Promotion and tenure review
Candidate dossier

Aaron Shaw
Department of Communication Studies
Northwestern University
Evanston, IL 60208
847-467-3803
aaronshaw@northwestern.edu
http://aaronshaw.org

Submitted Fall, 2018

Table of Contents

Part I: Personal Statement p. 3

Part II: Curriculum Vitae p. 16

Part III: Research Materials p. 32

Part IV: Teaching Materials p. 435

Part I: Personal Statement

Research Statement

Overview

Diverse forms of social interaction, collective action, and collaboration increasingly happen in online communities and open digital commons like Wikipedia and Reddit. In these settings, millions of people create, organize, share, and access information resources of all kinds. Some of these communities solve information provision and integrity problems at unprecedented scale. They also upend traditional models of the organization of collective action and the production of public information goods.

I research organization and participation in open, online collaborative environments. In particular, I analyze peer production – a form of community-based online collaboration used to create public information goods like Wikipedia, OpenStreetMap, and the Linux operating system (Benkler et al., 2015). Peer production has transformed knowledge production and innovation, inspiring millions of attempts to build open, collaborative communities online. However, the vast majority of these efforts to start communities never get far. Why do some attempts to mobilize open online collaborations create valuable information goods while most do not? When does peer production advance inclusive, accessible, and high quality information production? How do the institutions of open collaboration interact with underlying social forces such as socioeconomic inequality?

I make three overlapping contributions in my work: (1) identify organizational factors that shape the growth, management, and survival of effective open collaboration systems; (2) extend understanding of the sources, mechanisms, and effects of participation inequalities online and in peer production communities; and (3) articulate agendas for the future of computational and empirical social scientific research into open collaboration, peer production, online communities, and related phenomena.

Overall, my research reveals that organizing effective peer production entails tradeoffs and shifts between openness and closure, mass participation and inequality, and egalitarianism and hierarchy. These results contradict prevailing scholarly and popular perspectives, which tend to portray open, online collectives as either radically democratic utopias or unruly anti-social mobs. I find that empirical reality lies beyond such dichotomies and situate online collaboration within long-running debates about organizations, communication, and the provision of public goods.

Below, I summarize the contributions and impact of my work.[1] Throughout, I tie the study of online participation to enduring topics of social scientific inquiry, including the organization of collective action, public goods production, and inequality. I pursue methodological rigor and innovation, as well as the dissemination of open access and reproducible research whenever possible in order to inform policy and policy. Increasingly, I advance these projects with students and collaborators in the Community Data Science Collective (CDSC), the research group I direct.

Organization of open collaboration and peer production

Open online collaboration offers extraordinary opportunities to harness communication networks and collective action for the production and distribution of valuable public information resources. I study **why and how some instances of online collective action and peer production mobilize large communities of contributors and sustain valuable information goods when the vast majority do not.** I focus on institutions and communication as determinants of effective peer production.

[1] As of November 2018, my research has generated over 30 publications which, according to Google Scholar, have been cited nearly 1,800 times and hold an h-index of 13.

My early studies in this area described governance and participation in U.S. political blogs. I designed a comparative analysis of elite U.S. political blogs, finding that the presence of participatory institutions differed across the left and right sides of the political spectrum (Shaw and Benkler, 2012).[2] I then investigated governance and leadership in the largest participatory U.S. political blog, *The Daily Kos*. Using mixed methods, I derived a novel theory and empirical evidence of "decentralized gatekeeping," the tendency for distributed interactions to create extreme status inequalities within a participatory media organization (Shaw, 2012a).[3] This finding — that even within the most open, informal organizations, micro-level social interactions generate persistent social structures — has become a key theme of my work.[4]

Seeking to elaborate and test organizational theories in peer production settings, I began a series of large-scale, comparative analyses in collaboration with Benjamin Mako Hill.[5] The first project built on my work on decentralized gatekeeping and tested whether peer production projects become more oligarchic over time. The findings, published in the *Journal of Communication* (Shaw and Hill, 2014), challenge conventional assumptions about peer production and suggest that while large communities may begin open, they become more hierarchical as they grow.

I expanded my research on the organization of peer production by forming the Community Data Science Collective (CDSC), a joint lab group I lead with Professor Hill that is based at Northwestern and the University of Washington.[6] In 2015, I was awarded a $500,000 grant by the National Science Foundation (NSF) to pursue this work together with Hill.[7] With this funding, I have completed multiple projects, many executed by graduate students under my mentorship. The work focuses on several facets of peer production organization summarized below: governance and leadership, newcomer socialization, organizational culture, and community boundaries.

Governance and leadership: Several projects build directly on my earlier work on oligarchy and gatekeeping. In a paper published this year in *The ACM Conference on Human Factors in Computing Systems* (*CHI*), we show striking patterns of leadership entrenchment and declining newcomer participation in a population of wikis (TeBlunthuis et al., 2018). This work, combined with my earlier studies, provides strong evidence that large peer production projects become more closed and hierarchical over time. In another recent paper published in *CHI*, we surveyed founders of wiki communities about their motives and goals for starting peer production projects (Foote et al., 2017a). Founders start communities for narrow reasons and with modest goals that do not align with the theories and models used to analyze their performance. A new project will evaluate whether the aims of project founders explain variations in founder behavior and subsequent project outcomes.

[2] The paper, co-authored with Yochai Benkler, received a Best Paper award from the Information Technology and Politics division of the American Political Science Association and honorable mention for Best Graduate Student Paper from the Communication and Information Technology section of the American Sociological Association (CITASA). It also attracted media coverage from National Public Radio and *The Nation*.

[3] The paper, published in *Politics & Society*, received honorable mention for Best Paper from CITASA in 2012.

[4] Both studies were part of my dissertation (Shaw, 2012b), which received honorable mention for Best Dissertation from the Communication and Technology Division of the International Communication Association.

[5] My collaborations with Hill began when we were both Ph.D. students. He is currently Assistant Professor of Communication at the University of Washington.

[6] Our group includes eight Ph.D. students, a post-doctoral fellow, and affiliated members. See wiki.communitydata.cc.

[7] The Project Summary and Project Description from the successful proposal are included with my published materials as part of this file.

Newcomer socialization: Even large peer production projects struggle to recruit and retain new contributors. I directed a CDSC collaboration with the Wikimedia Foundation[8] on a large-scale field experiment to evaluate a gamified tutorial, "The Wikipedia Adventure," for new Wikipedia contributors. We published the results in *The ACM Conference on Computer Supported Cooperative Work and Social Computing* (*CSCW*): contradicting prior literature and a user survey, we find that the system had no effect on newcomers' subsequent activity (Narayan et al., 2017). We are now preparing a submission to *CHI* evaluating an alternative approach to newcomer socialization based on lowering barriers to interpersonal communication (rather than a structured tutorial). Easier communication seems to have a positive, but limited impact on the activity and retention of newcomers. Both studies break new ground by finding that interaction matters more than formal training for mobilizing participation in peer production.

Organizational culture: Groups develop "organizational cultures" that encompass both particular ways of working as well as shared justifications for collective routines. I have directed two of the first studies of variations in organizational culture in online groups. One, forthcoming at *CSCW* 2018, reports a parallel ethnography of two pairs of online communities undergoing mergers. We find that the (mis)alignment of organizational cultures between the merging groups explains why one merger fails and the other succeeds (Kiene et al., 2018).[9] The second project compares over twenty of the largest language editions of Wikipedia (e.g., German, Japanese, and Hebrew Wikipedias among others) and shows that divergent communication practices explain variations in collaborative production.[10] These studies provide the first evidence of how organizational culture impacts online groups. In future work, I will investigate how particular cultural attributes shape community outcomes.

Community boundaries: Theories of collective action, public goods, and online communities make incommensurate predictions about the effects of open community boundaries. In a manuscript that has been conditionally accepted by *Communication Research*, we take advantage of an unannounced change that occurred on over 130 wikis that suddenly required would-be participants to log-in before they could contribute (Hill and Shaw, 2018a). We synthesize a theory of tradeoffs around community boundaries and compare the rate of high and low quality contributions before and after the design change. We find that the account requirement causes decreased participation, but that the number of lost high quality contributions far outweighs the number of deterred low quality contributions. This novel result suggests that communities should open boundaries when either the volume or costs of poor quality participation are low.

I am now developing new models of why and how open peer production organizations effectively sustain digital commons. During a 2017-2018 fellowship at the Center for Advanced Study in the Behavioral Sciences (CASBS) at Stanford, I started a project to explain why peer production projects may start open and unstructured only to become more hierarchical and closed over time. A separate manuscript examines whether ecological forces of competition and mutualism affect online petitions.[11] Another manuscript assesses whether communication network structures explain the growth and survival of nascent peer production communities. Most recently, following a successful concept note submission,

[8] The Wikimedia Foundation (WMF) is a non-profit organization that helps to manage the technologies, intellectual property, and communities of volunteers who maintain Wikipedia and related projects. See wikimedia.org for details.

[9] The paper, published with Charles Kiene and Benjamin Mako Hill, received honorable mention for a Best Paper award at *CSCW* 2018.

[10] A non-archival draft of this project appeared as Maddock et al. (2017).

[11] A non-archival draft of this project appeared as TeBlunthuis et al. (2017).

I am in discussions with the Ford and Sloan Foundations about a $195,000 grant to the CDSC for a study of the dynamics of peer production in areas of critical digital infrastructure. We will develop models and practical strategies to avert failures among peer production projects, some of which are relied upon for millions of transactions and network connections daily.

Participation inequalities online

Extreme participation inequalities are a persistent, widely-observed attribute of online environments. A small proportion of participants account for a massive proportion of total activity — whether the activity is edits, tweets, posts, "likes," or something else. Often, participation inequalities reproduce macro-social and structural inequalities along lines of gender, class, or socioeconomic status. However, in other cases, online communities attract and mobilize historically marginalized groups. **What are the sources and impacts of differential patterns of inequality in online participation?**

Several of my early projects revealed the importance of digital inequalities. The first, a large-scale experiment comparing incentive designs in Amazon's Mechanical Turk (AMT), demonstrated positive effects of social comparisons and financial punishment on worker performance (Shaw et al., 2011).[12] A secondary finding, however, showed that workers' web use skills and country of residence explained their performance far better than incentives. A second experiment compared motivations of AMT workers in India and the United States and found many differences between the two groups (Antin and Shaw, 2012). Around the same time, I conducted the aforementioned study of gatekeeping on Daily Kos, finding large, persistent, and consequential inequalities of participation and status (Shaw, 2012a).

Such broad evidence led me to investigate the sources and effects of digital inequalities more deeply. First, I conducted an analysis of the relationship between young adults' online participation, web use skills, and political engagement with Professor Eszter Hargittai. We found some of the first evidence that online participation and web use skills predicted inequalities in civil and political engagement, publishing the results in the *Journal of Broadcasting and Electronic Media* (Hargittai and Shaw, 2013).

Next, I sought to understand socioeconomic inequalities in peer production communities like Wikipedia. A widely-publicized survey indicated that perhaps fewer than 10% of Wikipedia editors were female. I explored this "gender gap" in two studies. The first, published in *PLOS ONE* (Hill and Shaw, 2013), revisited the original survey and applied a novel weighting technique to construct more precise estimates of demographic attributes. We found that the true proportion of female editors was likely 20-25%, more than double the original estimate. The second study, published in *Information, Communication & Society*, further refined this story by finding that the Wikipedia gender gap arose only among highly-skilled internet users (Hargittai and Shaw, 2015). In other words, controlling for numerous background attributes and internet use variables, young males and females with low internet skills are equally *unlikely* to edit Wikipedia. These two projects have reshaped the debate on the Wikipedia gender gap and have influenced the policy interventions of the Wikimedia Foundation. My work shifted understanding of the sources of the gender gap and suggests that overcoming participation divides will require lowering the barriers that prevent less-skilled participants from getting involved in the first place.

Most recently, I elaborated and tested a more comprehensive model of how individuals participate online through a similar process to other "pipelines" of skills and career development. I designed and

[12] The paper, published with John Horton and Daniel Chen when we were all Ph.D. students, received honorable mention for a Best Paper award at *CSCW* 2011 and was covered by the popular technology blog BoingBoing.

commissioned an original survey of adult internet users in the United States in Summer 2016 in collaboration with Professor Hargittai. Our first findings, published in *Journal of Communication* earlier this year, support a model of a pipeline of digital inequalities applied to the case of Wikipedia editing (Shaw and Hargittai, 2018). Gaps of education and web use skills explain variations at every stage of the pipeline of awareness and behaviors necessary to become a Wikipedia editor. Other factors including gender, income, and race and ethnicity matter at specific stages, such as knowing that Wikipedia can be edited in the first place. These results contextualize the Wikipedia gender gap debate and set the stage for more robust inquiry into the underpinnings of online stratification. I conclude that expanding participation in online knowledge production requires disseminating the knowledge that makes such participation possible. I have recently presented this work to Wikimedia Foundation staff.

I will continue this research through several new projects. I am preparing a manuscript for submission with Professor Hargittai that compares our national sample of Internet users with a sample of survey participants from AMT. In response to prior literature that draws general behavioral inferences from AMT worker samples we demonstrate two key areas where AMT workers diverge: their online participation and their prosocial attitudes. I am also designing a follow-up project to evaluate the role of internet skills as an intervening factor that mediates online prosocial behavior. Finally, I have planned a study with Professor Hargittai to extend the pipeline model to other participatory websites.

Research agenda setting

I am also defining agendas and standards for research on peer production, computational communication, and online organizations. Here, my contributions include the dissemination of rigorous empirical tools, analysis, and data for reproducible research as well as the synthesis of prior literature. I have pursued this work through several published articles and invited book chapters, as well as editorial and advisory service roles with several organizations that I discuss later in my Service Statement.

In several pieces, I synthesized prior literature on crowd work and collective action in the context of human computer interaction research. The first, published in *CSCW* (Kittur et al., 2013), looks at crowd-based systems of work and collaboration. My co-authors and I advocate for a more empirically-informed and humane future of crowd work. My contributions to the piece focused on initial conceptualization and co-authoring a section on worker motivation and behavior. This paper has made a particularly large impact in the fields of human computation and crowdsourcing, becoming one of the most cited contributions to *CSCW* in recent years.

The second project summarized design challenges involved in mobilizing online collective action. I worked with collaborators to identify key challenges that confront systems builders who seek to mobilize, organize, and sustain participatory collective action in online environments. In a paper published in the AAAI *International Conference on Weblogs and Social Media* (Zhang et al., 2014), we describe barriers to online collective action and deployments of a prototype system, "WeDo," intended to overcome them. We then published a short, invited companion piece in the ACM magazine *Interactions* that lays out the core theoretical argument in a more concise, accessible manner (Shaw et al., 2014).[13]

I have also developed novel research datasets and tools for reproducible peer production research. I began with the aforementioned Wikipedia gender gap study (Hill and Shaw, 2013), in which the propensity score adjustment technique we applied had appeared almost exclusively in the private sec-

[13] I was awarded grants totaling $50,000 by Microsoft Research's Fuse Labs (one grant jointly with Professors Liz Gerber and Haoqi Zhang) to support this work and two empirical studies of crowd-based resource sharing systems at *The AAAI Workshop on Human Computation* (*HCOMP*) and *CSCW* (Kim et al., 2016, 2017).

tor, resulting in no published, empirical applications. We used two publicly available survey datasets and released our analysis code and the paper itself under open access licenses. In response, several subsequent studies have used these resources and peer production organizations (including the Wikimedia Foundation) have sought our advice on survey designs and analysis.

I published two related measurement projects in the *Proceedings of the ACM Symposium on Open Collaboration* (*OpenSym*) (Hill and Shaw, 2014, 2015). Both introduce central dimensions of wiki-based collaboration systems that had either been ignored or systematically mischaracterized in previous research. We describe new software that enables researchers to identify these overlooked dimensions of wikis as well as demonstration datasets and analysis revealing the impact of the new tool and its implications for future research. We released all the materials from these studies under open access licenses to facilitate reproducibility and adoption by other researchers.

Finally, I have published several invited agenda setting and instructional handbook chapters. In one, published by MIT Press, I review the state of social science research on peer production with Yochai Benkler and Benjamin Mako Hill (Benkler et al., 2015). In a second, I provide an informal orientation to research on online crowds (Shaw, 2015). In a new contribution to the *Sage Handbook of Social Media*, I conduct a fully reproducible analysis of social media research with Jeremy Foote and Professor Hill that introduces popular computational methods as well as novel findings about the development of the literature (Foote et al., 2017b).[14] Most recently, I wrote a chapter forthcoming in *The Oxford Handbook on Networked Communication* with Professor Hill identifying key challenges and opportunities for organizational analysis of data from populations of online communities (Hill and Shaw, 2018b). All of these chapters advance the diffusion, accessibility, and reproducibility of my research expertise.

I aim to broaden the accessibility and impact of research on peer production in future work. I am writing a book proposal titled *A Concise Introduction to Peer Production* together with Professor Hill and Christian Pentzold. We will submit the proposal to MIT Press for review for their "Essential Knowledge Series" later this Fall. The book will synthesize nearly two decades peer production research for non-specialist and non-scholarly readers. It will also summarize the implications for the future of online organization, public goods production, and open collaboration in digital commons.

Summary and future directions

Open organizations, peer production, and participatory digital commons have become critical spaces of social interaction, communication, and collaboration. The challenges of understanding, designing, and governing effectively in these settings are more urgent than ever. I will continue to pursue research that advances the contribution areas described above by extending prior findings and pursuing new, innovative discoveries. My projects integrate theory-driven analysis and rigorous, reproducible, methods with a growing emphasis on novel conceptual frameworks. In new projects, I also aim to disseminate theories, findings, and implications from research on digital commons beyond existing specialist audiences. I am increasingly directing this work by mentoring student members of the Community Data Science Collective, with whom I will continue to explore new directions.

[14] This chapter received coverage in the popular technology press (e.g., O'Reilly Radar and BoingBoing).

Teaching

Through teaching, I seek to facilitate students' mastery of skills that enable their intellectual, civic, and professional development. I create inclusive and interactive classroom environments, in which students pursue group and project-based learning and participate in multiple forms of assessment. I also apply insights from my research in my teaching and cultivate deeply collaborative mentoring relationships that support students publishing independent research. My pedagogical approach has benefited from a Searle Center for Teaching Junior Faculty Fellowship during the 2013-2014 Academic Year.

Students have embraced my teaching, giving consistently positive feedback on my classes[15] and naming me to the 2014 Associated Student Government Faculty and Administration Honor Roll in recognition of my teaching and mentorship. Both undergraduate and graduate students describe my teaching style as "enthusiastic", "vibrant", and "participatory." The students' outcomes have also been extremely positive, with many completing independent projects and publications and taking positions in the academic, non-profit, and private sectors. Recent placements among students I have worked with include positions at Facebook and the Wikimedia Foundation, as well as a Fulbright Scholarship and a tenure-track faculty position at an elite liberal arts college.

I have developed five courses since coming to Northwestern.[16] At the undergraduate level, I completely re-designed Communication Studies 378, "Online Communities and Crowds," which I have offered three times. The course is part of the Digital Media and the Organizational and Strategic Communication Modules in the Communication Studies Department. The centerpiece of the course is a sequence of project-based learning assignments in which students work in groups to write Wikipedia articles on topics of their choosing; engage in peer feedback and evaluation of their performance; and assess the experience through an iterative cycle of written reflection.[17] These assignments have a huge impact beyond the classroom, resulting in encyclopedia content that has been viewed hundreds of thousands of times. The class has also led to multiple undergraduate student research mentorships, including a research assistantship and an independent study that generated a peer-reviewed publication with a then-undergraduate student as the lead author (Burgess and Shaw, 2016).

At the graduate level, I have created two elective seminars and updated the curriculum for two required seminars. The first elective seminar, "Online Collective Action" introduces students to a mix of classic and contemporary research on collective action and organization, with a focus on open communities on the internet. The second elective seminar, "Research Design for Causal Inference," provides advanced methodological training in quantitative social scientific research methods for identifying causal effects. To my knowledge it was the first class on these methods taught in any communication department in the United States. As part of this course I provide students with a hands-on, applied introduction to the R statistical computing language.

Since 2015, I have served as Director of Graduate Studies for the Media, Technology and Society (MTS) Ph.D. program. In my role as program director, I updated and taught the two core seminars required of both MTS students and those in the Ph.D. program on Technology and Social Behavior

[15] Across all the CTEC course evaluations of all the classes I have taught, the "overall rating of the instruction" averages 5.4 out of 6.

[16] Links to the most recent website for each course are available at aaronshaw.org/teaching. I have included a copy of the syllabus from each course as part of this dossier.

[17] I designed these assignments and first implemented them during my Searle Center Fellowship with guidance from the Searle Center staff. The final report I submitted as part of my Searle Fellowship is included among my teaching materials as part of this file.

(TSB). The first of these, "Introduction to graduate research in TSB and MTS" (MTS 501) introduces new Ph.D. students to a wide range of professional skills and perspectives for building a successful and satisfying research career in graduate school and beyond. The second, "The Practice of Scholarship" (MTS 503) supports second-year Ph.D. students developing the skills to design, conduct, and publish independent scholarly research.

Both of my elective seminars have enjoyed steady enrollment from the MTS and TSB Ph.D. programs as well as graduate students from SESP, McCormick, Weinberg, and Kellogg programs. Student work in these classes has led to multiple research proposals, qualifying exams, dissertation chapters, and peer-reviewed publications. At least two students from my MTS 501 class have developed materials from that class into successful NSF Graduate Research Fellowship Program (GRFP) proposals, resulting in multi-year external awards.

I also mentor graduate students to become active, independent scholars and instructors through my research group, the Community Data Science Collective (CDSC)[18], which I direct with Professor Benjamin Mako Hill of the University of Washington. In the CDSC, I co-supervise eight doctoral students at Northwestern and the University of Washington. I chair three dissertation committees and have served on five others. Nine students (five in MTS and four in TSB) have completed qualifying exams under my supervision. The majority of these mentoring activities have already yielded, or are in the process of yielding research projects of publishable quality, including multiple co-authored pieces (e.g. Kim et al., 2017; Twyman et al., 2017). Most recently, my first doctoral advisee has accepted and is scheduled to begin a tenure-track faculty position in Computer Science at Carleton College in 2019.

In the coming years, I plan to extend my teaching and mentorship activities in multiple ways. Together with my advisee Jeremy Foote, I am developing new undergraduate and public workshops in computational and data science research methods. I am also introducing a new graduate seminar in statistics and statistical computing in Spring 2019. Together with Professor Daniel Immerwahr (History), I am designing an interdisciplinary course on theories and uses of information. In addition, I have just initiated a process to refine the graduate curriculum of the MTS program in consultation with the MTS faculty members, students, and other leaders in the School of Communication (more on this below).

[18] https://communitydata.cc

SERVICE

I embrace leadership and service opportunities at Northwestern as well as in my professional communities. In the Department of Communication Studies at Northwestern, I have served on multiple committees, including the Undergraduate Curriculum Committee, a faculty search committee in Computation and Communication, the Steering Committee for the undergraduate Digital Media Module, as well as the Advisory Committee to the Master of Science in Communication (MSC) program.

In the School of Communication, I have served as Director of Graduate Studies of the Media, Technology, and Society (MTS) Ph.D. program since 2015. MTS is the largest Ph.D. program in the School of Communication with 40 students currently enrolled. In this role, I have chaired the program's admissions committee three times and coordinated its speaker series for several years. My responsibilities also include directing the annual student review process, chairing regular faculty meetings, and holding an annual town hall open to all current MTS students. I also initiated the creation of a new Graduate Student Steering Committee which operates jointly across the MTS and Technology and Social Behavior (TSB) Ph.D. programs to support student professional development and community engagement.

The MTS program has sustained its already impressive record of success under my leadership. In a 2017-2018 program review of the Communication Studies Department, external reviewers praised MTS for training interdisciplinary and innovative scholars. Last year, the School of Communication conducted an independent survey of graduate students across all of its Ph.D. programs and MTS performed well, with students noting the flexibility and interdisciplinarity of the program. Both the SoC survey and the program review also identified opportunities for improvement in MTS and, as mentioned above, I have initiated conversations among the MTS faculty and other leaders in the School of Communication to develop a set of proposals to this end. The goal of this process is to reinforce the program's strengths while also adapting to needs that have emerged as the program has grown and evolved over the past ten years. In particular, I aim to articulate common learning objectives and procedures to support more effective advising and more consistent assessment of doctoral student milestones.

Beyond Northwestern, I am an active member of multiple professional networks in which I also pursue leadership roles. I currently serve on the applied and public policy award committee for the International Communication Association (ICA) at the invitation of ICA President Patricia Moy. I am also on the Fellows Selection Committee for the Center for Advanced Study in the Behavioral Sciences at Stanford. I am a founding editorial board member for two journals, *Social Media and Society* and *Computational Communication Research*, and have served as a Program Committee member or Program Committee Chair for multiple Association for Computer Machinery (ACM) conferences and workshops. I serve as a reviewer for numerous journals and conferences in communication, human-computer interaction, and sociology; scholarly presses including MIT Press and Sage; private foundations, as well as multiple directorates of the U.S. National Science Foundation.

I also regularly advise some of the largest online communities based on the findings from my research. In this capacity, I am a member of the Advisory Board to the Zooniverse Project (a citizen science project based at Oxford University in the United Kingdom and the Adler Planetarium in Chicago). I regularly present to and collaborate with research, communication, and community management staff at the Wikimedia Foundation (the organization that maintains Wikipedia and related projects), wikiHow Inc., and Wikia Inc. These commitments have all grown out of my research work and help to ensure the broader impact of my scholarship.

References

Antin, J. and **Shaw, A.** (2012). Social desirability bias and self-reports of motivation: A study of Amazon Mechanical Turk in the US and India. In *Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2925–2934, New York, NY, USA. ACM.

Benkler, Y., **Shaw, A.**, and Hill, B. M. (2015). Peer production: A form of collective intelligence. In Bernstein, M. and Malone, T., editors, *The Handbook of Collective Intelligence*. MIT Press, Cambridge, MA.

Burgess, E. R. and **Shaw, A.** (2016). Evaluating open collaboration opportunities in the fire service with FireCrowd. In *Proceedings of the 12th International Symposium on Open Collaboration*, OpenSym '16, pages 22:1–22:4, New York, NY, USA. ACM.

Foote, J., Gergle, D., and **Shaw, A.** (2017a). Starting online communities: motivations and goals of wiki founders. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 6376–6380, New York, NY, USA. ACM.

Foote, J., **Shaw, A.**, and Hill, B. M. (2017b). A computational analysis of social media scholarship. In Burgess, J., Marwick, A., and Poell, T., editors, *The Sage Handbook of Social Media*. Sage, Thousand Oaks, CA, USA.

Hargittai, E. and **Shaw, A.** (2013). Digitally savvy citizenship: The role of internet skills and engagement in young adults' political participation around the 2008 presidential election. *Journal of Broadcasting & Electronic Media*, 57(2):115–134.

Hargittai, E. and **Shaw, A.** (2015). Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information, Communication & Society*, 18(4):424–442.

Hill, B. M. and **Shaw, A.** (2013). The Wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PLOS ONE*, 8(6):e65782.

Hill, B. M. and **Shaw, A.** (2014). Consider the redirect: A missing dimension of Wikipedia research. In *Proceedings of The International Symposium on Open Collaboration*, OpenSym '14, pages 28:1–28:4, New York, NY, USA. ACM.

Hill, B. M. and **Shaw, A.** (2015). Page protection: Another missing dimension of Wikipedia research. In *Proceedings of the 11th International Symposium on Open Collaboration*, OpenSym '15, New York, NY, USA. ACM.

Hill, B. M. and **Shaw, A.** (2018a). The hidden costs of requiring accounts online: Quasi-experimental evidence from peer production. Conditionally accepted to *Communication Research*.

Hill, B. M. and **Shaw, A.** (2018b). Studying populations of online communities. Forthcoming in González-Bailón, S., and Foucault Welles, B., editors, *The Oxford Handbook of Networked Communication*, Oxford University Press, Oxford, UK.

Kiene, C., **Shaw, A.**, and Hill, B. M. (2018). Managing organizational culture in online group mergers. *Proceedings of the ACM on Human-Computer Interaction*. Vol. 2, CSCW Article 89 (November 2018). ACM, New York, NY, USA.

Kim, Y., Harburg, E., Azria, S., **Shaw, A.**, Gerber, E., Gergle, D., and Zhang, H. (2016). Studying the effects of task notification policies on participation and outcomes in on-the-go crowdsourcing. In *Fourth AAAI Conference on Human Computation and Crowdsourcing, HCOMP '16*, Palo Alto, CA, USA. AAAI.

Kim, Y., **Shaw, A.**, Zhang, H., and Gerber, E. (2017). Understanding trust amid delays in crowdfunding. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 1982–1996, New York, NY, USA. ACM.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., **Shaw, A.**, Zimmerman, J., Lease, M., and Horton, J. (2013). The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 1301–1318, New York, NY, USA. ACM.

Maddock, J., **Shaw, A.**, and Gergle, D. (2017). Talking about talk: Coordination in large online communities. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 1869–1876, New York, NY, USA. ACM.

Narayan, S., Orlowitz, J., Morgan, J., Hill, B. M., and **Shaw, A.** (2017). The Wikipedia Adventure: Field evaluation of an interactive tutorial for new users. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 1785–1799, New York, NY, USA. ACM.

Shaw, A. (2012a). Centralized and decentralized gatekeeping in an open online collective. *Politics & Society*, 40(3):349–388.

Shaw, A. (2012b). *An Interactional Account of Online Collective Action*. Dissertation, University of California, Berkeley, Berkeley, CA.

Shaw, A. (2015). Hired hands and dubious guesses: Adventures in crowdsourced data collection. In Hargittai, E. and Sandvig, C., editors, *Digital Research Confidential*. MIT Press, Cambridge, MA.

Shaw, A. and Benkler, Y. (2012). A Tale of Two Blogospheres: Discursive Practices on the Left and Right. *American Behavioral Scientist*, 56(4):459–487.

Shaw, A. and Hargittai, E. (2018). The pipeline of online participation inequalities: The case of Wikipedia editing. *Journal of Communication*, 68(1):143–168.

Shaw, A. and Hill, B. M. (2014). Laboratories of oligarchy? How the iron law applies to peer production. *Journal of Communication*, 64(2):215–238.

Shaw, A., Zhang, H., Monroy-Hernández, A., Munson, S., Hill, B. M., Gerber, E., Kinnaird, P., and Minder, P. (2014). Computer Supported Collective Action. *Interactions*, 21(2):74–77.

Shaw, A. D., Horton, J. J., and Chen, D. L. (2011). Designing incentives for inexpert human raters. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, CSCW '11, pages 275–284, New York, NY, USA. ACM.

TeBlunthuis, N., **Shaw, A.**, and Hill, B. M. (2017). Density dependence without resource partitioning: Population ecology on Change.org. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17 Companion, pages 323–326, New York, NY, USA. ACM.

TeBlunthuis, N., **Shaw, A.**, and Hill, B. M. (2018). Revisiting "The rise and decline" in a population of peer production projects. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 355:1–355:7, New York, NY, USA. ACM.

Twyman, M., Keegan, B. C., and **Shaw, A.** (2017). Black lives matter in Wikipedia: Collective memory and collaboration around online social movements. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 1400–1412, New York, NY, USA. ACM.

Zhang, H., Monroy-Hernández, A., **Shaw, A.**, Munson, S. A., Gerber, E., Hill, B. M., Kinnaird, P., Farnham, S. D., and Minder, P. (2014). WeDo: End-to-end computer supported collective action. In *Eighth International AAAI Conference on Weblogs and Social Media, ICWSM '14*, Palo Alto, CA, USA. AAAI.

Part 2: Curriculum Vitae

Aaron Shaw

Department of Communication Studies
Northwestern University

2240 Campus Drive
Evanston, IL 60660

http://aaronshaw.org
aaronshaw@northwestern.edu

Appointments

Northwestern University, 2012-Present
Assistant Professor, Department of Communication Studies, 2013-Present
Director of Graduate Studies, Media, Technology & Society (MTS), 2015-Present

Assistant Professor (by courtesy), Sociology
Faculty Affiliate, Buffett Institute for Global Studies
Faculty Associate, Northwestern Institute for Policy Research
Faculty Affiliate, SONIC Lab
Visiting Assistant Professor, Department of Communication Studies, 2012-2013
Visiting Assistant Professor (by courtesy), Sociology, 2012-2013

Harvard University, 2008-Present
Berkman Klein Center for Internet & Society
Faculty Associate, 2013-Present
Fellow, 2008-2013

Education

University of California, Berkeley
2012. Ph.D., Sociology.
Dissertation: *An Interactional Account of Online Collective Action*
2010. Master of Arts, Sociology.

Stanford University,
2005. Master of Arts, Humanities
2004. Bachelor of Arts (with honors), Humanities

Awards, Fellowships, and Grants

2018. ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2018), Best Paper Honorable Mention for *Managing Organizational Culture in Online Group Mergers*.

2017-2018. Lenore Annenberg and Wallis Annenberg Fellow in Communication. Center for Advanced Study in the Behavioral Sciences (CASBS) at Stanford University.

2016. National Science Foundation, IIS:CHS: Pathways to Community Success: Advancing a Comparative Science of Online Collaborative Organization, Co-PI with Benjamin Mako Hill, University of Washington ($499,684).

2016. Summer Institute on Organizations and Their Effectiveness. Center for Advanced Study in the Behavioral Sciences (CASBS) at Stanford University.

2014. Faculty and Administrator Honor Roll, Northwestern University Associated Student Government (ASG).

2014. Microsoft Research, FUSE Labs. Sharing Economy Grant, Co-investigator with Haoqi Zhang and Liz Gerber, Northwestern University ($25,000).

2013. Microsoft Research, FUSE Labs Junior Faculty Research Award ($25,000).

2013-2014. Junior Faculty Fellow, Searle Center for Advancing Teaching & Learning, Northwestern University.

2013. Honorable Mention, Dordick Dissertation Award, Communication & Technology Division, International Communication Association (ICA).

2013. Information Technology and Politics Section of the American Political Science Association (APSA). Best Paper Award for *A Tale of Two Blogospheres: Discursive Practices on the Left and Right*.

2012. Post-doctoral Fellow, Center for Information Technology Policy (CITP), Princeton University (declined).

2012. Communication and Information Technology Section of the American Sociological Association (CITASA), Best Paper Honorable Mention for *Centralized and Decentralized Gatekeeping in an Open Online Collective*.

2012. Communication and Information Technology Section of the American Sociological Association (CITASA), Best Graduate Student Paper Honorable Mention for *A Tale of Two Blogospheres: Discursive Practices on the Left and Right*.

2011. ACM Conference on Computer Supported Cooperative Work (CSCW2011), Best Paper Honorable Mention for *Designing Incentives for Inexpert Raters*.

2010-2011. Berkman Center Law Lab Fellowship supported by Ewing Marion Kauffman Foundation.

2009-10. Hornaday Graduate Summer Fellow, Greater Good Science Center, UC Berkeley.

2009. UC Berkeley Sociology Department Research Grant.

2008-2009. Berkman Center Fellowship supported by the Ford Foundation.

2008. UC Office of the President, Labor & Employment Research Fund (LERF) Fellowship.

2007. U.S. Department of Education, Foreign Language and Area Studies (FLAS) Summer Fellowship.

2007. Andrew Mellon Summer Fellowship for Latin American Sociology.

2006-07. U.S. Department of Education, Foreign Language and Area Studies (FLAS) Graduate Fellowship.

Publications

Refereed Articles

2018 (conditionally accepted). Hill, Benjamin Mako and Aaron Shaw. Hidden costs of requiring accounts online: Quasi-experimental evidence from peer production. *Communication Research.*

2018. Kiene, Charles, Aaron Shaw, and Benjamin Mako Hill. Managing organizational culture in online group mergers. *Proceedings of the ACM Human-Computer Interaction* 2, CSCW, Article 89 (November 2018), ACM, New York, NY.

2018. Aaron Shaw and Eszter Hargittai. The pipeline of online participation inequalities: The case of Wikipedia editing. *Journal of Communication* 68(1): 143–168.

2018. Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. Revisiting "The rise and decline" in a population of peer production projects. *Proceedings of the 2018 ACM Conference on Human Factors in Computing Systems (CHI '18).*

2017. Foote, Jeremy, Darren Gergle, and Aaron Shaw. Starting online communities: Motivations and goals of wiki founders. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17).*

2017. Kim, Yongsung, Aaron Shaw, Haoqi Zhang, and Elizabeth Gerber. Understanding trust amid delays in crowdfunding. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work (CSCW '17).*

2017. Narayan, Sneha, Jake Orlowitz, Jonathan Morgan, Benjamin Mako Hill, and Aaron Shaw. The Wikipedia Adventure: Field evaluation of an interactive tutorial for

new users. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work (CSCW '17).*

2017. Twyman, Marlon, Brian C. Keegan, and Aaron Shaw. Black Lives Matter in Wikipedia: Collaboration and collective memory around online social movements. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work (CSCW '17).*

2016. Burgess, Eleanor R. and Aaron Shaw. Evaluating open collaboration opportunities in the fire service with FireCrowd. *Proceedings of the 12th International Symposium on Open Collaboration* (*OpenSym '16).*

2016. Kim, Yongsung, Emily Harburg, Shana Azria, Aaron Shaw, Elizabeth Gerber, Darren Gergle, and Haoqi Zhang. Studying the effects of task notification policies on participation and outcomes in on-the-go crowdsourcing. *Proceedings of thje Fourth AAAI Conference on Human Computation and Crowdsourcing (HCOMP '16).*

2015. Hargittai, Eszter and Aaron Shaw. Mind the skills gap: The role of internet know-how and gender in contributions to Wikipedia. *Information, Communication, and Society* 18(4): 424-442.

2015. Hill, Benjamin Mako and Aaron Shaw. Page protection: Another missing dimension of Wikipedia research. *Proceedings of the 11th International Symposium on Open Collaboration* (*OpenSym '15).*

2014. Hill, Benjamin Mako and Aaron Shaw. Consider the redirect: A missing dimension of Wikipedia research. *Proceedings of the International Symposium on Open Collaboration (OpenSym '14).*

2014. Shaw, Aaron and Benjamin Mako Hill. Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication* 64(2): 215–38.

2014. Zhang, Haoqi, Andres Mónroy-Hernandez, Aaron Shaw, Sean Munson, Elizabeth Gerber, Benjamin Mako Hill, Peter Kinnaird, and Shelley Farnham. WeDo: Exploring computer supported collective action. *Proceedings of the International Conference on Weblogs and Social Media (ICWSM).*

2013. Hargittai, Eszter and Aaron Shaw. Digitally savvy citizenship: The role of internet skills and engagement in young adults' political participation around the 2008 presidential election. *Journal of Broadcasting and Electronic Media (JoBEM) 57*(2):115-134.

2013. Hill, Benjamin Mako and Aaron Shaw. The Wikipedia gender gap revisited: Characterizing survey response bias in peer production communities. *PLOS ONE.* 8(6):e65782.

2013. Kittur, Aniket, Jeffrey V. Nickerson, Michael S. Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John J. Horton. The future of crowd work. *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work (CSCW '13).*

2012. Shaw, Aaron. Centralized and decentralized gatekeeping in an open online collective. *Politics & Society. 40*(3): 349-388.

2012. Antin, Judd and Aaron Shaw. Social desirability bias and self-reports of motivation: A cross-cultural study of Amazon Mechanical Turk in the US and India. *Proceedings of the 2012 ACM Conference on Human Factors in Computing Systems (CHI '12).*

2012. Shaw, Aaron and Yochai Benkler. A tale of two blogospheres: Discursive practices on the left and right. *American Behavioral Scientist. 56*(4): 459-487.

2011. Shaw, Aaron. Insurgent expertise: The politics of free/livre and open source software in Brazil. *Journal of Information Technology & Politics.* 8 (3): 253-272.

2011. Shaw, Aaron, John Horton and Daniel Chen. Designing incentives for inexpert human raters. *Proceedings of the 2011 ACM Conference on Computer Supported Cooperative Work (CSCW '11).*

2010. Hester, Vaughn, Aaron Shaw, and Lukas Biewald. Scalable crisis relief: Crowdsourced sms translation and categorization with Mission 4636. *Proceedings of the 2010 ACM Conference on Development (ACM DEV 2010).*

Book Chapters and Invited Articles

2018 (*forthcoming*). Hill, Benjamin Mako and Aaron Shaw. Studying populations of online communities. In *Oxford Handbook of Networked Communication*, edited by Brooke Foucault Welles and Sandra González-Bailón. Oxford University Press.

2017. Foote, Jeremy, Aaron Shaw, and Benjamin Mako Hill. A computational analysis of social media scholarship. In *Sage Handbook of Social Media*, edited by Jean Burgess, Thomas Poell, and Alice Marwick, Russell Sage Foundation.

2015. Benkler, Yochai, Aaron Shaw, and Benjamin Mako Hill. Peer production: A form of collective intelligence. In *The Handbook of Collective Intelligence*, edited by Michael Bernstein and Thomas Malone. MIT Press.

2015. Shaw, Aaron. Hired hands and dubious guesses: Adventures in crowdsourced data collection. In *Digital Research Confidential*, edited by Eszter Hargittai and Christian Sandvig. MIT Press.

2014. Shaw, Aaron, Haoqi Zhang, Andrés Monroy-Hernández, Sean Munson, Liz Gerber, Benjamin Mako Hill, Peter Kinnaird, and Patrick Minder. Computer supported collective action. *Interactions*, Social Media Forum. 21(4):74-77.

2009. Shaw, Aaron. The access divide: Brazil and the United States struggle to balance innovation and equality in the information age. *Americas Quarterly,* Winter Issue.

2008. Shaw, Aaron. The problem with the Anti-Counterfeiting Trade Agreement (and what to do about it). *Knowledge Ecology Studies*, Vol. 2. Available at: http://kestudies.org/ojs/index.php/kes/issue/view/3/showToc.

2008. Shaw, Aaron. Cultura livre, movimentos e humanização do Capital, *Le Monde Diplomatique Brasil*, 12 December, Available at: http://diplo.uol.com.br/2008-12,a2690

2008. Shaw, Aaron. Why politics and institutions (still) matter for ICT4D, Publius Project, Berkman Center for Internet & Society, Available at: http://publius.cc/2008/09/26/shaw-why-politics-and-institutions-still-matter-for-ict4d/.

