7]. During this same period, algorithmic tools such as "bots" became important parts of Wikipedia's quality control systems [10], and the proportion of edits that were rejected increased [12]. Building on this prior work, Halfaker et al.'s "The Rise and Decline of an Open Collaboration System" [11] found evidence in support of the theory that three elements of Wikipedia's quality control system—newcomer rejection, algorithmic tools, and norm entrenchment—could explain the transition from growth to decline. However, despite the impact and influence of RAD's explanation, it has not been replicated beyond Wikipedia until now.

Replication in Social Computing Research

Although HCI research prizes novelty and provocation, it also seeks to build scientifically rigorous, replicable, and generalizable knowledge [27]. Replicability refers to how well results hold up when other researchers follow reported procedures. Hornæk et al. define replication studies as attempts "to confirm, expand, or generalize an earlier study's findings" [15]. Generalizability (external validity) refers to the degree to which results hold up across different populations [4]. Replication studies thus assess whether details of context or methodological choice explain results.

Although comparative analysis of peer production communities has emerged as an important means to understand the life cycles and dynamics of social computing systems [20, 22, 25], there have been few efforts to establish whether findings from Wikipedia and other large communities replicate or generalize [3, 14]. In one important exception, Kittur and Kraut [16] examine the prevalence of social mechanisms related to conflict and coordination in Wikipedia among nearly 7,000 wikis from Wikia. Their work found both similarities and differences between these communities and Wikipedia.

Replicating RAD

Do the relationships described in RAD generalize to other peer production communities? The evidence on project life cycles, stabilization, and entrenchment suggests that similar patterns may occur beyond Wikipedia. However, Wikipedia's scale and popularity make it a unique outlier among these communities. It is likely unusual in other ways as well. In their conclusion, the RAD authors note: "Wikipedia's [growth and quality assurance] challenges may seem unique to its status as one of the largest collaborative projects in human history." Nevertheless, they suggest that their analysis of "sociotechnical gatekeeping and its consequences" has general applicability. Indeed, their conclusions have informed analyses

of crowdsourced fund-raising [1], social media [5], and online collaborative mapping [21].

This paper assesses the replicability and external validity of RAD to provide an empirical foundation for such generalization. In doing so, we also evaluate several alternative explanations that RAD could not rule out. In particular, the simultaneous decline and entrenchment RAD observes in Wikipedia could be driven by external factors related to time, such as the rise of other online communities (e.g., Facebook) that might compete for newcomers. By studying a population of communities whose trajectories start at different points in time, we can model wiki age accounting for calendar time in ways RAD could not. By studying many communities, we can also better understand the scope of RAD's generalizability by measuring variation between wikis.

METHODS

We attempt to follow RAD's measures and methods to the fullest extent possible. In some places, we are forced to make changes to accommodate differences between English Wikipedia and Wikia and the fact that our analysis includes multiple wikis. To describe our methodology, we briefly summarize RAD's techniques and note several ways that we diverge. Additional detail on operationalization is provided in RAD. We also provide access to the complete R source code that we used to complete our analysis in the supplementary material that accompanies this paper. In the description that follows, variable names are italicized.

The RAD authors present three interdependent analyses. The first tests whether the rejection of edits made by newcomers causes decreased newcomer retention which in turn leads to a decline in the number of active editors. To support this claim, the RAD authors use data from English Wikipedia to plot three trends: the number of active contributors, the rate of newcomer survival, and the rate of newcomer rejection. The first plot shows the rise and decline in active contributors (i.e., individuals who make at least 5 edits in a given month). The second plot shows that the proportion of good-faith newcomers who "survive" falls over time. The third plot shows that the proportion of good-faith newcomers "rejected" in their first edit session rises over time. RAD considers a newcomer to have *survived* if the newcomer edits during the period between 60 days and 180 days after their first edit session (i.e., sequence of consecutive edits less than one hour apart) and to have been *rejected* if a change the newcomer makes to an article in their first edit session is undone. Using data from all newcomers drawn from a set of Wikia wikis, we replicate these plots in our Study 1.