Selected Lectures and Presentations

Keynotes, Invited Lectures, and Non-Archival Conference Presentations

Openness, closure, and the organization of online commons

2018. CASBS Seminar Series, Stanford University, Stanford, CA, January.

2018. HCI Lunch Speaker Series, Computer Science Department, Stanford University, Stanford, CA, February.

2018. International Conference on Complex Systems, Crowdsourcing and Collective Intelligence Workshop, Cambridge, MA, July.

2018. Annual Meeting of the American Political Science Association (APSA), Boston, MA, September.

The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing (with Eszter Hargittai)

2018. Wikimedia Research Showcase, October.

2017. Annual Meeting of the American Sociological Association (ASA), Montreal, Canada, August.

2017. Annual Meeting of the International Communication Association (ICA), San Diego, CA, May.

2017. Colloquium Series, Department of Media and Communications, The London School of Economics & Political Science (LSE), May.

The Hidden Costs of Requiring Account Creation: Evidence from Peer Production
(with Benjamin Mako Hill)

2017. Management of Organizations (MORS) Colloquium Series, Haas School of Management, UC Berkeley, Berkeley, CA, October.

2017. Law and Technology Speaker Series, The Information Society Project at Yale Law School, New Haven, CT, January.

2016. Workshop and Lecture Series on the Law & Economics of Innovation, ETH Zurich, Center for Law & Economics, Department of Humanities, Social and Political Science, November.

2016. Voice, Technology & Impact Workshop, Department of Communication, University of Washington, Seattle, May.

2015. ANN-SONIC-NICO 7th Annual International Workshop on Network Theory, Northwestern University, October.

Social Structures of Productive Online Volunteer Communities
(with Jeremy Foote and Benjamin Mako Hill)

2016. Organizational Communication Mini-Conference (OCMC), Evanston, IL, October.

2016. Annual Meeting of the American Sociological Association (ASA), Seattle, WA, August.

2016. Annual Meeting of the International Communication Association (ICA), Fukuoka, Japan, June.

The Wikipedia Adventure: Field Evaluation of an Interactive Tutorial for New Users
(with Sneha Narayan, Jake Orlowitz, Jonathan Morgan, and Benjamin Mako Hill)

2016. Workshop and Lecture Series on the Law & Economics of Innovation, ETH Zurich, Center for Law & Economics, Department of Humanities, Social and Political Science, November.

2016. Annual Meeting of the International Communication Association (ICA), Fukuoka, Japan, June.

Beyond Access: How User Research in Online Communities Can Empower Libraries

2015. Keynote Presentation, "Library in the Life of the User," OCLC Research partners' meeting, Chicago, IL, October.

Learning from Populations of Online Organizations

(with Benjamin Mako Hill)

2015. International Communication Association (ICA) Pre-conference: Communication in a Networked Age. San Juan, Puerto Rico, May.

2015. International Communication Association (ICA) Pre-conference: Computational Approaches to Advance Communication Research. San Juan, Puerto Rico, May.

2015. Computational Social Science Summit, Kellogg School of Management, Northwestern University, May.

Freedoms of Access and Engagement?

2015. Center for Diaspora and Transnational Studies, University of Toronto, January.

2014. Hermanns Lecture, English and Philosophy Departments, University of Texas at Arlington, October.

Laboratories of Oligarchy? How the Iron Law Extends to Peer Production

(with Benjamin Mako Hill)

2014. Collective Intelligence Conference, MIT, Cambridge, MA, June.
2014. Annual Midwest Political Science Association Conference, Chicago, April.
2013. Computational Social Science Workshop, University of Chicago, November.
2013. Chinese University of Hong Kong, C-Centre, August.
2013. MIT Economic Sociology Working Group, May.
2013. Cooperation Group, Berkman Center, Harvard, April.
2013. Center for Information Technology Policy, Princeton, April.
2013. European Consortium of Political Researchers (ECPR), Joint Workshop Sessions. Mainz, Germany, March.

Status, Social Signaling and Collective Action in a Peer Production Community

(with Benjamin Mako Hill and Yochai Benkler)

2012. American Sociological Association Annual Meeting, Denver, August.
2011. Laboratory for Social Research Seminar, UC Berkeley, October.
2011. Wikimedia Foundation, San Francisco, CA, August.
2011. Open and User Innovation Workshop. Vienna, Austria, June.
2011. MIT Economic Sociology Working Group, May.

Internet Skills and Wikipedia's Gender Inequality

(with Eszter Hargittai)

2014. Annual Conference of the International Communication Association, Seattle, May.
2014. Wikimania, Barbican Center, London, August.

2014. Berkman Center for Internet & Society Luncheon Series, Harvard University, January.

Intentional Crowds: Using goal-aligned microtasks to mobilize groups

2013. Human Crowdsourcing Lunch Series, Carnegie Mellon University, Pittsburgh, March.

Hired Hands and Dubious Guesses: Adventures in Crowdsourced Data Collection

2012. Annual Meeting of the American Sociological Association. Denver, August.

The Internet, Young Adults, and Political Engagement in the 2008 Presidential Election (with Eszter Hargittai)

2012. Annual Conference of the International Communication Association. Phoenix, May.

2011. Annual Midwest Political Science Association Conference, Chicago, March.

Centralized and Decentralized Gatekeeping in an Open Online Collective

2012. Northwestern University, Communication Studies Department, Evanston, January.

2011. Annual Meetings of the American Sociological Association (ASA). Las Vegas, August.

2011. Universidade Estadual de Campinas (UNICAMP), Laboratorio de Jornalismo (LabJor). Campinas, Brazil. June.

A Tale of Two Blogospheres: Discursive Practices on the Left and Right (with Yochai Benkler)

2011. International Communication Association (ICA) Annual Conference. Boston, May.

Panels and Public Talks

2014. Welcoming new voices vs. defending the world's knowledge: The future of Wikipedia's open community. National Communication Association (NCA), Human Communication and Technology Division Panel. Chicago, November.

2013. Wiki ecology: Understanding wiki-success through comparative analysis. Wikimania 2013, Hong Kong Polytechnic University, Hong Kong, China, August.

2012. Fail whaling: Designing from deviance and failures in social computing. ACM Conference on Human Factors in Computing Systems, (with Michael Bernstein, Michael Conover, Benjamin Mako Hill, Andrés Monroy-Hernández, Brian Keegan, Sarita Yardi, R. Stuart Geiger, and Amy Bruckman), Austin, TX, March.

2011. Using social awards build better free culture and free software projects: Evidence from three field studies." Presented to the Wikimedia Foundation. San Francisco, CA. August.

2011. Using social awards to build better free software & culture projects (with Benjamin Mako Hill). 12th Annual Forúm Internacional Software Livre (FISL), Pontifíca Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil. July.

2010. Introductory remarks. CrowdConf 2010: The Future of Distributed Work, Research Track. San Francisco, CA. October.

2010. Invited panelist, Current Scholarship on the Political Blogosphere, Netroots Nation. Las Vegas, NV. July.

Guest Lectures and Workshops

2017. Research Workshop: Online communities & crowds. Department of Media and Communications, The London School of Economics and Political Science (LSE), May 3.

2016. Learning for Life Lectures. Northwestern University, November 9.

2016. Teaching and Learning with Small Group Projects. Searle Center Faculty Fellows Presentation, Northwestern University, March 10.

2015. Teaching with Wikipedia. Searle Center for for Advancing Teaching & Learning Faculty Round table, Northwestern University, Spring.

2014-2016. Guest Lecture: Peer Production. Communication Studies 227, "Communication and Technology," Northwestern University, Fall.

2014. Guest Lecture: Learning from Peer Production: Tough Problems for Open Online Communities. Linguistics 363 / Slavic 322, "Making a Dictionary," Northwestern University, Spring.

2014. Guest Lecture: Crowds and Human Computing. COMM 587B, "Innovation Communities," University of Washington, Spring.

2014. Accounting for Transaction Costs in Collective Action Online, Management and Organizations (MORS) Group Workshop on Social Movements and Social Enterprise, Kellogg School of Management, Nortwestern University, Spring.

2011. Designing for Cooperation with Social Incentives (with Benjamin Mako Hill and Yochai Benkler). Harvard Law School, Introduction to Internet Law (iLaw) Course. Cambridge, MA. September.

Teaching

Graduate Advisor (Dissertation Chair)
Jeremy Foote, MTS, Northwestern (2014-Present)
Thomas Rousse, MTS, Northwestern (2014-Present)
Sneha Narayan, TSB, Northwestern (2013-2018 *expected graduation*)

Graduate Dissertation Committee
Nathan TeBlunthuis, Communication, University of Washington
Yongsung Kim, TSB, Northwestern
Ilya Gokhman, MTS, Northwestern
Michael Greenberg, TSB, Northwestern
Erin Klawitter, MTS, Northwestern

Graduate General Exam Supervision
Ilya Gokhman, MTS, Northwestern (2017)
Yongsung Kim, TSB, Northwestern (2016, 2017)
Ashley Niler, MTS, Northwestern (2017)
Ashley Walker, MTS, Northwestern (2017)
Miriam Boon, TSB, Northwestern (2016)
Marlon Twyman, TSB, Northwestern (2016)
Sabrina Connell, MTS, Northwestern (2015)
Michael Greenberg, TSB, Northwestern (2014)
Erin Klawitter, MTS, Northwestern (2013)

Undergraduate Research Supervisor
Bria Marie Royal, Northwestern Communication Studies ‘16 (2015-2016)
Amiko Krisa Bunag Lagrimas, Northwestern Communication Studies '15 (2015)
Eleanor Burgess, Northwestern Communication Studies '15, Fulbright (2013-2015)

Courses Taught
Online Communities and Crowds (undergraduate, 2013, 2014, 2016).
http://aaronshaw.org/teaching/2016/occ
Northwestern University, Communication Studies 378.
Online communities and crowds represent the most innovative and ground-breaking organizations today. Distributed groups collaborate over the Internet to write free encyclopedias (Wikipedia), launch social movements (Avaaz, MoveOn), create software (GNU/Linux), share music and films (the Pirate Bay), develop innovative products (CHDK), and conduct advanced scientific research (Zooniverse). When and why do these efforts succeed? What motivates participants to join, contribute, and sustain these communities? How can online communities' and crowds' successes be harnessed and reproduced? What can be learned from their shortcomings? This course presents an

intensive and interdisciplinary introduction to the study of online communities and crowds through a combination of readings, discussion, and group-based projects. We will analyze how and why some of these systems are so wildly effective at mobilizing and organizing people in ways that seem to have been impossible a few decades ago.

Introduction to Graduate Research in TSB and MTS (graduate, 2016).
http://aaronshaw.org/teaching/2016/mts501
Northwestern University, Media, Technology & Society 501.
The goal of this seminar is to introduce first-year students in the MTS and TSB Ph.D. programs to the fields they cover and how to be a productive and responsible graduate student. Participants in the seminar will achieve the following goals: (1) understand, interpret, discuss, and assess academic career strategies and development; (2) develop, articulate, and apply effective academic career development strategies in the context of your own research program; (3) reflect on (assess) and iteratively improve upon your own work and that of your peers in light of the concerns analyzed in class. We will do this through a combination of readings, writings, activities, in-class discussions, and guest visits from MTS and TSB faculty.

The Practice of Scholarship (graduate, 2016)
http://wiki.communitydata.cc/Practice_of_scholarship_(Spring_2016)
Northwestern University, Media, Technology & Society 503.
The goal for this course is simple: submit a piece of academic research for publication by the end of the quarter. The piece should (obviously) be original. You should be the primary person responsible for the research and should be the lead author of the submission. The course and assignments are structured to help you cultivate (more of) the skills, wisdom, and experience necessary to publish independent, original, and high-quality scholarship in relevant venues for your work. There are several milestones to help you measure your progress towards manuscript submission at the end of the quarter. The seminar will run as a workshop in which you will produce written work and provide feedback on each other's work every week. Most weeks, we will also read and discuss materials related to the crafts of designing, conducting, writing, submitting, reviewing, revising, and publishing scholarly research. The experience will probably feel like a combination of a writing boot-camp and an extended group therapy session.

Online Collective Action (graduate, 2013, 2014, 2015).
http://aaronshaw.org/teaching/2015/peer
Northwestern University, Media, Technology & Society 525.
Large scale online collective action such as peer production encompasses the most significant organizational innovations that have emerged from Internet-mediated social practice. The best known examples of peer production include free/libre and open source software (FLOSS) such as the GNU/Linux operating system, and free culture projects such as Wikipedia. This seminar introduces key issues in the study of peer

production and prepares students to contribute to this interdisciplinary field of research. The first part of this seminar surveys some of the most influential threads of empirical and theoretical research on collective action, online collaboration, and peer production. The second part focuses on readings and discussions driven by student interests and projects.

Research Design for Causal Inference (graduate, 2013, 2014, 2015).
http://aaronshaw.org/teaching/2015/causal
Northwestern University, Media, Technology & Society 525.
Inquiries into cause and effect lie at the heart of almost all social and behavioral research. For example: What are the effects of media exposure on children's educational attainment? Which kinds of outreach and incentives promote participation in voluntary organizations? Can social pressure increase political engagement? This course is a Ph.D. seminar focused on analyzing, designing, and conducting empirical causal inference using quantitative methods. Through readings, problem sets, independent empirical projects, and research design exercises, students in the class will develop the skills needed to understand, design, implement, analyze, and disseminate research that addresses causal questions.

Service and Engagement

Professional Service

2019-2021. Applied/Public Policy Research Award Committee Member, International Communication Association.
2019. Fellows Selection Committee, Center for Advanced Study in the Behavioral Sciences (CASBS) at Stanford University.
2019. Program Committee Member (Associate Chair), ACM Conference on Human Factors in Computing Systems (CHI).
2018-Present. Founding Editorial Board Member, *Computational Communication Research*.
2017. Panelist, United States National Science Foundation, Cyber-Human Systems Program, Information and Intelligent Systems Division.
2017. Program Committee Member, Collective Intelligence Conference.
2017. Program Commitee Member, AAAI Conference on Human Computation and Crowdsourcing (HCOMP).
2017. Senior Program Committee Member, ACM Symposium on Open Collaboration.
2016-2017. Program Committee Member, International Conference on Computational Social Science (IC2S2).
2016. Faculty discussant, Research Escalator Session, Organizational Communication Division, Annual Conference of the International Communication Association.
2016. Program Committee Co-Chair, ACM Symposium on Open Collaboration, Wikipedia Research Track.

2015. Organizer, ANN-SONIC-NICO 7th Annual International Workshop on Network Theory, Fall.
2015-Present. Founding Editorial Board Member, *Social Media & Society*.
2015. Program Committee Member (Associate Chair), ACM Conference on Computer Supported Cooperative Work (CSCW).
2012-2015. Program Committee Member, SoHuman Workshop on Crowdsourcing and Human Computation (SocInfo).
2010-2012. Program Committee and Conference Co-Chair. *CrowdConf: The Future of Crowdsourcing and Distributed Work*. San Francisco, CA.
2006-2007. Business Manager and Editorial Collective Member, *Berkeley Journal of Sociology*.

Ad hoc reviewer for MIT Press, Sage Press, Wiley-Blackwell, the National Science Foundation of the United States (Directorate for Computer and Information Sciences and Engineering, Directorate for Social, Behavioral, and Economic Sciences), as well as publications and conferences including:

Journal of Communication
Journal of Computer Mediated Communication
Social Media & Society
Communication Theory
International Journal of Communication
ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)
ACM Conference on Human Factors in Computing Systems (CHI)
ACM Collective Intelligence Conference
International Conference on Computational Social Science (IC2S2)
Hawaii International Conference on System Sciences (HICCS)
AAAI Workshop on Human Computation (HCOMP)
ACM Transactions on Computer-Human Interaction (ToCHI)
ACM Transactions on Social Computing.

University Service

Northwestern University

2015-2017; 2018-Present. Director of Graduate Studies, Media, Technology & Society.
2015-2017. Media, Technology & Society Graduate Admissions Committee (Chair).
2015-2016. Member, Computation and Communication Faculty Search Committee.
2015-2016. Advisory Committee, Master of Science in Communication; Communication Studies Department.
2014-2015; 2016-2017; 2018-2019. Coordinator, Media, Technology & Society Speaker Series.
2014-2015. Communication Studies Undergraduate Curriculum Committee.
2013-Present. Communication Studies, Digital Media Module Steering Committee.
2013. Co-organizer, Wiki Research Hackathon.

Professional Affiliations
International Communication Association
Association of Computing Machinery
American Sociological Association
American Political Science Association

Industry and Non-Profit Affiliations
2014-Present. Citizen Science Advisory Board Member, Adler Planetarium & Oxford University.
2014-2018. President, Board of Directors, Tamarindo Foundation, Inc.
2009-2011. Consulting Research Scientist. CrowdFlower Inc., San Francisco, CA.

Part 3: Research Materials

This section contains all of the peer reviewed papers and book chapters published since my arrival at Northwestern University. I have also included the Project Description from my successful National Science Foundation grant proposal (funded in 2016). All materials are included in chronological order by year and alphabetized by first author's last name within year. Please note that any of the materials discussed in Part I or listed on my CV that do not appear below are available from my website (http://aaronshaw.org) or upon request.

Politics & Society
http://pas.sagepub.com/

Centralized and Decentralized Gatekeeping in an Open Online Collective

Aaron Shaw

Politics & Society 2012 40: 349
DOI: 10.1177/0032329212449009

The online version of this article can be found at:
http://pas.sagepub.com/content/40/3/349

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

Additional services and information for *Politics & Society* can be found at:

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

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

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

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

>> Version of Record - Aug 2, 2012

What is This?

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Politics & Society
40(3) 349–388
© 2012 SAGE Publications
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0032329212449009
http://pas.sagepub.com

Centralized and Decentralized Gatekeeping in an Open Online Collective

Aaron Shaw[1]

Abstract

This paper presents a study of gatekeeping in the U.S. political blog "Daily Kos." Open online collectives like Daily Kos use relational mechanisms, such as gatekeeping, to manage organizational boundaries and filter the contributions of participants. However, neither prior theories of gatekeeping nor the existing analyses of open online collectives account for the character or implications of gatekeeping in the Daily Kos community. Using qualitative evidence as well as statistical analysis of a large sample of comment threads on the site from 2008, I argue that gatekeeping on Daily Kos takes centralized and decentralized forms, and that both modes depend critically on relational boundary work among site participants. Centralized gatekeeping proceeds through actions by high-status members of the community. Decentralized gatekeeping, by contrast, consists of more numerous and small-scale interactions between community members, who filter and moderate each other's participation. Both forms of gatekeeping enhance the ability of site leaders and incumbent community members to regulate access to privileges and agenda-setting responsibilities on the site. These findings imply that the egalitarian ethos of open online collectives exists in tension with the mechanisms through which participation and status inequalities emerge among participants. How collectives engaged in mobilization and discursive production resolve this tension will shape the long-term impact of online participation and blogs on the political and public spheres.

Keywords

politics, Internet, gatekeeping, organizations, social movements

[1]Northwestern University, Evanston, IL, USA

Corresponding Author:
Aaron Shaw, Department of Communication Studies, Northwestern University, Evanston, IL, 60208, USA
Email: aaronshaw@northwestern.edu

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Introduction

Online modes of political participation have spurred overlapping debates about the Internet's potential to transform political engagement and the public sphere. Some observers herald the democratizing potential and egalitarian character of large-scale online communities and collaborative platforms, such as political blogs or social network sites (SNS).[1] Others have questioned the idea that networked communication over the Internet has altered or overcome underlying social inequalities that ultimately determine who gets to participate in the public and political spheres.[2] In these debates, the institutional and organizational dynamics internal to online communities engaged in political action have not received sustained analytical attention. The extensive body of existing research on democratic engagement, media, and the public sphere has demonstrated that such organizational and institutional dynamics play a central role in determining both the success or failure of movements as well as the sociopolitical implications of particular forms of organization.[3] The relationship between democratic movements, digital media, and online modes of organization requires further consideration.

In this paper, I conduct a mixed-methods analysis of the dynamics of interaction and inequality within a large-scale open political community online. Specifically, I focus on practices of gatekeeping in the U.S. political blog, Daily Kos.[4] Arguably the largest and most prominent participatory political blog in the United States, Daily Kos embodies several core aspects of the practices of online engagement and participation occurring in the blogosphere.[5] The massive scale of participation on Daily Kos creates a set of coordination and information filtering problems, which the site, as a community organization, "solves" through a distributed system of content moderation and filtering. In contrast with earlier modes of media production and with more hierarchical, exclusive forms of blogging, this system does not rely exclusively on the actions of individual elites in key choke-points of bureaucratic authority to perform centralized gatekeeping roles. Rather, Daily Kos also relies on the collective wisdom of its community of participants to filter the content contributed by their peers. In this sense, Daily Kos relies on formal, organizational hierarchies as well as less formal inequalities that emerge through discursive persuasion.

The key organizational dynamic of Daily Kos's system of distributed filtering and moderation emerges as a pattern of decentralized gatekeeping, whereby practices of boundary work and social closure take on a more collective aspect than in previous forms of media production and political organization. In contrast with "traditional," centralized gatekeeping, which proceeds through actions by elite members of a collective or organization, decentralized gatekeeping consists of more diffuse, small-scale interactions between community members. Once aggregated, these small-scale interactions feed back into the status hierarchy of participants, enhancing the ability of high-status and more experienced members of the community to regulate access to privileges and agenda-setting responsibilities on the site. At the same time, decentralized

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

gatekeeping also functions as a means by which regular members of the community can assume limited authority and legitimately contest norms. In both kinds of gatekeeping, site participants tend to negotiate status and influence on a relational, discursive basis, rather than on the basis of bureaucratic hierarchy or formal organizational structure alone. As a result, not only content, but also people become the targets of gatekeeping. In this way, decentralized gatekeeping practices contribute to the reproduction and contestation of status inequalities within Daily Kos that reinforce the site's mechanisms of information filtering and mobilization.

Such patterns of organization and gatekeeping behavior matter for two reasons. First, the surge in online modes of political engagement and tools suggests that Daily Kos and other formally open, online-centered movements represent new "laboratories of democracy" in the Tocquevillian sense. As an increasing number of social movement organizations, political parties, and private firms adopt online tools for collaboration and collective action, the dynamics of online collectives become more relevant for the study and practice of public engagement. If the adoption of these tools brings with them novel organizational governance practices or disciplinary mechanisms, these phenomena portend a broader transformation of the organizational basis of democratic politics. They also shed light on other, offline environments—such as political parties or social movement organization meetings—in which relational, discursive mechanisms play a role in determining hierarchies of status and influence as a collective seeks to mobilize consensus around a common objective.

The second reason why online gatekeeping matters has to do with the character of the sort of partisan movement organization—engaged in media production, mobilization, and information dissemination—that Daily Kos exemplifies. The flow of attention, influence, and status in these new media organizations will shape the networked public sphere as well as the future of democratic politics. As the blogosphere has grown and become a stable part of the political ecosystem in the United States over the past five years, the processes by which ideas and individuals achieve visibility within blogs and related online movements remain opaque. This study examines gatekeeping processes within one of the largest, most dynamic political blogs with the objective of contributing to a wider debate about the politics of news and information production in the contemporary era.

The U.S. political blogosphere emerged during the 2004 presidential campaign. Since that time, the landscape of political blogs has undergone several dramatic shifts as well as a series of somewhat more subtle evolutionary changes. The most significant, ongoing transformation has been the formalization and professionalization of the most prominent individuals and organizations involved in political blogging. As part of this process, the political blogosphere has become integrated into the loose organizational networks of the press, political parties, nonprofits, consulting firms, and political action committees that make up the wider sphere of American politics.[6] As an organizational field, the blogosphere has stabilized to a large degree; however, important differences characterize elite blogs on the left and right, corresponding to distinct models of democratic political organization and discursive production.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

The first political blogs were highly amateur affairs, and usually consisted of the writings of a well-informed, outspoken political outsider with passionate views. Some blogs, especially MyDD and Daily Kos on the left, made an effort to incorporate multiple contributors and voices into the conversation, but these were the exception rather than the rule.[7] This began to change around the time of the 2004 presidential campaign as the blogosphere emerged as a viable medium of opinion-generation, news diffusion, muckraking, and mobilization.[8] On the left, bloggers played a crucial role driving the Howard Dean campaign in the Democratic primaries and upending South Carolina Senator Trent Lott in response to a racist remark at a fundraiser.[9] On the right, bloggers helped reveal that CBS News and Dan Rather had used fraudulent documents about George W. Bush's record in the Texas Air National Guard, resulting in a subsequent investigation and hastening Rather's retirement.[10] In each of these cases, "A-list" bloggers proved themselves at least equal to journalists in more traditional formats and organizations. In general, they benefited from the speed of publication and transparency norms that characterized the blogosphere from its earliest days. Without the burdens of hierarchical organizations or editorial oversight, the bloggers framed issues and pursued stories in a provocative way that many print, radio, and television journalists were simply unprepared or unwilling to do. As a result the bloggers accrued credibility as well as the attention of the public, the media, and political elites.

During these early years, political bloggers were derided as pajama-clad voices from the political wilderness.[11] However, the elite bloggers resembled their peers in the media and political institutions in terms of educational credentials, class, race, and gender.[12] Furthermore, in the years between 2004 and 2008, numerous "first-wave" political bloggers received book deals or were hired as columnists by national publications looking to build online traffic and advertising revenue.[13] Over time, more A-list bloggers could be found on Sunday morning political talk shows or authoring op-eds in major news outlets. As the bloggers professionalized, substantive differences between them and news producers in partisan broadcast print or television media became less salient.

Blogging organizations also became more formal during this time. Most of the early blogs used off-the-shelf blogging software and were authored by individuals. For example, of the fourteen unique sites counted as top blogs in 2004 by Drezner and Farrell, only four were either (part of) an incorporated organization or had a formal organizational hierarchy.[14] Of those same fourteen blogs and bloggers, all of them are now either independent corporate entities or part of incorporated organizations. This pattern holds across a larger sample of top blogs as well. Data collected by Shaw and Benkler in summer 2008 show that at that time, 93 out of 155 (60 percent) of top political blogs were either (part of) an incorporated entity or involved a formal organizational hierarchy.[15]

Bloggers and blog communities have also integrated themselves into the broader field of established political organizations, networks, and actors. While this process has followed an uneven pace, a number of bloggers on the left and right have become prominent figures within mobilizations and campaigns, signaling their growing role as

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

power brokers and agenda setters. On the right, bloggers at first integrated into existing organizations or movements by working with major partisan media outlets (e.g., the *National Review* and Fox News), or by forming new media strategy and campaign consultancies.[16] As a result, the right political blogosphere did not claim leadership of a clearly defined grassroots movement or constituency, but some of its members became an active force in shaping the voice and agenda of conservative politics. In contrast, the left blogosphere elites have more aggressively sought to transform the landscape of political mobilization through the creation of new organizations and constituencies under the banner of a "netroots" movement. These efforts have proceeded by means of conferences, fundraising campaigns, and new advocacy organizations agitating for change within the Democratic Party.[17]

The processes of professionalization, formalization, and political integration that occurred in the blogosphere between 2004 and 2008 did not follow an even trajectory across all blogs, but they nevertheless resulted in an overall stabilization and institutionalization of the field.[18] Many of the elite blogs established consistent styles and regular communities of reader-participants despite the cyclical ebb and flow of attention around national elections. This pattern of stabilization has also manifested in the distribution of attention across the blogosphere as a whole. Using secondary data gathered by Karpf, a comparison of the mean monthly rank of top fifty left and right blogs along several metrics of authority and attention from June 2009 through January 2011, shows the distribution of ranks across left and right blogs to be stable, despite some turnover within the two groups.[19] Correlations between blog rankings for all sites on all measures in the first and last months of Karpf's data collection are likewise positive and significant.[20] The ecosystem of blogs, while still new and innovative relative to the wider field of political organizations, may not be as volatile as casual observers would believe.

Nonetheless, important differences have emerged between the left and right of the blogosphere. Most significantly, cross-ideological variation in the adoption of participatory blogging platforms has resulted in "two blogospheres" characterized by distinct democratic affordances.[21] Similarly, as implied above, the elite bloggers on the two major sides of the political spectrum have taken divergent approaches when it comes to integrating their discursive production into broader projects of mobilization and movement building. The significance of these cross-ideological differences hinges on whether they facilitate distinct pathways of influence and engagement, as well as the extent to which they do or do not become institutionalized over time.

Dynamics of Control In Open (Online) Collectives

Those who claim that blogs and networked movements hold the potential to democratize political discourse and participation ground their arguments in the view that online collectives do not simply reproduce existing, offline inequalities, but rather enable new publics to coalesce and mobilize in a more egalitarian fashion than was historically feasible.[22] Counterarguments have underscored the persistence of socioeconomic and other sociostructural inequalities as predictors of participation.[23]

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Scholars have also criticized the design of technical systems that algorithmically reproduce preexisting inequalities of attention and prestige.[24] However, the fact that online tools might make enhanced collaboration and participation possible does not determine their effects. The question of what sorts of mechanisms of influence and control prevail within online collectives remains unresolved in these debates. Likewise, even if the mechanisms driving selection into participatory, collaborative online movements tend to reproduce offline inequalities of access or attention, that does not foreclose the possibility that those online movements could still embrace a more open and democratic character than their predecessors. Answering these concerns more precisely requires closer investigation of the internal dynamics of control within online collectives.

In organizational terms, Daily Kos and participatory blogs have a great deal in common with other "open" network and community organizations.[25] Some research has analyzed dimensions of control within these organizational types in the context of online collectives such as the groups that produce free and open source software.[26] However, most explanations of how these online communities work draw primarily on institutional and transaction cost economics as well as social psychological theories of motivation. Specifically, Benkler, Lerner, and Tirole, as well as Weber have all argued that distinct features of digitally networked information production enable large-scale, "nonmarket" systems to overcome the obstacles to collective action, public goods creation, and sharing identified by classical economic theory.[27] These claims originated as rejoinders to longstanding debates on "the tragedy of the commons" and collective action failures in the context of public goods creation.[28] As such, their authors view the fundamental puzzle of commons-based production online as a twofold question: Why do individuals make contributions to online collective goods in the absence of financial incentives and how do large numbers of individuals coordinate and sustain their contributions in the absence of either formal organizational structures or markets? Their answers to these questions emphasize relatively static sets of norms, incentives, and motivational profiles. As a result, they overlook the importance of both the structural dynamics within online communities as well as the interactions between community members.

Dynamics of social reproduction, governance, and institutionalization within online collectives all entail a complex set of relational processes that emerge through the interactions of community members, who actively manage organizational boundaries. For example, the work of O'Mahony and Ferraro shows that, over time, the growth of the Debian Linux community has led the community members to negotiate and implement a steadily more and more complex boundary-management process, leading to increasingly formalized and hierarchical governance structures.[29] Somewhat paradoxically, these formal and exclusionary structures are combined with direct democratic institutions that work to preserve the project's formal openness. The preservation of certain kinds of direct democracy thus appears to support the cultivation of formal organizational structure. The pattern adheres loosely to Michels' "Iron Law of Oligarchy" and reproduces a similar sort of emergent hierarchy to that identified in

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Freeman's critique of the tyrannical "structurelessness" of the 1970s U.S. feminist movement.[30]

Such questions of organizational governance and democracy reconnect this line of research with analyses of online social movement organizations and democratic political mobilization. Several previous studies have argued over whether the emergence of networked, politically engaged collectives collaborating over the Internet have transformed the structure and dynamics of the public and political spheres.[31] However, among this body of research, only Karpf offers a typology of networked political organizations or a theoretical framework for thinking about the evolution of such communities in relation to existing political movement organizations.[32] For Karpf, political organizations with strong online organizing components hold much in common with their pre-Internet counterparts, but the lower communication costs enabled by digital technologies have facilitated less hierarchical intra-organizational structures. The transformation of infrastructure has in turn brought about a generational shift as new advocacy organizations on the U.S. left have embraced these possibilities.

The literature on movement organizations' use of networked technologies and strategies has a blind spot when it comes to providing more general accounts of how these supposedly "new" practices of less hierarchical mobilization proceed. For example, Karpf theorizes about the "phases" of growth that the new generation of Internet-savvy organizations pass through, but he does not analyze this process in much depth, nor does he explain the relationship of the organizations' internal dynamics to their apparent "product": a complex social system that generates and disseminates a vast amount of information. Online collectives engaged in political mobilization constitute novel sorts of institutions, movement organizations, and fields of power with characteristics that resemble their purely "offline" predecessors.[33] Closer consideration of the microsocial dynamics of networked movement organizations and communities can therefore speak to the mechanisms of influence and agenda setting in these environments as well as the means by which open online collectives function as efficient information processing systems.

Gatekeeping and Boundary Work in the Networked Public Sphere

Gatekeeping represents one particularly salient mechanism that has not received adequate attention in the context of online collectives engaged in (political) discursive production. Along with Google and sites that sort and filter information by means of algorithms, Daily Kos embodies a key aspect of the broader shift toward distributed, social information processing online. This shift, in addition to altering the terrain of social movements and political mobilization, has also transformed the structure and dynamics of the public sphere.[34] In a context where attention becomes a scarce resource subject to intense competitive pressures, social mechanisms of agenda setting and influence acquire enhanced importance. Explaining how social processes such as gatekeeping may or may not undergo transformations in online movements

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

and news organizations can clarify what the rise of networked communication means in relation to the organizational and informational dynamics that characterized the mass-mediated public sphere.

Early studies of gatekeeping focused on individuals who held extraordinary control over flows of goods, ideas, and attention within families, groups, or society as a whole.[35] Over time, gatekeeping research turned to the role of elites within newspapers, television, and other information-producing professions. Practices of gatekeeping by editorial staff inside news-making organizations have historically drawn special attention as the quintessential examples of how institutional, cultural, and organizational dynamics influence what in fact becomes "news" in the first place.[36]

More recent gatekeeping research has likewise focused on journalism and media production, but has shifted to consider the role of institutions, processes, and the structural dimensions of social relations in driving the movement of information and access to resources.[37] A few studies have specifically examined gatekeeping in the context of online collectives and communities.[38] In addition, several studies have explored analogous processes to gatekeeping that structure the dissemination of information in networked environments. Boczkowski's ethnography of networked newsrooms in Argentina illustrates how the adoption of online publication, content aggregators, and the intensified competition for reader attention in a flooded information marketplace have transformed the organizational practice of newspaper production, resulting in less diverse content.[39] Hindman's analysis of "Googlearchy" argues that the combined effects of power laws of attention together with many Internet users' increasing reliance on algorithmic information filtering produces an extremely small elite capable of dominating the networked information ecosystem.[40] In contrast, Benkler and collaborators have argued that the presence of such power laws do not eliminate the possibility that the Internet could be used to democratize political communication, but that the actually existing conditions of online discursive production require further study.[41]

From a theoretical perspective, none of the existing research has explained the ways gatekeeping or related processes form part of the diffuse dynamics of contention and negotiation that go into managing open online collectives. In these environments, the struggle for attention and influence among the numerous participants means that gatekeeping is as much about inequalities in the attention and influence that accrue to particular people as well as to particular types of content. In other words, gatekeeping in open collectives becomes a means of constructing normative boundaries around legitimate discourse and action, and restricting the voice of those who do not adhere to the norms. In this aspect, gatekeeping constitutes a specific form of relational boundary work in the service of elite "status closure."[42] Such relational work encompasses the diverse repertoire of practices through which individuals and groups define, imitate, contest, and reconstruct social categories. These practices also serve as the everyday mechanisms through which categorical inequalities, social movements, and economic exchanges cohere into larger structures.[43] As part of a broader repertoire of organizational governance practices, relational boundary work among the participants

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

in online collectives simultaneously drives the emergence of formal hierarchy at the same time that it enables the preservation of open, democratic institutions.

At its core, decentralized gatekeeping consists of numerous, microlevel interactions between individuals engaged in a particular collective endeavor. Through the aggregation of distributed, relational exchanges that draw on existing rhetorics, norms, and codes of behavior, these individuals participate in the stabilization and reproduction of larger scale social dynamics. Over time, this process results in wider patterns of path dependency and creates institutionalized impediments to sudden shifts in the social order. Thus, with or without the central points of control through which traditional practices of gatekeeping proceed, organizations or communities constituted through distributed social interaction have a tendency to generate the sorts of deeply entrenched hierarchies and structures identified in earlier work on democratic, open, or "structureless" organizations.[44]

The dynamics of coordination and organization in large-scale online communities that generate, filter, and disseminate information on a massive scale offer a compelling arena for research into networked gatekeeping. Despite the growing number of sites on the Internet that fit this description—examples include Daily Kos, Reddit, Digg, and Slashdot—only a few studies have attempted to characterize the social dynamics of information filtering within these sorts of networked communities.[45] My work on Daily Kos contributes to this body of research by extending theories of centralized, elite-level gatekeeping to incorporate an analysis of the decentralized, relational practices of social information filtering and production pursued in large networked communities. A relational perspective focused on diffuse microlevel interactions expands existing theories of gatekeeping beyond the traditional focus on central choke points of control.[46] The central points within organizational hierarchies or networks traditionally identified as the locus of gatekeeping activities remain significant, but (in the context of open or network organizations) gatekeeping practices also occur throughout a collective. In this sense, some kinds of gatekeeping may function as a distributed form of social control rather than a form of top-down coercion.

Research Design and Methods

In order to analyze the gatekeeping practices in Daily Kos, this study combines qualitative observation with a statistical analysis of a large sample of comment threads on the site from 2008. The next section of the paper introduces the site and presents an overview of some of the formal rules and technologies that govern interactions between participants. This section establishes a baseline understanding of the system of community governance that operates throughout the site as well as a sense of the site history and its prevailing culture. In particular, I focus on the mechanisms by which the community engages in distributed content moderation and filtering, an activity that accounts for a large proportion of user-to-user interactions on the site.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

From the overview, I use qualitative evidence to build the case that gatekeeping happens through centralized and decentralized mechanisms on Daily Kos. I conducted six months of qualitative observation of the Daily Kos site between March and November, 2008. During this time, I maintained a user account that I used to access posts and comment threads. I also contributed several posts of my own, comments, and diary recommendations. At regular intervals on almost every day during this period, I logged into the site, read the most recent front page stories, skimmed user comments and recommendations, and reviewed several of the recent and recommended user diaries. Field notes provided the basis of a preliminary analysis, which I later refined and corroborated through the use of the site's searchable archive. From this data, I analyze a series of events surrounding the 2008 Democratic Party presidential primary elections. During this period, the divide between supporters of Hillary Clinton and Barack Obama within the Democratic Party as a whole gave rise to numerous arguments and conflicts among Daily Kos participants. Such an intense period of contentious debate thus provided an ideal opportunity to observe how the community enforced norms and constructed boundaries. In several ways, these events illustrate that gatekeeping practices emerge through everyday user-to-user interactions, giving rise to inequalities and hierarchies among the participants.

Following the inductive elaboration of the idea of decentralized and centralized gatekeeping, I construct a statistical test for relationships between measures of commenting and recommendation behavior that would signal the presence of decentralized gatekeeping. The data consist of a large sample of comment threads and recommendations taken from a subset of posts on the site, a daily political humor series called "Cheers & Jeers" (or C & J, henceforth). C & J represents one of the most stable and well-established subcommunities within Daily Kos. Therefore, although it is not necessarily representative of the site as a whole, it occupies a privileged position within the community culture and provides a window into the sorts of norms and values that prevail among an exceptionally active group of Daily Kos participants.[47] I use the full set of C & J user comments and recommendations from 2008. I collected this data in April 2010 using a script that parsed and stored records from the site archives via the built-in search functionality.[48] The resulting data set allows me to describe the network created by the recommendations between users and to examine whether or not relational boundary work among Daily Kos community members produces systematic gatekeeping effects.

Overall, the research question that drives the quantitative analysis asks whether or not the system of distributed moderation and filtering used on Daily Kos functions by means of decentralized gatekeeping. For the purposes of this analysis, I define gatekeeping as the systematic reproduction of an unequal and regular flow of valued resources—especially influence—to an incumbent group or organization and take decentralized gatekeeping to entail the production of inequalities along these lines by means of distributed, collective behavior.[49] Through a series of hypothesis tests, I test whether large-scale patterns of status accumulation consistent with decentralized gatekeeping exist.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

My definition of gatekeeping hinges on the identity and role of "incumbent" groups or individuals. In the context of a large online community focused on the production of political information and mobilization, the possession of privilege, status, or power may lead to comparative advantages in terms of articulating a perspective, having that perspective heard, and eliciting some sort of response or action from other members of the community.[50] For the purposes of this analysis, I focus on two overlapping forms of incumbency: elite status and expertise. The test of gatekeeping effects is therefore a test of whether or not elite status or expertise associate with increased reputational returns to comments.[51] I break this down into three specific hypotheses, each of which operationalizes a distinct component of incumbency:

Hypothesis 1a: More experienced users will receive more recommendations per comment.

Hypothesis 1b: More active users will receive more recommendations per comment.

Hypothesis 1c: Formal elite users will receive more recommendations per comment.

The fact that comment recommendations between Daily Kos users form a directed graph in which practices of reciprocity may drive the flow of attention and status (and thereby generate gatekeeping effects) among community members also suggests a secondary hypothesis:

Hypothesis 2: Users who give more recommendations will receive more recommendations per comment.

If I find support for hypotheses 1a-c and 2, I could conclude that comment recommendation practices among Daily Kos participants contain patterns of behavior consistent with the presence of decentralized gatekeeping that reinforce the privileges of incumbent users of the site. Support for some subset of hypotheses would still imply the presence of behavior consistent with decentralized gatekeeping, but would also suggest that certain forms of incumbency associate with status advantages whereas others do not. In contrast, a result that failed to reject the null hypothesis of no effects for all four hypotheses would indicate that comment recommendation does not follow a pattern consistent with any of the forms of decentralized gatekeeping described here.