Additionally, RAD provides evidence that newcomer rejection is a mechanism for declining newcomer retention by estimating logistic regression models predicting newcomer survival. According to these models, newcomers are less likely to survive both when rejected and when the community was older. RAD presents separate models for good-faith newcomers and for all newcomers. The variables in their model are *year* to model time, *session edits* (the number of edits made in the first session) to account for the newcomer's early activity level, *messaged* specifying if the newcomer was messaged during

the newcomer's first 60 days, *reverted* indicating if the newcomer had an edit to an existing page rejected, and *deleted* to specify whether the newcomer created a new page which was deleted. We replicate these findings in our Study 2.

RAD's second analysis builds closely on the same logistic regression to test their theory that the rise of algorithmic quality control tools are an additional cause of Wikipedia's transition from rise to decline. They follow the methods of Geiger et al. [9] to track a number of different tools, including bots, to create a variable, *tool reverted*, that indicates whether the newcomer was reverted by a bot or human using an algorithmic tool. Plots in their paper show that tool use increased greatly and that desirable newcomers were increasingly likely to be reverted by tools. RAD argues that tool use may decrease newcomer retention over-and-above other forms of rejection, because tool users are less likely to practice a norm thought to mitigate discouragement following rejection known as the "BOLD, revert, discuss cycle" (BRD). BRD prescribes that reverting editors reciprocate discussion with those they revert. A negative coefficient for *tool reverted* in the logistic regression model described above provides evidence that algorithmic tool use may be a mechanism for declining newcomer retention. Because tool-based rejection is extremely rare on Wikia, we do not attempt to replicate RAD's finding that tool use is associated with lower levels of "BOLD, revert, discuss." The rest of their analysis is replicated in our Study 2.

In their third analysis, the RAD authors seek to measure the entrenchment of norms on Wikipedia. Norms are formed at many sites on Wikipedia, including three different kinds of norm pages analyzed by RAD: official policy pages, less formal guidelines, and informal essays. As evidence that norm entrenchment may be a cause of the decline, they plot the number of edits to these different kinds of pages over time. Edits to policies and guidelines began decreasing in 2006. Edits to essays slowed during the transition from rise to decline in 2008, decreasing thereafter.

RAD once again uses a logistic regression predicting whether an edit to a norm page is *reverted* to provide evidence of norm entrenchment: norm pages become more difficult to edit over time, measured as *year*, and those with greater *editor tenure* have their contributions to norm pages reverted less often than newer editors. They also model whether or not the norm page was an *essay*, the interaction between *editor tenure* and *essay*, and the interaction between *essay* and *year*. They find that essays had calcified substantially less than policies. Norm pages categories do not exist systematically on Wikia, so we are not able to reproduce the analysis of different levels of formality in norm pages. We replicate the other analyses from RAD's regression analysis in our Study 3.

As we have suggested, there are several parts of RAD that we do not attempt to replicate. RAD makes considerable effort to address a threat arising from high prevalence of vandalism on Wikipedia—because vandals may not intend to continue contributing, they may be unaffected by rejection. Additionally, a decline in the number of desirable newcomers may also be a cause of the decline in active contributors. To address these potential confounds, they hand-code a sample of "good-faith"

newcomers and report that the proportion of newcomers classified as "good-faith" fell during the period of rapid growth, about one year before the transition to decline. Their results for their sample of good-faith newcomers and for all newcomers are substantively similar.

Our work is only able to replicate RAD in a sample of all newcomers and does not attempt to create a sub-sample of desirable contributors. As experienced Wikipedians, the RAD authors and their volunteer coders were qualified to judge the quality of newcomers. Many Wikia wikis are about subject matter we are not familiar with. In many cases, they are written in languages we cannot read. We are confident in our results despite this omission for two reasons. First, RAD found very similar estimates in the models restricted to good-faith newcomers and the models that include all newcomers. To the extent that their good-faith-only estimates represented a robustness check that their analysis passed, we are comfortable forgoing it. Second, exploratory analysis of our data suggests that rates of vandalism are lower on Wikia than on Wikipedia, which should lessen the underlying threat.

Additionally, our analysis deviates from RAD in several ways that reflect the challenges and threats associated with studying a population of communities. Most importantly, we diverge from RAD by using estimation techniques and additional control variables appropriate to data nested within multiple communities. Because we consider multiple communities, it is possible that a single person might be a "newcomer" in our dataset more than once. To avoid analytic problems with repeated measures of users, and because individuals with experience in other wikis are likely not newcomers in the way RAD conceptualized them, our analysis identifies newcomers as individuals who have not edited any wiki in our sample and who are not marked as bots. In results available in our supplement, we fit models that include newcomers with prior experience in other wikis in our sample. Our results are not substantively different.

Of course, RAD itself has limitations. In particular, the quality of the evidence for the proposed causes of newcomer retention hinges on the assumption that other contemporaneous factors did not drive the decline. However, external events such as the rise of social media sites such as Facebook, as well as cultural changes in how the Internet was used and popularly understood, overlap with the transition from growth to decline. Studying multiple wikis that began at different points in time allows us to partially address this limitation in RAD. Our analysis inherits other limitations from RAD that we do not address. Importantly, entrenchment is theorized to contribute to declining newcomer retention, although this relationship is not modeled explicitly.

Data

Our dataset consists of page, user, and revision history data from 740 wikis publicly hosted on Wikia, the largest peer production wiki platform in terms of number of communities. Our initial dataset included all public edits to all Wikia wikis between February 24th 2002 and April 9th 2010. The 740 wikis whose data we use to replicate RAD include the top 1% of Wikia wikis by number of unique registered article editors.
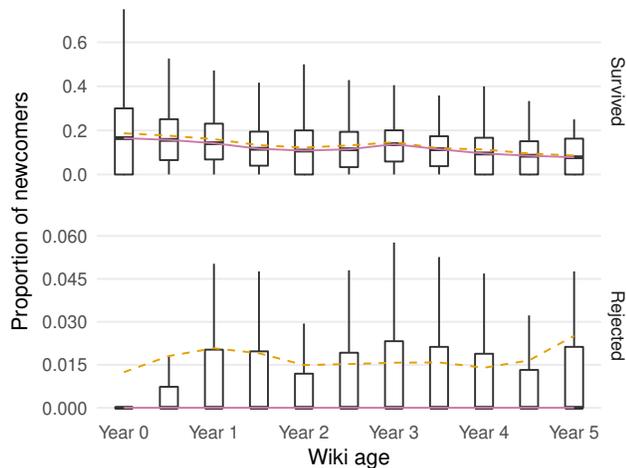
Figure 2. Newcomer survival and rejection over time. The orange dashed line shows the mean and the pink solid line shows the median. Years with data from at least 20 wikis are shown.

We include only these wikis because they have newcomer and governance activity appropriate for replicating RAD. We follow the RAD authors by only including newcomers we can observe for 180 days. We also obtain records identifying bot and administrator accounts from the Wikia API. We exclude 31 wikis where the API is unavailable because the wiki had been deleted since the XML archives we used were created.

Our dataset includes substantial variation between wikis, including linguistic diversity, activity level, and organizational complexity. Wikis vary in size, with numbers of unique contributors ranging from 98 to 309,694 (median 766). Some wikis in our sample produce collections of facts about popular culture, video games, and fandom. Others, such as the *Althistory Wiki*,[2] write collaborative fiction. Still others, such as *Uncyclopedia*,[3] parody Wikipedia. These 740 wikis also vary along our measures. For example, quality control practices vary, and the number of reverts within our communities ranges from 0 to 28,713 (median 57). Only 9.3% use bots to revert edits and the number of bot reverts among wikis with any ranges from 1 to 865 (median 5). The communities also vary in terms of policy making activity. A "namespace" is a high-level category used on all wikis. The project namespace is typically used for policy and documentation and governance activity. The number of edits to the project namespace ranges from 0 to 71,958 (median 85).