A Brief Overview of Daily Kos

Markos Moulitsas Zúñiga founded Daily Kos as a solo-authored weblog on May 26, 2002. Since then, Daily Kos has become one of the most heavily trafficked, collaborative political blogs on the Internet. Currently, the site attracts roughly 10 million page views from more than 600,000 unique visitors in the United States per month, placing at the top of the political blogosphere and on the same plane as other large

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

participatory websites like slashdot.com.[52] Sometimes characterized as radically left-wing, the site has become the preeminent symbol of the so-called netroots movement.[53] While many online-only news and discussion sites can match the quantity of traffic on Daily Kos, few have harnessed user participation and content to a comparable degree.[54]

Between 2003 and 2006, the Daily Kos website underwent several transformations. First, in October 2003, Moulitsas implemented a technical migration to a software platform designed to incorporate more dynamic forms of participation and interaction among users of the site.[55] The new platform incorporated reader contributions on a larger scale. In addition, during 2004, Moulitsas formalized a system of contributing editors who shared posting responsibilities with him. As a result of these technical and organizational changes, the density of content and participation on the site exploded during the 2004 presidential election cycle, contributing to the insurgent Howard Dean campaign.[56] As a result, Daily Kos became a symbol of a new kind of networked political movement on the American left.

Between 2006 and 2011, the Daily Kos site and community assumed a relatively stable form.[57] The volume of participation, influence, and attention of Daily Kos has made the site a benchmark by which other large-scale political discussion sites are measured. As of April 2010, the site had more than 200,000 registered users, out of which several thousand actively participated on the site every day.[58] The extensive technical and social system by which users filter and moderate each other's contributions to the site represents a massive proportion of activity on the site as well as a large-scale system of distributed community governance.

Distributed Content Filtering and Moderation

Within the Daily Kos community, the practice of distributed content filtering and moderation provides a basis for evaluating and prioritizing certain kinds of information over others. To facilitate this process, the software platform that makes up the site delimits four categories of content and four tiers of participants. The content categories are front-page stories; user diaries; comments; and ratings. The participant tiers are site elites; trusted users; registered users; and readers. In general, the platform incentivizes content contributions that earn the approval of other users.

The program provides all registered users with the ability to construct a public identity tied to a unique username. Once registered and logged in, users can comment, post diary entries, and customize a personal page. They do this with the knowledge that their work will be visible to any site visitor and (potentially) rated by other users. The multiple forms of user-generated content then serve as a foundation for other key social features on Daily Kos: content rating and reputation building. Together, these features make up the institutional landscape within which the site's participants interact. They also provide the means for community members to manage norms and rules. Every step in this process entails interactions between users.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Site elites on Daily Kos are a relatively small group of individuals with formal positions, status, and extraordinary privileges as identified on the site's masthead page.[59] These include Markos Moulitsas himself as well as the technical staff and editors of the site. The site elites are distinguished from other participants by their ability to post front-page stories and perform other managerial operations on the site as a whole. Daily Kos's formal organizational structure manifests itself through the presence and roles of the site elites. Moulitsas directly selected the earliest site elites. More recently, a clearer organizational hierarchy has been created whereby some elites are listed on the masthead as holding specific editorial responsibilities.

The procedures by which new site elites are chosen remain opaque, but Moulitsas and others have made statements on the topic that suggest that he and the other editors manage a system of semi-informal selection.[60] Given the site's status as a prominent movement premised on advancing Democratic Party interests, it is noteworthy that the positions are not determined through any sort of site-wide electoral process. At the same time, site elites make it clear in public statements on the site that they view themselves as ultimately accountable to and dependent on the readership for their continued legitimacy as community leaders.[61] Whether or not these statements are reliable accounts of elite community members' status matters less than the fact that the elites justify their presence on such populist terms. This sort of justificatory logic signals the elites' ideologically informed desire to understand themselves as part of a (small "d") democratic movement, even though their positions are not acquired or preserved through any sort of well-defined democratic process.[62]

Front-page stories appear as the primary content on the site's main URL and are therefore the most visible and accessible content on the site. The site's editors and elites usually write the front-page stories, but they are also sometimes user diaries that the elites have "promoted" to the front page. Through front-page stories, site elites thus anchor the community's textual production in a steady stream of reporting and analysis that meet high-quality standards. As a result, diary promotion to the front page confers prestige as well as formal reputation gains to diary authors.

Diaries are personal blogs that any registered user of the site may write and post. They vary widely in terms of length, quality, and content. In general, they mimic the tone and length of front-page stories while attempting to contribute original evidence or insights on an issue of interest to the site's participants. Formal and informal guidelines for writing the diaries exist, but many actual diaries deviate from these norms.

Immediately after posting, every diary authored by a registered user who is not a site elite appears in a Recent Diaries sidebar along the front page, increasing its visibility among other users of the site for a brief period. Other registered users can recommend user diaries, and the number of recommendations appears together with the diary title in the Recent Diaries sidebar. If a diary accumulates a large number of recommendations, it may automatically appear in the more prominently placed Recommended Diaries sidebar on the front page. Placement in the Recommended

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Diaries sidebar persists longer than in the Recent Diaries sidebar and signals that a diary has achieved an exceptional amount of popularity among the site's users.

Comments are the most ubiquitous and varied category of content on the site. Any registered user can comment on any story or diary. The comments on each story and diary appear as a threaded discussion below the story or diary in question. In general, front-page stories attract a few hundred comments each. Comments function as one of the primary vehicles through which Daily Kos participants negotiate their views and engage in sustained interactions with other users of the site.

Ratings, like comments, are ubiquitous on Daily Kos. With the exception of front-page posts, all content submitted to the site immediately becomes subject to user ratings, through which any registered user can contribute feedback. This user feedback is then used to facilitate the reading and filtering of information on the site. In the case of diaries, the accumulation of positive ratings (recommendations) from numerous and/or prominent users results in advantageous placement and more widespread dissemination on the site through the Recommended Diaries sidebar. With comments, the accumulation of positive ratings adds to the reputation score of the user who posted the comment. As a result, ratings often become a mechanism of social exchange, whereby users will engage in reciprocal support of one another's contributions. Registered users who accumulate many recommendations from their peers acquire "trusted user" status. Trusted user status provides access to additional recommending features on the site platform. Registered users that have not achieved trusted status are only capable of submitting positive ratings (also known as recommendations). Trusted users can provide negative or "troll" ratings as well. Any comment that receives a sufficiently low sum of ratings is hidden from untrusted registered users of the site and from readers. By convention, users tend to contribute only positive ratings of each other's comments; however, in the case of spam or inappropriate contributions, negative ratings may be used by the "trusted" members of the community to collaboratively remove content.

The boundary that separates trusted from untrusted users is very porous and is managed technically in an automated fashion by the site's platform.[63] A registered user may shift between trusted and untrusted status, depending on the regularity and extent to which their contributions to the site receive recommendations from other users. Maintaining trusted status is a largely symbolic achievement with relatively narrow material benefits. At the same time, the process of becoming trusted on the site offers participants an avenue to enhanced forms of commitment, and those users who are trusted sometimes refer to their status as an indication of their commitment to the community.

The interactions of individual users of the site produce a dynamic flow of information, reputation, attention, and influence. The accumulation of comments and positive ratings contribute to the visibility of diaries or front-page stories, as well as to the reputations of users. Users with strong reputation scores acquire additional privileges on the site and, if they continue to attract positive attention through recommendations and

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

comments on their content, the potential to acquire enhanced influence among the other members of the site. At the same time, contributions that are uninteresting or otherwise inconsistent with the standards, norms, and views of a sufficient proportion of the site's participants fall from public view. In this way, the technical and social application of the moderation and filtering system takes advantage of the fact that readers can quickly process information to identify whether something is relevant to their interests or not. By relying on the aggregation of numerous judgments about other's contributions, the site's users collectively categorize and identify interesting, relevant, or controversial content. Over time, the content that rises to the top acquires status and becomes more likely to attract further attention.

The Daily Kos platform was designed to accommodate extensive interactions among users and to facilitate a process by which certain contributions acquire heightened visibility and impact over others. In this way, the system accommodates varying levels of commitment and activity, and it also aims to harness the unequal distribution of popularity and influence among users to promote content and discussion deemed interesting by the larger pool of participants.

It is important to underscore that although Moulitsas and the administrators of the site retain control over the precise metrics used by the site's reputation algorithm, this is not the same as directly determining users' reputations or their patterns of adopting the site's features. In this sense, the architectural choices that Moulitsas made in selecting the platform and customizing it should be understood in the context of the norms and interactive practices through which the community of Daily Kos participants has brought the site's technologies "to life."

Two Varieties of Gatekeeping on Daily Kos

The foregoing overview suggests that gatekeeping practices on Daily Kos can proceed through several different mechanisms. The site offers a range of tools and venues through which participants contribute content to the site, evaluate each others' contributions, and engage in formal as well as informal processes of rule or norm enforcement. In addition, the different tiers of user roles that are built into the Scoop software platform presume a hierarchy of privileges and status. The result is a system in which more and less formalized as well as more and less centralized processes of gatekeeping take place. This gatekeeping can apply to participants as well as to content, both of which may be classified as undesirable.

I distinguish between two kinds of gatekeeping—centralized and decentralized—both of which reproduce structural advantages of elites or incumbents. Centralized gatekeeping occurs at high-status positions within the community by site administrators or elite users. Decentralized gatekeeping, on the other hand, entails more diffuse processes that require the participation of numerous site users. In both cases, the effects are comparable: site elites and incumbent users (those who are more experienced, active, or comparatively empowered) play a privileged role establishing, interpreting, and enforcing the ground rules, norms, and frameworks within which

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Table 1. Types of Gatekeeping on Daily Kos

Centralized	Decentralized
Formal rule & policy-making	Emergent social norms
User-training (FAQs)	Norm dissemination (user-to-user)
Rule enforcement (site elites)	Norm enforcement (user-to-user)
Status closure (site elites)	Status closure (incumbent users)
Agenda setting (front page)	Agenda setting (diaries/comments)

discursive production and political mobilization occur on the site. Table 1 shows a side-by-side comparison of examples of the two types.

Compared to previous types of gatekeeping, the key distinction of decentralized gatekeeping concerns the extent to which the gatekeeping not only entails the active participation of "the gated," but also the way distributed actions by the members of a collective can generate similar effects to the concerted efforts of a single individual.[64] Below I introduce examples of both kinds of behavior.

Centralized Gatekeeping in the Clinton Supporters' Strike

Overall, centralized gatekeeping on Daily Kos emerges as a byproduct of the social position and status of site elites. Consistent with the role of news editors in the print media and other professions, these elites utilize their status and position within the Daily Kos community to reproduce their own authority and restrict access to privileges.[65] As a result, visible mechanisms of centralized gatekeeping on Daily Kos tend to be more formalized than mechanisms of decentralized gatekeeping. Site elites—and in particular Markos Moulitsas—possess the resources to convert their perspectives into formal elements of site governance system more than any other members of the site community. Some of the ways they exercise these privileges therefore become very visible and very public, while others less so.

The period around the Democratic Party presidential primaries of 2008 provided numerous example of centralized gatekeeping in response to some contentious conflicts within the community. During the lead up to "Super Tuesday" (February 4, 2008) the polarization between supporters of Hillary Clinton and Barack Obama on the site grew increasingly tense. Moulitsas and the site's editors had not endorsed Obama by that point, but some of them opposed some of the strategies pursued by Clinton's campaign as she sought to divide the party after losing momentum in the popular vote.[66] The conflict bled over into the comments and diaries, where Obama and Clinton supporters clashed. Resentful of the widespread pro-Obama sentiment among their peers, many Clinton advocates claimed to be the victims of unfair treatment. On March 14, "Allegra," one of the Clinton supporters and a prominent diarist, proclaimed a "Writer's Strike," urging Hillary supporters and their sympathizers to immediately cease visiting, reading, and contributing to the site:

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

> I've been posting at DailyKos for nearly 4 years now and started writing diaries in support of Hillary Clinton back in June of last year. Over the past few months I've noticed that things have become progressively more abusive toward my candidate and her supporters.
>
> I've put up with the abuse and anger because I've always believed in what our online community has tried to accomplish in this world. No more. DailyKos is not the site it once was thanks to the abusive nature of certain members of our community.
>
> I've decided to go on "strike" and will refrain from posting here as long as the administrators allow the more disruptive members of our community to trash Hillary Clinton and distort her record without any fear of consequence or retribution. I will not be posting at DailyKos effective immediately. I will not help drive up traffic or page-hits as long as my candidate—a good and fine DEMOCRAT—is attacked in such a horrid and sexist manner not only by other diarists, but by several of those posting to the front page.[67]

Responses to the so-called strike were mixed. However, Moulitsas and some of the other senior members of the community took an aggressive stance against the strikers, mocking them and arguing that they had no right to expect consensus on the site, especially given that Clinton opposed some of the central tenets of the community. Three days after Allegra's post, Moulitsas responded directly, arguing that Clinton's willingness to split the Democrat's super-delegate vote against the popular vote would lead to a civil war within the party:

> Clinton knows this, it's her only path to victory, and she doesn't care. She is willing—nay, eager to split the party apart in her mad pursuit of power.
>
> If the situations were reversed, and Obama was lagging in the delegates, popular vote, states won, money raised, and every other reasonable measure, then I'd feel the same way about Obama. (I pulled the plug early on Dean in 2004.) But that's not the case.
>
> It is Clinton, with no reasonable chance of victory, who is fomenting civil war in order to overturn the will of the Democratic electorate. As such, as far as I'm concerned, she doesn't deserve "fairness" on this site. All sexist attacks will be dealt with—those will never be acceptable. But otherwise, Clinton has set an inevitably divisive course and must be dealt with appropriately.[68]

Moulitsas situates his critique of Clinton and the strikers in reference to the founding ideals of the site, reiterating his authority on the blog and underscoring the distinction between formal community rules (e.g., "sexist attacks...will never be acceptable") and the political vision behind his creation of the site. He shows no sympathy for those who disagree with his perspective.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

The example demonstrates how Moulitsas uses his position and status as leader to affirm and reinforce community standards in the wake of important conflicts. These actions contribute to centralized gatekeeping on the site insofar as they reinscribe boundaries of acceptable behavior and political beliefs. Moulitsas can, in theory, kick out any users whose beliefs he disagreed with, but such actions could undermine his (and the community's) claims to enact democratic values and discourse.[69] Instead, he tends to engage in moral, political, and intellectual persuasion of the site's users, arguing for a particular vision of Democratic party empowerment and electoral strategy. In casting the Clinton supporters as misguided and worthy of mockery, Moulitsas symbolically sets the boundaries of what he sees as a legitimate political position and classifies a subset of the community as beyond the pale.

The episode illustrates most of the mechanisms of centralized gatekeeping identified in Table 1. Moulitsas draws on his capacity as site founder to make and enforce formal rules; lay down guidelines about acceptable behavior; and set the community political agenda of the site through his own posts (all of which appear on the front page).

Whatever his status or administrative privileges, Moulitsas cannot mandate the views of the site's user community by fiat. His agenda-setting and norm-enacting capacity hinges, in part, on the fact that he engages other site elites and experienced users in such discussions and seems to only take drastic action (like banning or publicly denouncing users' views) under extraordinary circumstances. In this way, his actions also contribute to an overall process of status closure, whereby he exercises influence over the selection process of new site elites and the privileges entailed by elite status.

"Everyday" and Decentralized Gatekeeping

The distinct aspects of decentralized gatekeeping boil down to particulars of scale, context, and scope. First, decentralized gatekeeping happens on a larger scale than centralized gatekeeping. It incorporates a wider range of individuals and consists of activities that are far more diffuse and numerous. Second, the contexts in which decentralized gatekeeping takes place are more widely accessible and less reliant on formal status divisions. In practice, this means that decentralized gatekeeping "happens everywhere," in contrast with centralized gatekeeping that necessarily occurs in settings and situations where site elites engage in the activities described above. In many cases, behaviors that contribute to decentralized gatekeeping are more banal than their centralized counterparts. Finally, the scope of decentralized gatekeeping tends to be narrower than that of centralized gatekeeping. This does not mean that decentralized gatekeeping behavior or its effects have less significant implications for the site than centralized gatekeeping, but merely that decentralized gatekeeping rarely entails the sort of claims to formalized authority and responsibility for the site as a whole that characterize centralized gatekeeping in many cases.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Despite these differences, the mechanisms of decentralized gatekeeping parallel those of centralized gatekeeping. In both cases, Daily Kos community members perform relational work to establish, negotiate, enforce, and adapt boundaries, norms, and standards that constitute the site. In the process of both, incumbent community members tend to reproduce their own access, privileges, and status. In other words, the effects of decentralized gatekeeping are also broadly consistent with those of centralized gatekeeping: the reproduction of social structure, authority, and privilege in a manner consistent with the participatory design of the Daily Kos community.

In May 2008, less than one month after the Clinton supporters' strike, two conflicts occurred in a "Cheers & Jeers" comment thread that illustrate how everyday community governance on Daily Kos contributes to an overall pattern of decentralized gatekeeping.[70] Cheers & Jeers, or "C & J," is written by Bill Harnsberger, who is known by his username, "Bill in Portland Maine." C & J began as a normal user diary, but its popularity led Markos Moulitsas to invite Bill in Portland Maine to make it a biweekly feature on the site in April 2005.[71] Since then, the column has become a mainstay and a source for many of the inside jokes on the site. It appears every weekday at approximately 9 a.m. Eastern Time, and typically attracts more comments than other posts. Every C & J post follows a predictable format and aesthetic, including satirical "cheers" and "jeers" from Harnsberger that incorporate light political commentary and pop-culture references. Many of the comments come from regular readers who greet Bill in Portland Maine directly or post their own "cheers and jeers" for the day. The threads focus more on sociable and friendly interactions than on contentious political debate.

In response to one of the first C & J comments on May 13, 2008, a long-time member of the site called "joan reports" posted a comment linking to an offsite blog about the Democratic presidential primaries in West Virginia.[72] A little further down the thread, joan reports also submitted an identical comment and within a few minutes, a particularly active C & J contributor called "Phil N DeBlanc" responded to joan reports' first comment by asking her to "quit spamming C & J." Beneath her second comment, other users posted comments mocking it as spam. In response, joan reports apologized on both subthreads. The apology earned joan reports a number of recommendations and positive comments from other users, one of whom remarked, "She's cool—First time I ever saw someone apologize for spamming."

Almost two hours later in the same thread, another user, "2Nurselady," wrote: **"CHEERS TO HILLARY CLINTON For fighting for the rights of ALL of the voters in this country to have their voices heard, including Florida and Michigan!"** (original capitalization and emphasis). In the wake of the writers' strike and amidst the amicable atmosphere of C & J, 2Nurselady's contribution looked like an intentional provocation, and thus met with immediate condemnation, drawing a number of negative ("troll") ratings. Bill in Portland Maine jumped to 2Nurselady's defense, arguing, "[The comment] Seems like an honest statement of opinion. No profanity. No name-calling. You may disagree with how Hillary wants to 'fight' for the as-yet-unseated delegates in FL and MI, but the statement isn't inflammatory."

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

In response, some of the most frequent commentators on C & J, including one user called "homogenius," responded to Bill in Portland Maine, supporting the negative ratings of 2Nurselady's comments (which would make them invisible to untrusted participants):

> Uhhhh Bill,
>
> Sweetie? Honey Pie?
>
> This isn't about site rules or the FAQ. It's about community norms in C & J. A significant number of us feel that candidate shit violates the spirit and intent of C & J. It doesn't matter whether he or she is inflammatory, insulting, or profane.
>
> I yield to the sense of the community on this, but that's my impression. But I'm sure as fuck not gonna hang out here if we're gonna be subjected to the same shit as the rest of this site for the duration of the primary season.
>
> However, I'm guessing that this nurse person isn't a [Trusted User] and can't see our comments once he or she gets [troll]'d so I would suggest we leave one unhidden for that purpose.
>
> What saith [sic] the rest of the rabble—do we open up C & J to unlimited candidate shit or do we maintain our oasis of insanity to titillate the snark gland and soothe the savage beast?

Bill in Portland Maine and homogenius went back and forth several times on the issue, but the negative ratings of many other users had effectively hidden 2Nurselady's comment in the meantime. Less than ten minutes after her first comment, 2Nurselady posted again:

> JEERS TO THE DAILY KOS For removing my posts because they were pro-Hillary Clinton. I actually complimented The Daily Kos yesterday for allowing variant points of view and now, because I commented in a positive way about Hillary Clinton this morning, you've removed those posts??? Is this CNN???

2Nurselady's response illustrates that she does not understand the site's comment filtering system. At this point, the participants responded by recommending her original post to ensure that it remained visible, and homogenius explained the situation:

> Dear Nurselady,
>
> We have unhidden this comment in case you are not a trusted user and can't read our responses to your comments which have been hidden.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

> In any other diary, your comments would not have received hide ratings (aka "donuts"). They would have been derided for being in bold face with too much in all caps (both are seen as shouting in cyberspace).
>
> However, Cheers and Jeers is generally a place of respite from the candidate wars and other trials and psychoses of daily life and Daily Kos. Bill in Portland Maine's proprietorship of C & J is supported directly by contributions from Kossacks. So, in addition to Mr. Kos's rules and regulations, we have some traditions and boundaries unique to C & J. I'm only one small contributor, so I defer to community sentiment...but I believe this is substantially correct.
>
> If you continue to post candidate comments in C & J, you will likely find them hidden. If you continue to post in bold face and overuse caps, you will be soundly (and justifiably) jeered.
>
> I hope this was helpful. Please feel free to ask for clarification from the community.

2Nurselady thanked homogenius for the information and never returned to comment in C & J again.

These interactions demonstrate some of the ways that participants who do not hold formal elite status on Daily Kos undertake gatekeeping roles in order to moderate and filter content on the site. In the case of joan reports' comments, site participants applied a set of standards without the intervention of site elites or the use of troll ratings. They used a combination of mockery, sympathy, and direct criticism, eliciting a public apology for behavior that broke with the unwritten rules of the C & J sub-culture.

In the 2Nurselady incident, site participants once again coordinated in response to a violation of the C & J community norms, but one of the site elites as well as the Scoop rating system also played an important role in shaping the interactions. Looking at 2Nurselady's user page, a number of users quickly recognized a pattern of behavior that was inconsistent with the standards of written communication on the site as a whole and C & J in particular. Several of them then negotiated with a site elite (Bill in Portland Maine) over the proper course of action. When 2Nurselady returned and demonstrated that she did not understand the informal boundaries of conversations on C & J, one user (homogenius) explicitly clarified the normative logic at work. Although 2Nurselady went on to amend her conversational tactics in subsequent comments and diaries, she ceased participating five days later.

In both cases, two well-known contributors to the C & J conversations, homogenius and Phil N DeBlanc, drew on their authority to speak on behalf of "the community" of C & J participants. Other experienced users joined as well, giving weight to the categorization of joan reports' and 2Nurselady's comments as unacceptable. In doing so, they utilized the site's technical infrastructure and drew upon existing institutions in

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Table 2. Summary Data: Cheers & Jeers, April 13, 2008

	total	unique users	mode	median	mean	max
comments made	772	146	1	2	5.3	73
recommendations received	9516	146	9	31	65.2	537
recommendations given	9516	179	1	7	53.2	700

order to preserve the flow of information they recognized as legitimate (in this case, casual socializing and political humor).

These examples demonstrate that the interactions between Daily Kos participants involve more than just discussion. The interactions also entail identifying and categorizing behavior in accordance with formal and informal standards. Hierarchies of status and identity become operationalized in these acts of categorization, which in turn reproduce boundaries of legitimate discourse and practice.

The application of these emergent norms contribute to both centralized and decentralized gatekeeping as site participants use them simultaneously to acquire and to signal "in-group" membership within experienced and visible subcommunities on Daily Kos. They also feed into the production of status inequalities between those site members who can successfully navigate and appropriate in-group codes and those who cannot. The practice of policing the boundary between trusted and untrusted users through comment rating thus functions as a form of status closure, whereby trusted users regulate access of untrusted users to privileges.

It is possible to quantify and visualize the relationship between comments and recommendations within the April 13, 2008 C & J thread involving joan reports and 2Nurselady. Table 2 contains a descriptive summary of all the comments and recommendations in that thread.

Note that the distributions of all the variables are extremely skewed. Figure 1 presents a scatterplot of comments made (log scale) against recommendations received (log scale) for all participants in the thread. Each point is also scaled to represent the (log) number of recommendations given.

Not surprisingly, the graph demonstrates a positive relationship between the number of comments made and the number of recommendations received. It also seems to show a positive relationship between comments, recommendations received, and recommendations given (thus the growing size of points along both the X and Y axes).

Table 3 contains summary data for each of the key participants discussed above. Here, the positive association between the different types of gatekeeping behavior depicted earlier and authority within the site's reputation system becomes transparent.

These examples show an association between status, gatekeeping behavior, and participation in a single comment thread. The visualization also illustrates that the centralized gatekeeping practices performed by high-status individuals like homogenius and Phil N DeBlanc also possess a collective, decentralized aspect insofar as their individual judgments are reinforced by dozens of other individuals who chose to recommend the comments in question or not.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

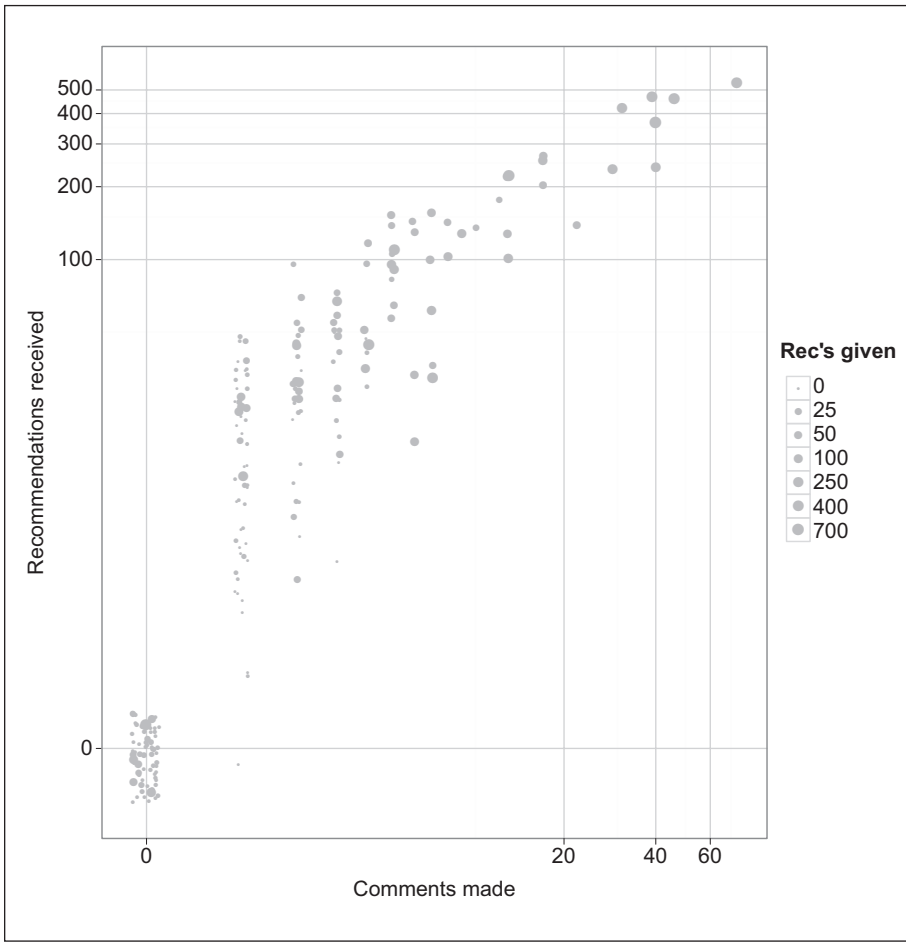


Figure 1. Summary Scatterplot: Cheers & Jeers, April 13, 2008

Table 3. Key Participants, Cheers & Jeers, April 13, 2008

	comments	recommendations received	recommendations given
Phil N DeBlanc	46	459	500
Bill in Portland Maine	13	219	103
homogenius	22	138	46
joan reports	4	47	0
2Nurselady	3	5	0

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Patterns of Decentralized Gatekeeping

In order to evaluate whether the relational work that happens in C & J comment threads produces decentralized gatekeeping at an aggregate level, I require more data and different methods. The evidence from the single comment thread discussed in the previous section suggests an effective means of testing for the presence of aggregated patterns of decentralized gatekeeping behavior. That example showed how site participants utilize the system of comments and recommendations to negotiate status relations, rewarding desirable forms of participation, experience, and commitment with higher numbers of recommendations. In this section, I extend this finding and test whether these associations persist at a much larger scale, using a sample of more than 5,000 site participants and their behavior in several hundred comment threads within "Cheers & Jeers" during all of 2008. First, I present a descriptive analysis of the users, comments, and recommendations in my sample, including attributes of the network created by user comment recommendations. Then, using the same data, I construct a deductive test of the hypothesis that the relational work of recommending comments among Daily Kos users produces an aggregate pattern of decentralized gatekeeping.

Data and Methods

The data set I use to conduct this analysis contains all of the comments and recommendations from Cheers & Jeers in 2008. As a frame for drawing a purposive sample of users, comments, and recommendations, C & J captures a very prominent and culturally significant subset of the Daily Kos population. It also makes for a strong test of decentralized gatekeeping, as participants in C & J threads tend to be, if anything, more experienced and more active than a random sample of site members would be. As a result, if decentralized gatekeeping occurs in C & J, that would suggest that it is also very likely to be present across the rest of the site, where experienced or elite users are even more likely to encounter and interact with new users less familiar with the norms and rules of the community.

As discussed above in the Research Design and Methods section, the outcome of interest in this analysis is the reputational status of individual participants in the C & J comment threads measured by the number of recommendations their comments receive. I operationalize this as two distinct dependent variables. First, for the purposes of hypothesis testing and modeling the relationship between incumbency and reputational status, I divide each individual's total number of recommendations received by their total number of comments made. The resulting variable is thus the individual user's average recommendations per comment in 2008. This measure approximates what Daily Kos users experience and see in the context of comment threads, where each comment appears alongside the number of recommendations it has received. Second, for the purposes of describing and analyzing the network formed by site users (vertices) and their recommendations of comments (directed edges),

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Table 4. Summary of activity on Cheers & Jeers, 2008

	total	unique users
stories	237	–
comments made	172920	4660
recommendations received	1973303	4165
recommendations given	1973303	4595

I calculate the Bonacich centrality score of each user within the 2008 comment-recommendation network.[73]

I include several independent variables in my analysis as measurements of distinct dimensions of user incumbency and controls. To measure experience, I use a count of the number of months elapsed between the month in which the user created his or her account and January 2008. I also include an indicator of whether or not the user appeared on the site's masthead and was thus formally a site elite. Another variable, "activity", consists of a count of the number of C & J comment threads in which the user participated in 2008 by either posting a comment or recommending someone else's comment. Finally, I also measure the total number of recommendations made by the user in C & J comment threads during 2008 (outdegree within the recommendation network).

My analysis proceeds in two steps. First, I describe the user-comment-recommendation network using a series of summary statistics as well as graph-level indices. For some of these indices, I conduct nonparametric tests to determine whether or not the network is more centralized or consists of more reciprocal ties than would be predicted by chance given its particular density and size.[74]

In the second stage of my analysis, I use ordinary least squares regression (OLS) and quantile regression models to test for evidence of an association between user incumbency and reputational status.[75] The rationale behind my use of quantile regression derives from the extremely skewed distribution of the dependent variable—recommendations received per comment per user. Given the extreme inequalities across the distribution, I have reason to expect that the processes driving status acquisition may vary across the distribution of the outcome. By dividing the distribution of this outcome into quantiles and then calculating coefficients for each quantile, I am able to estimate more precisely the relationship between each of the independent variables and my outcome measure. Following Koenker, I also calculate bootstrapped standard errors and p-values for each of these coefficients.

Results

Table 4 presents summary data about the sample of comments, recommendations given, and recommendations received among Cheers & Jeers users in 2008. Descriptive statistics for all of the variables, including pairwise correlations, are provided in Table 5.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Table 5. Summary Statistics, Cheers & Jeers, 2008

	Summary Values				Pairwise Correlations					
	min	median	mean	max	(1)	(2)	(3)	(4)	(5)	(6)
(1) comments made	0	2	29.93	11347						
(2) recommendations received	0	3	341.52	77751	0.76*					
(3) recommendations received per comment	0	1	7.65	7248	0.00	0.09*				
(4) Bonacich centrality score	0	0.01	0.14	31.83	0.00	0.01	0.00			
(5) recommendations given	0	17	341.52	76136	0.94*	0.78*	0.01	0.00		
(6) active stories	1	2	8.95	237	0.52*	0.63*	0.08*	0.01	0.64*	
(7) months of experience†	0	21	20.21	51	0.03*	0.02	–0.01	0.08*	0.05*	0.08*

†Interval scale with January, 2008 = 0
*p-value ≤0.05

Table 6. Network Variables and Graph Level Indices, Cheers & Jeers, 2008

Measure	Value
vertices (active users)	5778
edges (recommendations)	1973303
mutual edges (dyads)	47698
asymmetric edges (dyads)	180070
null edges (dyads)	16461985
network density	0.01
edgewise reciprocity	0.35
log-odds reciprocity ratio†	3.74

†Calculated against the baseline probability of an edge within any graph with the same number of nodes and edges.

As is typical for measures of traffic, recommendation, and attention in online communities, all of the different measures of reputational status follow extremely skewed distributions, with median values close to zero. Indeed, plotting and fitting curves to the distribution of user recommendations received, Bonacich centrality, and recommendations-per-comment reveals that they all follow a "parabolic fractal" distribution, one of several "power law" distributions that recur frequently in natural and social phenomena.[76] Given such extreme distributions, I use the natural logarithm for reputational status measures as well as activity and recommendations given in the correlation tables below as well as all subsequent models.

Table 6 includes descriptive and graph-level analyses of the comment recommendation network. Overall, these results reveal a fairly centralized graph in which the

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Table 7. Regression Results, Daily Kos Cheers & Jeers, 2008
Dependent Variable: Recommendations Received per Comment†

		Quantile Regression					
	OLS	0.1	0.25	0.5	0.75	0.9	0.95
(Intercept)	.202***	−.123***	−.305***	−.523***	.265***	1.212***	1.913***
Recommendations given†	−.110***	0	0	0	−.078***	−.157***	−.244***
Active stories†	.733***	.179***	.441***	−.757***	.865***	.962***	.997***
Months' experience	.003***	0	0	0	.006***	.008***	.011***
Elite user	−.083	.001	−.001	0	−.047	−.031	−.352

$N = 4655$
Adjusted $R^2 = 0.330$
†Indicates logged variable. Dependent variable is also logged.
***p-value ≤0.01

vast majority of dyads contain mutual ties and where reciprocal ties are much more likely than would be predicted by chance.

Given that the graph is relatively sparse, with only a moderate portion of the total possible dyadic ties between users realized, the nonparametric calculation of the log-odds ratio of reciprocity (LRR) is noteworthy. This measure uses the size and density of the network to calculate the probability that any tie in the graph is reciprocated given the underlying probability of an edge existing in the first place. The relatively high odds of a reciprocal tie suggest that C & J users who were active as both commenters and recommenders during 2008 were likely to engage in reciprocal recommendations.

The results of the models, presented in Table 7, show evidence of a complex, differential relationship between users' experience, activity, and reputational status.[77] Overall, the results suggest a significant, positive association between user activity and the number of recommendations-received-per-comment. The association varies depending on the outcome, growing steadily through the fiftieth percentile of the distribution before stabilizing around a maximum level.

In addition, I find a significant, positive association between user experience and the number of recommendations-received-per-comment. This association follows a distinct pattern of growth across the different quantiles of the dependent variable, remaining relatively insignificant below the fiftieth percentile and then growing rapidly over the rest of the distribution.

A significant, negative relationship between recommendations-given and recommendations-received-per-comment also emerges at the fiftieth percentile of the dependent variable, dropping steeply across the rest of the distribution. I find no evidence of a significant relationship between elite user status and recommendations-per-comment.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Discussion: Gatekeeping and Reputational Status

Returning to the hypotheses listed earlier, the results suggest that experience and activity level have a significant association with recommendations-received-per-comment (supporting *H1a* and *H1b*). At the same time I find no evidence of an association between elite status and recommendations-received-per-comment (failing to reject the null for *H1c*), as well as a significant negative association between recommendations-given and recommendations-received-per-comment (reversing the expected relationship from *H2*).

There appear to be strong associations between reputational status and experience and activity within C & J comment threads. Overall, these findings support the idea that patterns of decentralized gatekeeping emerge in the data. Users' relative experience and activity levels within comment threads associate with reputational payoffs. At the same time, as the negative association of recommendation-giving and comment-receiving illustrates, not all forms of activity associate with the same kinds of reputational outcomes. User incumbency predicts reputational status more effectively depending on the sort of activity users tend to pursue.

These findings have several important limitations, the most important of which stems from potential for endogeneity in the models. The analysis I have presented here cannot distinguish between situations where Daily Kos users participate on the site longer because they receive more recommendations versus situations where Daily Kos users receive more recommendations because they have participated on the site longer. Nevertheless, the possibility of reverse causality along these lines does not undermine the validity of my finding that "experience pays" for Daily Kos users insofar as peer-generated status measures on the site are concerned. The strong association between experience and status implies that decentralized gatekeeping effects—the systematic accrual or retention of privileges by incumbent individuals within the community—occur irrespective of the direction of the causal processes involved. Indeed, I would argue that it is likely that a process of self-selection unfolds whereby the community members who decide to return to the site over the course of many weeks, months, and years may do so both because they receive positive reinforcement and because they seek to acquire greater standing among their peers. In concluding that more experience is associated with greater reputational status, it is not necessary to separate the one story from the other. As I argued earlier, decentralized gatekeeping, unlike its centralized variants, need not be the product of self-conscious filtering procedures or acts of exclusion. Instead, decentralized gatekeeping effects encompass the range of mechanisms by which both self-selection and exclusion give rise to and reinforce incumbent advantages within information-generating systems. Longitudinal analysis of the same data could not reasonably distinguish between these competing causal storylines either, as there is no meaningful way to control for the (presumably variable) differences in intrinsic quality of comments or commenters. This is a critical limitation because without such controls, no statistical test can separate users who stay on the site longer because they accrue greater status from those who accrue greater

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

status because they stay on the site longer. Future research should design more nuanced tests that can disentangle these respective explanations of my findings.

Given the culturally significant role of C & J within the Daily Kos site as a whole, as well as the relatively egalitarian social norms that prevail among C & J participants, I conclude that similar patterns of behavior are likely present across the rest of the site, although they may be more difficult to observe or measure in a valid way.

Strong benefits to experience and influence make intuitive sense—indeed, such patterns are exactly what the existing empirical and theoretical literature on online informational production and exchange that I reviewed above would predict. However, the evidence that Daily Kos participants' highly unequal levels of experience, activity, and reputation associate so strongly suggests that the dynamics of social reproduction among the site's users are quite stable. In this sense, the case illustrates Tilly's notion of the relational foundations of inequality:

> In a relational view, inequality emerges from asymmetrical social interactions in which advantages accumulate on one side or the other, fortified by the construction of social categories that justify and sustain unequal advantage. As a rough analogy, consider a conversation involving initially equal partners in the course of which (through wit, guile, knowledge, or loudness) one conversationalist gradually gains the upper hand.[78]

Tilly's "rough analogy" of relational work to a conversation helps map this claim onto the context of Daily Kos. Among Daily Kos participants, stable, extreme inequalities of status and participation have emerged in a number of distinct arenas. In general, site participants who are more active contributors and have a greater amount of experience tend to accrue greater status. This association between experience, participation level, and status suggests that the interactions among site participants play an important role in generating and sustaining social hierarchies.

Conclusions

The results presented here suggest the presence of two kinds of gatekeeping on Daily Kos, both of which may have contradictory effects on the site. Collective action in a large and "open" online community entails more than the boundary work traditionally conceived as centralized gatekeeping by site elites and administrators: it also relies on participants' decentralized interactions, which give rise to decentralized gatekeeping and status inequalities of their own. These inequalities, in turn, feed back into the complex of norms, practices, and standards that prevail among site users, contributing to the community's overall political activities as well as its social dynamics.

Gatekeeping practices by participants on Daily Kos thereby contribute to the negotiation and reproduction of inequalities on the site. My examples illustrate how high-status contributors to Cheers & Jeers performed boundary work in moderating the comment threads. Sometimes—as with 2NurseLady—they did so through disagreement

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

with site elites like Bill in Portland Maine; however, the aggregated evidence taken from the entire archive of 2008 C & J comment threads suggests that the effects also reproduce diffuse status inequalities across the user population as a whole. On the basis of this qualitative and quantitative evidence, future research should pursue more precise identification strategies, which can test for the presence or effects of decentralized gatekeeping.

In the incidents involving joan reports and 2NurseLady, the ability of community members to moderate successfully depended, in part, on their ability to draw on and deploy appropriate codes of behavior established on the site. In deploying these codes, they enacted informal community norms that validated contributions consistent with the overarching goals articulated by Moulitsas, Harnsberger, and other site elites elsewhere. When necessary, they also adapted existing standards to suit the needs of particular situations.

Across these examples, the distinction between centralized and decentralized gatekeeping is porous, and there are many ways in which Moulitsas' and other site elites' actions influence the patterns of gatekeeping that prevail across the rest of the community. The effects of decentralized gatekeeping are, in some sense, the cumulative result of many small-scale examples of relational boundary work, each instance of which involves much smaller numbers of people in specific interactions. Among the site elites and leaders, the tendency toward social closure and exclusion must continually be balanced against the ideological and organizational exigencies to egalitarian and democratic ideals. As 2Nurselady demonstrated, appeals to such ideals—even when made by an outsider who neither understands nor has a long-term commitment to the site—are taken seriously.

The mechanisms of gatekeeping described here only reflect a snapshot of a single dimension of participation (commenting) on the Daily Kos site during a specific period of time. Patterns of decentralized gatekeeping and path dependency reinforced through comment recommendation are not likely to determine the social structure among the site's community as a whole. Making comments and receiving recommendations is far from the only means by which members of the site might participate and acquire status in the eyes of their peers. There are also user-blogs (diaries); participation in offline and advocacy events (such as the annual Netroots Nation Conference or unaffiliated social events); as well as work within the formal Daily Kos organization or other organizations that make up the netroots political movement. In other words, the practices of comment recommendation and the corresponding patterns of path dependency that go along with these practices only constitute a single dimension of the multiplex social processes through which status hierarchies may emerge, rise, and fall within Daily Kos as a whole. Reputational gains through commenting thus do not guarantee that a given user of the site will become well known or achieve broader influence, although they are an index of a certain kind of reputational standing. Status achieved or measured through comments is not determinative of other sorts of status within the community. However, comment recommendations do provide a visible and objective indicator of one dimension of the community's status relations. More fine-grained

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

longitudinal analyses of the patterns of comment posting and recommendation in relation to other modes of status acquisition would be necessary to establish the precise mechanisms by which Daily Kos participants become more or less influential members of the community overall. Also, further analysis will be necessary to establish whether or not the oligarchic tendencies revealed by this analysis persist across these other forms of behavior.