STUDY 1: TRAJECTORIES
We first replicate RAD's Figure 2 to determine whether the "rise and decline" pattern generalizes. To compare across wikis of vastly different size, we divide the number of monthly active contributors to each wiki in a given month by the standard deviation of that measure within that wiki. Figure 1 plots our results and shows a trajectory of growth and decline similar to English Wikipedia. While Wikipedia's exponential growth

---

[2] `http://althistory.wikia.com/` (`https://perma.cc/4EPQ-FW6Q`)
[3] `http://uncyclopedia.wikia.com/` (`https://perma.cc/L9CC-KN5A`)

was more explosive and lasted longer, the average active Wikia wiki follows a similar pattern. These communities begin small, tend to grow for 3-4 years, and then transition from growth to decline. Because few wikis in our dataset have existed for more than 5 years, we visualize only months with at least 43 active wikis (the 90th percentile). Although the downward trend continues after this threshold, the estimates become noisier.

Next we replicate RAD's Figures 3 and 4 to visualize the average trajectories in newcomer survival and rejection. Our results are shown in Figure 2. Lines connect the mean and median rates for all wikis active in each period to show the overall trend. Box plots visualize the variation between wikis. The top panel of Figure 2 corresponds to RAD's Figure 3 and shows box plots for the proportion of newcomers who survive in each year. As in Wikipedia, newcomer retention declines over time in the average wiki in our dataset. The trend is statistically significant (Spearman's $\rho = -0.16$, $p < 0.001$).

The bottom panel of Figure 2 corresponds to RAD's Figure 4 and shows box plots for the proportion of newcomers who are rejected over time. Although rejection is much less common in our wikis than in English Wikipedia, wikis in our sample exhibit increasing rates of newcomer rejection. The trend is statistically significant (Spearman's $\rho = 0.1$, $p < 0.001$). Although our estimates point in the same direction as RAD's, the average trajectory is qualitatively different. Rates of newcomer rejection are initially very low, increase over the first year, and remain level for most of the wiki's lifetime. They begin increasing again in the 4th year, when the number of active editors tends to decline.

STUDY 2: NEWCOMER SURVIVAL

Methods
We replicate RAD's first logistic regression model predicting whether a newcomer *survived* to test whether being *reverted* or *tool reverted* in the first edit session makes newcomers less likely to survive. RAD includes a single variable capturing the linear effect of time which reflects both the age of Wikipedia and the passage of calendar time. Having multiple wikis, we can tease apart wiki age and calendar time by measuring *wiki age* as the time since the first edit to each wiki in years. We include a linear specification of this variable following RAD. We also add a control for calendar time by adding a categorical variable for *quarter* that includes dummy variables for each 90-day calendar period. We also include *wiki*, a categorical variable with 740 levels to account for variation in baseline level of newcomer retention between wikis and to address issues of serial correlation in our standard errors. We do not report the results for these categorical variables, both for the sake of clarity and because they control for variation in the dataset that does not relate to the core theoretical concerns.

We cannot replicate several facets of this part of RAD's analysis in this study. RAD considers two kinds of newcomer rejection. Although the first of these corresponds with our measure for *reverted*, the RAD authors also consider whether an article created by a newcomer in their first session is *deleted*.

	β	SE	p value
Intercept	−2.83	0.88	< 0.001
Reverted	−0.72	0.04	< 0.001
Messaged	0.68	0.01	< 0.001
Tool reverted	−0.22	0.28	0.43
Session edits	0.33	0.01	< 0.001
Wiki age	−0.37	0.07	< 0.001
Deviance	262447		
Num. obs.	329636		

Table 1. Coefficient and standard errors estimated from our fitted logistic regression predicting newcomer survival. Coefficients for wiki and quarter are omitted from this table but available in the supplementary material.

We do not have information about deleted pages. Additionally, RAD considers two kinds of algorithmic tools: fully automated "bots" and semi-automated editing interfaces that automatically alert human users to suspected vandalism. These interfaces are either very rare or invisible on Wikia, so our measure of *tool reverted* only includes rejection by bots. Summary statistics for our analytic variables are available in the supplementary material.

Results
Table 1 shows our fitted regression model. This table closely mirrors the first column of RAD's Table 1. Like RAD, we find that newcomers reverted in their first edit session are less likely to survive ($\beta = -0.72$, $SE = 0.04$). The magnitude of our coefficient for *reverted* is very close to that reported by RAD ($\beta = -0.68$, $SE = 0.04$). According to our model, a newcomer who is reverted in their first session has 0.49 times the odds of continuing to contribute of a newcomer who is not reverted. We also find a negative relationship between *wiki age* and newcomer survival ($\beta = -0.37$, $SE = 0.07$). Again, this is very close to that reported by RAD ($\beta = -0.40$, $SE = 0.012$). Our parameter estimate for *tool reverted* ($\beta = -0.22$, $SE = 0.28$) suggests that newcomers who are rejected by a bot might be less likely to survive. However, the magnitude of this coefficient is too small relative to its standard error to support confidence in this conclusion.[4]

STUDY 3: ENTRENCHMENT

Methods
Finally, we replicate RAD's second model that predicts whether or not an edit to a policy page will be reverted. Although RAD carefully distinguishes between official policy pages and essays, Wikia contributors do not systematically label policy pages in this way. Therefore, we follow Shaw and Hill [25] and analyze all edits to the project namespace. This departure presents a substantial threat to validity, as the project namespace may be used for purposes besides documenting norms. Despite this limitation, we believe this measure provides the best available opportunity to study norm entrenchment in Wikia. Not all of the wikis in our sample utilize the

[4]A post-hoc power analysis suggests that, even if the true relationship is the same as that observed in RAD, we may have been unable to observe it because only 113 newcomers were reverted by bots in our dataset. See the supplementary materials for details.

	β	SE	p value
Intercept	−18.40	500.60	0.97
Editor tenure	−0.44	0.02	< 0.001
Wiki age	0.68	0.11	< 0.001
Deviance	96585		
Num. obs.	703614		

Table 2. Fitted estimated from our fitted logistic regressions predicting reverts of edits to project namespace pages. Categorical variables for wiki and quarter are omitted from this table but available in the supplementary material.

namespace, so we use only the subset of 662 wikis that do for this analysis. Summary statistics for our analytic variables are available in the supplementary material.

Results
Table 2 shows the fitted model results and replicates RAD's Table 2. Like RAD, we find that contributors with greater *editor tenure* are less likely to have their edits to policy pages reverted ($\beta = -0.44$, $SE = 0.016$). Our model predicts that, everything else equal, an editor with a 1-week tenure faces about 1.5 times the odds of having their edit reverted compared to an editor with a 1-year tenure. RAD reported an odds ratio of 1.3. Also consistent with RAD, we find that project page edits become more likely to be reverted as *wiki age* increases ($\beta = 0.68$, $SE = 0.11$). According to our model, an edit to the project namespace on a wiki that is 1 year old has about 2 times the odds of rejection as when the wiki is 1 week old.

DISCUSSION
We find that the patterns of community entrenchment documented in English Wikipedia also occur in comparable Wikia wikis. Wikis in our dataset experience growth in active contributors over about three years and then decline. Newcomer survival tends to decline over time, and newcomers who are rejected are less likely to survive. Older editors have more influence over norms, and norms become more difficult to change.

By studying these dynamics outside Wikipedia, we can rule out potential explanations of RAD's results linked to unique characteristics of Wikipedia, such as its specific culture. We can also rule out explanations linked to the specific time at which English Wikipedia experienced its decline. The diversity and size of our sample support both more precise estimation of the observed relationships in the data as well as stronger confidence in the validity of our inferences.

Our work has important limitations. Our data are observational, our sample may have unknown biases, and our measures may contain hidden sources of error. For example, wiki editors may change accounts, bots may be unreported, and the project namespace may include material unrelated to norms. Omitted variables may also bias our results. Readers should be careful not to draw causal conclusions from our findings.

The units of analysis in our regression models are newcomers and edits to project namespaces. Because the wikis in our sample have different numbers of each, the average effects we report could disproportionately reflect the experience of

users in the communities that contribute the most observations to our sample. As a robustness check, we fit another set of regression models where each wiki is given equal weight. Our substantive conclusions are robust to this change. Indeed, the re-weighted models suggest that the relationships reported in RAD may even be stronger in smaller or less active communities. In one unsurprising exception, we find that norm pages do not appear to become more difficult to edit over time in wikis that make very little use of the project namespace. These preliminary findings suggest analysis of the relationship between the size of a community and governance systems as a promising direction for future work. Details are available in the supplementary material.