To the extent that the patterns of decentralized gatekeeping through comment recommendation analyzed here constitute a path dependent system of status relations, they imply several conclusions relevant to theories of gatekeeping, democratic organizations, and the networked public sphere. First of all, the presence of decentralized gatekeeping complicates the "myth of digital democracy" perspective, elaborated by Hindman and others, which views the online public sphere as nothing more than a new setting in which old elites can exert their influence. Just as traditional social movement organizations cannot be defined in reference to the identities of their leaders alone, it does not make sense to characterize Daily Kos as an extension of the personality traits of Markos Moulitsas. In moments of conflict, Moulitsas may draw on his monopolistic control over the site infrastructure as a rhetorical justification for the legitimacy of his perspectives, but both he and the other site elites ultimately rely on the persuasiveness of their rhetoric and their ability to build discursive consensus. In this sense, the processes and practices of decentralized gatekeeping reinforce the democratic basis of the site—only the best contributions and the most compelling contributors receive broad support—at the same time as they serve the interests of those who already possess status and influence within the community.

Indirectly, decentralized gatekeeping dynamics partially reinforce Hindman's view that hierarchies, status, and social reproduction emerge through the network dynamics of communication and participation in the open and nonbureaucratic organizations of the blogosphere. In this sense, even though Daily Kos may be unique in terms of its scale and recognition in the political blogosphere, the site exemplifies patterns of attention, influence, and participation found within open source software development communities, Wikipedia, networked social movement organizations, and other sorts of open collective action projects. If decentralized gatekeeping prevails across these other kinds of environments, the putative link between openness and egalitarian outcomes is indeed a myth. Decentralized gatekeeping may represent yet another pathway toward inequality production within the larger sphere of democratic and online participation, while at the same time introducing a means for a new elite to achieve prominence within particular sites or movement organizations.[79]

This analysis of gatekeeping in Daily Kos also reveals some of the variations that characterize gatekeeping in networked environments in contrast with their offline counterparts. Previous studies of offline gatekeeping have focused on bureaucratic settings with fairly clear boundaries. These settings have given rise to modes of gatekeeping behavior largely consistent with the organizational affordances and constraints particular to each environment. For example, newspaper editors exercise individual and collective authority over what sort of news makes the front page.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Likewise, tenured faculty at research universities review grant proposals to determine which projects will receive funding and institutional support. In both cases, incumbents or elites establish and enforce norms that enable them to manage the boundaries of a particular field. In addition, the gatekeeping might proceed at either an individual or a collective level. A single editor makes choices about articles that go into her particular section and newspaper, and she also attends conferences and social events where fellow editors discuss industry-wide standards.

Gatekeeping mechanisms in online collectives like Daily Kos have similar dimensions to offline settings. Moulitsas, the other site elites, and some of the well-recognized incumbent users may hold the authority to directly shape and enforce the sorts of behavior that are considered legitimate on the site. In particular, Moulitsas' unique role as the site's figurehead as well as his absolute authority over the site's infrastructure provide him with broad dictatorial powers, although he seems to avoid using them to their full extent.[80] At the same time, the collective gatekeeping practices also take on a more decentralized character as the practice of deploying categories of behavior and enforcing norms is distributed more widely across the community. Although this study has characterized these decentralized gatekeeping effects and found quantitative evidence of their presence on an aggregate scale, my findings cannot speak to the effects of decentralized gatekeeping on either the character of political discourse or political engagement. Other studies have addressed these issues,[81] but more research will be necessary to evaluate whether and how such transformations matter for the future of political engagement, news production, and collective action.

A theoretical and empirical account of decentralized gatekeeping thus represents a useful counterpoint to previous work emphasizing the salience of participatory affordances within the networked public sphere in general and the political blogosphere in particular.[82] Institutionalized status inequalities within open online communities contradict some of the radical egalitarian ideals that make these sites attractive to many people at the same time as they facilitate long-term movement-building goals and the continued commitment of community insiders.

The establishment and preservation of an elite minority within open online collectives may serve an analogous purpose to the stabilization of management structure and roles in more traditional organizational forms. Elite community members can provide continuity as well as a baseline of contributions to the site at the same time as they play an agenda-setting role relative to their peers. They can also incorporate less-experienced peers into the community through the transmission of norms and training.[83] In the context of building an effective movement organization, this kind of leadership can promote long-term movement survival, the achievement of advocacy goals, and other forms of success. In open online collectives, gatekeeping and similar practices may therefore be necessary for the success and survival of the community as they allow for the cultivation of a high signal-to-noise ratio in what would otherwise be a cacophonous, chaotic environment. According to such a view, the exclusion of certain people and perspectives could serve a productive function inasmuch as it allows for the site to achieve growth and coherence and the leaders to achieve their objectives. This would imply a networked corollary to Michel's "Iron Law of Oligarchy." As findings from a

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

single case do not provide an empirical foundation for settling such debates, however, future work should elaborate these claims and test their applicability to a wider range of open online collectives and network organizations.

In terms of the field of U.S. politics and political organizations, the patterns of intra-blog stabilization exemplified by Daily Kos coupled with the stabilization of the organizational field of political blogs discussed earlier suggest that the initial period of disruptive innovation that characterized the blogosphere between 2004 and 2008 may already have ended. This does not mean that the influence of the political blogosphere and associated movements such as the netroots is waning or diminished, but rather that the place of blogs within the field of U.S. politics and the networked public sphere may no longer be as volatile as it initially seemed. Blogs are now an established piece of the political information and movement ecosystem and it makes intuitive sense that their internal dynamics would likewise assume a relatively stable form. Within the Daily Kos community, the stabilization of a set of norms, hierarchies, and elites imply that the community has, in some sense, solved a core problem of movement-building and social reproduction. At the same time, this opens up related questions about the mechanisms of organizational transformation. The dynamics of path dependency discussed here already are playing an important role in shaping the future of the site and its ability to mobilize its membership. It will be important to see whether or not this stabilization encompasses the netroots movement and the role of the Internet in political organization and behavior.

Acknowledgments

The author (alphabetically) thanks Abigail Andrews, Judd Antin, Yochai Benkler, Fred Block, Nora Broege, Laura Clawson, Peter Evans, Archon Fung, Benjamin Mako Hill, Dave Karpf, James Katz, Matthew Keller, Andrew Korzhuev, Colleen Ortegon, Charles Pearson, and Andrew Schrank, all of whom contributed valuable input and advice on this paper at various stages. Thanks also to the members of *Politics & Society* editorial collective, as well as to audiences and discussants at UC Davis, Rutgers, Northwestern, the 2011 Annual Meeting of the American Sociological Association, and the Berkman Center for Internet and Society.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Support for this research was provided by the Ford Foundation, the Ewing Marion Kauffman Foundation, and the Berkman Center for Internet and Society.

Notes

1. Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven: Yale University Press, 2006); Aaron Shaw and Yochai Benkler, "A Tale of Two Blogospheres: Discursive Practices on the Left and Right," *American*

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Behavioral Scientist 56, no. 4 (2012): 459-487. David Karpf, *The MoveOn Effect: The Unexpected Transformation of American Political Advocacy* (New York: Oxford University Press, 2012). David Karpf, "Understanding Blogspace," *Journal of Information Technology & Politics* 5, no. 4 (2008): 369–385.

2. Eszter Hargittai, "Digital Na(t)ives: Variation in Internet Skills and Uses among Members of the 'Net Generation'," *Sociological Inquiry* 80, no. 1 (2010): 92–113; Matthew Hindman, *The Myth of Digital Democracy* (Princeton, NJ: Princeton University Press, 2008); Kay Lehman Schlozman, Sidney Verba, and Henry E. Brady, "Weapon of the Strong? Participatory Inequality and the Internet," *Perspectives on Politics* 8, no. 2 (2010): 487–509; Cliff Zukin et al. *A New Engagement? Political Participation, Civic Life, and the Changing American Citizen*. (New York: Oxford University Press, 2006).
3. Kenneth T. Andrews et al., "Leadership, Membership, and Voice: Civic Associations That Work," *American Journal of Sociology* 115, no. 4 (2010): 1191–1242; Rodney Dean Benson and Érik Neveu, *Bourdieu and the Journalistic Field* (New York: Polity, 2005); Dana R. Fisher, *Activism Inc.: How the Outsourcing of Grassroots Campaigns Is Strangling Progressive Politics in America* (Stanford: Stanford University Press, 2006); Michael Schudson, "The Sociology of News Production," *Media, Culture & Society* 11, no. 3 (1989): 263–282; Michael Schudson, "The News Media as Political Institutions," *Annual Review of Political Science* 5, no. 1 (2002): 249–269; Theda Skocpol, *Diminished Democracy: From Membership to Management in American Civic Life* (Norman: University of Oklahoma Press, 2003).
4. "Daily Kos," http://dailykos.com.
5. Shaw and Benkler, "A Tale of Two Blogospheres."
6. Daniel Drezner and Henry Farrell, "The Power and Politics of Blogs," *Public Choice* 134 (2008): 15–30; Matthew R. Kerbel, *Netroots: Online Progressives and the Transformation of American Politics* (Boulder, CO: Paradigm Publishers, 2009).
7. Chris Bowers and Matthew Stoller, *Emergence of the Progressive Blogosphere: A New Force in American Politics*, New Politics Institute (2005), http://www.newpolitics.net/node/87?full_report=1; Drezner and Farrell, "The Power and Politics of Blogs"; Karpf, "Unexpected Transformations"; and Kerbel, *Netroots*.
8. Drezner and Farrell, "The Power and Politics of Blogs."
9. Jerome Armstrong and Marcos Moulitsas Zúñiga, *Crashing the Gate: Netroots, Grassroots and the Rise of People-Powered Politics* (White River Junction, VT: Chelsea Green, 2006); Benkler, *The Wealth of Networks*.
10. Jonathan V. Last. "What Blogs Have Wrought," *The Weekly Standard* 10, no. 3 (Sept. 2004), http://www.weeklystandard.com/Content/Public/Articles/000/000/004/640pgolk.asp?nopager=1.
11. Ibid.
12. Hindman, *The Myth of Digital Democracy*.
13. Examples of books include Armstrong and Moulitsas Zúñiga, *Crashing the Gate,* and Glenn Reynolds, *An Army of Davids: How Markets and Technology Empower Ordinary People to Beat Big Media, Big Government, and Other Goliaths* (Nashville: Thomas Nelson, 2006); Columnists include Ross Douthat (*The New York Times*); Mickey Kaus (Slate.com); Andrew Sullivan (*The Atlantic Monthly*); and Kevin Drum (*Mother Jones*).

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

14. Data collected through the examination of archived copies of these blogs on the Internet Archive's "Way Back Machine," http://wayback.archive.org.
15. These were originally two separate questions, the first about incorporation and the second about organizational structure. In total, 62 out of 155 (40 percent) were (part of) an incorporated organization and 77 (50 percent) involved a formal hierarchy of some sort. Details of the sampling, coding, and analysis techniques can be found in Shaw and Benkler, "A Tale of Two Blogospheres." For both questions, Krippendorff's α reliability coefficient was ≥ .7.
16. Subsequently, several bloggers also played a role in supporting and building momentum around the Tea Party movement, although it is not clear that the political blogosphere has constituted a more important component of the Tea Party than broadcast media, informal social networks, or existing advocacy organizations.
17. Matt Bai, *The Argument: Billionaires, Bloggers, and the Battle to Remake Democratic Politics* (New York: Penguin Press, 2007); Karpf, "The MoveOn Effect"; Kerbel, *Netroots*.
18. Such dynamics of stabilization following disruptive technological innovations appear consistent to those observed in other organizational fields. See Neil Fligstein, *The Architecture of Markets: An Economic Sociology of Twenty-First-Century Capitalist Societies,* (Princeton, NJ: Princeton University Press, 2001).
19. The data comes from Karpf's Blogosphere Authority Index (BAI), described in David Karpf, "Measuring Influence in the Political Blogosphere: Who's Winning and How Can We Tell?" *Politics and Technology Review* (George Washington University: Institute for Politics, Democracy and the Internet March 2008), 33–41. http://www.asiaing.com/politics-and-technology-review-march-2008.html. Using Wilcoxon Rank Sum tests, I found that all differences between left and right on every metric are significant $p \leq 0.001$. The BAI is available at: http://blogosphereauthorityindex.com.
20. Correlations calculated using Pearson's r, all $p \leq 0.01$. For the aggregate BAI correlation test between first and last month rank for all sites, $r = 0.75$.
21. Shaw and Benkler, "A Tale of Two Blogospheres"; Karpf, "The MoveOn Effect."
22. For example, see chapters 6 and 7 of Benkler, *The Wealth of Networks*, and Clay Shirky, *Here Comes Everybody: The Power of Organizing Without Organizations* (New York: Penguin Press, 2008).
23. E.g., Schlozman, Verba, and Brady, "Weapon of the Strong?" and Hargittai, "Digital Na(t)ives."
24. Hindman, *The Myth of Digital Democracy*.
25. Paul S. Adler. "Market, Hierarchy, and Trust: The Knowledge Economy and the Future of Capitalism," *Organization Science* 12, no. 2 (2001): 215–234; Gina Neff and David Stark, "Permanently Beta: Responsive Organization in the Internet Era," *Society Online: The Internet in Context*, eds., Philip N. Howard and Steve Jones (Thousand Oaks, CA: Sage, 2004):173–188; William G. Ouchi, "Markets, Bureaucracies, and Clans," *Administrative Science Quarterly* 25, no. 1 (1980): 129–141; Walter W. Powell. "Neither Market Nor Hierarchy: Network Forms of Organization," *Research in Organizational Behavior,* eds., B. M. Staw and L. L. Cummings, Vol. 12. (Greenwich, CT: JAI Press, 1990): 295–336.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

26. E.g. Siobhán O'Mahony and Fabrizio Ferraro, "The Emergence of Governance in an Open Source Community," *Academy of Management Journal* 50, no. 5 (2007):1079–1106.
27. Yochai Benkler, "Coase's Penguin, or, Linux and the Nature of the Firm," *Yale Law Journal* 112, no. 3 (2002); Benkler, *The Wealth of Networks*; Josh Lerner and Jean Tirole, "Some Simple Economics of Open Source," *The Journal of Industrial Economics* 112, no. 3 (2002): 197–234; and Steven Weber, *The Success of Open Source* (Cambridge, MA: Harvard University Press, 2004). The description of these phenomena as "nonmarket" comes from Benkler and is problematic given the prominence of multinational corporations and individuals residing within wealthy, capitalist democracies in these new modes of production.
28. Garret Hardin, "The Tragedy of the Commons" *Science* 162 (December 13, 1968): 1243–1248; Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, MA: Harvard University Press, 1965). Explicitly or not, such arguments drew on the strategy pioneered by Ostrom in her analysis of institutional solutions to managing common-pool resources (CPRs). Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge, UK: Cambridge University Press, 1990).
29. O'Mahony and Ferraro, "The Emergence of Governance in an Open Source Community"; and Siobhán O'Mahony and Fabrizio Ferraro, "Managing the Boundary of an Open Project," *Market Emergence and Transformation,* eds., W.W. Powell and J. Padgett (Forthcoming).
30. Robert Michels, *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*, trans. Eden Paul and Cedar Paul (New York: Hearst International Library Co., 1915); Jo Freeman. "The Tyranny of Structurelessness," *Berkeley Journal of Sociology* 17 (1973): 151–165.
31. Benkler, *The Wealth of Networks*; Pablo J. Boczkowski, *News at Work: Imitation in an Age of Information Abundance* (Chicago: University of Chicago Press, 2010); Hindman, *The Myth of Digital Democracy*; Shirky, *Here Comes Everybody*.
32. Karpf, "The MoveOn Effect."
33. Benson and Neveu, *Bourdieu and the Journalistic Field*; Schudson, "The News Media as Political Institutions."
34. Benkler, *The Wealth of Networks*, Chapters 6 and 7; Boczkowski, *News at Work*; Hindman, *The Myth of Digital Democracy*.
35. Kurt Lewin, "Frontiers in Group Dynamics: Concept, Method and Reality in Social Science; Equilibria and Social Change," *Human Relations* 1, no. 1 (1947): 5–41.
36. David Manning White, "The 'Gatekeeper': A Case Study in the Selection of News," *Journalism Quarterly* 27 (1950): 383–390. Schudson,"The Sociology of News Production" speaks to the broader significance of this issue.
37. Karine Barzilai-Nahon, "Toward a Theory of Network Gatekeeping: A Framework for Exploring Information Control," *Journal of the American Society for Information Science and Technology* 59, no. 9 (2008): 1493–1512; Karine Barzilai-Nahon, "Gatekeeping: A Critical Review," *Annual Review of Information Science and Technology* 43 (2009): 1–79; Steven E. Clayman and Ann Reisner, "Gatekeeping in Action: Editorial Conferences and

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Assessments of Newsworthiness," *American Sociological Review* 63, no. 2 (1998): 178–199; Pamela J. Shoemaker and Tim P. Vos, *Gatekeeping Theory* (New York: Routledge, 2009).

38. For example, see Barzilai-Nahon, "Toward a Theory of Network Gatekeeping"; Barzilai-Nahon, "Gatekeeping: A Critical Review"; and Brian Keegan and Darren Gergle. "Egalitarians at the Gate," *Proceedings of the 2010 ACM conference on computer supported cooperative work - CSCW '10* Savannah, Georgia (2010): 131–134.
39. Boczkowski, *News at Work*.
40. Hindman, *The Myth of Digital Democracy*.
41. Benkler, *The Wealth of Networks;* Shaw and Benkler, "A Tale of Two Blogospheres."
42. On relational work and boundary work, see: Charles Tilly, *Durable Inequality* (Berkeley, CA: University of California Press, 1999); Charles Tilly, "Relational Origins of Inequality." *Anthropological Theory* 1, no. 3 (Sept. 2001): 355–372; Charles Tilly, *Why?* (Princeton, NJ: Princeton University Press, 2006); Viviana A. Zelizer, *The Purchase of Intimacy* (Princeton, NJ: Princeton University Press, 2005), 34–35. On status closure, see Max Weber, *Economy and Society* (Berkeley, CA: University of California Press, 1978).
43. See Tilly, *Why?* especially 19–21.
44. Freeman, "The Tyranny of Structurelessness"; Michels, *Political Parties*; O'Mahony and Ferraro, "The Emergence of Governance in an Open Source Community."
45. Exceptions include Cliff Lampe and Paul Resnick. "Slash(dot) and Burn: Distributed Moderation in a Large Online Conversation Space." *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI),* Vienna: ACM Press, (2004): 543–550; as well as Cliff Lampe, Erik Johnston, and Paul Resnick, "Follow the Reader: Filtering Comments on Slashdot," *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI),* San Jose, CA: ACM Press, (2007), both of which look at practices of distributed moderation within Slashdot. In addition, Karpf, "Understanding Blogspace"; and David Karpf, "Macaca Moments Reconsidered: Electoral Panopticon or Netroots Mobilization?" *Journal of Information Technology & Politics* 7, no. 2 (2010): 143–162 offer a number of salient insights particular to the internal dynamics of Daily Kos. See also, Barzilai-Nahon, "Toward a Theory of Network Gatekeeping"; Barzilai-Nahon, "Gatekeeping: A Critical Review"; and Keegan and Gergle, "Egalitarians at the Gate."
46. Barzilai makes a related point, but she is more concerned with shifting the focus of analysis (to "the gated") rather than the idea that gatekeeping itself may have decentralized as well as centralized forms. See Barzilai-Nahon, "Toward a Theory of Network Gatekeeping" and Barzilai-Nahon, "Gatekeeping: A Critical Review."
47. As I explain below, using C & J comment threads as a data source also fixes a number of potential confounding factors, such as posting time of day, post content, and subcommunity norms.
48. The script was written in the Python 3 programming language by Andrew Korzhuev, a student in the Information Sciences Department at the Saint Petersburg State Institute of Technology, in March 2010. The script is licensed under the GNU GPL v3 and I am happy to share it along with the data it collected. This data collection was possible because the

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Daily Kos site maintains an open, publicly searchable archive of posts, comments, and recommendations, all of which are licensed by the site owners for reproduction and reuse.

49. This operational definition builds on the arguments of Clayman and Reisner, who argued that gatekeeping encompassed both the process by which gatekeepers managed the flow of a given resource (e.g., publication in a newspaper) as well as the criteria that determined patterns of selection (e.g., "newsworthiness"). See: Clayman and Reisner, "Gatekeeping in Action," especially 179-180.
50. Shaw and Benkler elaborate on this point as well as its implications for political blogging in the networked public sphere in "A Tale of Two Blogospheres." See also, Benkler, *The Wealth of Networks*.
51. Note that I define each of these variables as well as the statistical methods used for this analysis in greater detail later in the paper.
52. This is substantially less than major news outlets like Huffington Post (980 million page views, 35 million unique visitors) or The *New York Times* (400 million page views, 2 million visitors), but neither of these sites offer anything close to the participatory infrastructure that Daily Kos does. What is more, in years of congressional and presidential elections, traffic to Daily Kos typically surges. Page view and visitor data taken from Google AdPlanner on October 2011, http://google.com/adplanner. Regarding the history of Daily Kos, see Armstrong and Moulitsas Zúñiga, *Crashing the Gate*.
53. For more on the netroots, see Kerbel, *Netroots*. Also, on Daily Kos, see Karpf, "The MoveOn Effect." It bears note that several of the site editors have publicly repudiated the idea that the site is truly radical. In a personal communication about this paper, Laura Clawson, a contributing editor to the site since 2006, underscored this point to me.
54. For a comparison of collaborative practices on the left and right of the U.S. political blogosphere, see Shaw and Benkler, "A Tale of Two Blogospheres."
55. The platform, called "Scoop," was originally developed by Rusty Foster for the "Kuro5in" (pronounced "corrosion"), discussion forum, which is geared towards computer programming and related interests, http://www.kuro5in.org, accessed April 3, 2010.
56. Armstrong and Moulitsas Zúñiga, *Crashing the Gate*.
57. At the beginning of 2011, Daily Kos deployed a completely transformed interface—dubbed "DK4"—that incorporates a number of new features. As this article was already under review at that time, I do not incorporate any discussion of the DK4 features into my analysis.
58. User data from http://dailykos.com. During 2008, the site generated between 500,000 and 2 million hits daily. Traffic data from aggregate statistics collected by "The Truth Laid Bear," http://www.ttlb.com and Nielsen's "Blogpulse" service, http://www.blogpulse.com. The mean daily traffic at this time was around 600,000 hits.
59. "About Daily Kos," accessed March 27, 2010, http://dailykos.com/special/about2.
60. Armstrong and Moulitsas Zúñiga, *Crashing the Gate*; and Marcos Moulitsas Zúñiga, *Taking on the System: Rules for Radical Change in a Digital Era* (New York: Penguin, 2008).
61. Several examples of this arise in the threads discussed below. See, in particular, Moulitsas' post in response to the writers' strike. Also, this kind of rhetoric emerges repeatedly in

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

Moulitsas' books: Armstrong and Moulitsas Zúñiga, *Crashing the Gate*; and Moulitsas Zúñiga, *Taking On the System*.

62. I am not aware of any means by which site participants can "recall" an editor from her position.
63. The Scoop platform computes whether users are trusted using an algorithm customized by Moulitsas and the site's chief technical officer, Jeremy Bingham. The details of this algorithm are secret to prevent users from gaming the system and acquiring trusted status too easily.
64. Definitions of gatekeeping can be found in Barzilai-Nahon, "Toward a Theory of Network Gatekeeping"; Barzilai-Nahon, "Gatekeeping: A Critical Review"; Clayman and Reisner, "Gatekeeping in Action"; Lewin, "Frontiers in Group Dynamics: Concept, Method and Reality in Social Science; Equilibria and Social Change"; Shoemaker and Vos, *Gatekeeping Theory*.
65. Thomas F. Gieryn, "Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists." *American Sociological Review* 48, no. 6 (1983): 781–795; Clayman and Reisner, "Gatekeeping in Action."
66. The site elites were far from unanimous in their support for Obama. Of the (at the time) 26 individuals identified as site elites, at least one publicly endorsed Hillary Clinton and another voted for John Edwards in the primaries. See: http://www.dailykos.com/storyonly/2008/1/1/428108/-My-Vote-1-8-08; and http://www.dailykos.com/storyonly/2008/2/5/450060/-Im-For-Hillary-Clinton. Thanks to Laura Clawson, one of the Daily Kos editors, who pointed me to this information in a personal email exchange.
67. "Writers' Strike at DailyKos," accessed December 17, 2010, http://www.dailykos.com/story/2008/3/14/20827/4727/132/476843.
68. "The Clinton civil war," accessed December 17, 2010, http://www.dailykos.com/storyonly/2008/3/17/12417/1285/527/478498.
69. See Karpf, "The MoveOn Effect," for an example where Moulitsas banned users for promoting the idea that the U.S. government had been behind the September 11 attacks.
70. The full comment thread, including all quotations I use below, can be found at: http://www.dailykos.com/story/2008/5/13/85118/7425 (accessed April 28, 2012).
71. "Cheers and Jeers: Tuesday," http://www.dailykos.com/storyonly/2005/4/5/104652/3850.
72. joan reports has participated since 2004, contributing thousands of comments and diaries.
73. I use Bonacich centrality in this case because it operationalizes a relational concept of influence where the value of someone's attention or recommendation is measured in turn by who recommends their contributions. For more on the measure, see: Phillip Bonacich, "Power and Centrality: A Family of Measures," *The American Journal of Sociology* 92, no. 5 (1987): 1170–1182. Approximating eigenvector centrality scores, I set the attenuation parameter equal to 3/4 the inverse of the maximum eigenvalue of the user-recommendation adjacency matrix. In this case, that value $\beta = 9.85 \times 10^{-6}$. The use of such small, positive β results in centrality scores that slightly reward individuals whose comments have more recommendations from individuals whose comments have also received many recommendations.
74. The logic of inquiry guiding these analyses follows Stanley Wasserman and Katherine Faust, *Social Network Analysis: Methods and Applications* (New York: Cambridge

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

University Press, 1995); as well as Carter T. Butts, "Social Network Analysis: A Methodological Introduction," *Asian Journal Of Social Psychology* 11, no. 1 (2008): 13–41.

75. On quantile regression, see: Roger Koenker, *Quantile Regression* (New York: Cambridge University Press, 2005); and Roger Koenker and Kevin F. Hallock, "Quantile Regression," *The Journal of Economic Perspectives* 15, no. 4 (2001): 143–156.
76. Lada Adamic and Bernardo Huberman, "Power Law Distribution of the World Wide Web," *Science* 287 (2000): 2115; and Jean Laherrère, "Distributions de type "fractal parabolique" dans la Nature," *Comptes Rendus de l'Académie des Sciences, Série II a: Sciences de la Terre et des Planétes* 7 (1996): 535–541.
77. For the purposes of interpretation, note that I use the natural logarithm of the dependent variable as well as several of the independent variables. These results are robust to the inclusion of interaction and quadratic terms.
78. Tilly, "Relational Origins of Inequality": 362. Tilly makes this argument in a discussion of how the emergence of credit made reputation "crucial to commercial viability" (356).
79. Eszter Hargittai and Aaron Shaw, "Weapon of the Strong Revisited? The Internet, Young Adults, and Political Engagement," *Annual Meeting of the Midwest Political Science Association (MPSA),* Chicago, 2011; and Jen Schradie, "The Digital Production Gap: The Digital Divide and Web 2.0 Collide," *Poetics* 39, no. 2 (2011): 145–168.
80. This sort of "benevolent dictatorship" is a common organizational structure in online collectives. See Weber, *The Success of Open Source* on the role of Linus Torvalds in history and organization of the Linux kernel development. As Weber argues, strong norms of meritocracy as well as the strict technical demands of creating a functional computer operating system are reasons this autocratic mode of governance retains legitimacy, despite its obvious contradictions with the egalitarian and communitarian ethos underpinning open source and online collaboration.
81. Eric Lawrence, John Sides, and Henry Farrell, "Self-Segregation or Deliberation? Blog Readership, Participation, and Polarization in American Politics," *Perspectives on Politics* 8, no. 1 (2010):141–157; Aaron Smith, *The Internet's Role in Campaign 2008*, The Pew Research Center, Internet & American Life Project (2009), http://www.pewinternet.org/Reports/2009/6--The-Internets-Role-in-Campaign-2008.aspx; and Schlozman, Verba, and Brady, "Weapon of the Strong?"
82. Benkler, *The Wealth of Networks*; Shaw and Benkler, "A Tale of Two Blogospheres."
83. Discussions of this process can be found in both Judd Antin and Coye Cheshire, "Readers are not free-riders," *Proceedings of the 2010 ACM conference on computer supported cooperative work - CSCW '10* (2010):127–130; and Jennifer Preece and Ben Shneiderman, "The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation," *AIS Transactions on Human-Computer Interaction* 1, no. 1 (2009): 13–32.

Bio

Aaron Shaw (aaronshaw@northwestern.edu) is an assistant professor in the Department of Communication Studies at Northwestern University and a fellow at the Berkman Center for Internet and Society at Harvard University. His research focuses on the dynamics of participation, mobilization, and organization in large collectives online, and includes studies of wikis, online political engagement, and crowdsourcing.

Downloaded from pas.sagepub.com at NORTHWESTERN UNIV LIBRARY on August 26, 2012

This article was downloaded by: [Northwestern University]
On: 08 February 2014, At: 13:39
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Broadcasting & Electronic Media
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hbem20

Digitally Savvy Citizenship: The Role of Internet Skills and Engagement in Young Adults' Political Participation around the 2008 Presidential Election

Eszter Hargittai [a] & Aaron Shaw [b]
[a] Communication Studies Department, Northwestern University and Fellow at Harvard's Berkman Center for Internet & Society
[b] Berkman Center for Internet and Society at Harvard University
Published online: 06 Jun 2013.

To cite this article: Eszter Hargittai & Aaron Shaw (2013) Digitally Savvy Citizenship: The Role of Internet Skills and Engagement in Young Adults' Political Participation around the 2008 Presidential Election, Journal of Broadcasting & Electronic Media, 57:2, 115-134, DOI: 10.1080/08838151.2013.787079

To link to this article: http://dx.doi.org/10.1080/08838151.2013.787079

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the

Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Downloaded by [Northwestern University] at 13:39 08 February 2014

BEA
Educating for Tomorrow's Media

Routledge
Taylor & Francis Group

Digitally Savvy Citizenship: The Role of Internet Skills and Engagement in Young Adults' Political Participation around the 2008 Presidential Election

Eszter Hargittai and Aaron Shaw

Popular narratives assume that digital media play a central role mobilizing voters and especially young adults. Based on unique survey data of a diverse group or young adults from Spring, 2009, we consider the relationship between differentiated internet uses, and online and offline political engagement around the time of the 2008 U.S. presidential election. Thanks to our rich data set, we are able to consider both online and offline activities while taking into consideration more traditional measures. Our findings suggest that online forms of political engagement complement offline engagement. The pathways to young adults' political participation remain relatively stable. We also find an association between Internet skills, social network site usage and greater levels of engagement. These findings imply that although Internet usage alone is unlikely to transform existing patterns in political participation radically, it may facilitate the creation of new pathways for engagement.

Popular narratives following the 2008 U.S. presidential election portray digital media as playing a central role in mobilizing voters and especially young adults to

Eszter Hargittai *(Ph.D. Sociology, Princeton University) is the Delaney Family Professor in the Communication Studies Department at Northwestern University and Fellow at Harvard's Berkman Center for Internet & Society. Her research interests include the role of Web-use skills in Internet use and how people may benefit from their online activities.*

Aaron Shaw *(Ph.D. Sociology, University of California, Berkeley) is a Fellow at the Berkman Center for Internet and Society at Harvard University and, beginning in Fall, 2013, will be an assistant professor of Communication Studies at Northwestern University. His research interests include collective action, collaboration and mobilization online.*

The authors are grateful for a generous grant (#87311) from the John D. and Catherine T. MacArthur Foundation that made this project possible. They thank participants in the workshops of Northwestern's Institute for Policy Research and Harvard's Berkman Center for Internet & Society for their valuable feedback. A previous version of this paper was presented at the Midwest Political Science Association 2011 meetings and the American Sociological Association 2011 meetings. The authors appreciate the assistance of Ericka Menchen-Trevino and Jessica Diamond with data collection, Cassi Saari with background research, and the undergraduate research assistants of the Web Use Project group in 2009 for data entry. The helpful support of Ann Feldman and Tom Moss is kindly acknowledged. The authors are also grateful to the Berkman Center for Internet & Society at Harvard University for its support.

© 2013 Broadcast Education Association
Journal of Broadcasting & Electronic Media 57(2), 2013, pp. 115–134
DOI: 10.1080/08838151.2013.787079 *ISSN: 0883-8151 print/1550-6878 online*

Downloaded by [Northwestern University] at 13:39 08 February 2014

Downloaded by [Northwestern University] at 13:39 08 February 2014

engage in the political process (Miller, 2008; Stirland, 2008). Despite the ubiquity of such expansive claims about the role of the Internet in driving political engagement among young Americans (Fisher, 2008; Melber, 2008; Ruffini, 2008), we know little about how variations in young people's Internet experiences and skills may link to voting or other forms of online and offline political engagement.

Research has shown that more than ever, Americans utilize the Internet to access political information and engage in political activities (Foot & Schneider, 2002; Kenski & Stroud, 2006; Smith, 2009). In the context of the 2008 U.S. presidential election that produced elevated participation rates among young and historically disenfranchised subsets of the population (Godsay, Nover & Kirby, 2010; Kirby & Kawashima-Ginsberg, 2009) we ask: Did the Internet reinforce existing patterns in political engagement or did it open up new avenues to participation? To what extent do Internet skills and online practices relate to engagement in comparison with established predictors?

We approach these questions using a broad, multi-dimensional concept of political engagement (Brady, 1999; Schlozman, Verbal, & Brady, 2010; Smith, Schlozman, Verba, & Brady, 2009; Verba, Schlozman, & Brady, 1995; Zukin, Keeter, Andolina, Jenkins, & Delli Carpini, 2006). Within this framework, political engagement consists of multiple dimensions including cognitive engagement, political voice, and political participation. All of these dimensions constitute important indicators of engagement in political life, despite the fact that they do not necessarily entail equivalent levels of commitment, effort, or impact (Zukin et al. 2006, pp. 49–87). We look at whether or not respondents' demographic attributes, socioeconomic status (SES), Internet experience and skill, online political information practices, civic engagement, political knowledge and political interest relate to voting and other types of political action.

Existing survey data lack nuanced measures of online behavior that would enable us to examine these relationships. Our data allow us to test for associations between SES, political engagement, and a highly detailed set of measures about online practices and skills among a diverse cohort of first-year college students in the Chicago area. Given the characteristics of our sample, we cannot determine whether or not our findings are representative of broader national patterns. However, we can make conjectures about the relevance of online practices and skills for youth political engagement.

The Changing Face of Political and Civic Engagement Among Young Americans

Our research enters a wide-ranging debate on the character and extent of young American citizens' political engagement in a changing technological and media environment. Within this research, two key narratives have emerged: one focuses on the idea of *disaffected citizenship* and the other on patterns of *cultural displacement* (Loader, 2007, pp. 1–5). The narrative of disaffected citizenship argues that the

Downloaded by [Northwestern University] at 13:39 08 February 2014

declining rates of youth participation in political and civic life as measured by indicators such as electoral participation and membership in associations indicate an alarming withdrawal from public life. By contrast, the narrative of cultural displacement underscores how young people's rejection of the modes of civic and political engagement favored by their parents and grandparents has been accompanied by the rise of new practices of public participation (Bennett, Wells, & Freelon, 2011; Bennett, Wells, & Rank 2009; Papacharissi, 2010; Weaver Lariscy, Tinkham, & Sweeter, 2011). Few studies fall exclusively into one or the other category of this schematic binary, but a brief elaboration of the two perspectives provides an effective introduction into the concepts and previous findings that frame our research questions.

On the disaffected citizenship side of the debate, Putnam's (2000) landmark study painted a picture of declining civic life in the United States and underscored the extent to which young Americans had not embraced their parents' and grandparents' willingness to participate in civic associations of various kinds. Putnam reported lower levels of political party and electoral participation among younger citizens in comparison with older cohorts and warned of an imminent threat to American democracy, culture, and civic life. He pointed to the rise of cable television and broadcast media as a partial cause of these trends, arguing that broadcast media consumption undermined deliberation in the public sphere and participation in civic life more generally. He drew no specific conclusions about the effects of the Internet, but underscored its potential to open up new avenues for collective action and associational activities.

Zukin and colleagues (2006) have used extensive survey evidence to re-contextualize the decline observed by Putnam. These authors find that Americans overall—and young Americans in particular—have not simply abandoned practices of political and civic participation, but have instead found new modes of engagement that earlier research had either overlooked or downplayed. Their analysis underscores the rise of novel forms of political action including participation in online communities such as social network sites to connect with like-minded peers around specific issues. These findings provide an example of the cultural displacement narrative, in which the erosion of electoral participation and civic associations does not herald the decline and fall of American democracy, but rather suggests a transmutation into a new, perhaps more digitally mediated form, which Bennett and colleagues (2009, 2011) call ''actualizing citizenship'' and through which Papacharissi (2010) defines characteristic modes of engagement in the networked ''private sphere.''

The above examples of the disaffected citizenship and cultural displacement narratives illustrate that the definition and operationalization of the concept of political engagement matters for the resulting analysis. Typically, the disaffected citizenship account relies on a relatively narrow set of indicators of political participation, such as voting in elections; donating to, or volunteering for parties and candidates; and communicating directly with political representatives (Putnam, 2000, 31–47). Research has argued that these indicators of ''dutiful citizenship'' have a generational

aspect (Bennett et al., 2009, 2011; Weaver Lariscy et al., 2011; Zukin et al., 2006). In contrast, the cultural displacement account necessarily requires a more expansive approach, incorporating a wider range of activities representing a multidimensional civic and political life (Zukin et al., 2006, pp. 49–87). In other words, youth political engagement varies greatly depending on how it is measured, and the decision to include one or another variable implies, to some extent, the acceptance of one or the other analytical paradigm.

While we do not have an *a priori* preference for either the disaffected citizenship or the cultural displacement narrative, we prefer the multidimensional approach to political engagement favored by adherents of the latter perspective. In this regard, we agree with previous research arguing that both the disaffected citizenship and cultural displacement narratives seem to contain a grain of truth (Bennett et al., 2009, 2011; Papacharissi, 2010). A broader definition of participation and engagement can incorporate a fuller range of activities that individuals may perceive, in some way or another, as political. This approach allows us to capture shifts in the nature of civic life as they happen and to compare our findings more rigorously to other research employing different approaches to the same questions.

We adopt the broad definition of political participation provided by Verba (1995) and colleagues as, ''activity that has the intent or effect of influencing government action—either directly by affecting the making or implementation of public policy or indirectly by influencing the selection of people who make those policies'' (p. 38). This definition encompasses both electoral and non-electoral forms of participation such as attention to political issues, the public expression of ideas (voice), and a diverse range of political actions (Brady, 1999).

The Internet and Youth Political Engagement

The increasing spread of information and communication technologies has led to questions about and research on how time spent online may be related to youth political engagement. Of particular interest has been the potential of the Internet for increasing youth political participation given that members of this age group are most likely to be online and spend more time on the Web and using social media tools than other cohorts (Lenhart, Purcell, Smith, & Zickhur, 2010). In this literature, an early emphasis on tests of potential *instrumental effects* of Internet usage on political engagement outcomes has largely given way to approaches analyzing the interdependence of online and offline engagement practices with other socioeconomic and attitudinal factors (Bimber, 2001, 2003; Xenos & Moy, 2007).

Analyses of the relationship between the Internet and political engagement have evolved alongside research documenting the decline of civic associationism and political participation among younger Americans (Fisher, 2006; Putnam, 2000; Skocpol, 2003). A study of Americans' social networks (McPherson, Smith-Lovin, & Brashears, 2006, 2009) argued that the rise of the Internet may have been responsible for

Downloaded by [Northwestern University] at 13:39 08 February 2014

Downloaded by [Northwestern University] at 13:39 08 February 2014

increasing social isolation on a national level, but was refuted by findings that Internet use, and especially the use of social network sites (SNS), associates strongly with more diverse networks, enhanced community engagement, and heightened public activity overall (Hampton, 2011; Hampton, Lee, & Her, 2011; Hampton, Sessions, Her, & Rainie, 2009). A meta-analysis of related research found little evidence to support the claim that the Internet had reduced civic engagement overall, and some indication that it may have had slight positive effects under certain conditions (Boulianne, 2009).

With regard to political communication and participation, no strong consensus has emerged as to whether or not the Internet alters political engagement, in part because it appears that the effects of Internet usage are mediated by a number of intervening factors such as existing social inequalities (Schlozman et al., 2010; Smith et al., 2009). Vromen (2007) found that SES and demographic factors predict political participation better than Internet use. Another study suggested that while Internet use may affect political knowledge and information seeking directly, more ''concrete'' forms of engagement such as voting, attending rallies, or talking about politics are contingent on education level, attention to, and interest in politics (Xenos & Moy, 2007). Experimental research conducted specifically with young respondents suggests that exposure to political information online has differential effects depending on prior interest and engagement levels (Xenos & Kyoung, 2008).

Other studies suggest that news media consumption and interactive communication may promote political participation among young people (Bachmann, Kaufhold, Lewis, & Gil de Zúñiga, 2010; Bakker & de Vreese, 2011; Bennett et al., 2009, 2011; Esser & de Vreese, 2007; Gil de Zúñiga & Valenzuela, 2011; Hampton, 2011; Kim & Kim, 2007). Some studies have focused specifically on the relationship of political practices and social network site uses (e.g., Ancu & Cozma, 2009), and have had mixed results about whether young SNS users are more ore less likely to engage politically than users of other forms of media (Baumgartner & Morris, 2010; Vitak et al., 2011; Zhang, Johnson, Seltzer, & Bichard, 2010). Socioeconomic factors do appear to play a moderating role in the relationship between SNS usage and political activity (Pasek, more, & Romer, 2009).