Despite our effort at generalization, we cannot know if our findings will generalize beyond the wikis in our sample. That said, we think the mechanisms driving the emergence of entrenchment on wikis are similar to mechanisms theorized to drive the emergence and centralization of authority in democratic organizations. For example, Michels' "iron law of oligarchy" predicts that bureaucracies arising in large democratic organizations will centralize authority [19, 25]. Similarly, Freeman's "The Tyranny of Structurelessness" describes how, even when activist groups deliberately avoid creating formalized rules and bureaucracies, informal structures arise [8]. Indeed, due to their opacity, informal structures can be more difficult for newcomers to navigate than formalized bureaucracies. Drawing from this earlier theoretical work and our own results, we believe that the patterns of increasing entrenchment and newcomer rejection we estimate will generalize beyond the wikis in our sample to other peer production projects and informal organizations. Understanding why some communities in our sample show more entrenchment than others remains a fascinating subject for further research.

CONCLUSION
Our study supports RAD's claim that quality control practices help explain increases in entrenchment and decreases in growth among peer production communities. Our work contributes to social computing and peer production research by providing evidence in support of the external validity of RAD, an influential empirical study. We also contribute to a small but growing literature on replication in HCI by demonstrating a replication study focused on generalizability. Our evidence in support of generalizability rests not only on the signs of our regression coefficients, but also on the similarity of our point estimates and visualizations. This work supports designers and community managers who are acting on the implications in RAD's earlier work.

REFERENCES
1. Ajay Agrawal, Christian Catalini, and Avi Goldfarb. 2014. Some Simple Economics of Crowdfunding. *Innovation Policy and the Economy* 14, 1 (2014), 63–97.

2. Yochai Benkler. 2002. Coase's Penguin, or, Linux and the Nature of the Firm. *Yale Law Journal* 112, 3 (2002), 369–446.

3. Yochai Benkler, Aaron Shaw, and Benjamin Mako Hill. 2015. Peer Production: A Form of Collective Intelligence. In *Handbook of Collective Intelligence*, Thomas W. Malone and Michael S. Bernstein (Eds.). MIT Press, Cambridge, Massachusetts, 175–204.

4. Kenneth Bollen, John T. Cacioppo, Robert M. Kaplan, Jon A. Krosnick, James L. Olds, and Heather Dean. 2015. *Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science*. Technical Report. National Science Foundation.

5. Kate Crawford and Tarleton Gillespie. 2016. What Is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint. *New Media & Society* 18, 3 (March 2016), 410–428. DOI: http://dx.doi.org/10.1177/1461444814543163

6. Kevin Crowston, Kangning Wei, James Howison, and Andrea Wiggins. 2008. Free/Libre Open-Source Software Development: What We Know and What We Do Not Know. *Comput. Surveys* 44, 2 (March 2008), 7:1–7:35. DOI:http://dx.doi.org/10.1145/2089125.2089127

7. Andrea Forte, Vanesa Larco, and Amy Bruckman. 2009. Decentralization in Wikipedia Governance. *J. Manage. Inf. Syst.* 26, 1 (2009), 49–72. DOI: http://dx.doi.org/10.2753/MIS0742-1222260103

8. Jo Freeman. 1972. The Tyranny of Structurelessness. *Berkeley Journal of Sociology* 17 (Jan. 1972), 151–164.

9. R. Stuart Geiger, Aaron Halfaker, Maryana Pinchuk, and Steven Walling. 2012. Defense Mechanism or Socialization Tactic? Improving Wikipedia's Notifications to Rejected Contributors. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*. AAAI Publications, Dublin, Ireland, 122–129.