Earlier studies have underscored the importance of demographic, socioeconomic, and political attention-related variables, all of which appear to moderate the relationship of Internet use and political engagement among young Americans. Additionally, some research suggests that Web-use skills may also be related to various forms of Internet usage (Correa, 2010; Hargittai, 2010; Hargittai & Hinnant, 2008; Wasserman & Richmond-Abbott, 2005). Those with more Internet skills are more likely to engage in a diverse set of online activities than those with less know-how even when holding other factors such as demographic and socioeconomic background constant. An independent effect of user sophistication on various types of Internet uses calls into question the stability of previous findings on the relationships between Internet use, SES, demographic background and political engagement.

Based on the literature reviewed above, we test the following interrelated hypotheses about young people's political engagement:

H_1: Greater levels of Internet experiences will associate with increased:
- H_{1a}: online political information practices
- H_{1b}: civic engagement
- H_{1c}: voting
- H_{1d}: political engagement behaviors other than voting

H_2: Greater levels of online political information practices will associate with increased:
- H_{2a}: civic engagement
- H_{2b}: voting
- H_{2c}: political engagement behaviors other than voting

H_3: Greater levels of civic engagement will associate with increased:
- H_{3a}: voting
- H_{3b}: political engagement behaviors other than voting

Data and Methods

While data sets exist with detailed information about people's political engagement, participation, knowledge and interests, such data sources are rarely coupled with nuanced measures of people's Internet uses, experiences and skills. To address this shortcoming of available resources, we collected our own survey data to test the hypotheses proposed above.

Data Collection

We administered a paper-pencil survey in February–April, 2009, just a few months after the 2008 U.S. presidential elections, to students in class in the one course at the University of Illinois, Chicago that is required for everybody: the First-Year Writing Program. Overall, counting all students who were enrolled in the course, the final response rate is 80.5% yielding valid data from 1,115 first-year students. Because voting in the presidential election is of central interest to this paper, we excluded people who reported not being eligible to vote and thus our analyses are based on the 1,001 first-year students who were eligible to participate in the elections in Fall, 2008.

We administered the survey in class on paper rather than on the Web so as not to bias against students who are online less frequently or who are less inclined to fill out Web-based forms (e.g., due to lack of enough private time spent online). Since having ample time online to engage in various activities is linked to the questions of interest in this study, it was important not to use a data-collection method that might be related to it.

Downloaded by [Northwestern University] at 13:39 08 February 2014

Independent Variables

Respondent Background

Table 1 presents the demographic and socioeconomic composition of our sample showing that participants are diverse on these measures.

Political Capital. The concept of political capital captures respondents' interest in, knowledge of, and investment in politics. Table 2 presents the basic descriptive statistics about these variables. The item that captured partisanship asked respondents to classify their political views on a five-point scale from ''very liberal'' to ''very conservative'' that also included an open-ended ''other'' response option. We recoded the answers along a 3-point scale where the endpoints of the original scale (i.e., *very conservative* and *very liberal*) were categorized as ''high partisanship'' and ''middle-of-the-road'' became ''low partisanship.'' We then coded the open-ended responses by hand using the three-point partisanship scale. The vast majority

Table 1
Background of Study Participants

Variable	Percent
Men	40.5
Women	59.5
Age	
18	67.2
19	31.9
20–24	0.9
Race and Ethnicity	
African American, non-Hispanic	10.8
Asian American, non-Hispanic	20.5
Hispanic	24.5
Native American, non-Hispanic	0.4
White, non-Hispanic	42.0
Parents' Highest Level of Education	
Less than high school	7.5
High school	15.7
Some college	23.3
College	34.2
Graduate degree	18.6

Downloaded by [Northwestern University] at 13:39 08 February 2014

Table 2
Respondents' Political Capital

Variable	Mean	St. Dev.
Partisanship (1–3)	1.8	(0.7)
Political interest (1–4)	2.5	(0.9)
Political knowledge (0–5)	3.6	(1.3)

Downloaded by [Northwestern University] at 13:39 08 February 2014

of these write-in responses expressed a general lack of interest in politics and were coded as ''low partisanship.''

To capture respondents' political interest, we asked students to rate themselves on a four-point scale ranging from ''not at all interested'' to ''very interested.'' Our political knowledge variable was based on a series of open-ended questions, each of which asked respondents about a different aspect of U.S. government and politics. Specifically, the ones we use here asked respondents the following with percent of correct responses indicated in parentheses: (1) what office Joe Biden holds (71%); (2) which party has the most members in the House of Representatives (64%); (3) which party is more conservative (74%); (4) whose responsibility it is to determine if a law is constitutional (this was a multiple-choice question with the following options: the President, Congress, or Supreme Court) (60%); and (5) how much of a majority is needed in the House and Senate to override a presidential veto (also asked as a multiple-choice question with ''one-half,'' ''two-thirds'' and ''three-quarters'' as possible responses) (93%). These questions were developed previously and have been validated as effective measures of Americans' political knowledge (Delli Carpini & Keeter, 1996). The resulting variable takes on values ranging from 0–5 depending on how many of the questions respondents answered correctly.

Internet Experiences. We include measures of veteran status (how long the respondent has been an Internet user), autonomy (access to the Internet at different locations indicating freedom to use the Internet when and where one wants to), frequency of use and skill. See Table 3 for descriptive statistics.

We measure veteran status in number of user years and find that on average respondents had been online for just under 6 years at the time of the study. We log this measure in the analyses due to the assumption that each additional year of experience with the medium has diminishing returns. We operationalize autonomy of use by counting up the number of locations, from among ten, where respondents report having access to the Internet. Students report 6.6 access locations on average. We log this measure for the same reason as use years. To determine frequency of use, we asked students how much time they spend on the Web (excluding email, chat and voice services) on an average weekday and an average Saturday/Sunday and then calculate weekly number of hours from these responses. On average, students report browsing Web sites for just over 17 hours weekly.

Table 3
Internet Use Context and Experiences

Variable	Mean	St. Dev.
Number of use years (0–9 scale)	5.7	(2.3)
Number of locations to access the Internet (0–10 scale)	6.6	(2.2)
Hours spent on the Web weekly	17.3	(10.0)
Skill index (1–5 scale)	3.2	(0.8)
Social network site usage (0–2)	1.0	(0.8)

Downloaded by [Northwestern University] at 13:39 08 February 2014

We use an established index measure made up of 27 Internet-related items (Cronbach's $\alpha = 0.94$) to measure users' Internet skills (Correa, 2010; Hargittai, 2010). We also include a measure (0–2) of social network site usage, based on two questions. One inquired how frequently respondents check other people's status updates, the other asked about posting their own updates. We coded both as a binary response, counting those who engaged in either activity ''a few times a week or more'' as a positive outcome and added the two together.

Dependent Variables

We constructed three dependent variables that correspond to online political information practices, civic engagement, and political participation. Online political information practices consist of the consumption and production of political information or discourse online. Civic engagement, by contrast, captures the extent to which individuals are active members of their communities in ways that are less clearly political. Finally, political participation incorporates more explicit forms of political action.

Online Political Information Practices. To determine whether respondents engage with political content online, we asked whether they: (a) use the Internet to look for, read or watch political content; (b) discuss, comment on, respond to, forward, post political content online; (c) visit/read blogs on ''politics, economics, law, policy''; and (d) comment on such blogs. We use a binary measure for online political information practices whereby this variable gets a 1 if the respondent reported engaging in any of the above four activities weekly or more often. It may be that people differ in drawing distinctions between blogs and other types of content, but this is not of concern to us since we created one universal binary measure based on all related experiences (see Table 4).

Civic Engagement. To measure civic engagement, we constructed a scale using four variables asking whether respondents had performed each of the following activities offline: (1) attended the meeting of a club or organization to which they

Table 4
Respondents' Online Political Information Practices (in Percentages)

Variable	Never	At Least a Few Times a Year, But Less than Weekly	Weekly or More Often
Looked for, read, watched content online about politics	20.0	37.6	42.4
Discussed, commented on, responded to, forwarded, posted content online about politics*	79.5	14.4	6.1
Visited/read blogs on politics, economics, law, policy	39.8	34.7	25.4
Commented on blogs about politics, economics, law, policy	84.1	7.7	8.3

Note. *For this question, the answer options were never, have done it once or twice, have done it more often.

Downloaded by [Northwestern University] at 13:39 08 February 2014

belong "every few weeks" or more; (2) ever organized a meeting or event of a club or organization to which they belong; (3) ever performed volunteer work; (4) talked to friends or family about current events and things they heard about in the news "a few times per week" or more.

The majority of respondents report doing volunteer work at least a few times a year. Just over half (54%) talk to friends or family about current events at least a few times a week. Close to half (42%) of respondents attend club or organization meetings with some regularity and a third reported organizing such meetings at least a few times a year. Overall, less than ten percent of the sample does not fall into any of these categories while close to 14% fall on the active end of the scale for all four with the rest in between. The variable for civic engagement ranges from zero to four.

Political Participation. We use three separate measures for political participation: (1) having voted in the 2008 presidential election; (2) having signed a petition; and (3) having engaged in various other types of political action. The level of voting by respondents (62.1%) mirrors the national level for young people enrolled in college (59.7%) (Godsay et al., 2010; Kirby & Kawashima-Ginsberg 2009). Over one hundred students (11.5%) reported having registered to vote, but not having voted, while the rest (26.4%) had not registered.

To assess petition signing as well as our other political action variable, we rely on a series of binary measures asking whether respondents had in the past two years ever (a) "signed a petition about a social or political issue (online or on paper)"; (b) "contacted a political official"; (c) "worked as a volunteer on a political

campaign''; (d) ''donated money to a political campaign''; or (e) been ''paid to work on a political campaign.'' Note that none of these phrasings specify whether the respondents performed any of the activities in question online or offline, resulting in a measure that captures engagement across both domains. Of these five actions, the first—signing a petition—requires much less involvement than the others and so we look at it separately. Almost two-thirds (64.9%) of respondents reported having done this at least once in the two years preceding the study.

We created a separate binary variable for more involved political action, which got a 1 if the respondent had engaged in any of the four activities (b–e on the list above) at least once in the past two years. An affirmative response to any of these five activities resulted in a 1 for the binary political action variable. Just over a fifth (21.5%) of participants had contacted a political official, 14.1% had donated money to a political campaign, just over a tenth (10.6%) had volunteered on a political campaign and 4.1% had been paid to work on a political campaign. Overall, about a third (34.0%) of respondents had engaged in at least one of these four activities.

Downloaded by [Northwestern University] at 13:39 08 February 2014

Methods of Analysis

We run a series of logistic and OLS regressions to explore what explains variation in online political information practices (binary outcome), civic engagement (interval outcome), voting in the presidential election (binary outcome) and engaging in political action (binary outcome for both signing a petition and being involved in another way). First, we look at how user background and political capital are related to the outcome variables and then we consider how Internet experiences mediate those relationships.

Results

Explaining Online Political Information Practices

Model 1 in Table 5 considers how respondents' demographic and socioeconomic background as well as political capital relate to their likelihood of visiting or interacting with political content online. Women are much less likely to do so than men and, not surprisingly, those who are more interested in politics are much more likely to do so than those who are less interested. Once we add information about Internet experiences to the model (Model 2), we find that time spent online as well as Web-use skills are both positively related to accessing or discussing political content on the Web. The addition of these variables barely changes the size of the coefficient for having an interest in politics, but reduces the coefficient and decreases the statistical significance for gender. This is in line with other work that has found that skill is a mediating factor when it comes to women's online engagement in

Downloaded by [Northwestern University] at 13:39 08 February 2014

Table 5
Logistic and OLS Regressions on Online Political Information Practices, Civic Engagement, Voting in the Presidential Election and Political Action (Standard Errors in Parentheses)

	Online Political Information Practices				Civic Engagement				Voting in the Presidential Election				Political Action I: Signing Petition				Political Action II: Other			
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8		Model 9		Model 10	
Independent Variables																				
Female	−.692***	(.17)	−.490***	(.18)	−.012	(.08)	.082	(.08)	.559***	(.15)	.635***	(.16)	.261#	(.15)	.429**	(.16)	.136	(.15)	.180	(.17)
Race/Ethnicity																				
African American	.116	(.27)	.057	(.28)	.010	(.13)	.026	(.13)	1.356***	(.31)	1.317***	(.31)	−.435#	(.24)	−.418#	(.25)	.337	(.24)	.387	(.25)
Asian American	.163	(.11)	−.102	(.11)	.207***	(.05)	.189***	(.05)	−.353***	(.09)	−.382***	(.10)	−.145	(.10)	−.159	(.10)	−.065	(.10)	−.053	(.10)
Hispanic	−.154	(.22)	−.110	(.23)	−.113	(.10)	−.054	(.10)	.371	(.20)	.451*	(.20)	−.138	(.19)	−.031	(.20)	−.136	(.20)	−.028	(.20)
Parental Education																				
Less than high school	−.080	(.37)	−.001	(.38)	−.450**	(.17)	−.329	(.17)	.304	(.35)	.334	(.37)	.182	(.33)	.374	(.34)	.704	(.33)	.805*	(.34)
High school	.005	(.27)	.102	(.28)	−.436***	(.13)	−.335**	(.13)	−.421	(.25)	−.412	(.26)	.073	(.25)	.269	(.25)	.024	(.25)	.123	(.25)
Some college	.123	(.25)	.154	(.26)	−.285*	(.12)	−.243	(.11)	−.313	(.23)	−.341	(.23)	−.147	(.22)	−.052	(.23)	−.226	(.22)	−.227	(.23)
College	.038	(.23)	.027	(.23)	−.171	(.10)	−.158	(.10)	−.292	(.21)	−.319	(.21)	−.009	(.20)	−.080	(.21)	−.257	(.20)	−.283	(.20)
Political Capital																				
Partisanship	.227	(.12)	.239	(.12)	−.030	(.06)	−.031	(.05)	.223*	(.11)	.211	(.11)	.136	(.10)	.109	(.11)	.226*	(.11)	.226*	(.11)
Interest in politics	1.554***	(.11)	1.543***	(.12)	.221***	(.04)	.119*	(.05)	.370***	(.09)	.333***	(.10)	.002	(.08)	−.059	(.10)	.486***	(.09)	.456***	(.10)
Political knowledge	.070	(.07)	.070	(.07)	.009	(.03)	.008	(.03)	.214***	(.06)	.215***	(.06)	.178***	(.06)	.167**	(.06)	.029	(.06)	.044	(.06)

(*continued*)

Downloaded by [Northwestern University] at 13:39 08 February 2014

Table 5
(*Continued*)

	Online Political Information Practices		Civic Engagement		Voting in the Presidential Election		Political Action I: Signing Petition		Political Action II: Other	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Civic engagement	—	—	—	—	.145* (.06)	.113 (.06)	.418*** (.06)	.352*** (.07)	.374*** (.06)	.341*** (.07)
Online political information practices	—	—	—	.260** (.09)	—	.137 (.18)	—	.111 (.17)	—	.072 (.18)
Internet Experiences										
Years of Internet use	—	.494 (0.56)	—	−.062 (.26)	—	−.011 (.52)	—	−.434 (.50)	—	.679 (.51)
Number of access locations	—	.244 (.63)	—	1.327*** (.28)	—	.171 (.57)	—	.481 (.55)	—	.739 (.57)
Time spent on Web weekly	—	.615** (.23)	—	.038 (.10)	—	.212 (.21)	—	−.271 (.21)	—	.092 (.21)
Social network site use	—	.127 (.11)	—	.097* (.05)	—	.023 (.10)	—	.356*** (.09)	—	.182# (.10)
Internet skill	—	.357*** (.12)	—	.132* (.05)	—	.101 (.11)	—	.433*** (.11)	—	.040 (.11)
Intercept	−3.968*** (−8.83)	−9.361*** (2.36)	1.729*** (.19)	−2.445* (1.04)	−2.03*** (.40)	−3.419 (2.12)	−1.07*** (.38)	−2.001 (2.04)	−3.233*** (.43)	−7.819*** (2.13)
N	978	966	978	966	978	966	975	963	978	966
Pseudo R2	.275	.298	.077	.132	.112	.119	.057	.087	.087	.097

Note. #$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

various realms (e.g., Correa, 2010; Hargittai & Walejko, 2008). Our results also support previous findings indicating a strong association between political interest and political information practices engagement more broadly (Zukin et al., 2006). These results lend support to H_{1a}.

Explaining Civic Engagement

Consistent with previous literature (e.g., McAdam & Brandt, 2009; Zukin et al., 2006), parental education, our proxy for socioeconomic status, exhibits a statistically significant relationship with traditional forms of civic engagement (see Model 3 in Table 5). Students who come from families where parents have lower levels of education (whether less than high school, high school or some college) are less likely to engage in civic activities than those who have at least one parent with a graduate degree. Our results also suggest that Asian American students are more likely to participate in such activities. Political interest also shows a strong relationship to this outcome.

In Model 4, we add online political information practices as well as other Internet experiences and Web-use skill to the model. Online political information practices are positively related to traditional forms of civic engagement. Additionally, autonomy of Internet use, engaging with social network sites and Internet skill also show a positive relationship to this outcome. The coefficient for Asian Americans decreased slightly, but remains statistically significant at the $p < 0.001$ level. The relationship of parental education, however, seems to be affected by some of the variables in the full model. This relationship merits further consideration as it suggests that skilled, autonomous Internet users who engage politically online and in social network site usage tend to pursue civic engagement to a greater extent than their socioeconomic peers. These results show support for H_{1b} and H_{2a}.

Explaining Voting in the Presidential Election

Consistent with the literature, results of the logistic regression looking at what explains voter participation in the 2008 presidential election (Model 5 in Table 5) suggest that female and African American students were more likely to vote while Asian American respondents show lower likelihood to have done so than Whites (Kirby & Kawashima-Ginsberg, 2009). There is no relationship with our proxy for socioeconomic status, parental education. Partisanship, political interest, and political knowledge are all positively related to voting. Our findings also highlight a positive relationship between traditional measures of civic engagement and having cast a vote in the election despite the fact that the underlying correlation between these measures was weak ($\alpha = .07$).

Interestingly, the addition of Internet experiences including online political information practices results in very moderate changes in the findings and only a slight improvement in model fit. Women are even more likely to have voted,

Downloaded by [Northwestern University] at 13:39 08 February 2014

Asian Americans are even less, but now Hispanic students exhibit a slightly higher likelihood of voter participation than Whites. Also, civic engagement is no longer related to voting. Notable is the absence of any relationship, whatsoever, between each of our Internet use variables and voting in the presidential election. That is, reading or discussing political content online, spending more time online, having been an Internet user longer, having more autonomy or higher skill, and engaging with social network sites are all unrelated to voting. Here we find no support for H_{1c} and H_{2b}, and only weak support for H_{3a}.

Explaining Political Engagement

Models 7 and 8 as well as 9 and 10 in Table 5 explore variations in non-voting forms of political participation, first for signing a petition, and then for more involved types of political action.[1] The results of Model 7 indicate that of the demographic and socioeconomic attributes, only gender and race have a relationship with signing a petition. Women were more likely to sign petitions while African American respondents were less likely to do so. As was the case for voting, political knowledge and civic engagement also have strong, positive associations with petition signing.

The addition of Internet experience measures in Model 8 reveals two additional explanatory factors of petition signing: social network site use and Internet skill. The strength and magnitude of these associations illustrate the salience of certain kinds of Internet experiences for one specific form of low-intensity political engagement.

For forms of political action that demonstrate more exceptional levels of engagement, Models 9 and 10 in Table 5 reveal some intriguing relationships in our data that deviate from our other outcome measures. Here we find no evidence of a relationship between gender, race or ethnicity and political action. However, we do see robust, positive associations between partisanship, political interest, civic engagement and political actions other than voting and petition signing. This is true despite the fact that, as with voting, our measures of civic engagement and political engagement were only weakly correlated ($\alpha = .22$ for signing a petition, $\alpha = .21$ for other political action). Adding Internet experience variables in Model 10 reveals that two additional factors help explain political action: social network site usage and low (less than high school) parental education. On the basis of these models, we find weak support for H_{1d}; no support for H_{2c}; and strong support for H_{3b}.

Taken together, the survey results indicate that several aspects of college students' Internet use help explain variations in their online political information practices, civic engagement, and some political engagement activities although not voting. Socioeconomic status, race, and gender, by contrast, matter to varying degrees for online political information practices, civic engagement, voting, and petition signing, but hardly at all for other forms of political engagement. As expected, political capital variables (i.e., political interest, knowledge and partisanship) play a significant role in explaining all of the outcomes of interest. Also, online political information practices help explain variations in civic engagement, which in turn

Downloaded by [Northwestern University] at 13:39 08 February 2014

shows a strong positive association with both voting and other forms of political engagement.

Discussion and Conclusion

The results support recent research indicating that more skillful Internet use and social network site use can facilitate social capital building and political engagement (Baumgartner & Morris, 2010; Hampton et al., 2009, 2011; Schlozman et al., 2010; Shah, Cho, Eveland, & Kwak, 2005; Smith et al., 2009; Vitak et al., 2011; Zhang et al., 2010). This finding is especially important in light of the strong associations between social capital and democratic participation. We do not find evidence that college-age youth in Chicago around the time of the 2008 U.S. presidential election demonstrated extraordinary patterns of voting behavior or online political engagement. The glaring absence of any significant associations between the Internet-related variables and voting lends strong support to previous findings that the Internet does not directly affect this particular form of political participation (Baumgartner & Morris, 2010; Bimber, 2001, 2003; Vromen, 2007; Xenos & Moy, 2007; Zhang et al., 2010).

However, our story becomes more complicated when we consider political activities that go beyond casting a ballot. Participation on social network sites and higher Web-use skills matter when it comes to explaining which people sign a petition. Furthermore, we see that social network site use also helps to explain more involved forms of political engagement such as volunteering, donation, and contacting political officials. In addition, we find support for an indirect relationship linking online political information practices through civic engagement to other forms of political engagement (including voting).

As a result, we conclude that the young people in our sample who use the Internet more actively do not appear less engaged nor less likely to participate in their communities than their counterparts. Nonetheless, more active Internet use alone is unlikely to overcome longstanding trends towards disengagement. We do find evidence that online social networking may bring about heightened political activity. We show that those individuals who possess the ability to navigate the Internet more skillfully are also more likely to engage in low-intensity forms of political action such as signing petitions. Inasmuch as Web-use skills are associated with other forms of socioeconomic status and privilege—and prior research has suggested that they are (Correa, 2010; Hargittai, 2010; van Deursen, 2010)—these results provide additional evidence that the mere presence of the Internet in society does not result in the transcendence of longstanding social inequalities. Indeed, our finding a relationship between Internet skills and petition signing arguably suggests that the facilitation of ''clicktivism''—relatively ''low quality'' forms of engagement that do not enhance the deliberative character of a democracy—would be limited to those Internet users with the ability to click more effectively in the first place (Shulman, 2009).

Downloaded by [Northwestern University] at 13:39 08 February 2014

This finding takes on added significance in light of previous work showing that young people's participation in civic, political and other sorts of engagement predicts their engagement levels later in life (McAdam & Brandt, 2009; McFarland & Thomas, 2006). Our results indicate that particular kinds of online behaviors—and in particular SNS usage—associate with enhanced democratic engagement. Future research should elaborate and test more specific mechanisms through which these online modes of participation may or may not contribute to the set of practices known to advance democratic engagement.

While our paper was able to address holes in the literature, important puzzles remain. More research needs to be done to understand what it is about social network site usage in particular that explains higher likelihood of engagement in political action and its relationship to traditional forms of civic engagement (Gil de Zúñiga & Valenzuela, 2011; Hampton, 2011; Vitak et al., 2011). It would be interesting to learn more about the processes that underlie the relationship between online political information practices and traditional forms of civic engagement and why they do not pertain similarly to political activities (Papacharissi, 2010). In addition, given the unique characteristics of the sample population included in this study and the timing of the survey administration following the 2008 presidential election, it would be important to revisit these questions in light of future data. Hopefully our findings encourage the collection of appropriately detailed, and longitudinal data on engagement and Internet experiences. Likewise, the question remains whether any of the associations we observe here may or may not also become national phenomena. Our results that Web skills and SNS usage associate strongly with civic engagement activities is particularly relevant in light of past findings that youth voluntary association participation contributes to engagement later in life (McFarland & Thomas, 2006). Future research can assess whether or not these associations are additive or causal. Furthermore, our findings that various measures of Internet use, Web-use skills, and Web autonomy help explain online political information practices, civic engagement, petition signing, and political engagement activities lend strong support to the claim that Internet use is a differentiated phenomenon whose variations matter for a wide range of social, economic, and political outcomes.

Downloaded by [Northwestern University] at 13:39 08 February 2014

Note

[1]While voting correlates with the other forms of political engagement ($\alpha = .12$ for signing a petition, $\alpha = .17$ for other political action), we chose to leave it out of Models 7–10 to preserve analytical and conceptual clarity of our results. Post-hoc testing revealed that the results reported in both of these models are robust to the inclusion of voting.

References

Ancu, M., & Cozma, R. (2009). MySpace politics: Uses and gratifications of befriending candidates. *Journal of Broadcasting & Electronic Media, 53*, 567–583.

Bachmann, I., Kaufhold, K., Lewis, S. C., & Gil de Zúñiga, H. (2010). News platform preference: Advancing the effects of age and media consumption on political participation. *International Journal of Internet Science*, *5*, 34–47.

Bakker, T. P., & de Vreese, C. H. (2011). Good news for the future? Young people, internet use, and political participation. *Communication Research, 38*, 451–470.

Baumgartner, J. C., & Morris, J. S. (2010). MyFaceTube politics: Social networking websites and political engagement of young adults. *Social Science Computer Review, 28*, 24–44.

Bennett, L. W., Wells, C., & Freelon, D. (2011). Communicating civic engagement: Contrasting models of citizenship in the youth web sphere. *Journal of Communication, 61*, 835–856.

Bennett, L. W., Wells, C., & Rank, A. (2009). Young citizens and civic learning: Two paradigms of citizenship in the digital age. *Citizenship Studies, 13*, 105–120.

Bimber, B. A. (2001). Information and political engagement in America: The search for effects of information technology at the individual level. *Political Research Quarterly, 54*, 53–67.

Bimber, B. A. (2003). *Information and American democracy: Technology in the evolution of political power*. New York, NY: Cambridge University Press.

Boulianne, S. (2009). Does internet use affect engagement? A meta-analysis of research. *Political Communication, 26*, 193–211.

Brady, H. E. (1999). Political participation. In J. P. Robinson, P. R. Shaver, & L. Wrightman (Eds.), *Measures of political attitudes* (737–801). San Diego, CA: Academic Press.

Correa, T. (2010). The participation divide among ''online experts'': Experience, skills and psychological factors as predictors of college students' web content creation. *Journal of Computer-Mediated Communication, 16*, 71–92.

Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.

Esser, F., & de Vreese, C. H. (2007). Comparing young voters' political engagement in the United States and Europe. *American Behavioral Scientist, 50*, 1195–1213.

Fisher, D. R. (2006). *Acitvism inc.: How the outsourcing of grassroots campaigns in strangling progressive politics in America*. Stanford, CA: Stanford University Press.

Fisher, D. R. (2008). From the bottom-up: Using the internet to mobilize campaign participation. *Publius Project*. Retrieved from http://publius.cc/2008/12/09/from-the-bottom-up-using-the-internet-to-mobilize-campaign-participation

Foot, K. A. & Schneider, S. M. (2002). Online action in campaign 2000: An exploratory analysis of the U.S. political web sphere. *Journal of Broadcasting & Electronic Media, 46*, 222–244.

Gil de Zúñiga, H., & Valenzuela, S. (2011). The mediating path to a stronger citizenship: online and offline networks, weak ties and civic engagement. *Communication Research, 38*, 397–421.

Godsay, S., Nover, A., & Kirby, E. (2010). *The minority youth vote in the 2008 presidential election*. Tufts University: The Center for Information & Research on Civic Learning & Engagement. Retrieved from http://www.civicyouth.org/featured-the-minority-youth-vote-in-the-2008-presidential-election/

Hampton, K. (2011). Comparing bonding and bridging ties for democratic engagement: Everyday use of communication technologies within social networks for civic and civil behaviors. *Information, Communication & Society, 14*, 510–528.

Hampton, K., Lee, C., & Her, E. J. (2011). How new media affords network diversity: Direct and mediated access to social capital through participation in local social settings. *New Media & Society, 13*, 1031–1049.

Hampton, K., Sessions, L., Her, E. J., & Rainie, L. (2009). *Social isolation and new technology: How the internet and mobile phones impact Americans' social networks*. Pew Research Center: Pew Internet & American Life Project. Retrieved from http://pewinternet.org/Reports/2009/18--Social-Isolation-and-New-Technology.aspx

Hargittai, E. (2010). Digital na(t)ives? Variation in internet skills and uses among members of the ''Net Generation,'' *Sociological Inquiry, 80*, 92–113.

Hargittai, E. & Hinnant, A. (2008). Digital inequality: Differences in young adults' use of the internet. *Communication Research*, *35*, 602–621.

Downloaded by [Northwestern University] at 13:39 08 February 2014

Hargittai, E. & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication and Society, 11*, 239–256.
Kenski, K., & Stroud, N. J. (2006). Connections between internet use and political efficacy, knowledge, and participation. *Journal of Broadcasting & Electronic Media*, *50*, 173–192.
Kim, K. S., & Kim, Y-C. (2007). New and old media uses and political engagement among Korean adolescents. *Asian Journal of Communication*, *17*, 342–361.
Kirby, E. H., & Kawashima-Ginsberg, K. (2009). *The youth vote in 2008*. Tufts University: The Center for Information & Research on Civic Learning & Engagement. Retrieved from http://www.civicyouth.org/quick-facts/youth-voting/
Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010). *Social media & mobile internet use among teens and young adults*. Pew Research Center: Pew Internet & American Life Project. Retrieved from http://pewinternet.org/Reports/2010/Social-Media-and-Young-Adults.aspx
Loader, B. (2007). Introduction: Young citizens in the digital age: Disaffected or displaced? In B. Loader (Ed.), *Young citizens in the digital age: Political engagement, young people and new media* (pp. 1–18). New York, NY: Routledge.
McAdam, D., & Brandt, C. (2009). Assessing the effects of voluntary youth service: The case of teach for America. *Social Forces, 88*, 945–970.
McFarland, D. A., & Thomas, R. J. (2006). Bowling young: How youth voluntary associations influence adult political participation. *American Sociological Review, 71*, 401–425.
McPherson, M., Smith-Lovin, L., & Brashears, M. E. (2006). Social isolation in America: Changes in core discussion networks over two decades. *American Sociological Review, 71*, 353–375.
McPherson, M., Smith-Lovin, L., & Brashears, M. E. (2009). Models and marginals: Using survey evidence to study social networks. *American Sociological Review*, *74*, 670–681.
Melber, A. (2008). Obama's web-savvy voter plan. *The Nation.* Retrieved from http://www.thenation.com/article/obamas-web-savvy-voter-plan
Miller, C. C. (2008). How Obama's internet campaign changed politics. *Bits: Business, Innovation Technology Society*. Retrieved from http://bits.blogs.nytimes.com/2008/11/07/how-obamas-internet-campaign-changed-politics/
Papacharissi, Z. (2010). *A private sphere: Democracy in a digital age*. Malden, MA: Polity Press.
Pasek, J., more, e., & Romer, D. (2009). Realizing the social internet? Online social networking meets offline civic engagement. *Journal of Information Technology & Politics, 6*(3–4), 197–215.
Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York, NY: Simon and Schuster.
Ruffini, P. (2008). Using the internet to win the election. *The Next Right*. Retrieved from http://www.thenextright.com/patrick-ruffini/using-the-internet-to-win-the-election
Schlozman, K. L., Verba, S., & Brady, H. E. (2010). Weapon of the strong? Participatory inequality and the internet. *Perspectives on Politics, 8*, 487–509.
Shah, D. V., Cho, J., Eveland, W. P., & Kwak, N. (2005). Information and expression in a digital age: Modeling internet effects on civic participation. *Communication Research*, *32*, 531–565.
Shulman, S. (2009). The case against mass e-mails: Perverse incentives and low quality public participation in U.S. federal rulemaking. *Policy & Internet, 1*. Retrieved from http://www.psocommons.org/policyandinternet/vol1/iss1/art2
Skocpol, T. (2003). *Diminished democracy: From membership to management in American civic life*. Norman, OK: University of Oklahoma Press.
Smith, A. (2009). *The internet's role in campaign 2008*. The Pew Research Center for the People and the Press: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/Reports/2009/6--The-Internets-Role-in-Campaign-2008.aspx
Smith, A., Lehman Schlozman, K., Verba, S., & Brady, H. E. (2009). *The internet and civic engagement*. The Pew Research Center for the People and the Press: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/Reports/2009/15--The-Internet-and-Civic-Engagement.aspx

Downloaded by [Northwestern University] at 13:39 08 February 2014

Stirland, S. L. (2008). Propelled by internet, Barack Obama wins presidency. *Threat Level*. Retrieved from http://www.wired.com/threatlevel/2008/11/propelled-by-in/

van Deursen, A. J. A. M. (2010). Internet skills: Vital assets in the information society. Dissertation. Enschede, The Netherlands: University of Twente.

Verba, S., Lehman Schlozman, K., & Brady, H. E. (1995). *Voice and equality: Civic voluntarism in American politics*. Cambridge, MA: Harvard University Press.

Vitak, J., Zube, P., Smock, A., Carr, C. T., Ellison, N., & Lampe, C. (2011). It's complicated: Facebook users' political participation in the 2008 election. *Cyberpsychology, Behavior, and Social Networking*, *14*(3), 107–114.

Vromen, A. (2007). Australian young people's participatory practices and internet use. In B. Loader (Ed.), *Young citizens in the digital age: Political engagement, young people and new media* (pp. 97–113). New York, NY: Routledge.

Wasserman, I. M., & Richmond-Abbott, M. (2005). Gender and the internet: Causes of variation in access, level, and scope of use. *Social Science Quarterly, 86*, 252–270.

Weaver Lariscy, R., Tinkham S. F., & Sweetser, K. D. (2011). Kids these days: Examining differences in political uses and gratifications, internet political participation, political information efficacy, and cynicism on the basis of age. *American Behavioral Scientist*, *55*, 749–764.

Xenos, M. A., & Kyoung, K. (2008). Rocking the vote and more: An experimental study of the impact of youth political portals. *Journal of Information Technology & Politics*, *5*, 175–189.

Xenos, M. A., & Moy, P. (2007). Direct and differential effects of the internet on political and civic engagement. *Journal of Communication*, *57*, 704–718.

Zhang, W., Johnson, T. J., Seltzer, T., & Bichard. S. L. (2010). The revolution will be networked: The influence of social networking sites on political attitudes and behavior. *Social Science Computer Review, 28*, 75–92.

Zukin, C., Keeter, S., Andolina, M., Jenkins, K., & Delli Carpini, M. X. (2006). *A new engagement? Political participation, civic life, and the changing American citizen*. New York, NY: Oxford University Press.

Downloaded by [Northwestern University] at 13:39 08 February 2014

OPEN ACCESS Freely available online

PLOS ONE

The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation

Benjamin Mako Hill[1,3*☯], **Aaron Shaw**[2,3☯]

1 Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **2** Department of Communication Studies, Northwestern University, Evanston, Illinois, United States of America, **3** Berkman Center for Internet & Society, Harvard University, Cambridge, Massachusetts, United States of America

Abstract

Opt-in surveys are the most widespread method used to study participation in online communities, but produce biased results in the absence of adjustments for non-response. A 2008 survey conducted by the Wikimedia Foundation and United Nations University at Maastricht is the source of a frequently cited statistic that less than 13% of Wikipedia contributors are female. However, the same study suggested that only 39.9% of Wikipedia readers in the US were female – a finding contradicted by a representative survey of American adults by the Pew Research Center conducted less than two months later. Combining these two datasets through an application and extension of a propensity score estimation technique used to model survey non-response bias, we construct revised estimates, contingent on explicit assumptions, for several of the Wikimedia Foundation and United Nations University at Maastricht claims about Wikipedia editors. We estimate that the proportion of female US adult editors was 27.5% higher than the original study reported (22.7%, versus 17.8%), and that the total proportion of female editors was 26.8% higher (16.1%, versus 12.7%).

Citation: Hill BM, Shaw A (2013) The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation. PLoS ONE 8(6): e65782. doi:10.1371/journal.pone.0065782

Editor: Angel Sánchez, Universidad Carlos III de Madrid, Spain

Received December 16, 2012; **Accepted** April 26, 2013; **Published** June 26, 2013

Copyright: © 2013 Hill, Shaw. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work has funded by the MIT Sloan School of Management, Northwestern University, the Berkman Center for Internet and Society and the Ford Visionary Leadership Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: BMH serves on the advisory board of the Wikimedia Foundation as an unpaid advisor. The Wikimedia Foundation is one of the co-creators of the dataset which the authors analyze in this paper. The authors were not involved, in any way, in the creation or collection of the survey and do not believe BMH's role as advisor had any effect on the nature of this research. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: mako@mit.edu

☯ These authors contributed equally to this work.

Introduction

Accurately describing the demographics of individuals who contribute to Wikipedia, the largest volunteer-written, free knowledge resource on the Internet, as well as other "peer production" communities [1], presents challenges to traditional sampling and survey methods [2]. The easiest means of recruiting subjects for such research is through the distribution of "opt-in" survey instruments that ask project contributors to voluntarily respond to public notices. However, the self-selection processes underpinning this sampling technique tend to produce biased and unreliable data.

One of the most well-known examples of such an opt-in web survey occurred between October 29 and November 3, 2008, when researchers at the Wikimedia Foundation (WMF) and the United Nations University at Maastricht (UNU-MERIT) used a notice on each Wikipedia web page to administer an opt-in survey to 179,192 Wikipedia users and contributors [3]. The WMF/UNU-MERIT survey's claim that less than 13% of Wikipedia contributors are female was widely reported in the press and prompted the Wikimedia Foundation to launch an initiative to raise the proportion of female contributors to 25% [4]. The WMF/UNU-MERIT survey relied on a non-random sample of self-selected participants. Self-selection is common to other surveys of Wikipedia contributors, which have shown similar results [5].

The response rate to the WMF/UNU-MERIT survey was very low. Using ComScore estimates of viewership in October 2008 [6], and assuming an even distribution across the month, respondents represent approximately 0.4% of the 45 million unique visitors to Wikipedia during the period in which the survey was administered. Although editors were overrepresented in the survey (33.2% of respondents described themselves as either contributors or ex-contributors), they represented only 6.8% of individuals who had ever contributed to Wikipedia at that point in time.

There are also concerns, well-documented in survey research, that self-selected samples may not be representative of the population of interest because certain sub-groups of respondents may be more likely to participate in the survey than others [2,7]. For example, Russian first-language readers represented 24.3% of the WMF/UNU-MERIT survey's respondents (the single largest language group) although the Russian language Wikipedia reflects only 2.5% of Wikipedia's global readership – a fact that WMF/UNU-MERIT researchers were aware of but unable to explain [3]. More systematic forms of bias are also a concern. For example, previous work has shown that women are less likely than men to respond to opt-in Internet surveys that focus on a topic in which women tend to have less interest [7].

No statistical process exists that would allow us to recover unbiased estimates of the true population values for Wikipedia editors on the basis of the WMF/UNU-MERIT survey alone or

any of the previous opt-in surveys conducted on Wikipedia. However, the fact that the WMF/UNU-MERIT survey includes data on Wikipedia *readers* allows us to take advantage of demographic data from a nationally representative phone survey of US adults conducted by the Pew Research Center's Internet & American Life Project [8] less than two months after the WMF/UNU-MERIT survey. While Pew estimated that Wikipedia's readership in the US was evenly split between males and females, only 39.9% of WMF/UNU-MERIT viewership sample was female.

Drawing on recent research in online survey methods [9,10], we combine the data from the Pew and WMF/UNU-MERIT surveys and construct a logistic "propensity score" [11] model to estimate the likelihood that a US adult Wikipedia reader participated in the WMF/UNU-MERIT survey. Using this model, we then calculate a correction for WMF/UNU-MERIT's estimation of the US adult Wikipedia editor population. Finally, we extend this correction to the population of Wikipedia editors as a whole, offering adjusted estimates for all of the shared covariates based on assumptions we make about consistent selection response bias in the WMF/UNU-MERIT survey instrument.

Methods

The procedures used in this research were reviewed by the Institutional Review Boards at the Massachusetts Institute of Technology (MIT) and Northwestern University. MIT determined the project to be exempt from review and Northwestern determined that it did not qualify as human subjects research.

We estimate bias by comparing the results from the WMF/UNU-MERIT sample [3] (an anonymized copy of the WMF/UNU-MERIT data is available via email from the UNU-MERIT researchers) with data on Wikipedia readership from the Pew Research Center's Internet & American Life Project [12]. The Pew data was gathered in a nationally representative phone survey of American adults conducted in December, 2008, several weeks after the WMF/UNU-MERIT survey. Details of the Pew survey methodology are available from Pew [8,13]. The survey asked respondents who either use the Internet or email whether they "use the Internet to look for information on Wikipedia." The phrasing of this question does not perfectly match the language of any of the WMF/UNU-MERIT survey questions but is comparable to that survey's "reader" category of Wikipedia users.

Based on the overlapping coverage of Wikipedia readers in the two surveys, we apply a propensity score adjustment technique developed to measure and correct bias in opt-in web surveys [9,10,14,15]. Propensity scores were originally used to model the likelihood of non-random selection into observational studies [11]. We adopt a propensity score procedure that estimates non-response bias using a representative "Reference Survey" population [9]. This procedure, described in detail by Valliant and Dever [9] has two steps: (1) using the opt-in and reference survey datasets to model the propensity of individuals in the universe of potential respondents to opt in to the survey; (2) using the results of the propensity score model to weight the opt-in survey data to generate adjusted estimates of population-level variables for which reference survey data does not exist.

Studies comparing the results of propensity-adjusted volunteer web surveys with both the results of other adjustment techniques as well as known population values have found evidence of important limitations. Some previous findings indicate that most propensity score adjustments improve the precision and bias of volunteer survey data, but diverge on the precise extent of the bias reduction [10,14]. The same research also indicates that some types of questions may be more or less resistant to correction through the propensity score adjustment approach. Related work has suggested that an association between the probability of volunteering and any of the other analysis variables in the survey can bias the results [9].

In line with Valliant and Dever's first step, we create a single dataset by combining data from the Pew study (the subset of the representative sample of US adults who had looked for information on Wikipedia) with the subset of WMF/UNU-MERIT respondents who indicated that they were both 18 years of age or older and US residents (approximately 7500 individuals). Included in the combined dataset are a series of covariates collected in both surveys (age, gender, education level, immigrant status, marital status, parental status, student status) as well as the original Pew survey weights. Following Valliant and Dever, we reweight the subset of the Pew dataset so that respondents represent the estimated population of US Wikipedia readers. We do so by multiplying the original Pew survey weights by 102,138 (equivalent to 230,118,00, the estimated US population in 2008 [16], divided by 2,253, the size of the Pew sample). Applying these weights, we then use logistic regression to model the likelihood that a member of the reference population of US adult Wikipedia readers opted to participate in the WMF/UNU-MERIT survey.

Our logistic regression model estimates the probability that a respondent in the combined dataset of US Wikipedia readers opted into the WMF/UNU-MERIT survey using a set of shared covariates. The formal model is as follows:

$$\mathrm{P}[i \in \mathrm{WMF/UNU}] = \beta + \beta age_i + \beta female_i + \beta married_i + \beta children_i + \beta immigrant_i + \beta student_i + \boldsymbol{\beta education_i} + \varepsilon$$

With the exception of *age*, all measures are coded as dummy variables. To combine datasets, *education* was coded as a vector of dummies that reflect whether the respondent's highest level of education was high school, college, or graduate school. Respondents without a high school diploma are the omitted category in the fitted model. Parameter estimates from the fitted regression model are shown in Table 1.

Valliant and Dever's second step suggests that the reciprocal of the predicted probabilities from the model estimated in Step 1 can act as a set of weights to recover unbiased estimates of observed covariates for the subset of the WMF/UNU-MERIT opt-in survey described by the reference survey (i.e., US adult readers of Wikipedia). To correct for bias in the WMF/UNU-MERIT editor subsample, we use the fitted logistic regression model created in Step 1, and described in Table 1, to estimate weights for editor respondents (a) in the US and (b) globally, by taking the reciprocal of the probability predicted by our fitted model for *every* individual in the WMF/UNU-MERIT dataset.

This process of estimating the demographic characteristics and attributes of the editor population entails an assumption that is empirically untestable. By applying the same weights from the original propensity model to the subsamples of editors, we are assuming that the covariance structure driving response in these populations is identical to the samples of US adult Wikipedia readers. We discuss this limitation below.

Results

After applying weights based on the propensity score model reported in Table 1, we estimate that females, married people, and individuals with children were underrepresented in the WMF/

Table 1. Logistic Regression Model of Participation in the WMF/UNU-MERIT Survey.

	Model 1
(Intercept)	−11.02*
	(0.30)
age	−0.04*
	(0.00)
female	−0.31*
	(0.10)
married	0.12
	(0.12)
children	−0.30*
	(0.11)
immigrant	0.16
	(0.16)
student	−0.07
	(0.14)
$\text{education}_{\text{secondary}}$	1.39*
	(0.28)
$\text{education}_{\text{college}}$	1.08*
	(0.27)
$\text{education}_{\text{graduate}}$	1.45*
	(0.29)
N	7771
AIC	20.19
BIC	298.52
$\log L$	29.90

Standard errors in parentheses.
*indicates significance at $p < 0.05$.
Weighted logistic regression model estimating the likelihood of a US adult Wikipedia reader responding to the WMF/UNU-MERIT survey along a set of covariates shared between the WMF/UNU-MERIT and Pew surveys. *age* is given in years and all other variables are dummy variables. Note that *education* is given a series of dummies with "less than high school diploma" as the omitted category.
doi:10.1371/journal.pone.0065782.t001

UNU-MERIT sample while immigrants and students were overrepresented. Our adjusted estimates suggest that the proportion of US adult female editors was 27.5% greater than the WMF/UNU-MERIT estimate (22.7%, versus 17.8%). Applying the same propensity score model to generate weights for the full sample of WMF/UNU-MERIT respondents, we estimate that the total proportion of female editors was 26.8% greater than the WMF/UNU-MERIT estimate (16.1% versus 12.7%). Adjusted estimates for other demographic variables are shown in Table 2.

Limitations

These adjusted estimates are limited by the precision of our propensity score estimates. They are also contingent on several assumptions. One assumption central to the propensity score technique and untestable with these data is that any selection bias in the WMF/UNU-MERIT survey occurred along the observed covariates shared between it and the Pew dataset. A second, untestable assumption is that selection pressures along observed covariates affecting the propensity of US adult Wikipedia *readers* to volunteer for the WMF/UNU-MERIT survey are identical to the selection pressures affecting the propensity of US adult Wikipedia *editors* (and, in our global estimates, all Wikipedia editors) to volunteer for the WMF/UNU-MERIT survey. We do not assume that the demographics of Wikipedia editors and readers are the same. Our key assumption is that the process of opting-in to the WMF/UNU-MERIT survey is biased in ways that under- and overrepresent respondents consistently across the sample.

There are reasons to suspect that this second assumption of identical selection pressures between readers and editors may have been violated, particularly in the global sample of editors. For example, there is evidence that contributors to online platforms like Wikipedia have different demographic profiles than those who merely use the Internet for information seeking [17]. Additionally, the disproportionate response rate of Wikipedia editors from Russia in the WMF/UNU-MERIT survey suggests the presence of sources of bias that cannot be estimated through the propensity score method using a reference population of US adults alone. Finally, the response rate for editors, while very low, is still higher than the response rate for readers.

For our adjusted estimates of gender inequality, the most problematic violation of the key assumption would occur if female editors responded to the opt-in survey at a higher rate relative to male editors than the relative rate at which female readers responded. In this case, the raw WMF/UNU-MERIT results would represent overestimates of female editors. One potential cause of such an outcome could be the fact that female editors face systematic barriers to participation in Wikipedia [18,19]. As a result of these barriers, it might be the case that active female editors would be more motivated to contribute than active male editors and that this increased motivation to edit might also translate into greater relative motivation to respond to the opt-in survey. We believe that this threat is mitigated by the fact that the WMF/UNU-MERIT definition of editors included individuals who had edited Wikipedia in the past but had then ceased to do so. Indeed, previous evaluations of the WMF/UNU-MERIT survey have studied these barriers by "deterred" female contributors [18]. Because these women are included in our sample of editors, our findings should not be driven by higher female attrition as long as these former editors do not also become less likely to answer the survey. Of course, we cannot fully reject this threat.

On the other hand, there are reasons to be confident that our assumption holds and that bias will follow similar patterns across the WMF/UNU-MERIT sample. The WMF/UNU-MERIT survey presented a single instrument to all Wikipedia visitors without distinguishing between readers and editors. Particularities of the survey instrument (e.g., the length, framing, wording, presentation, etc.) may have appealed to some demographic groups over others. Additionally, there is a extensive history of results in psychology that suggests that there are consistent gender differences in self-confidence – both in general and, especially, in regards to attitudes toward computers [20–22]. These psychological gender differences may drive underrepresentation of women in the context of opt-in surveys. Although it does not speak to differences between the editor and contributor populations, research by Chang and Krosnick has shown that women are less likely than men to respond to opt-in Internet surveys and that this bias is particularly present when the survey focuses on a topic in which women tend to have less interest [7]. Although we cannot reject all potential threats, we believe that, on balance, there is a theoretical justification for believing our results represent improvements over than the uncorrected WMF/UNU-MERIT estimates.

Table 2. Unadjusted and Adjusted Covariate Distributions for Survey Results.

Variable	Readers US (Pew)	Readers US (UNU)	Editors US (UNU)	Editors US Adj.	Editors (UNU)	Editors Adj.
female	49.0	39.9	17.8	22.7	12.7	16.1
married	60.1	44.1	30.9	36.3	33.2	38.4
children	36.0	29.4	16.4	27.6	14.4	25.3
immigrant	10.1	14.4	12.1	9.8	8.2	7.4
student	17.7	29.9	46.0	38.5	47.7	40.3

Comparison of results for the proportion of Wikipedia readers and editors from the nationally representative Pew survey and the WMF/UNU-MERIT survey (UNU) for a series of dichotomous variables in both surveys. Adjusted numbers for editors assume that response bias for editors is identical to observed response bias for readers and, in the rightmost column, that bias is stable for editors outside the United States.
doi:10.1371/journal.pone.0065782.t002

Discussion

Opt-in surveys of online communities like the WMF/UNU-MERIT study are widespread and persistent despite their well-known limitations. Using a nationally representative sample of Wikipedia readers, we apply the method of propensity score adjustment described by Valliant and Dever to estimate the survey-response bias for the subpopulation of US adult Wikipedia readers. We then extend the propensity score adjustment method by using results of the model to create new estimates for the demographic characteristics of other subpopulations of Wikipedia users. Contingent on explicit assumptions and on the precision of the propensity score estimates, we suggest that this extension of propensity score adjustment techniques represents a novel method of characterizing self-selection bias in surveys of contributors to online communities.

In the case of Wikipedia and the WMF/UNU-MERIT survey, we find evidence that the proportions of editors who are female, married, or parents, have been underestimated, while the proportions of immigrants and students have been overestimated. We find support for the substantive finding that female editors are underrepresented – but less than previous surveys have suggested. Although the basic takeaways in regards to the underrepresentation of women in the WMF/UNU-MERIT survey remain intact, certain policy decisions, like the Wikimedia Foundation's strategic goal to increase female editorship to 25%, may want to be raised in light of these adjusted estimates. In addition, future surveys of Wikipedia readers and editors should attempt to address the underlying sources of bias identified by this study.

Because the WMF/UNU-MERIT survey was presented to all visitors to Wikipedia, we were able to use Pew's data on Wikipedia readership to estimate opt-in response bias of the instrument. All other Wikipedia surveys that we are aware of have surveyed only editors. Unfortunately, this means that the propensity score adjustment techniques we adapt here cannot be applied to subsequent surveys despite the fact that Pew has continued to produce representative samples of US adult Wikipedia readers (see http://pewinternet.org/Data-Tools/Explore-Survey-Questions/Roper-Center.aspx? k = wikipedia Accessed May 7, 2013). We would urge the administrators of future Wikipedia editor surveys to consider surveying at least a random sample of Wikipedia readers with the same instrument. Doing so would allow these surveys to be adjusted using the method described in this paper.

Although we urge caution, we believe that the assumptions underlying our approach can be tested and, once refined, applied to other web communities. While high-quality, nationally representative data from sources like Pew is unlikely to exist for most online communities, reliable demographic data for many popular websites is available through market research firms like QuantCast and ComScore. Any web community running a survey of its contributor base can also survey its readership using the same instrument. By targeting *all* of a website's visitors with an opt-in survey, demographic data from market research and advertising firms can play a similar role to the Pew data in our analysis to generate comparable estimates of the impact of self-selection.

Supporting Information

File S1 Includes the R source code used in our analysis. The source code uses a publicly available dataset from the Pew Research Center's Internet & American Life Project [8] and an anonymized version of the WMF/UNU-MERIT survey available upon request from the UNU-MERIT researchers [3].
(R)

Acknowledgments

The authors acknowledge the support of the Berkman Center for Internet & Society, and in particular the center staff. Members of the Berkman Center Cooperation Research Group as well as three anonymous reviewers provided valuable feedback on earlier versions of this piece. We also thank the Wikimedia Foundation, UNU-MERIT, and the Pew Research Center's Project on the Internet & American Life for publicly sharing the data from their surveys.

Author Contributions

Conceived and designed the experiments: BMH AS. Performed the experiments: BMH AS. Analyzed the data: BMH AS. Contributed reagents/materials/analysis tools: BMH AS. Wrote the paper: BMH AS.

References

1. Benkler Y (2006) The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press.
2. Dillman DA, Smyth JD, Christian LM (2009) Internet, mail, and mixed-mode surveys : the tailored design method. Hoboken, N.J.: Wiley & Sons, 3rd edition.
3. Glott R, Ghosh R, Schmidt P (2010) Wikipedia survey. Technical report, UNU-MERIT, Maastricht, Netherlands. Available: http://wikipediasurvey.org/. Accessed 2011 April 4.
4. Cohen N (2011) Wikipedia ponders its gender-skewed contributions. The New York Times.
5. Wikimedia Foundation (2011) Editor survey 2011. Technical report, San Francisco, CA. Available: https://meta.wikimedia.org/wiki/Editor_Survey_2011. Accessed 2012 Dec 4.
6. ComScore (2008). My matrix: Wikimedia foundation sites. Available: http://upload.wikimedia.org/wikipedia/meta/b/ba/ComScore_trend_data_on_WMF_Sites,_as_of_Nov_08.pdf. Accessed 2013 May 7.

7. Chang L, Krosnick JA (2009) National surveys via rdd telephone interviewing versus the internet comparing sample representativeness and response quality. Public Opinion Quarterly 73: 641–678.
8. Fox S, Jones S (2008) The social life of health information. Technical report, Pew Internet & American Life Project. Available: http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information.aspx. Accessed 2011 April 4.
9. Valliant R, Dever JA (2011) Estimating propensity adjustments for volunteer web surveys. Sociological Methods & Research 40: 105–137.
10. Schonlau M, Soest Av, Kapteyn A, Couper M (2009) Selection bias in web surveys and the use of propensity scores. Sociological Methods & Research 37: 291–318.
11. Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70: 41–55.
12. Pew Research Center (2008). December 2008 - health. Available: http://www.pewinternet.org/Shared-Content/Data-Sets/2008/December-2008-Health.aspx. Accessed 2012 May 7.
13. Princeton Survey Research Associates International (2008). Fall tracking 2008 questionnaire. Available: http://www.pewinternet.org/~/media/Files/Data%20Sets/2008/December_2008_Health_Questionnaire.doc. Accessed 2013 March 31.
14. Schonlau M, Zapert K, Simon LP, Sanstad KH, Marcus SM, et al. (2004) A comparison between responses from a propensity-weighted web survey and an identical RDD survey. Social Science Computer Review 22: 128–138.
15. Lee S, Valliant R (2009) Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. Sociological Methods & Research 37: 319–343.
16. US Census Bureau (2010) Statistical abstract of the united states. Technical report, Washington, DC. Available: http://www.census.gov/compendia/statab/2010/2010edition.html. Accessed 2012 Dec 4.
17. Hargittai E, Walejko G (2008) The participation divide: Content creation and sharing in the digital age. Information, Communication and Society 11: 239–256.
18. Collier B, Bear J (2012) Conflict, criticism, or confidence: an empirical examination of the gender gap in wikipedia contributions. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. New York, NY, USA: ACM, CSCW '12, pp. 383–392. doi: 10.1145/2145204.2145265.
19. Morgan JT, Bouterse S, Walls H, Stierch S (2013) Tea and sympathy: crafting positive new user experiences on wikipedia. In: Proceedings of the 2013 conference on Computer supported cooperative work. New York, NY, USA: ACM, CSCW '13, p. 839–848. doi:10.1145/2441776. 2441871.
20. Busch T (1995) Gender differences in self-efficacy and attitudes toward computers. Journal of Educational Computing Research 12: 147–158.
21. Lundeberg MA, Fox PW, Punccohar J (1994) Highly confident but wrong: Gender differences and similarities in confidence judgments. Journal of Educational Psychology 86: 114–121.
22. Schumacher P, Morahan-Martin J (2001) Gender, internet and computer attitudes and experiences. Computers in Human Behavior 17: 95–110.

The Future of Crowd Work

Aniket Kittur[1], Jeffrey V. Nickerson[2], Michael S. Bernstein[3], Elizabeth M. Gerber[4], Aaron Shaw[4], John Zimmerman[1], Matthew Lease[5], and John J. Horton[6]

[1]Carnegie Mellon University, [2]Stevens Institute of Technology, [3]Stanford University, [4]Northwestern University, [5]University of Texas at Austin, [6]oDesk
{nkittur, johnz}@cs.cmu.edu, jnickerson@stevens.edu, msb@cs.stanford.edu, {egerber, aaronshaw}@northwestern.edu, ml@ischool.utexas.edu, john_horton@odesk.com

ABSTRACT
Paid crowd work offers remarkable opportunities for improving productivity, social mobility, and the global economy by engaging a geographically distributed workforce to complete complex tasks on demand and at scale. But it is also possible that crowd work will fail to achieve its potential, focusing on assembly-line piecework. Can we foresee a future crowd workplace in which we would want our children to participate? This paper frames the major challenges that stand in the way of this goal. Drawing on theory from organizational behavior and distributed computing, as well as direct feedback from workers, we outline a framework that will enable crowd work that is complex, collaborative, and sustainable. The framework lays out research challenges in twelve major areas: workflow, task assignment, hierarchy, real-time response, synchronous collaboration, quality control, crowds guiding AIs, AIs guiding crowds, platforms, job design, reputation, and motivation.

Author Keywords
Crowdsourcing; crowd work; organization design; research vision

ACM Classification Keywords
H.5; K.4.3

INTRODUCTION
Crowdsourcing rapidly mobilizes large numbers of people to accomplish tasks on a global scale. For example, volunteer-based collective projects such as Wikipedia owe their success and longevity to the ongoing efforts of thousands of individual contributors around the world.

Complementing volunteer-based crowdsourcing, a paying *crowd work* industry is now quickly growing in scope and ambition. Crowd work today spans a wide range of skill and pay levels, with commercial vendors providing access to a range of workers and focused support for various task. For example, anyone with access to the Internet can perform *micro*-tasks on the order of seconds using platforms such as Amazon's Mechanical Turk, while more skilled workers can complete multi-hour tasks on professional online marketplaces such as oDesk or work for months to solve R&D challenges on open innovation platforms (e.g. Innocentive). Incentives and work structures also vary tremendously, ranging from crowdsourcing contests awarding prizes to winners (e.g. programming tasks on Topcoder) to micro-labor platforms that pay workers per task.

While not all jobs are amenable to being sent down a wire, there are portions of almost any job that can be performed by the crowd [63]. We foresee a world in which crowd work continues to expand, unlocking an incredible number of opportunities for careers and skilled work in online marketplaces. However, we also foresee a serious risk that crowd work will fall into an intellectual framing focused on low-cost results and exploitative labor. With diminished visibility and communication channels vis-a-vis traditional workplaces, workers may be treated as exchangeable and untrustworthy, having low or static skill sets and strong motivations to shirk. Workers may become equally cynical, having fewer bonds, enforceable contracts, and power than with traditional workplaces [130]. Such concerns may grow ever sharper unless this trajectory is somehow altered.

This work originally emerged from the question: "*Can we foresee a future crowd workplace in which we would want our children to participate?*" We suggest that this question has a number of attractive properties as a banner around which to rally research, as well as serving as an anchor to ground speculation. It is simple enough to convey concisely, involving an evaluative component that everyone, with or without children, can make. We intentionally keep the "we" ambiguous so that readers with different values and cultural contexts may "try on" the question, providing a conceptual lens easily refocused, and encourage discussion regarding the challenges this question poses to deciding what is valuable and pride-worthy work. What new services, systems or features are needed for a future of crowd work that the reader would be proud to see

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CSCW '13, February 23–27, 2013, San Antonio, Texas, USA.
Copyright 2013 ACM 978-1-4503-1331-5/13/02...$15.00.

his or her children take on as their livelihood? Is this a desirable path for the next generation?

Looking toward that positive future, in this paper we contribute an analytic framework for research in crowd work. We lay out a vision for a possible future for crowd work that entails:

- **Worker considerations**, such as motivation, feedback, and pay. These may be addressed by mechanisms to maintain reputation, provide better interaction with requesters, and increase motivation.
- **Requester considerations,** such as coordination, task decomposition, and quality control. These may be addressed through workflow mechanisms including electronically mediated collaboration.

Our analytic framework is organized around a multidisciplinary survey of the literature that speaks to these challenges and helps to envision a positive future. We also include specific comments from crowd workers we surveyed in order to elicit their thoughts and suggestions. We translate our findings into a set of pragmatic design considerations that we believe are crucial in guiding design and motivating research in this field. We are following in the tradition of a set of research contributions which delineated design principles as part of a call for action [69,98,102,114,123].

Crowd Work

A variety of terminology is currently used in regard to crowds, e.g. crowdsourcing, collective intelligence, human computation, serious games, peer production, and citizen science [2,12,90,105]. We focus this paper on *paid, online crowd work*, which we define here as the performance of tasks online by distributed *crowd workers* who are financially compensated by *requesters* (individuals, groups, or organizations). In this sense, crowd work is a *socio-technical work system* constituted through a set of relationships that connect organizations, individuals, technologies and work activities [142]. Online crowd work takes place in marketplaces that allow requesters to seek workers and support workers in finding work. For this paper, we surveyed a number of contemporary, popular crowd work platforms. These platforms include general-purpose marketplaces (e.g., Mechanical Turk, oDesk, Freelancer, Crowdflower, MobileWorks, ManPower) as well as markets for specific expertise (e.g., TopCoder, uTest, 99Designs). While these platforms are intended for legitimate tasks, these and other platforms are sometimes appropriated for illegal or nefarious purposes (e.g., gold farming, CAPTCHA solving, and crowdturfing) [35].

Through our definition, we necessarily omit a wide range of voluntary crowd work, such as wikis [22], games with a purpose [2], captchas [3], and citizen science [31,106,122]. Much has already been written about these systems (e.g., [12,108]). However, not only is paid work the cornerstone of our existing economy and labor markets, but even volunteers typically engage in some form of paying work in order to sustain themselves. Moreover, we believe there will always be some forms of work needed by society which are not amenable to gamification and volunteering, and for which demand will outstrip supply via unpaid channels. As such, we are interested in developing a future of paid crowd work that extends paid work into the online environment. In addition, we exclude offline crowd work, such as day labor, as it does not possess the same opportunities for distribution and global scalability as online work. For the remainder of the paper, we will use the term *crowd work* to refer to the performance of online tasks by crowd workers who are financially compensated by requesters.

We intentionally focus coverage on areas that may be of greater interest to the CSCW community, especially issues related to computer science, psychology, and organization science. We also draw on other important areas where appropriate (e.g., labor economics, ethics, law) and acknowledge these as critical to the future economy. Many aspects of these issues lie beyond the traditional purview of scientists or designers (e.g., labor regulations); however, we recognize that addressing them will be necessary for a positive future of crowd work.

Pros and Cons

Crowd work has the potential to support a flexible workforce and mitigate challenges such as shortages of experts in specific areas (e.g., IT work) or geographical locations. For individuals, crowd work also creates new opportunities for income and social mobility in regions of the world where local economies may be stagnant and local governmental structures discourage investment.

However, crowd work can be a double-edged sword, equally capable of enhancing or diminishing the quality of workers' lives. We may see echoes of past labor abuses in globally distributed crowd work: extremely low pay for labor, with marketplaces such as Amazon's Mechanical Turk reported to effectively pay an average of $2/hour [65][126] with no benefits or worker protections. The per-task payment structure used in most crowd work markets is akin to piecework compensation in manufacturing [118], and can offer an invitation for gaming behavior which can negatively influence quality [78]. Moreover, crowds can be deployed in the service of questionable goals: to break captchas, to mine gold in games, and even, potentially, to locate dissidents [158]. The recent film "In Time" (2011) provided a pop culture depiction of how such a society might function where continual performance of menial tasks was literally required for worker survival. Many writers have painted similarly bleak pictures [40,136,137].

Crowd work may also displace current workers and has the potential to replace some forms of skilled labor with unskilled labor as tasks are decomposed into smaller and smaller pieces. Tasks such as speech transcription and copyediting are increasingly being accomplished with

crowd labor, and researchers have found that even some complex and expert tasks such as writing, product design, or translation may be amenable to novice crowd workers with appropriate process design and technological support [75,77,152,153].

This displacement is coupled to a new form of Taylorism [88,141], in which organizations optimize cognitive efficiency [157] at the expense of education and skill development. Taylorism yielded to more enlightened job design after several decades (and protracted struggles by workers), but given the short time commitment between crowd worker and requester, it is easy to imagine heightened exploitation and dehumanization.

As scientists, engineers, and designers, we can propose and evaluate new structures for crowd work and help imagine and bring about more positive futures. We can do so both through the intentional creation of desirable work environments as well as the cultivation of increased demand for work and workers. In particular, we suggest a role for researchers in conceptualizing and prototyping new forms of crowd work that go beyond the simple, independent, and deskilled tasks that are common today, with the goal of blazing a trail for organizations and platforms that will form the foundation of future crowd work.

ENVISIONING FUTURE CROWD WORK

How can we move towards a future of crowd work that is more attractive for both requesters and workers than existing systems? Even more ambitiously, can we design a future of crowd work that is more attractive and more effective than traditional labor systems?

Current crowd work typically consists of small, independent, and homogenous tasks, as shown in Figure 1. Workers are paired with an instance of each task to produce an output. Such simple, small-scale work has engendered low-pay, piece rate reward structures, in part due to the perception that workers are homogenous and unskilled. The current model is also insufficient to support the complexity, creativity, and skills that are needed for many kinds of professional work that take place today. Nor can it drive factors that will lead to increased worker satisfaction, such as improved pay, skill development, and complex work structures.

Theories of Organizational Behavior and Distributed Computing

Much professional work consists of complex sets of interdependent tasks that need to be coordinated across individuals with different expertise and capabilities [89]. For example, producing a book, an academic paper, or a new car all may involve many individuals working in structured teams, each with different skills and roles, collaborating on a shared output. To address these more complex goals we draw on concepts from both the organizational behavior [89,97,143,148] and the distributed computing literatures [9,132]. We propose that

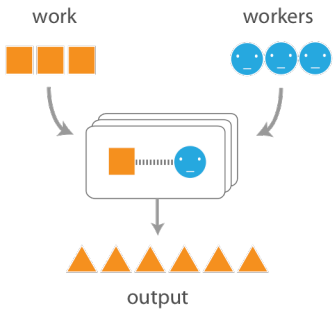


Figure 1: Current crowd work processes.

crowdsourced labor markets can be viewed as large distributed systems in which each person, such as a worker on Mechanical Turk, is analogous to a processor that can solve a task requiring human intelligence. In this way a crowdsourcing market could be seen as a loosely coupled distributed computing system [9]. Fleshing out this analogy, we develop here the beginnings of a framework for the future of crowd work that integrates the human aspects of organizational behavior with the automation and scalability of the distributed computing literature.

Both distributed organizations and computing systems face many common fundamental challenges in accomplishing complex work. Key challenges in distributed computing include partitioning computations into tasks that can be done in parallel, mapping tasks to processors, and distributing data to and between processors [9,25,96,132]. Many of these challenges map to coordination dependencies identified by Malone & Crowston [89] that also apply to human organizations. Below we discuss two categories of overlap between coordination dependencies discussed in organizational science, their analogs in distributed computing, and their implications for the beginnings of a framework for the future of crowd work.

Managing shared resources

Whenever a limited resource needs to be shared, coordinating how that resource is allocated becomes important. Allocating a fixed pool of workers to multiple tasks that must be completed under a deadline is a classic example of managing shared resources. Malone & Crowston [89] suggest a number of examples of task allocation mechanisms, ranging from first come/first serve, to markets, to managerial decisions. In distributed computing systems, managing shared resources is of similarly vital importance. Tasks must be mapped to processors, requiring functions to govern task partitioning. Reorganization of this mapping must be possible as well, for example if a processor fails or takes a long time to return results (e.g., MapReduce [34]).

Managing producer/consumer and task/subtask relationships

In many situations, one activity produces something required as input for another activity. For example, the structure of an article needs to be decided on before the

sections can be written. These same requirements exist in distributed computing, in which tasks need to be scheduled so that they can be completed in the correct sequence and in a timely manner, with data being transferred between computing elements appropriately. Deciding how to divide a task into subtasks and managing those subtasks is also a challenging problem, especially for complex and interdependent tasks [61,89]. This is true whether a manager in an organization is trying to plan a large project or a programmer is trying to parallelize a complex task. Furthermore, top-down approaches in which a single person (e.g., the task creator) specifies all subtasks a priori may not be possible, or subtasks may change as the task evolves.

Crowd-Specific Factors

Unlike traditional organizations in which workers possess job security and managers can closely supervise and appropriately reward or sanction workers, or distributed computing systems in which processors are usually highly reliable, crowd work poses unique challenges for both workers and requesters ranging from job satisfaction to direction-setting, coordination, and quality control. For example, organizations can maintain high quality work through management, worker incentives, and sanctions. While some of these methods are available in crowd work (e.g., how much to reward workers, whether to reject their work, or impose a reputation penalty) their power is attenuated due to factors such as lack of direct supervision and visibility into their work behavior, lack of nuanced and individualized rewards, and the difficulty of imposing stringent and lasting sanctions (since workers can leave with fewer repercussions than in traditional organizations, such as to reference letters or work histories). The worker's power is also limited: requesters do not make a long-term commitment to the worker, and endure few penalties if they renege on their agreement to pay for quality work. In distributed computing systems, by contrast, requesters (programmers) have fewer problems with motivating and directing their workers (computers). However, machines cannot match the complexity, creativity, and flexibility that human intelligence manifests. Combining ideas from human and computer organization theories may thus provide complementary benefits and address complementary weaknesses over using either alone.

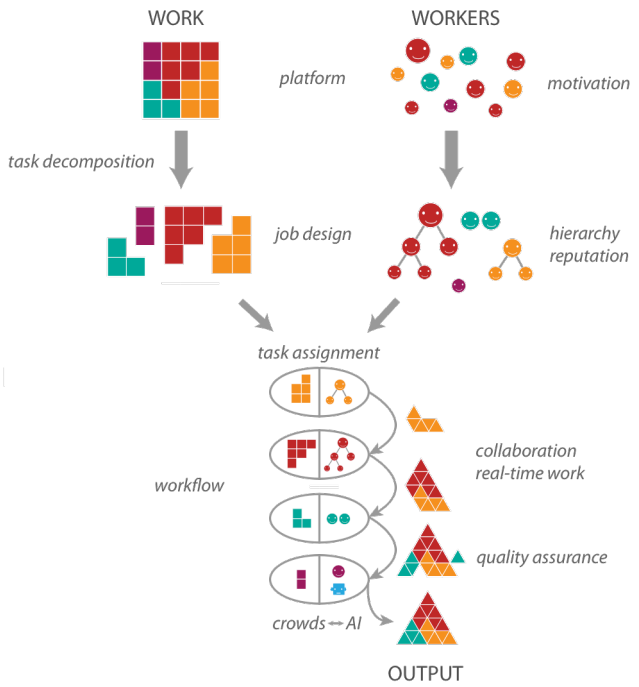


Figure 2: Proposed framework for future crowd work processes to support complex and interdependent work.

Framework

Figure 2 presents a framework that integrates the challenges posed by managing shared resources (such as assigning workers to appropriate tasks), managing producer-consumer relationships (such as decomposing tasks and assembling them into a workflow), and crowd-specific factors (such as motivation, rewards, and quality assurance). Many of its elements combine insights from organizational behavior and distributed computing: for example the task decomposition and task assignment functions use both human and computational processes.

The goal of this framework is to envision a future of crowd work that can support more complex, creative, and highly valued work. At the highest level, a **platform** is needed for managing pools of tasks and workers. Complex tasks must be **decomposed** into smaller subtasks, each **designed** with particular needs and characteristics which must be **assigned** to appropriate groups of workers who themselves must be properly **motivated**, selected (e.g., through **reputation**), and organized (e.g., through **hierarchy**). Tasks may be structured through multi-stage **workflows** in which workers may **collaborate** either synchronously or asynchronously. As part of this, **AI** may guide (and be guided by) crowd workers. Finally, **quality assurance** is needed to ensure each worker's output is of high quality and fits together.

Because we are concerned with issues of design – the technical and organizational mechanisms surrounding crowd work – we highlight in the process model twelve specific research foci (Figure 2) that we suggest are necessary for realizing such a future of crowd work. These foci are grouped into three key dimensions: foci relevant to the work process; the computation guiding, guided by, and underlying the work; and the workers themselves. Our 12 foci overlap each other in places. However, in total they provide a wide-ranging multidisciplinary view that covers current and prospective crowd work processes. For example, workflow techniques may be useful for handling the flow of documents through a set of tasks [111], but the effectiveness of these techniques can be amplified through clever job design that divides tasks and allocates incentives in a way that benefits both workers and requesters (cf. [62]).

Our model is based on empirical as well as theoretical input. In forming this model, we gathered feedback from both task requesters and workers. The authors of this paper have been requesters, have designed crowd workflows, and have worked for platform companies, and so requester and platform company issues are represented. We wished to also represent the voices of workers. We chose one popular crowdsourcing platform, Amazon Mechanical Turk, and asked questions of workers in the two countries with the largest number of crowd workers – the United States and India – who had completed and had approved more than 500 tasks (as enforced by the platform). Workers were paid USD $2 to comment on and add design suggestions for the 12 foci discussed below. Fifty-two workers responded from the US and 52 from India. Four of the responses were removed from the sample because of incomplete or inconsistent answers. Workers in India had a mean age of 27 (σ=6). Workers in the US had a mean age of 33 (σ=11). In India, 27% of workers were female, and in the US, 58% were female. In both countries responders had considerable experience: the mean total lifetime tasks were 6562 (σ=15292) and 9019 (σ=22460) for Indian and U.S. workers, respectively. The purpose of the survey was to provide workers a vehicle through which they could contribute their own insights. In general, the workers' responses were thorough and insightful, and we have integrated their ideas into this paper, quoting their answers where appropriate. Even though our survey is informal and relatively small in scale, we believe that it enriches this paper by providing a variety of workers' perspectives on the 12 topics we discuss next.

RESEARCH FOCI

In the sections below, we survey and analyze the 12 research foci that comprise our model. First, we consider the future of the work processes and how the work is organized and accomplished. Second, we consider the integration of crowd work and computation, including the symbiosis between human cognition, artificial intelligence (AI), and computationally-enabled crowd platforms. Finally, we consider crowd workers and how we can develop jobs, reputation systems, motivations, and incentives that will benefit them. In each subsection we state our motivation and goals, briefly review related work, and propose research issues and themes.

The Future of Crowd Work Processes

Increasing the value and meaning of crowd work will require that it move beyond simple deskilled tasks to complex professional work. In this section, we focus on the key challenges that must be met in order to enable complex crowd work processes: designing workflows, assigning tasks, supporting hierarchical structure, enabling real-time crowd work, supporting synchronous collaboration, and controlling quality.

Workflow

Motivation/Goals. Complex crowd work cannot be accomplished using the simple parallel approaches that are common today, such as aggregating multiple independent judgments through voting or majority rule. Complex tasks have dependencies, changing requirements, and require multiple types of expertise. Instead, *workflows* are needed that facilitate decomposing tasks into subtasks, managing the dependencies between subtasks, and assembling the results. While initial research has shown that enabling more complex workflows can result in large differences in output quality even with small differences in instructions, rewards, and task order [72,127], we have barely begun to understand the broader design space of crowd workflows.

Related Work. Traditional organizations have developed expertise in workflow design and management. The division of labor is a core tenet of task coordination; Adam Smith in his classic *The Nature and Causes of the Wealth of Nations* [133] described the associated efficacy benefits. Via division of labor, a greater pool of agents can work in parallel, specialize for the tasks they perform, and complete an assignment with less time lost to switching tasks [10]. Coordination is difficult among distributed workers, but organizational coordination techniques can be profitably applied to crowd work (e.g., [72,74,90,97,143]). Systems and formal languages support workflows in traditional firms [39], ranging from purely computational [151] to hybrid approaches where tasks are self-selected and then automatically routed onward [138].

In the context of crowds, workflow could involve a much larger scale of operation, and a much more heterogeneous set of actors. The design space ranges from massively redundant, independent tasks (e.g., contests that choose one entry [8,19,20]) to highly serial processes with work passed from one worker to the next (e.g., passing a task from worker to worker for improvements [87]). Recent systems and toolkits pursued a "flare and focus" approach for complex work by exploring a space of options and then drilling down to flesh out those options [15,75,87,150]. Crowd workers can guide workflows as well [1,75,80,154].

Research Proposal. Crowd workflows are still quite brittle, and are most successful with highly targeted tasks. To improve existing workflows, we must experiment and iterate on a large space of parameters, instructions, incentives and decompositions. Costs of doing so may be reduced through models of worker behavior [120] or by encapsulating and reusing proven design patterns [1][73]. Then, we must push crowd workflows toward more general tasks and wicked problems that have no clearly defined solution [112]. Rather than edit text, for example, crowd workflows should be able to support complex goals such as creativity and brainstorming, essay writing, music composition, or civic planning. Crowd workers reminded us in our survey responses that they also need help managing their own workflows as they juggle tasks from different requesters.

Task Assignment

Motivation/Goals. Sharing limited resources requires

coordination — allocating a fixed pool of workers to multiple tasks with deadlines is a classic example [117]. Ideally, requesters will see their tasks completed quickly, while workers are continuously employed with tasks that match their interests. In the worst case, workers are matched to tasks that are uninteresting or too difficult, and won't make the income they want or deserve.

Related Work. Management scientists have developed techniques such as first come/first serve task queues, markets, and managerial decisions [117]. Computer science research adds useful abstractions drawn from data partitioning, OS scheduling, and failover (e.g., [34]). Currently, workers are usually forced to sort queues by task volume and recency [29]. But some algorithms form teams dynamically based on expertise [5].

Research Proposal. Task assignment has typically involved either a first-come/first-served model (e.g., the ESP game, Galaxy Zoo [2,106]) or a market model (e.g., oDesk, Mechanical Turk). In either case, task designers must guess at the right combination of incentives and iterate until success. This process is both time-consuming and expensive. Better theoretical models, markets or automatic computational matching processes (e.g., [14]) could drastically reduce development costs and address search friction [4], an important issue in labor economics. The assignment of tasks in relation to individuals' abilities has also been studied as part of business workflow research [110,128]. Several workers in our survey complained that they spent too much energy finding appropriate tasks; one suggested that platforms provide automatic recommendations of possible next tasks based on workers' previous task choices. These comments suggest a research question: are the workers or the platforms better-suited to manage task assignment? That is, should tasks be pulled by workers or pushed by platforms?

Hierarchy

Motivation/Goal. Hierarchy has become the primary management strategy in traditional organizations. It benefits coordination, decision making, quality control, and assigns incentives and sanctions [30,92]. Hierarchies decompose large and complex tasks such as developing and manufacturing an automobile by clarifying legitimate authority and workflow across organizations. Hierarchies in crowd work could enable groups of workers to tackle new classes of complex work, increase efficiency, and support consistency and integration. Hierarchies may also allow workers to act more like teams, for example developing accountability standards, decision-making and conflict resolution processes, and review policies.

Related Work. Volunteer crowdsourcing platforms have evolved their own hierarchies and decision-making processes [104,156], appropriating techniques from other online communities where appropriate [101]. Most paid approaches have workers make hierarchical decisions collectively: for example, task decomposition and integration [75,80], quality oversight of each others' contributions [78,100], and leader elections to represent collective opinions [83]. oDesk and MobileWorks identify and empower workers to serve in leadership roles [79]. No comparative analyses of the effectiveness of these approaches exist yet.

Research Proposal. The fluid nature of crowd work opens the door to new kinds of hierarchy where workers transition between roles continuously. Crowd management could entail a layered tree made up of worker-leaders, requesters, machine learning systems, and algorithms. In such a setup, participants might be leaf nodes (workers) in one job but managers in another. Realizing this vision will require improvements to the platforms (e.g., oDesk teams) and the creation of systems that take advantage of it. Alternatively, organized groups of workers might begin applying to jobs as a single entity. But there may be resistance to hierarchy: one worker wrote, "I like the way it is. There does not seem to be a hierarchy. In fact, this is one of the most satisfying aspects of Mturk. Anyone can be their own boss." This comment suggests the value of further empirical study of the willingness of crowd members to manage or be managed by each other. Perhaps current forms of organizational structure will yield to new ones, in which the processes of managing and being managed will be more intertwined and are conceptually different than what we currently experience.

Realtime Crowd Work

Motivation/Goal. For work with tight completion-time constraints, we will need to create *flash crowds*: groups of individuals who arrive moments after a request and can work synchronously. Any application that wants to embed on-demand crowdsourcing (e.g., [15,17]) is limited by the problem of crowd latency, but current crowd tasks can take hours [15] or days [72].

*Related Work***.** Fast recruitment has been the major research thrust in realtime crowdsourcing so far. Early attempts were motivated by time-limited tasks such as searching for a missing person [59] and timed "competitions" [99,140]. In paid crowdsourcing, researchers began e-mailing a set of workers the night before the study and announced a time for the experiment. Keeping workers busy with old tasks brought wait times down to a half minute to one minute [17], and paying workers a small wage to stay on call is enough to draw a crowd together roughly two to three seconds later [13,14]. This technique can be modeled using queuing theory and adapted to bring together crowds in as little as 500 milliseconds.

*Research Proposal***.** The two core challenges for realtime crowdsourcing will be 1) scaling up to increased demand for realtime workers, and 2) making workers efficient enough to collectively generate results ahead of time deadlines. What happens as more tasks need such responses, and as the size of the crowd increases? Is it

possible to support a large number of realtime crowdsourcing tasks competing for workers' attention [14]?

Workers who arrive quickly can still be slow at completing work [13]. It is possible to design algorithms and workflows to guide workers to complete synchronous tasks quickly. So far, these techniques are restricted to particular domains [13], but general approaches may be possible. What would it take to begin with a sketch on a napkin, find a designer to mock up interface alternatives, a usability analyst to test those prototypes, and a front-end engineer to implement the best one, all in a single afternoon? Workers seem interested in such tasks: one suggested providing "More communications options with the requester, something better than just emailing them, like some sort of immediate chat." Another wrote: "I'd like to see some sort of worker alert system on the dashboard for such events."

Synchronous Collaboration

*Motivation/Goal***.** Many tasks worth completing require cooperation – yet crowdsourcing has largely focused on independent work. Distributed teams have always faced challenges in cultural differences and coordination [60], but crowd collaboration now must create rapport over much shorter timescales (e.g., one hour) and possibly wider cultural or socioeconomic gaps.

Related Work. Many attempts at collaborative crowd work highly structure the communication between participants. For example, crowds can solve distributed problems by observing the behavior of neighbors [91], letting the system choose which suggestions to focus on [13], continuously elect new leaders [83] and pass on knowledge to new members [84]. These techniques reduce the damage that a single poor or malicious crowd worker can do, while also limiting the types of collaboration possible. They also set up opportunities for feedback and learning [42].