10. R. Stuart Geiger and David Ribes. 2010. The Work of Sustaining Order in Wikipedia: The Banning of a Vandal.

In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, New York, 117–126. DOI: `http://dx.doi.org/10.1145/1718918.1718941`

11. Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. 2013. The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist* 57, 5 (May 2013), 664–688. DOI: `http://dx.doi.org/10.1177/0002764212469365`

12. Aaron Halfaker, Aniket Kittur, and John Riedl. 2011. Don't Bite the Newbies: How Reverts Affect the Quantity and Quality of Wikipedia Work. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*. ACM, New York, New York, 163–172. DOI: `http://dx.doi.org/10.1145/2038558.2038585`

13. Michael T. Hannan and John Freeman. 1977. The Population Ecology of Organizations. *Amer. J. Sociology* 82, 5 (1977), 929–964. DOI: `http://dx.doi.org/10.2307/2777807`

14. Benjamin Mako Hill and Aaron Shaw. 2017. Studying Populations of Online Communities. In *Oxford Handbook of Networked Communication*, Sandra González-Bailón and Brooke Foucault Welles (Eds.). Oxford University Press, Oxford.

15. Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is Once Enough?: On the Extent and Content of Replications in Human-Computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, New York, 3523–3532. DOI: `http://dx.doi.org/10.1145/2556288.2557004`

16. Aniket Kittur and Robert E. Kraut. 2010. Beyond Wikipedia: Coordination and Conflict in Online Production Groups. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, New York, 215–224. DOI: `http://dx.doi.org/10.1145/1718918.1718959`

17. Piotr Konieczny. 2009. Governance, Organization, and Democracy on the Internet: The Iron Law and the Evolution of Wikipedia. *Sociological Forum* 24, 1 (2009), 162–192. DOI: `http://dx.doi.org/10.1111/j.1573-7861.2008.01090.x`

18. John W. Meyer and Brian Rowan. 1977. Institutionalized Organizations: Formal Structure as Myth and Ceremony. *Amer. J. Sociology* 83, 2 (Sept. 1977), 340–363.

19. Robert Michels. 1915. *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*. Hearst, New York, NY.

20. Felipe Ortega. 2009. *Wikipedia: A Quantitative Analysis*. Ph.D. dissertation. Universidad Rey Juan Carlos, Madrid, Spain.

21. Leysia Palen, Robert Soden, T. Jennings Anderson, and Mario Barrenechea. 2015. Success & Scale in a Data-Producing Organization: The Socio-Technical Evolution of OpenStreetMap in Response to Humanitarian Events. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, New York, 4113–4122. DOI: `http://dx.doi.org/10.1145/2702123.2702294`

22. Camille Roth, Dario Taraborelli, and Nigel Gilbert. 2008. Measuring Wiki Viability: An Empirical Assessment of the Social Dynamics of a Large Sample of Wikis. In *Proceedings of the 4th International Symposium on Wikis (WikiSym '08)*. ACM, New York, New York, 27:1–27:5. DOI:`http://dx.doi.org/10.1145/1822258.1822294`

23. Charles M. Schweik and Robert C. English. 2012. *Internet Success: A Study of Open-Source Software Commons*. MIT Press, Cambridge, Massachusetts.

24. W. Richard Scott and Gerald F Davis. 2006. *Organizations and Organizing: Rational, Natural and Open Systems Perspectives*. Pearson Prentice Hall, Upper Saddle River, New Jersey.

25. Aaron Shaw and Benjamin M. Hill. 2014. Laboratories of Oligarchy? How the Iron Law Extends to Peer Production. *Journal of Communication* 64, 2 (2014), 215–238. DOI:`http://dx.doi.org/10.1111/jcom.12082`

26. Bongwon Suh, Gregorio Convertino, Ed H. Chi, and Peter Pirolli. 2009. The Singularity Is Not Near: Slowing Growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration (WikiSym '09)*. ACM, New York, New York, 1–10. DOI: `http://dx.doi.org/10.1145/1641309.1641322`

27. Max L. Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, Dan Russell, and Harold Thimbleby. 2011. RepliCHI - CHI Should Be Replicating and Validating Results More: Discuss. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, New York, 463–466. DOI: `http://dx.doi.org/10.1145/1979742.1979491`