Unstructured collaboration also shows promise, for example by giving workers a task and placing them in collaborative text authoring software [77]. These techniques draw on synchronous collaboration research (e.g., [53,66]).

*Research Proposal***.** To shift from independent workers to teams of on-demand collaborators, we must revisit and extend traditional CSCW work on distributed teamwork. Short periods of intense crowd collaboration call for fast teambuilding and may require the automatic assignment of group members to maximize collective intelligence [5,149]. Finally, it will be a major research undertaking to invent and describe the tasks and techniques that succeed with synchronous collaboration.

Quality Control

Motivation/Goal. Quality problems are a serious challenge to the mass adoption of crowd work. The most appealing aspects of crowd work — such as high throughput, low transaction costs, and complex/subjective tasks — also make it susceptible to quality control issues. Workers *satisfice*, minimizing the amount of effort they expend, and in the extreme cheat or game the system. For example, 30% or more of submissions on Mechanical Turk may be low quality [15,72]. One worker warned us: "People collude to agree on wrong answers to trick the system." That is, quality filters based on consensus may be fooled by workers who agree to coordinate answers. Workers who have low expertise and requesters who provide unclear instructions also contribute to subpar responses. Problems arise even for workers who are highly motivated: these "eager beavers" often make well-intentioned but counterproductive contributions [15]. In our survey, workers saw quality control as a major issue that affected their compensation, and they expressed a dislike for their peers who lowered quality standards through misbehavior. However, many complained about requesters: one said: "Too often the job itself is badly designed or is messed up and there is a degree of misunderstanding between the worker and the job engineer."

Related Work. Of the research foci, quality control has arguably received the most attention so far. Approaches for quality control largely fall into two camps: up-front task design and post-hoc result analysis. Task design aims to design tasks that are resistant to low-quality work. For example, requesters can split work into fault-tolerant subtasks [15,75,103], apply peer-review or agreement filters [2,15,17,42,63,75], optimize instructions [43,72,127], and manipulate incentives [26,115,127].

Worker output approaches filter out poor quality work after results have been submitted. Workers' results can be compared to gold standard data on a pre-labeled set of the tasks [24,43,85]. Including gold standards can prevent workers' inherent biases from dominating the results [37]. However, authoring gold data can be burdensome, and gold standards may not be possible for subjective or generative tasks (e.g., writing an essay). Other common methods scale the influence of a submission according to how well that worker agrees with others [24,36,64,134] or according to the workers' votes. However, recruiting multiple workers costs more, agreement may not be possible for subjective or generative tasks, and the approach is susceptible to collusion [41]. False identities are increasingly being created as an attack on quality assurance methods.

A promising approach that addresses some worker output issues examines the way that workers do their work rather than the output itself, using machine learning and/or visualization to predict the quality of a worker's output from their behavior [119,120]. Similar but simpler approaches provide requesters more visibility into worker behavior, such as oDesk's Worker Diary which periodically takes snapshots of workers' computer screens. While powerful, such techniques must address privacy and autonomy concerns if widely deployed.

Research Proposal. While quality control is improving for tasks with a closed set of possible answers, we still have few techniques for open-ended work and highly skilled

tasks. Is it possible to robustly gauge workers' skills at tasks such as audio engineering, art critique, or poetry? Should we rely on peer evaluation, or data mine low-level activity to predict output quality? Do quality metrics map well across different marketplaces?

In the long term, we must move from reducing poor quality work to scaffolding truly excellent work. To do so, we can optimize workflows for creativity, innovation, and discovery. This vision will involve recruiting experts; metrics that enable us to evaluate such factors; and reward structures that value high-quality artifacts.

The Future of Crowd Computation

Crowd labor is already mediated by computation. However, computation has the opportunity to step into a much more active role in helping recruit and manage workers as well as contribute directly to work processes. Hybrid human-computer systems could tap into the best of both human and machine intelligence. We structure our discussion around the potential for reciprocal benefits between crowd workers and computational systems. *Crowds guiding AIs* considers how crowd intelligence can help train, supervise, and supplement automation. *AIs guiding crowds* considers how machine intelligence can help make the crowd more efficient, skilled, and accurate. Finally, we also consider the design and evaluation of crowdsourcing platforms.

Crowds Guiding AIs

Motivation/Goal. In human computation, people act as computational components and perform the work that AI systems lack the skills to complete [2]. By tapping into crowd intelligence, computational systems can support a much broader set of tasks. It should be understood that this area, out of all areas, least excites crowd workers, and perhaps with good reason: crowd workers may end up training machines to replace them.

*Related Work***.** Paid crowds have gathered large amounts of data to train algorithms. For example, crowds can: 1) match expert annotations on natural language processing tasks such as word sense disambiguation [23,134], 2) generate speech corpora for spoken language research [23,95], 3) annotate objects and people in images [135], and 4) help with graphics tasks such as identifying depth layers [51]. Crowds can also solve algorithmic problems such as graph coloring [68,91].

Research Proposal. While algorithms will continue to benefit from crowd-generated training data, there are opportunities to integrate crowds more deeply into algorithms. Rather than treating crowd data as ground truth labels, it may be profitable to understand and model the biases and intuitions that human cognition brings [147]. It may be possible to design machine learning algorithms that more deeply understand the human nature of these labels. Algorithms may also more directly model the tradeoff between cost and performance: for example, by using a combination of active learning and semi-supervised learning to collect the most informational labels [155].

AIs Guiding Crowds

Motivation/Goal. While large groups are increasingly adept at completing straightforward parallel tasks [54,139], they can struggle with complex work. Participants have varied skill levels, even well-intentioned contributions can introduce errors, and errors are amplified as they propagate through the crowd. It is possible to integrate crowds directly inside of software [15,17,46], and use the software to help guide crowd work; for example, a machine learning model can determine which work products may still be improved, and then assign workers most likely to make such improvements [32,33]. These systems may also be able to predict their expertise needs in advance, then train and adapt workers in an online fashion via automated tutoring [47,113] or peer learning [18].

Related Work. Computational approaches for designing and integrating workflow, incentive, and instruction patterns have shown promise [32,125], as well as techniques that trade off the strengths of crowds and artificial intelligence [67,146]. AIs could also serve as a reflective aids, encouraging the crowd to learn by pointing out what others have done in similar contexts (e.g., [109]).

Research Proposal. The research community should examine whether algorithmic management is an improvement over traditional organization management techniques. In an algorithmic organization crowds will need to be able to raise exceptions, as well as halt and restart processes. And AIs will need to know when then can proceed, when they need human help, and when they need to help the workers. In addition, workers should be able to modify their supporting AIs as needed. In the end, both workers and AIs should improve. To accomplish this, we will need to move from a setting where simple AIs completely determine a workflow to a richer, mixed-initiative setting where crowds and AIs jointly teach each other, and jointly control the work process.

Crowdsourcing Platforms

Motivation/Goal. Crowdsourcing platforms provide the central nexus between requesters and workers. As a result, platform design offers the opportunity to change our perceptions and understanding of crowd work in general, as well as shape the relationships and practical interactions between workers and requesters in practice. While several platforms already exist, limiting ourselves to existing platforms greatly restricts the scope and nature of change we can enact. By contrast, novel platforms can help drive the diffusion of new designs and techniques for crowd work.

Related Work. Platform research tends to optimize existing processes or reach out to new populations. For example, CrowdFlower experiments on its own gold standard metrics to minimize the number of times a question is asked [129].

Platforms may also manage a large number of real time requests or route tasks to ensure that all tasks have steady streams of incoming workers [14]. MobileWorks promotes its own workers to management positions based on their performance [79]. To find new populations, MobileWorks and mClerk engage with the developing world through specifically-tailored mobile phone interfaces [56,100]. Incentives can be used to recruit local experts as well: for example, a vending machine served up exam grading tasks and dispersed candy as a reward [58].

Research Proposal. While designing and building a new platform may seem daunting, examples like Brin and Page's Google [21] show that two graduate students can disrupt the commercial landscape and dramatically change the way people work. We challenge the community to similarly revolutionize our conception of what a crowdsourcing platform is and can achieve. Innovation should articulate a vision for crowd work that is effective, efficient, and fair. Beyond mere technology, negotiating the balance of power between interested parties is central to platforms and markets [11,130]. As crowd work already faces challenges related to power inequalities similar to those encountered in offline labor markets, future platforms may shape or be constrained by future regulatory intervention, on, for example, the use of independent contracting for regular, recurring work [45]. Beyond money, platforms might also support labor exchanged for virtual goods or performed in virtual environments, which raises additional policy issues [44]. Privacy is another issue: how can platforms disclose enough information to be trusted as a source of worker quality while also maintaining privacy? Experience with markets such as eBay and Amazon suggests that greater transparency may be helpful, but such mechanisms must be carefully managed to avoid abuse [38].

Security concerns will also continue to grow in importance. For example, recognition of trusted workers may lead to identity theft and fraudulent use of compromised accounts [41]. Security research must consider both new attacks on platforms and use of platforms for launching attacks [144].

The Future of Crowd Workers

Crowd work involves a partnership between requesters and workers. Thus, when designing the future of crowd work, it is important to develop tools to support not only the work itself but also those performing the work. Below we identify and discuss three important research challenges for supporting the crowd workers of the future: job design, reputation and credentials, and motivation and rewards.

Job Design

Motivation/Goals. Some tasks that need to be done are just dull. Motivating workers to accomplish such tasks can be challenging, and may lead to reduced engagement with the system: "It would be better if some of the task assignments weren't so monotonous…I don't see the long-term payoff and it discourages me." While dressing up such tasks as games may reduce boredom, entertainment represents a fairly superficial form of work satisfaction. We believe the future of crowd work depends on creating jobs that achieve both organizational performance and worker satisfaction.

Related Work. In traditional firms, when managers design jobs that provide skill variety, task identity, and task significance, workers find their jobs more meaningful [57]. When workers have autonomy in their jobs and receive feedback, they experience increased responsibility for and understanding of the impact of their work. Combined, these lead to increased performance for the firm and reduced employee problems such as absenteeism and turnover [57].

Unfortunately, many paid crowd work platforms do not provide as much skill variety, task identity, and task significance as volunteer platforms such as Wikipedia. There are direct payoffs when requesters convey the identity and significance of tasks to crowd workers [26,115]. Timely and task-specific feedback from peers and requesters, as well as opportunities for self-assessment, help workers to learn, persevere, and produce better results [42].

Research Proposal. An ideal crowd work system would allow workers to complete a whole and identifiable piece of work in a way that is satisfying and measurable (cf. [62]). The system would explain the significance of the job, offer peer-to-peer and expert feedback, and encourage self-assessment. These systems could offer a variety of ways to complete the task, and thereby not only provide autonomy for the worker, but also reduce errors [55]. Achieving this vision will require communication with workers about the scope and impact of their work, as well as verification that workers understand their impact.

Providing more context has tradeoffs, however. On the one hand, more context enables workers to better judge how the fruits of their effort will be used so they can make informed decisions about whether or not to perform work. On the other hand, reducing context may streamline work, leading to greater efficiencies for both requesters and workers. Moreover, requesters may not always want to fully disclose the context of work due to privacy, security, or intellectual property concerns. This suggests a need to balance distinct and competing concerns: how much information a worker needs before consenting to provide labor, how much information should be shared to motivate and retain workers, how to share context without introducing inefficiency and how much information may be legitimately withheld to protect interests of requesters.

Reputation and Credentials

Motivation/Goals. In traditional firms, reputation and credentials (e.g., letters of reference, certifications, work history) are critical to recruiting, creating lasting rewards and sanctions, and managing work quality. For example, institutions with brand names such as Google and Apple are likely to attract software engineers to apply for jobs, and, in turn, such names on the resumes of software engineers

attract other prospective employers. However, many existing crowd work systems have only coarse reputation and credentialing systems. For example, a worker's history on Mechanical Turk primarily measures the percentage of work that has been approved. In contrast, traditional employers judge a prospective employee's education and work history through a variety of instruments, including interviews, transcripts, and references. Likewise, applicants can investigate employers' reputation (and crowd workers in our survey wanted better reputation rankings for employers to be established within the platform [130]).

Related Work. Robust mechanisms for supporting trust, assurance, and accountability can support volunteer-based crowds and collaboration online [28,76]. Likewise, in the context of monetary rewards, reputation can have serious financial consequences [111], motivating people to manipulate systems for their own benefit. For example, workers may build networks of Sybil identities — that is, create pseudonyms [41] — to enhance their reputations and to foil quality control methods. Workers may also boost their ratings by agreeing implicitly or explicitly to recommend each other reciprocally [38]. The design and evaluation of future crowd work reputation systems must address these issues.

Research Proposal. The core challenge with reputation is trading off the benefits of pseudonymous, low transaction cost hiring with the richer but higher transaction cost hiring decisions most firms make today. For example, employers know little about workers on Mechanical Turk, but such workers can be hired and engaged in work nearly instantaneously. Meanwhile, platforms such as oDesk with richer reputation systems incur correspondingly higher transaction costs in hiring workers, involving negotiation and handshakes between workers and employers.

To address this challenge reputation systems will need to be robust to cheating and gaming while preserving the benefits of pseudonymity and supporting low-transaction cost hiring. One possible solution would be to create a web of trust in which requesters and workers validate each other as trustworthy [79]. However, malicious workers and requesters can infiltrate the community and spread their own web of trust. Interfaces may support detection of bad actors: for example, by highlighting topology [121,144], statistical patterns [93] and behavior [120].

The creation of technical tools for sharing information about workers should be coupled with more robust systems for monitoring and reporting requester abuses [131]. Lastly, these initiatives towards enhanced reputation systems should be balanced with the need to preserve privacy, as well as the potential payoffs of anonymous collaboration for both workers and requesters [16].

Motivation and Rewards

Motivation/Goals. Requesters often envision crowd workers as either anonymous individuals who are motivated to complete piecework by small monetary rewards or highly-skilled professionals who work on large, higher-paid tasks without management oversight. However, crowd workers are a diverse and multifaceted population with a range of motives and experiences. Many workers in micro-task environments are ambitious individuals, desiring to be CEOs or top-rate school teachers [48]. Yet few researchers have grappled with the diversity and richness of the motives of the individuals comprising the crowd. One worker reminded us that requesters, as well as workers, need to be motivated:

> *We could definitely use more motivation, we perform task[s] for mere pennies. Mturk should encourage and reward requesters that provide clear instructions, quick payment, and higher pay. Rewarding them would create more worthwhile tasks that we would take more seriously and work hard on. Good, credible HITs are few and far between.*

*Related Work***.** Research in psychology, sociology, management, and marketing provide insights into human motivation that are applicable to crowd work. Management research illustrates the challenge of clearly understanding, communicating, and rewarding desired behavior [70]. Workers seek to understand which activities are rewarded and then tend to do those activities to the virtual exclusion of others. Other studies find mixed results on the effect of financial incentives on the quality of workers' outputs, and underscore the performance and satisfaction benefits of harnessing intrinsic motivations in task design such as non-financial awards and recognition, meaningfulness of tasks, and the feeling of contributing towards the greater good [26,72,86,94,115,127].

Past research suggests requesters should (1) clearly understand and communicate desired behaviors, (2) understand and align worker motivations and incentives with these desired behaviors, and (3) design the requests and incentive structures in order to achieve both effective task completion and worker satisfaction. This also requires requesters to understand variations in crowd workers' motivations (e.g., [7,11,71,116]), for example in competence, enjoyment, connectedness, prosocial orientation, and autonomy [49].

Research Proposal. The future of crowd work requires that requesters and platform designers consider a broad set of motivations, not just financial incentives. We must create frameworks that acknowledge the dynamic nature of motivation and its dependence on context. For example, it is not clear that payment alone will be the optimal motivator for expert crowdsourcing markets. Such frameworks should enable us to move from analysis to design of new motivational schemes. Research should overcome the dichotomous emphases on dehumanizing piece-work and frictionless virtual collaboration in order to provide a more holistic framework within which to

understand and build systems that support workers' diverse motivations (e.g., [22,50,52,81,82,107,124]).

NEXT STEPS

Many of the really difficult problems that workers and requesters face will require advances on multiple foci at the same time. Below we describe three design goals demonstrating how the integration of multiple foci can lead to concrete next steps and calls to actions that will create a better environment for crowd workers, and a better set of human resources for requesters.

Create Career Ladders - *[motivation, job design, reputation, hierarchy]*

Crowd work today is largely a dead-end job, offering few opportunities for career advancement and economic mobility. As workers demonstrate aptitude and diligence, it is to the advantage of both requesters and workers to recognize workers' potential to take on new tasks requiring greater knowledge, effectiveness, and responsibility. As all organizations benefit from making the best use of available talent and workers' diverse abilities, more skilled workers should be paid for their expertise, and encouraged to train less skilled workers: for example, MobileWorks promotes workers to management jobs based on performance [79].

As workers demonstrate proficiency, they might be invited to create gold labels for verifying the quality of work from other, less established workers, which could be used to build webs of trusted workers. Proficient, trusted workers could also manage other workers, respond to reported issues, and provide first-pass triage of new task designs (to catch problems or make suggestions before new tasks go live). Crowd workers could eventually become employees themselves, or develop the skills needed to launch their own business using crowd work. For example, a career trajectory might proceed as: 1) entry-level, untrusted worker; 2) trusted worker; 3) hourly contractor; 4) employee. First steps toward building such a ladder include studying worker *motivations* in order to develop better *job designs*; creating lasting and transferable mechanisms for *reputation and credentialing* for workers; and building greater support for *hierarchy* in the form of structured teams that provide training for novices by skilled workers.

Improve Task Design through Better Communication - *[quality assurance, job design, task assignment, realtime crowd work, synchronous collaboration, platform]*

There is a popular myth that the poor quality of some crowd work stems largely from workers being lazy, stupid, or deceitful. In practice, both we and our surveyed workers have observed many cases where poor quality work instead arises from poorly designed crowdsourcing tasks. For example, a requester might assume a task obvious to them should be equally obvious to everyone else. However, even highly-educated workers may have difficulty understanding exactly what the requester actually wants. Task instructions are often incomplete or ambiguous, do not address boundary cases, and do not provide clarifying input-output examples of what is expected. Task interfaces may be poorly designed or even have bugs that make it impossible to complete a task. Gold labels used as trap questions may be far more subjective than requesters realize, leading to mistaken rejection of work.

What might we do to address these problems? Designers can make it easier and faster for requesters to create effective tasks. For example, platforms might provide task templates showing examples of proven task designs [27] to encourage shared mental models and improve *quality assurance*. Platforms could also help educate requesters about the impact of *job design* and *task assignment* on resulting quality, best practices, and common errors to be avoided. Platforms might even offer a "first pass" service in which a set of trusted workers (known to the platform) test out the task and report any issues encountered.

Platforms might also provide a wider array of communication channels between requesters and workers supporting *synchronous collaboration* and *real-time crowd work*. Workers we surveyed were adamant that the perception of poor crowd work quality was due, at least in part, to unclear instructions and insufficient feedback, and that they need more guidance to better understand what is expected. They suggested instant chat with requesters to clarify jobs, though we note this would require the continual presence of requesters while work is being performed. They also requested feedback during or just after a task. Both would be consistent with the practices of good managers and workers in other labor contexts, and so more experimenting with channels of communication could potentially have a large effect on both worker satisfaction and job quality. As a first step, requesters might provide ways for workers to clarify tasks in real time – with the requester, or with a more experienced worker. And, when work is in progress, they might provide informal feedback through these same channels.

Facilitate Learning - *[quality assurance, AIs guiding crowds, crowds guiding AIs, task assignment, reputation and credentials, platform]*

Crowd work naturally involves learning and assessment. Workers may need to acquire new skills to perform unfamiliar tasks, before or in the midst of performing the actual work. Workers may also polish and refine existing skills while completing more familiar tasks. Requesters must continually engage in *quality assurance*. Such a training-assessment cycle of work offers potentially exciting synergies with online, education by-doing. For example, DuoLingo (duolingo.com) explores this direction for foreign language learning. This idea can be generalized much further; for example, content generation tasks could be designed to better assess and enhance writing skills. A self-sustaining cycle might involve *AIs guiding crowds* on which tasks to complete (*task assignment*) depending on the worker's and requester's skill development and quality assurance goals, and then using the crowd-generated data to automate some of the simpler tasks (*crowds guiding AIs*).

The potential of crowd work-based education is enormous and multi-faceted, benefiting all parties by producing more skilled and employable workers. Online tutoring systems, perhaps augmented with human tutoring, could provide a path toward delivering more scalable education to the public at large [6,146]. Moreover, tracking and mining of work history could support personalized instruction and feedback, as well as recommending new tasks and learning modules. As workers master new skills and are assessed, badges or *credentials* could document this proficiency so that others can recognize and utilize this enhanced skill set. *Platforms* themselves can also be an important element of learning: data from crowd work can reveal what kinds of requests attract talented workers, patterns of learning and skill building among workers over time, the valuation and interaction between extrinsic and intrinsic rewards, and which types of tasks are most appropriate for which types of workers.

CONCLUSION

Crowd work may take place in the scale of minutes, but the impact of crowd work may be felt for generations. We have asked: what will it take for us, the stakeholders in crowd work – including requesters, workers, researchers – to feel proud of our own children when they enter such a work force? Answering this question has led to a discussion of crowd work from a longer-term perspective.

Specifically, we have synthesized a framework for the future of crowd work based on theory from organizational behavior and distributed computing, and informed by the concerns of crowd work stakeholders. Our hope is that this framework and the corresponding research challenges will spur discussion, experiment, and further insight. In sum, we envision a future where many cognitively complex, large-scale tasks can be decomposed into workflows and executed by crowds consisting of novices, experts and algorithms, and that the crowd work environment can be designed in such a way that satisfies the needs of both workers and requesters. After identifying the twelve major research foci that constitute the framework, we probed their current state, identified gaps, and created a call for action demonstrating ways they can be profitably integrated.

Calls for action must be made responsibly. Crowd work marketplaces are complex socio-technical systems, composed of many people and a changing technical infrastructure, with emergent organizational forms, new incentives being offered, and shifting labor pools. As a result of this complexity, there is the possibility of unpredictable side effects. For example, innovations that effectively train workers through micro-tasks may have ramifications for the world's educational institutions, and thereby for society as a whole. Hybrid combinations of workers and artificial intelligence that seek to build collective intelligence may instead lead to mechanized workers or human-imitating machines.

Our call for action, then, calls for both exciting innovation and also close observation of its effects. In crowd work, we have two important affordances: an ability to constitute new forms of organization in short amounts of time, and an ability to situate these organizations in an experimental context. While organization science has been built slowly based on observation, the proliferation of crowd work makes large-scale organizational experiments comparing distinct management strategies and task designs possible. These comparisons will help us understand how to improve crowd platforms, workers' skills, and requesters' assignments. Perhaps instead of a hadron collider, the field of crowd work needs a "social collider" in which different forms of organization can be tested. The goal should be better systems, better requests, better work, and better experience. We hope the community's observational, experimental, design and technical skills will play a vital role in shaping the future of crowd work and the next generation of workers.

ACKNOWLEDGEMENTS

We thank all the participants in the CHI 2012 CrowdCamp Workshop for their suggestions, and in particular Paul Andre who spearheaded the effort. We are pleased to acknowledge the crowd workers of the present: for their opinions and creative suggestions in response to our queries, and, more generally, for their pioneering spirit. Finally, we acknowledge Ashima, the nascent inspiration for this paper.

This work was supported under National Science Foundation awards IIS-0968561, IIS-0855995, OCI-0943148, IIS-0968484, IIS-1111124, IIS-1149797, IIS-1217096, and IIS-1217559. This work was also supported by a DARPA Young Faculty Award N66001-12-1-4256, a Temple Fellowship, Northwestern University, and the Center for the Future of Work, Heinz College, Carnegie Mellon University.

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

1. Ahmad, S., Battle, A., Malkani, Z., and Kamvar, S. The jabberwocky programming environment for structured social computing. *Proc. UIST '11*, (2011).
2. Von Ahn, L. and Dabbish, L. Labeling images with a computer game. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2004), 319–326.
3. Ahn, L.V., Blum, M., Hopper, N.J., and Langford, J. CAPTCHA: Using hard AI problems for security. *Proceedings of the 22nd international conference on Theory and applications of cryptographic techniques*, (2003), 294–311.
4. Albrecht, J. The 2011 Nobel Memorial Prize in Search Theory. *Department of Economics,*

Georgetown University.[http://www9. georgetown. edu/faculty/albrecht/SJE Survey. pdf], (2011).
5. Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., and Leonardi, S. Online team formation in social networks. *Proceedings of the 21st international conference on World Wide Web*, ACM (2012), 839–848.
6. Anderson, M. Crowdsourcing Higher Education: A Design Proposal for Distributed Learning. *MERLOT Journal of Online Learning and Teaching 7*, 4 (2011), 576–590.
7. Antin, J. and Shaw, A. Social desirability bias and self-reports of motivation: a study of amazon mechanical turk in the US and India. *Proc. CHI '12*, (2012).
8. Archak, N. and Sundararajan, A. Optimal Design of Crowdsourcing Contests. *ICIS 2009 Proceedings*, (2009), 200.
9. Bal, H.E., Steiner, J.G., and Tanenbaum, A.S. Programming languages for distributed computing systems. *ACM Computing Surveys (CSUR) 21*, 3 (1989), 261–322.
10. Becker, G.S. and Murphy, K.M. The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics 107*, 4 (1992), 1137–1160.
11. Bederson, B.B. and Quinn, A.J. Web workers unite! addressing challenges of online laborers. *Extended Abstracts CHI '11*, (2011).
12. Benkler, Y. *The wealth of networks: How social production transforms markets and freedom*. Yale Univ Pr, 2006.
13. Bernstein, M.S., Brandt, J., Miller, R.C., and Karger, D.R. Crowds in two seconds: Enabling realtime crowd-powered interfaces. *Proc. UIST '11*, (2011).
14. Bernstein, M.S., Karger, D.R., Miller, R.C., and Brandt, J. Analytic Methods for Optimizing Realtime Crowdsourcing. *Proc. Collective Intelligence 2012*, (2012).
15. Bernstein, M.S., Little, G., Miller, R.C., et al. Soylent: A Word Processor with a Crowd Inside. *Proc. UIST '10*, (2010).
16. Bernstein, M.S., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., and Vargas, G. 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community. *Fifth International AAAI Conference on Weblogs and Social Media*, AAAI Publications (2011).
17. Bigham, J.P., Jayant, C., Ji, H., et al. VizWiz: Nearly Real-time Answers to Visual Questions. *Proc. UIST '10*, (2010).
18. Boud, D., Cohen, R., and Sampson, J. Peer learning and assessment. *Assessment & Evaluation in Higher Education 24*, 4 (1999), 413–426.
19. Boudreau, K.J., Lacetera, N., and Lakhani, K.R. Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science 57*, 5 (2011), 843.
20. Brabham, D.C. Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application. *First Monday 13*, 6 (2008), 1–22.
21. Brin, S. and Page, L. The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems 30*, 1-7 (1998), 107–117.
22. Bryant, S.L., Forte, A., and Bruckman, A. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. *GROUP '05*, ACM Press (2005), 1–10.
23. Callison-Burch, C. and Dredze, M. Creating speech and language data with Amazon's Mechanical Turk. *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, Association for Computational Linguistics (2010), 1–12.
24. Callison-Burch, C. Fast, cheap, and creative: evaluating translation quality using Amazon's Mechanical Turk. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, (2009), 286–295.
25. Casavant, T.L., Braun, T.A., Kaliannan, S., Scheetz, T.E., Munn, K.J., and Birkett, C.L. A parallel/distributed architecture for hierarchically heterogeneous web-based cooperative applications. *Future Generation Computer Systems 17*, 6 (2001), 783–793.
26. Chandler, D. and Kapelner, A. Breaking monotony with meaning: Motivation in crowdsourcing markets. *University of Chicago mimeo*, (2010).
27. Chen, J.J., Menezes, N.J., and Bradley, A.D. Opportunities for Crowdsourcing Research on Amazon Mechanical Turk. *Interfaces 5*, (2011), 3.
28. Cheshire, C. Online Trust, Trustworthiness, or Assurance? *Daedalus 140*, 4 (2011), 49–58.
29. Chilton, L., Horton, J., Miller, R.C., and Azenkot, S. Task search in a human computation market. *Proc. HCOMP '10*, (2010).
30. Coase, R.H. The Nature of the Firm. *Economica 4*, 16 (1937), 386–405.
31. Cooper, S., Khatib, F., Treuille, A., et al. Predicting protein structures with a multiplayer online game. *Nature 466*, 7307 (2010), 756–760.
32. Dai, P., Mausam, and Weld, D. Decision-theoretic control of crowd-sourced workflows. *Proc. AAAI '10*, (2010).
33. Dai, P., Mausam, and Weld, D.S. Artificial intelligence for artificial artificial intelligence. *Twenty-Fifth AAAI Conference on Artificial Intelligence*, (2011).
34. Dean, J. and Ghemawat, S. MapReduce: Simplified Data Processing on Large Clusters. *To appear in OSDI*, (2004), 1.
35. Debeauvais, T., Nardi, B.A., Lopes, C.V., Yee, N.,

and Ducheneaut, N. 10,000 Gold for 20 Dollars: An exploratory study of World of Warcraft gold buyers. *Proceedings of the International Conference on the Foundations of Digital Games*, (2012), 105–112.
36. Dekel, O. and Shamir, O. Vox populi: Collecting high-quality labels from a crowd. *Proc. 22nd Annual Conference on Learning Theory*, (2009).
37. Della Penn, N. and Reid, M.D. Crowd & Prejudice: An Impossibility Theorem for Crowd Labelling without a Gold Standard. *Collective Intelligence*, (2012).
38. Dellarocas, C. Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms. *Proceedings of the 3rd ACM Conference on Electronic Commerce*, (2001), 171–179.
39. Van Der Aalst, W.M.P., Ter Hofstede, A.H.M., Kiepuszewski, B., and Barros, A.P. Workflow patterns. *Distributed and parallel databases 14*, 1 (2003), 5–51.
40. Doctorow, C. *For the Win*. Voyager, 2010.
41. Douceur, J. The sybil attack. *Peer-to-peer Systems*, (2002), 251–260.
42. Dow, S., Kulkarni, A., Klemmer, S., and Hartmann, B. Shepherding the crowd yields better work. *Proc. CSCW '12*, (2012).
43. Downs, J.S., Holbrook, M.B., Sheng, S., and Cranor, L.F. Are your participants gaming the system?: screening mechanical turk workers. *Proceedings of the 28th international conference on Human factors in computing systems*, (2010), 2399–2402.
44. Felstiner, A. Sweatshop or Paper Route?: Child Labor Laws and In-Game Work. *Proceedings of CrowdConf*, (2010).
45. Felstiner, A. Working the Crowd: Employment and Labor Law in the Crowdsourcing Industry. *Berkeley J. Emp. & Lab. L. 32*, (2011), 143–143.
46. Franklin, M., Kossmann, D., Kraska, T., Ramesh, S., and Xin, R. CrowdDB: answering queries with crowdsourcing. *Proc. SIGMOD '11*, (2011).
47. Gal, Y., Yamangil, E., Shieber, S., Rubin, A., and Grosz, B. Towards collaborative intelligent tutors: Automated recognition of users' strategies. *Intelligent Tutoring Systems*, (2008), 162–172.
48. Gerber, E. and Dontcheva, M. *Career Aspirations for Crowdworkers*. In preparation.
49. Gerber, E., Hui, J., and Kuo, P. *Crowdfunding: Why creators and supporters participate*. Segal Design Institute, Evanston, IL, 2012.
50. Ghosh, R. and Glott, R. *Free/Libre and Open Source Software: Survey and Study*. European Commission, 2002.
51. Gingold, Y., Shamir, A., and Cohen-Or, D. Micro Perceptual Human Computation. *To appear in ACM Transactions on Graphics (TOG)*, (2012).
52. Glott, R., Ghosh, R., and Schmidt, P. *Wikipedia Survey*. UNU-MERIT, Maastricht, Netherlands, 2010.
53. Greenberg, S. and Bohnet, R. GroupSketch: A multi-user sketchpad for geographically-distributed small groups. *Proc. Graphics Interface'91*, (1991).
54. Grier, D.A. *When Computers Were Human*. Princeton University Press, 2005.
55. Grier, D.A. Error Identification and Correction in Human Computation: Lessons from the WPA. *Proc. HCOMP '11*, (2011).
56. Gupta, A., Thies, W., Cutrell, E., and Balakrishnan, R. mClerk: Enabling Mobile Crowdsourcing in Developing Regions. *Proc. CHI '12*, (2012).
57. Hackman, J.R. and Oldham, G.R. Motivation through the design of work: Test of a theory. *Organizational behavior and human performance 16*, 2 (1976), 250–279.
58. Heimerl, K., Gawalt, B., Chen, K., Parikh, T.S., and Hartmann, B. Communitysourcing: Engaging Local Crowds to Perform Expert Work Via Physical Kiosks. *Proc. CHI '12*, (2012).
59. Hellerstein, J.M. and Tennenhouse, D.L. Searching for Jim Gray: a technical overview. *Communcations of the ACM 54*, 7 (2011), 77–87.
60. Hinds, P. *Distributed work*. The MIT Press, 2002.
61. Von Hippel, E. Task partitioning: An innovation process variable. *Research policy 19*, 5 (1990), 407–418.
62. Holmstrom, B. and Milgrom, P. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org. 7*, (1991), 24.
63. Horton, J.J. and Chilton, L.B. The labor economics of paid crowdsourcing. *Proceedings of the 11th ACM conference on Electronic commerce*, (2010), 209–218.
64. Ipeirotis, P.G., Provost, F., and Wang, J. Quality management on amazon mechanical turk. *Proc. HCOMP '10*, (2010).
65. Ipeirotis, P.G. *Mechanical Turk, Low Wages, and the Market for Lemons*. http://www.behind-the-enemy-lines.com/2010/07/mechanical-turk-low-wages-and-market.html, 2010.
66. Ishii, H. and Kobayashi, M. ClearBoard: a seamless medium for shared drawing and conversation with eye contact. *Proc. CHI '92*, (1992), 525–532.
67. Kamar, E., Hacker, S., and Horvitz, E. Combining Human and Machine Intelligence in Large-scale Crowdsourcing. *Proc. AAMAS '12*, (2012).
68. Kearns, M., Suri, S., and Montfort, N. An experimental study of the coloring problem on human subject networks. *Science 313*, 5788 (2006), 824–827.
69. Kensing, F. and Blomberg, J. Participatory Design: Issues and Concerns. *Computer Supported Cooperative Work (CSCW) 7*, 3 (1998), 167–185.
70. Kerr, S. On the folly of rewarding A, while hoping for B. *Academy of Management Journal*, (1975), 769–783.
71. Khanna, S., Ratan, A., Davis, J., and Thies, W.

Evaluating and improving the usability of Mechanical Turk for low-income workers in India. *Proc. ACM Symposium on Computing for Development '10*, (2010).
72. Kittur, A., Chi, E.H., and Suh, B. Crowdsourcing user studies with Mechanical Turk. *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, (2008), 453–456.
73. Kittur, A., Khamkar, S., André, P., and Kraut, R.E. CrowdWeaver: visually managing complex crowd work. *Proc. CSCW '12*, (2012).
74. Kittur, A., Lee, B., and Kraut, R.E. Coordination in collective intelligence: the role of team structure and task interdependence. *Proceedings of the 27th international conference on Human factors in computing systems*, (2009), 1495–1504.
75. Kittur, A., Smus, B., Khamkar, S., and Kraut, R.E. Crowdforge: Crowdsourcing complex work. *Proceedings of the 24th annual ACM symposium on User interface software and technology*, (2011), 43–52.
76. Kittur, A., Suh, B., and Chi, E.H. Can you ever trust a wiki?: impacting perceived trustworthiness in wikipedia. *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, (2008), 477–480.
77. Kittur, A. Crowdsourcing, collaboration and creativity. *XRDS 17*, 2 (2010), 22–26.
78. Kochhar, S., Mazzocchi, S., and Paritosh, P. The anatomy of a large-scale human computation engine. *Proceedings of the ACM SIGKDD Workshop on Human Computation*, (2010), 10–17.
79. Kulkarni, A., Gutheim, P., Narula, P., Rolnitzky, D., Parikh, T.S., and Hartmann, B. MobileWorks: Designing for Quality in a Managed Crowdsourcing Architecture. *IEEE Internet Computing To appear*, 2012.
80. Kulkarni, A.P., Can, M., and Hartmann, B. Turkomatic: automatic recursive task and workflow design for mechanical turk. *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, (2011), 2053–2058.
81. Kuznetsov, S. Motivations of contributors to Wikipedia. *SIGCAS Comput. Soc. 36*, 2 (2006), 1.
82. Lakhani, K. and Wolf, B. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. In J. Feller, B. Fitzgerald, S.A. Hissam and K.R. Lakhani, eds., *Perspectives on Free and Open Source Software*. MIT Press, 2005, 3–22.
83. Lasecki, W.S., Murray, K.I., White, S., Miller, R.C., and Bigham, J.P. Real-time crowd control of existing interfaces. *Proc. UIST '11*, ACM Press (2011).
84. Lasecki, W.S., White, S.C., Murray, K.I., and Bigham, J.P. Crowd memory: Learning in the collective. *Proc. Collective Intelligence 2012*, (2012).
85. Le, J., Edmonds, A., Hester, V., and Biewald, L. Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution. *Proc. SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation*, (2010), 21–26.
86. Lewis, S., Dontcheva, M., and Gerber, E. Affective computational priming and creativity. *Proceedings of the 2011 annual conference on Human factors in computing systems*, (2011), 735–744.
87. Little, G., Chilton, L.B., Goldman, M., and Miller, R.C. Turkit: human computation algorithms on mechanical turk. *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, (2010), 57–66.
88. Littler, C.R. Understanding Taylorism. *British Journal of Sociology*, (1978), 185–202.
89. Malone, T.W. and Crowston, K. The interdisciplinary study of coordination. *ACM Computing Surveys (CSUR) 26*, 1 (1994), 87–119.
90. Malone, T.W., Yates, J., and Benjamin, R.I. Electronic markets and electronic hierarchies. *Communications of the ACM 30*, 6 (1987), 484–497.
91. Mao, A., Parkes, D.C., Procaccia, A.D., and Zhang, H. Human Computation and Multiagent Systems: An Algorithmic Perspective. *Proc. AAAI '11*, (2011).
92. March, J.G. and Simon, H.A. Organizations. (1958).
93. Marcus, A., Karger, D.R., Madden, S., Miller, R.C., and Oh, S. Counting with the Crowd. *In Submission to VLDB*, (2012).
94. Mason, W. and Watts, D.J. Financial Incentives and the "Performance of Crowds". *Proc. HCOMP '09*, ACM Press (2009).
95. McGraw, I., Lee, C., Hetherington, L., Seneff, S., and Glass, J.R. Collecting voices from the cloud. *Proc. LREC*, (2010).
96. Milner, R. *Communicating and mobile systems: the [symbol for pi]-calculus*. Cambridge Univ Pr, 1999.
97. Mintzberg, H. An emerging strategy of" direct" research. *Administrative science quarterly 24*, 4 (1979), 582–589.
98. Moran, T.P. and Anderson, R.J. The workaday world as a paradigm for CSCW design. *Proceedings of the 1990 ACM Conference on Computer-supported Cooperative Work*, (1990), 381–393.
99. Nalimov, E.V., Wirth, C., Haworth, G.M.C., and Others. KQQKQQ and the Kasparov-World Game. *ICGA Journal 22*, 4 (1999), 195–212.
100. Narula, P., Gutheim1, P., Rolnitzky, D., Kulkarni, A., and Hartmann, B. MobileWorks: A Mobile Crowdsourcing Platform for Workers at the Bottom of the Pyramid. *Proc. HCOMP '11*, (2011).
101. Nickerson, J.V. and Monroy-Hernandez, A. Appropriation and creativity: User-Initiated contests in scratch. *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, (2011), 1–10.
102. Norman, D.A. and Draper, S.W. *User centered*

system design; new perspectives on human-computer interaction. L. Erlbaum Associates Inc., 1986.
103. Noronha, J., Hysen, E., Zhang, H., and Gajos, K.Z. Platemate: crowdsourcing nutritional analysis from food photographs. *Proc. UIST '11*, (2011).
104. Preece, J. and Shneiderman, B. The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. *AIS Transactions on Human-Computer Interaction 1*, 1 (2009), 13–32.
105. Quinn, A.J. and Bederson, B.B. Human computation: a survey and taxonomy of a growing field. *Proc. CHI '11*, (2011).
106. Raddick, J., Lintott, C., Bamford, S., et al. Galaxy Zoo: Motivations of Citizen Scientists. *Bulletin of the American Astronomical Society*, (2008), 240.
107. Rafaeli, S. and Ariel, Y. Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia. In A. Barak, ed., *Psychological Aspects of Cyberspace*. Cambridge University Press, New York, NY, 2008.
108. Raymond, E. The cathedral and the bazaar. *Knowledge, Technology & Policy 12*, 3 (1999), 23–49.
109. Redmiles, D. and Nakakoji, K. Supporting reflective practitioners. *Software Engineering, 2004. ICSE 2004. Proceedings. 26th International Conference on*, (2004), 688–690.
110. Reijers, H., Jansen-Vullers, M., Zur Muehlen, M., and Appl, W. Workflow management systems+ swarm intelligence= dynamic task assignment for emergency management applications. *Business Process Management*, (2007), 125–140.
111. Resnick, P. and Zeckhauser, R. Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. *Advances in Applied Microeconomics 11*, (2002), 127–157.
112. Rittel, H.W.J. and Webber, M.M. Dilemmas in a general theory of planning. *Policy sciences 4*, 2 (1973), 155–169.
113. Ritter, S., Anderson, J.R., Koedinger, K.R., and Corbett, A. Cognitive Tutor: Applied research in mathematics education. *Psychonomic bulletin & review 14*, 2 (2007), 249–255.
114. Rogers, Y. Exploring obstacles: integrating CSCW in evolving organisations. *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, (1994), 67–77.
115. Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., and Vukovic, M. An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media: Barcelona, Spain*, (2011).
116. Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., and Tomlinson, B. Who Are the Crowdworkers? Shifting Demographics in Amazon Mechanical Turk. *alt.chi '10*, ACM Press (2010).
117. Roth, A.E. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica 70*, 4 (2002), 1341–1378.
118. Roy, D.F. Work satisfaction and social reward in quota achievement: An analysis of piecework incentive. *American Sociological Review 18*, 5 (1953), 507–514.
119. Rzeszotarski, J. and Kittur, A. CrowdScape: interactively visualizing user behavior and output. *Proceedings of the 25th annual ACM symposium on User interface software and technology*, (2012), 55–62.
120. Rzeszotarski, J.M. and Kittur, A. Instrumenting the crowd: using implicit behavioral measures to predict task performance. *Proc. UIST '11*, (2011).
121. Sadlon, E., Barrett, S., Sakamoto, Y., and Nickerson, J.V. The Karma of Digg: Reciprocity in Online Social Networks. Paris, December 2008. *Proceedings of the 18th Annual Workshop on Information Technologies and Systems (WITS)*, (2008).
122. Savage, N. Gaining wisdom from crowds. *Communications of the ACM 55*, 3 (2012), 13–15.
123. Schmidt, K. and Bannon, L. Taking CSCW seriously. *Computer Supported Cooperative Work (CSCW) 1*, 1 (1992), 7–40.
124. Schroer, J. and Hertel, G. Voluntary Engagement in an Open Web-Based Encyclopedia: Wikipedians and Why They Do It. *Media Psychology 12*, 1 (2009), 96.
125. Shahaf, D. and Horvitz, E. Generalized task markets for human and machine computation. (2010).
126. Shamir, B. and Salomon, I. Work-at-Home and the Quality of Working Life. *The Academy of Management Review 10*, 3 (1985), 455–464.
127. Shaw, A.D., Horton, J.J., and Chen, D.L. Designing incentives for inexpert human raters. *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, (2011), 275–284.
128. Shen, M., Tzeng, G.H., and Liu, D.R. Multi-criteria task assignment in workflow management systems. *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on Social Sciences*, (2003).
129. Sheng, V.S., Provost, F., and Ipeirotis, P.G. Get another label? improving data quality and data mining using multiple, noisy labelers. *Proc. KDD '08*, ACM (2008), 614—622.
130. Silberman, M.S., Irani, L., and Ross, J. Ethics and tactics of professional crowdwork. *XRDS 17*, 2 (2010), 39–43.
131. Silberman, M.S., Ross, J., Irani, L., and Tomlinson, B. Sellers' problems in human computation markets. *Proc. HCOMP '10*, (2010).
132. Skillicorn, D.B. and Talia, D. Models and languages for parallel computation. *ACM Computing Surveys (CSUR) 30*, 2 (1998), 123–169.

133. Smith, A. The Wealth of Nations (1776). *New York: Modern Library*, (1937), 740.
134. Snow, R., O'Connor, B., Jurafsky, D., and Ng, A.Y. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. *Proc. ACL '08*, (2008).
135. Sorokin, A. and Forsyth, D. Utility data annotation with Amazon Mechanical Turk. *Proc. CVPR '08*, (2008).
136. Sterling, B. *A good old-fashioned future: stories*. Spectra, 1999.
137. Sterling, B. *Shaping things*. 2005.
138. Stohr, E.A. and Zhao, J.L. Workflow automation: Overview and research issues. *Information Systems Frontiers 3*, 3 (2001), 281–296.
139. Surowiecki, J. *The Wisdom of Crowds*. Random House, New York, 2005.
140. Tang, J.C., Cebrian, M., Giacobe, N.A., Kim, H.-W., Kim, T., and Wickert, D. "Beaker". Reflecting on the DARPA Red Balloon Challenge. *Communications of the ACM 54*, 4 (2011), 78.
141. Taylor, F.W. *The principles of scientific management*. Harper & Brothers, New York, 1911.
142. Trist, E.L. *The evolution of socio-technical systems: A conceptual framework and an action research program*. Ontario Quality of Working Life Center.
143. Van de Ven, A.H., Delbecq, A.L., and Koenig Jr, R. Determinants of coordination modes within organizations. *American sociological review*, (1976), 322–338.
144. Viswanath, B., Mondal, M., Clement, A., et al. Exploring the design space of social network-based Sybil defenses. *Proceedings of the 4th International Conference on Communication Systems and Network (COMSNETS'12)*, (2012).
145. Wang, G., Wilson, C., Zhao, X., et al. Serf and Turf: Crowdturfing for Fun and Profit. *Arxiv preprint arXiv:1111.5654*, (2011).
146. Weld, D.S., Adar, E., Chilton, L., Hoffmann, R., and Horvitz, E. Personalized Online Education—A Crowdsourcing Challenge. *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*, (2012).
147. Welinder, P., Branson, S., Belongie, S., and Perona, P. The multidimensional wisdom of crowds. *Neural Information Processing Systems 6*, 7 (2010), 1—–9.
148. Williamson, O.E. The economics of organization: The transaction cost approach. *American journal of sociology 87*, 3 (1981), 548–577.
149. Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N., and Malone, T.W. Evidence for a collective intelligence factor in the performance of human groups. *science 330*, 6004 (2010), 686–688.
150. Yu, H., Estrin, D., and Govindan, R. A hierarchical proxy architecture for Internet-scale event services. *Enabling Technologies: Infrastructure for Collaborative Enterprises, 1999.(WET ICE'99) Proceedings. IEEE 8th International Workshops on*, (1999), 78–83.
151. Yu, J. and Buyya, R. A taxonomy of scientific workflow systems for grid computing. *ACM Sigmod Record 34*, 3 (2005), 44–49.
152. Yu, L. and Nickerson, J.V. Cooks or cobblers?: crowd creativity through combination. *Proceedings of the 2011 annual conference on Human factors in computing systems*, (2011), 1393–1402.
153. Yu, L. and Nickerson, J.V. An Intenet-Scale Idea Generation System. *ACM Transactions on interactive Intelligent Systems 3*, 1 (2013).
154. Zhang, H., Law, E., Miller, R.C., Gajos, K.Z., Parkes, D.C., and Horvitz, E. Human Computation Tasks with Global Constraints. *Proc. CHI '12*, (2012).
155. Zhao, L., Sukthankar, G., and Sukthankar, R. Robust Active Learning Using Crowdsourced Annotations for Activity Recognition. *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*, (2011).
156. Zhu, H., Kraut, R.E., Wang, Y.-C., and Kittur, A. Identifying shared leadership in Wikipedia. *Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM (2011), 3431–3434.
157. Zittrain, J. Ubiquitous human computing. *Philosophical Transactions of the Royal Society A 366*, (2008), 3813–3821.
158. Zittrain, J. Human Computing's Oppenheimer Question. *Proceedings of Collective Intelligence*, (2012).

Consider the Redirect: A Missing Dimension of Wikipedia Research

Benjamin Mako Hill
University of Washington
Department of Communication
Seattle, WA 98195
makohill@uw.edu

Aaron Shaw
Northwestern University
Department of Communication Studies
Evanston, IL 60208
aaronshaw@northwestern.edu

ABSTRACT

Redirects are special pages in wikis that silently transport visitors to other pages. Although redirects make up a majority of all article pages in English Wikipedia, they have attracted very little attention and are rarely taken into account by researchers. This note describes redirects and illustrates why they play an important role in shaping activity in Wikipedia. We also present a novel longitudinal dataset of redirects for English Wikipedia and the software used to produce it. Using this dataset, we revisit several important published findings about Wikipedia to show that accounting for redirects can have important effects on research.

Categories and Subject Descriptors

H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work—*web-based interaction*

General Terms

Wikipedia, Peer Production, Redirects, Wikis

1. INTRODUCTION

A feature of most wikis, "redirects" are special pages that silently transport visitors to other pages. In Wikipedia, the only indication that one has visited a redirect is that the page title and the URL in the browser are different and a very small hyperlinked message appears near the article title (see Figure 1). Clicking on this link will take the user to the redirect page itself. Redirects in Wikipedia are normal pages that begin with "`#redirect [[Target]]`" where "`Target`" is the page to which visitors will be redirected. Although redirect pages can contain extensive text, their content is almost never viewed and very rarely edited. Despite their near-invisibility, redirects play an important role in shaping activity in Wikipedia. Redirects are a majority of all article pages in English Wikipedia and are viewed millions of times each month. They represent a central form of the encyclopedia's "hidden order" [7] and contribute to wikis' usability and user experience.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
OpenSym '14 , Aug 27-29 2014, Berlin, Germany
ACM 978-1-4503-3016-9/14/08.
http://dx.doi.org/10.1145/2641580.2641616



Figure 1: An example of a redirect in English Wikipedia. Note the small redirect notice below the title.

That said, redirects have attracted very little attention from researchers studying Wikipedia and are, with rare exceptions (e.g., [3, 8]) rarely discussed explicitly in the analysis of Wikipedia data. In this note, we make several contributions: First, we introduce a longitudinal database that makes it easier to study redirects in English Wikipedia over time and use this database to characterize the enormous volume of activity around redirects. Then, we use the database to illustrate the importance of considering redirects in two relationships of central interest to many researchers: (1) the distribution of edits over articles and (2) the relationship between views and edits. We conclude with guidance for how researchers should account for redirects in future work.

2. DESCRIBING REDIRECTS

Redirects constitute a key piece of Wikipedia's invisible infrastructure that supports the broader usability of the site. They are also an example of how some of the least visible activities and elements of wikis lend structure and salience to visitors' experience.

Redirects connect many types of pages. For example, redirects link alternative spellings or transliterations (e.g., from "Mohammed" → "Muhammad"), common abbreviations (e.g., "NYC" → "New York City"), translations (e.g., "Nueva York" → "New York City") or common misspellings (e.g., "Barrack Obama" → "Barack Obama"). Plural nouns will often redirect to the singular form (e.g., "Frogs" → "Frog"). Like many other wiki platforms, MediaWiki (the software that runs Wikipedia and other Wikimedia Foundation (WMF) Wikis) treats articles with differences in case and spacing as distinct. As a result, many redirects exist between articles that only differ in capitalization and white-space (e.g., "Newyorkcity" and "New york city" → "New York City").

Redirects perform several critical functions within Wikipedia [3]. They facilitate easy hyperlinking by creating targets from alternative names and spellings. They facilitate moving pages and other maintenance by ensuring that the flow of traffic throughout the wiki is uninterrupted. They provide common shortcuts for discussing content or policies. They help search engines – including Wikipedia's internal search engine – locate the correct information for a particular search term. They also contribute to the options in Wikipedia's auto-complete search box.

The creation and maintenance of redirects are examples of Wikipedia's hidden order because, like other pages in Wikipedia, redirects are products of collaborative effort. Wikipedia editors create redirects and then update them as pages move and redirects' targets change. Controversial redirects are discussed in a "Redirects for Discussion" page,[1] where both likely outcomes (i.e., decisions to delete or to change the target of a redirect) will be invisible. When an article is renamed, MediaWiki automatically creates a redirect in place of the old article so that existing hyperlinks still function. Like other pages, redirects are placed in categories, discussed on talk pages, and tagged with templates.

Similarly invisible aspects of Wikipedia and work in wiki communities have attracted scholarly attention. Early research into Wikipedia treated the encyclopedia's "hidden order" as the behind the scenes coordination and discussion that editors engaged in to create the public-facing pages that most visitors to the site interact with [7, 6]. Related work examined invisible and peripheral forms of participation in Wikipedia and other online communities, demonstrating that these forms of labor are nonetheless legitimate modes of collaboration (e.g., [1, 2]). More recent studies have examined how different kinds of labor can give rise to distinct "social roles" and types of "wikiwork" [5, 9]. Redirects offer another example of how invisible and seemingly peripheral modes of participation in social computing systems perform valuable socio-technical functions.

3. LONGITUDINAL REDIRECT DATASET

One challenge with studying redirects in Wikipedia is that common sources of Wikipedia data treat redirects differently. For example, many requests to the MediaWiki API[2] will automatically follow redirects. Although the API indicates the presence of a redirect in metadata, it can effectively hide redirects from researchers. Other widely used data sources, like WMF's database dumps and page view data,[3] present redirects as normal pages, inviting different methodological and substantive problems [3]. The WMF has, at least since 2013, published a database of redirects as an SQL "dump" file for every wiki they host. These data reflect a cross-sectional snapshot of the redirect network within the wiki at the time that the dump is produced. Unfortunately, cross-sectional data on redirects is insufficient for many purposes because Wikipedia's redirect network changes over time as editors make changes to redirects and their targets.

For example, the Wikipedia article on "NA" was initially created as a disambiguation page in January, 2004, edited 26 times, and then turned into a redirect to "Na" in February, 2006 when the "NA" and "Na" disambiguation pages were merged. Similarly, redirects can be turned into normal articles. For example, articles about characters in popular television series will often begin as redirects to the article about the show. When coverage of characters in the show's article grows, the redirects can be turned into stand-alone articles. Even within articles that are stable as redirects, targets can change over time (see Table 1 for an example).

Comprehensive dynamic data on redirects in Wikipedia is difficult to create because it requires looking within the text of the revisions of all articles. This full text history is currently several terabytes of uncompressed XML data and can require hundreds of hours of computer time to process. Using a research computing cluster at the Harvard MIT Data Center (HMDC), we parsed the full text revision history of English Wikipedia published in October, 2012, to create a list of all revisions that redirect. For each redirect, we recorded the revision timestamp as well as the page id, title, and the target.

From this process, we created a dataset of *redirect spells*. Each "spell" includes the title of the redirect or (i.e., the page that is being redirected from) the target of the redirect (i.e., the page being redirected to) and the start and end timestamps for the spell. This dataset includes 9,277,563 observations. Many redirect spells (67%) did not end (i.e., pages remained redirects at the time that the data was collected) so the "end" date is censored and marked as missing for these spells. Examples of spells for one page are shown in Table 1. In the example spells in the table, a single redirect was (1) created, (2) changed to redirect to a different target, and finally (3) changed back to the first target until at least the point of data collection.

We have published our dataset freely for other researchers under the same license used for all Wikipedia content. Because WMF continues to publish new dumps of Wikipedia – and because all other MediaWiki wikis use redirects – we have also made all of our software available for free under the GNU GPLv3. Both code and dataset are available on our website.[4]

4. REDIRECTS IN WIKIPEDIA

Our dataset contains details on 6,201,873 redirect pages representing 22% of pages hosted in English Wikipedia. This includes all of Wikipedia's "namespaces" including pages for administrative work, discussion, and file hosting where redirects are not widely used. The vast majority of redirects (5,318,869 or 86%) are in the public-facing article or "main" namespace. Within the article namespace, a majority of pages (55%) are redirects. As shown in Figure 2, the number of redirects in Wikipedia has grown in a nearly linear pattern over time.

The dataset suggests that redirects are indeed dynamic. Redirect spells last as little as a fraction of a second and as long as 10 years with a mean of 363 days (Median: 122, SD:

[1] http://enwp.org/WP:RFD
[2] https://www.mediawiki.org/wiki/API
[3] http://dumps.wikimedia.org/
[4] http://networkcollectiv.es/wiki-redirects/

	Page ID	Page Title	Target	Start	End
1	4151445	Kirtland Perky	Kirtland I. Perky	2006-02-21 17:18:48	2010-10-19 06:14:26
2	4151445	Kirtland Perky	Kirtland Irving Perky	2010-10-19 06:14:26	2011-06-03 05:11:13
3	4151445	Kirtland Perky	Kirtland I. Perky	2011-06-03 05:11:13	<NA>

Table 1: Example redirect spells for one article in our dataset. <NA> indicates that the data is right censored because the spell was ongoing at the point of data collection.

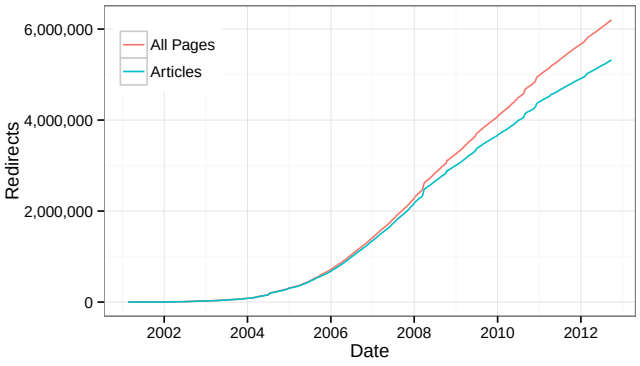


Figure 2: Count of redirects in English Wikipedia over time for all pages and within the article namespace only.

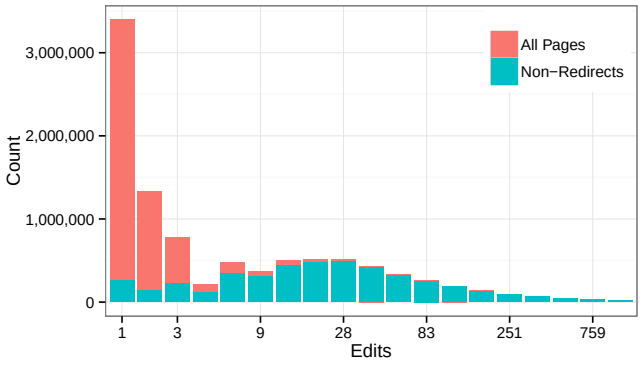


Figure 3: Histograms of pages in the article namespace based on number of edits on a log scale. The red histogram includes all pages in the article namespace ($N = 9,729,989$). The blue graphs includes only pages that were not redirects at the point of data collection ($N = 5,327,561$).

512). As we have noted, 67% of spells in the dataset were ongoing at the point of data collection. The dynamic dataset also shows that redirects are frequently shifting. 3.2% of articles (1.4% of all pages) were once redirects but were not at the time of our data collection. 24% of articles that have redirected once have experienced multiple spells and one administrative page experienced 713 different spells.

5. IMPACT OF REDIRECTS

The fact that such a large proportion of pages in English Wikipedia are redirects has important implications for Wikipedia research. That said, little previous research has explicitly taken redirects into account. For example, work examining the distribution of edits across different types of pages (e.g., [4]) should exclude or adjust for redirects because they are substantively different than other pages and systematically less visible to editors. Likewise, research analyzing the distribution of links or edits over pages (e.g., [8]) should adjust for the presence of redirects because they are systematically less likely to incorporate links or be edited.

In general, since redirects are both unfamiliar and invisible to the vast majority of wiki readers, it may make sense to either remove redirects entirely from analyses or to collapse activity on redirects with activity on their target pages. While this can be difficult for longitudinal analyses (due to the dynamic nature of redirects), it yields substantively different results under a wide variety of circumstances. We illustrate this in Figure 3 where we plot the distribution of edits per article as of October, 2012 with and without redirects. Because most articles are redirects and most redirects are edited very infrequently, the two distributions suggest radically different distributions of edits across pages. This has an important impact both on the methods researchers use to model Wikipedia activity and on our substantive understanding of the ways editors distribute their work across pages.

A growing body of studies using Wikipedia viewership data presents another area in wiki research where redirects are important but overlooked. Because viewers don't see redirects, viewing a redirect is substantively different from viewing a normal page. For example, if a user visits the article on "Seattle, Washington", this will be recorded as a view to the redirect even though the target article "Seattle" is displayed. In this sense, views of redirects will tend to be overcounted while views of target articles will tend to be undercounted. If a visitor subsequently presses the edit button on the top of the page, they will edit the target.

In a highly-cited article, Priedhorsky et al. combine Wikipedia view data with edit data and do not mention accounting for redirects [6]. Although not a central finding in their paper, the authors note with surprise that they find a weak correlation between edits and views. Because redirects are edited infrequently but "viewed" as often as millions of times per month each, redirects may be contributing to the surprisingly low correlation between edits and views noted by Priedhorsky et al. and others.

We use the closest full month of public viewership data to the period studied by Priedhorsky et al. (January, 2008) to revisit their finding with redirects in mind. First, we create cumulative tabulations of views (over January, 2008) and edits (over all time) for every page in English Wikipedia

Type of Views	Correlation statistic	
	Pearson (log)	Spearman
Articles		
Raw	0.67	0.64
Redirect-adjusted	0.81	0.79
All pages		
Raw	0.61	0.5
Redirect-adjusted	0.74	0.65

Table 2: Correlation between edits and views for pages in English Wikipedia in the article namespace ($n = 9,729,989$) and for all pages ($n = 28,397,622$). Raw edits reflects the views to articles as recorded in public page view data. Adjusted edits "shift" views from redirects to their targets. Pearson correlations are of log transformed edits and views.

with at least one edit. We then calculate the proportion of each month that each page redirected to each of its targets. Next, we multiply these proportions with counts of views for the redirects to create an estimate of the views made to pages while they were redirects. Finally, we subtract these products from the views tabulated for each redirect page and add them to the redirects' targets. For example, if the article named "Seattle, Washington" was viewed 100 times in January, 2008 and was a stable redirect to the article named "Seattle" for the entire month, we would shift all of the redirect's views to "Seattle." However, if the redirect was a stand-alone article for two weeks out of the month, we would shift only half of the views.

Table 2 shows that adjusting views for redirects substantially increases the correlation between edits and views. The Pearson correlation between log views in January, 2008 and log edits made to articles through that period increases from $\rho = 0.67$ to $\rho = 0.81$ when we consider redirects. We see a similar pattern and magnitude of increase in the strength of correlations using the full dataset of pages in Wikipedia and the non-parametric rank-based Spearman correlation.

6. CONCLUSIONS

Redirects constitute an important component of the structure of wikis. Researchers must reckon with redirects more carefully and explicitly in order to accurately represent the structure of wikis as well as the flow of work and attention of contributors and viewers. As a hidden but central aspect of Wikipedia's infrastructure, redirects deserve deeper and more sustained attention in their own right. For example, redirects are an untapped source of dynamic network data within Wikipedia. They are also a distinct form of contribution and viewership behavior that previous approaches have not explored in depth.

Future research can utilize the database of redirects we have created and the code used to generate it. For example, these data could be used in network analyses involving relational models of pages, articles, edits, or views to account for redirects. Likewise, studies considering the dynamic relationship between viewership and editorship will gain precision by incorporating redirects into their analysis. In addition, the large body of work mining Wikipedia's link and knowledge structure may uncover novel relationships and dynamics by considering redirects. This list is not exhaustive, but suggests situations in which including redirects in quantitative analysis may alter the magnitude, if not the direction, of research findings. Above all, we hope this note inspires future research to handle redirects in a more explicit and consistent manner. Given how little-analyzed and poorly-understood redirects have been, further work will be necessary to discover the places where redirects enhance our understanding of Wikipedia and other wikis.

7. REFERENCES

[1] J. Antin and C. Cheshire. Readers are not free-riders. *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW '10)*, pages 127–130, 2010.

[2] S. L. Bryant, A. Forte, and A. Bruckman. Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, GROUP '05, page 1–10, New York, NY, USA, 2005. ACM.

[3] B. J. Hecht. *The Mining and Application of Diverse Cultural Perspectives in User-Generated Content*. Ph.D. dissertation, Northwestern University, Evanston, IL, 2013.

[4] A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi. He says, she says: conflict and coordination in wikipedia. In *Proc. CHI '07*, 2007.

[5] D. W. McDonald, S. Javanmardi, and M. Zachry. Finding patterns in behavioral observations by automatically labeling forms of wikiwork in barnstars. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, page 15–24, New York, NY, USA, 2011. ACM.

[6] R. Priedhorsky, J. Chen, S. T. K. Lam, K. Panciera, L. Terveen, and J. Riedl. Creating, destroying, and restoring value in wikipedia. In *Proceedings of the 2007 international ACM conference on Supporting group work*, GROUP '07, page 259–268, New York, NY, USA, 2007. ACM.

[7] F. B. Viégas, M. Wattenberg, M. M. McKeon, and D. Schuler. The hidden order of wikipedia. In *Online Communities and Social Computing*, HCII, page 445–454. Springer-Verlag, Berlin, Heidelberg, 2007.

[8] J. Voss. Measuring wikipedia. In *Proceedings of the International Conference of the International Society for Scientometrics and Informetrics*, Stockholm, Sweden, 2005.

[9] H. T. Welser, D. Cosley, G. Kossinets, A. Lin, F. Dokshin, G. Gay, and M. Smith. Finding social roles in wikipedia. In *Proceedings of the 2011 iConference*, iConference '11, page 122–129, New York, NY, USA, 2011. ACM.

JOURNAL OF COMMUNICATION

Journal of Communication ISSN 0021-9916

ORIGINAL ARTICLE

Laboratories of Oligarchy? How the Iron Law Extends to Peer Production

Aaron Shaw[1] & Benjamin M. Hill[2]

1 Department of Communication Studies, Northwestern University, Evanston, IL 60208, USA
2 Department of Communication, University of Washington, Seattle, WA 98195, USA

Peer production projects like Wikipedia have inspired voluntary associations, collectives, social movements, and scholars to embrace open online collaboration as a model of democratic organization. However, many peer production projects exhibit entrenched leadership and deep inequalities, suggesting that they may not fulfill democratic ideals. Instead, peer production projects may conform to Robert Michels' "iron law of oligarchy," which proposes that democratic membership organizations become increasingly oligarchic as they grow. Using exhaustive data of internal processes from a sample of 683 wikis, we construct empirical measures of participation and test for increases in oligarchy associated with growth in wikis' contributor bases. In contrast to previous studies, we find support for Michels' iron law and conclude that peer production entails oligarchic organizational forms.

doi:10.1111/jcom.12082

Commons-based peer production—the distributed creation of freely accessible information resources through the mass aggregation of many contributions—represents a model of collective action and public goods production that integrates the use of digital communication networks and information technologies (Benkler, 2006). Peer production has generated public goods of enormous economic, cultural, and political value—such as Wikipedia and GNU/Linux—and has transformed the way that firms in many industries do business, shifted how politicians campaign for office, and changed the way that individuals share information (Benkler, 2006; Karpf, 2012; Shirky, 2008). Much of the excitement around peer production stems from the egalitarian principles linked to its organization, to the underlying network communication technologies at its core, and to suggestions that "communal information goods" (Fulk, Flanagin, Kalman, Monge, & Ryan, 1996) will tend to occur in groups with relatively equitable patterns of contribution and less hierarchical organizational forms (Connolly & Thorn, 1990; Rafaeli & LaRose, 1993).

According to various accounts, peer production projects function as novel forms of participatory organization with a broad democratizing potential (Benkler, 2006;

Corresponding author: Benjamin M. Hill; e-mail: makohill@uw.edu, mako@atdot.cc

Bennett & Segerberg, 2013; Castells, 2012; Fuster Morell, 2012). In addition, peer production communities have inspired a wave of associations, movements, and theorists to embrace networked technologies as tools for creating participatory democratic organizations. However, we also know that many successful peer production projects exhibit strong inequalities of participation and deeply entrenched leadership. Contributions to peer production follow a "power law" distribution where a small group of participants make an enormous proportion of contributions (Healy & Schussman, 2003; Ortega, 2009; Viégas, Wattenberg, McKeon, & Schuler, 2007) and many peer production projects, such as Linux and Ubuntu, are explicitly undemocratic, going so far as to describe their leaders as "benevolent dictators for life" (Hill, Burger, Jesse, & Bacon, 2008; Ingo, 2006). Although unequal participation and entrenched leaders do not necessarily mean that peer production projects are undemocratic, this evidence suggests that they may reproduce a pattern of behavior more consistent with Robert Michels' "Iron Law of Oligarchy" (1915), which proposes that as voluntary movements and membership organizations become large and complex, a small group of early members consolidate and exercise a monopoly of power within the organization as their interests diverge from the collective's.

We offer an empirical test of the assumption, central to previous theory, that large peer production projects resist increasing levels of organizational oligarchy. Using exhaustive longitudinal data of internal processes drawn from 683 wikis that have grown large and complex, we adapt Michels' iron law to the context of peer production communities and construct a series of hypothesis tests. In contrast to previous studies of communal public goods production in online environments (Rafaeli & LaRose, 1993) as well as ethnographic findings on Wikipedia from Konieczny (2009), we present quantitative evidence in strong support of Michels' iron law in peer production. As the peer production communities in our sample attract contributors, a small group of leaders, present at the beginning, tend to consolidate power as their interests diverge from those of other participants. These findings deviate from much previous literature on online organization, peer production, and communal information goods.

Wikis and Wikia Inc.

Our empirical setting is a large sample of peer production communities engaged in the collaborative creation of *wikis*. The term "wiki" refers to software designed to facilitate the collaborative, asynchronous creation and distribution of textual content. It also refers to the communities that use wiki software and to the products created by these groups (Leuf & Cunningham, 2001). Wikipedia is the most famous example of a wiki, but there are hundreds of thousands of other wikis with different goals, topics, and scopes. Through the enormous success of Wikipedia—one of the five most popular websites in the world—wikis are perhaps the most visible and successful modality of peer production.

Along with free/libre/open-source software (FLOSS), wikis figured prominently into Benkler's (2002) original conceptualization of commons-based peer production

and are widely cited as an archetype of peer production's new model of economic production, political participation, and information sharing. Despite their novel form, peer production communities like wikis encompass many characteristics that make them comparable to other kinds of voluntary associations. Many of them articulate explicitly democratic organizational ideals: They attempt to maintain member governance, they rely on attracting new members and leaders from within their membership, and they showcase effective solutions to collective action problems.

We analyze a population of peer production wikis hosted by the for-profit firm *Wikia*. Several aspects of Wikia make it an ideal setting in which to compare organizational governance in peer production communities. Wikia, founded in 2004, sought to apply the Wikipedia model of peer production beyond the education-based scope of the Wikimedia Foundation (the organization that supports Wikipedia). Wikia was founded by Jimmy Wales, Wikipedia's founder, and Angela Beesley, one of the most active and respected contributors to Wikipedia in its early years. Wikia's policies, structures, and technologies have been heavily influenced by Wikipedia. Although many firms host wikis (e.g., PBWiki, WikiSpaces, and SocialText), Wikia is unique in that it only hosts publicly accessible, volunteer-produced, peer production projects and never restricts viewership.

Wikia relies on peer production to create the content of its websites. As a result, Wikia does not restrict participation in content contribution except to combat spam or vandalism. Like Wikipedia, anybody can create an account on any Wikia wiki. The vast majority of these wikis allow contributions even without accounts. Like Wikipedia, FLOSS, and other peer production projects, all wiki content is distributed freely to the public.[1] These factors help ensure that Wikia wikis remain open and accessible, allowing communities to adopt more or less participatory and democratic organizational behavior.

In terms of the content and scope, Wikia wikis vary enormously, addressing popular culture and "fan culture" topics as well as subjects such as subcultures, software, food, fashion, and more. Some of the largest wikis host information about massive multiplayer online video games such as Halo or World of Warcraft, television shows such as *Lost*, movies and novels such as *Lord of the Rings*, and information about the academic job market. Taken together, Wikia wikis constitute one of the largest, broadest, and most inclusive populations of peer production communities.

Peer production as democratic organization

A growing body of communication research seeks to evaluate the possibilities for participatory democracy and collective action in an era of digital networks and online organization (Bennett & Segerberg, 2012; Bimber, Flanagin, & Stohl, 2012; Castells, 2012; Polletta, 2013). A chorus of scholars have described networked collectives, including wikis, as deeply democratic (Benkler, 2006; Earl & Kimport, 2011; Fuster Morell, 2010, 2012; Hess & Ostrom, 2011; Karpf, 2012; Kollock, 1999; Tufekci & Wilson, 2012; Weber, 2004). These claims look to peer production projects as a digital vanguard among networked organizations. One popular narrative proposes

that technologically savvy early adopters using the Internet as a tool for encyclopedia writing and software production discovered powerful new mechanisms for avoiding top-down bureaucratic control without sacrificing the quality of their products or processes (Benkler, 2006; Castells, 2012; Shirky, 2008).

This optimistic view of peer production finds support in several bodies of theoretical work in communication. First, research has linked the democratizing effects of new communication technologies to the production of specific types of "communal" and "connective" public goods such as discretionary databases of public information, libraries, census records, and community bulletin boards (Fulk et al., 1996). By the logic of these claims, the production of communal public goods in digital environments occurs more effectively and efficiently in large, egalitarian collectives (Connolly & Thorn, 1990; Fulk et al., 1996; Rafaeli & LaRose, 1993).[2] Second, although focused on a social movement context, Bennett and Segerberg's "logic of connective action" suggests that networked modes of collaborative, voluntary association driven by personalized used of technology-mediated communication—including those pursued in peer production communities—will entail flatter, less hierarchical organizational forms (Bennett & Segerberg, 2013). Finally, within the "collective action space" of Bimber et al. (2012), the members of peer production communities embrace "entrepreneurial" (i.e., less bureaucratic and routinized) modes of engagement along with more "personal" (i.e., less formal and instrumental) modes of interaction. Bimber et al. contend that such entrepreneurial-personal organizations tend to be relatively flat and are managed through informal, discursive norms.

Across the organizational communication literature, these theories predict that peer production communities ought to possess participatory, democratic characteristics.

Peer production and Michels' iron law of oligarchy

At the same time, empirical research on leadership, governance, and participation in peer production communities suggests a much more complicated reality than narratives emphasizing nonhierarchical and nonbureaucratic organization. First, nearly every population of peer production projects studied follows a "power law" distribution of contributions across contributors (Ortega, 2009; Schweik & English, 2012; Wu, Wilkinson, & Huberman, 2009). Second, there is some evidence that the informal hierarchies of attention, status, and influence that arise in large, successful, peer production projects cohere into formal and increasingly rigid bureaucratic structures that may not facilitate participatory democracy. For example, radical inequalities of status and participation are pervasive in Wikipedia (Kittur, Chi, Suh, & Mytkowicz, 2007; Loubser, 2010; Panciera, Halfaker, & Terveen, 2009; Priedhorsky et al., 2007; Viégas et al., 2007) and studies of the internal governance of online collectives indicate that hierarchical institutions exist, even if those institutions are not accompanied by formal bureaucratic structures (Black, Welser, Cosley, & DeGroot, 2011; Butler, Joyce, & Pike, 2008; Forte, Larco, & Bruckman, 2009; O'Mahony & Ferraro, 2007;

Shaw, 2012; Zhu, Kraut, Wang, & Kittur, 2011). Although extreme hierarchy, bureaucracy, and inequality do not rule out the possibility that peer production projects are democratic, these attributes are not typical of participatory organizations.

A related set of predictions derive from the literature on voluntary organizations as well as some of the previous studies on peer production. The "iron law of oligarchy," one of the most influential theories of governance in voluntary organizations that was developed by the German sociologist and political theorist Robert Michels (1915), encapsulates these perspectives. Michels sought to explain why democratically run political parties in Italy around the turn of the 20th century became less democratic as they grew larger and more complex.[3] His "iron law" proposes that two patterns drive increasing oligarchy within voluntary organizations as they grow: (a) structurally, the group becomes an increasingly formal and complex organization with a small group of professional leaders who exercise a monopoly over the mechanisms of authority and (b) these leaders develop independent interests in the preservation of the organization itself, resulting in the transformation of the goals and activities of the organization in ways that diverge from the interests of members. Following previous work, we refer to these distinct dynamics as the "structural" and "goal transformation" components of the iron law (Jenkins, 1977; Leach, 2005; Voss & Sherman, 2000).[4]

Previous studies of movement and voluntary organizations — in particular Lipset, Trow, and Coleman (1956), Piven and Cloward (1977), Rucht (1999), and Voss and Sherman (2000) — guide our operationalization of Michels' two components of oligarchy in the context of peer production projects. Structurally, the governance and leadership of oligarchic organizations must reside in the hands of a stable, entrenched minority that exercises dominant control over organizational resources and policy (Lipset et al., 1956). Oligarchic goal transformation occurs when organizational leaders and elites develop interests that diverge from those of other members and impose their interests on the rest of the organization. If it applies, Michels' iron law requires that both patterns hold across peer production projects as they grow over time.

Although distinct from oligarchy, broad evidence of hierarchy, bureaucracy, and inequality in peer production suggests oligarchic leadership may prevail. Additionally, unlike political parties or labor unions, wikis, in general, are not formed with the objective of facilitating political or economic representation of their members and, as a result, may be less committed to democratic governance. Finally, given the ease with which volunteer contributors may choose to "exit" the community, some previous work on voluntary associations would also predict that less democratic governance arrangements would prevail (Barakso & Schaffner, 2008).

On the other hand, the organizational form and governance of peer production projects in general, and wikis in particular, possess few technical or physical constraints. Not only does wiki software support widely distributed contributions, it also supports widely distributed authority and governance. For example, there is no limit on the number of formal leaders ("administrators") a wiki can have and no material cost to adding them. Indeed, in previous research, wikis have been considered some of the most egalitarian and democratic peer production projects in terms of their

opportunities for lateral authority and resistance to domination by a small minority (Konieczny, 2009). If peer production projects are, as some suggest, intrinsically supportive of participatory democratic organizational forms, there is reason to believe that wikis will be among peer production's most democratic organizations.

Testing for the iron law in wikis

Despite the broad interest in peer production as democratic organization, little previous empirical research on peer production has formally tested such claims about projects' organizational governance and several studies point to the need for more comparative analysis (Butler, 2001; Fuster Morell, 2010; O'Mahony & Ferraro, 2007; Shaw, 2012). Also, even though theoretical and empirical work asserts that organizational democracy and egalitarian participation will prevail in communities engaged in communal public goods production, concepts of organizational democracy are often left unspecified. For our purposes, organizational democracy consists in the active participation of multiple constituencies in the negotiation and exercise of legitimate authority within an organization. This definition derives from previous research on voluntary, membership, and movement organizations engaged in collective action of various kinds (Andrews, Ganz, Baggetta, Han, & Lim, 2010; Lipset et al., 1956; Ostrom, 1990; Voss & Sherman, 2000) and builds on foundational studies of public goods production and common pool resource management (Hardin, 1968; Olson, 1965; Ostrom, 1990). As with other translations of the iron law to organizations beyond political parties, oligarchy and democracy have distinct meanings in the context of peer production communities. Like previous work, we adapt and operationalize it to suit our empirical setting.

The structural component of the iron law concerns the distribution of formal authority within an organization. Wiki software facilitates role differentiation and systems of formal leadership. In most wikis, members can hold special privileges that include the ability to delete accounts and content, to lock and unlock pages for editing, and to block users. Other special privileges include access to special pages and to tools for removing spam and vandalism. Users with additional privileges are referred to as "administrators" within wiki communities. As individuals occupying structural positions of authority, administrator status provides an avenue to test for the structural component of the iron law. Our first hypothesis tests for structural oligarchy by considering the relationship between the distribution of administrator status and membership size: *The probability of adding new administrators declines as wikis' contributor bases grow (H1)*.

Although administrator status confers authority and special privileges, the practice of governance and leadership in wikis consists largely of the exercise of informal authority and the performance of tasks that loosely resemble the work of research editors and debate moderators (Black et al., 2011; Forte et al., 2009; Zhu et al., 2011). So although only an administrator can delete a page or ban a user, nonadministrators can, and frequently do, contribute to administrative pages in ways that may traditionally be considered part of bureaucratic administration of the

community (Butler et al., 2008; Joyce, Pike, & Butler, 2012). Much of this governance activity happens in dedicated administrative pages that, in most cases, any wiki user can create or edit (Kriplean, Beschastnikh, McDonald, & Golder, 2007). Although not every wiki uses these administrative pages, they act as dedicated spaces for crafting rules, policies, and guidelines (e.g., rules for deleting articles) as well as an area in which policies are implemented (i.e., discussion and voting on the deletion of particular pages). Our second hypothesis looks to these pages to consider the relationship between the administrative elite and participation in administrative activity: *Controlling for the total number of contributions to administrative pages, administrators will contribute more to administrative pages as wikis' contributor bases grow (H2)*. This hypothesis also tests for the structural component of oligarchy, and in particular the idea that oligarchic leaders monopolize the mechanisms of organizational authority.

Of course, the presence of a consolidated elite may not be problematic when the elite represents the interest of the broader membership. Structural indicators of oligarchy, in this sense, only become a problem in the presence of goal transformation when leaders use their positions of authority to alter the priorities of the organization to suit their own agendas over those of the organization's members. Support for H2 may reflect a shift of activity of the elite into governance and away from the interests of the membership as expressed by how they contribute to the wiki. For more direct evidence of goal transformation, we can examine a subset of "reverts" (i.e., complete removals of contributions) performed by administrators to other editors. Reverts, a technique used to reject low-quality, spam, or otherwise unwanted contributions, reflect a way that administrators impose their vision of legitimate and valuable participation on others.[5] More importantly, reverts provide a clear indication of divergence between the interests of the reverters and the reverted. If a contributor makes a good faith edit and that edit is undone, this signals a difference between the reverting and reverted users in terms of what they believe the content on the website should be. Formally, we predict that: *Controlling for the total number of contributions made by experienced contributors, the number of reverts by administrators of such contributions will increase as wikis' contributor bases grow (H3)*.

Method

Data and measures

In contrast with previous studies assessing either the iron law or the organizational dynamics of communal public goods production, our analysis draws on the full observational records of intraorganizational behavior within a population of organizations. This type of "big data" affords us an opportunity to test hypotheses in ways that previous studies on these topics could not. Our initial dataset includes data on 76,473 wikis created before April 2010. It includes the full history of every revision, from both registered and nonregistered users, for every wiki hosted by Wikia from the time of Wikia's founding in 2004 until the point of data collection in April 2010. Because

Table 1 Summary Statistics for All the Wikis Included in Our Analysis. ($n = 683$)

Variable	Minimum	Median	Mean	Maximum	*SD*
Edits	644	13,438	53,306	2,303,248	161,652
Pages	183	3,167	11,152	1,270,640	53,025
Editors	69	218	787	68,222	3,457
Reverted edits	0	285	1,441	122,950	5,886
Administrators	0	7	11	247	18
Age (months)	6	46	50	74	11
Project edits	0	55	622	59,726	3,224
Experienced user edits	418	12,270	49,020	2,020,925	149,606

Wikia does not host private wikis or private data, these datasets were systematically made public for every Wikia wiki. Data for all wikis in our sample remain public and searchable on the web and new datasets continue to be available, upon request, for every wiki. Wikia wikis all run the same software (MediaWiki), which is developed in collaboration with the Wikimedia Foundation. The fact that all the wikis in the population use identical software makes it possible to derive a set of comparative metrics for analyzing organizational governance and activity.

Although scholars often treat all wikis as examples of peer production, there is a highly skewed distribution in activity among wikis; the vast majority of wikis, such as FLOSS projects (Healy & Schussman, 2003), are small and uncollaborative. Considering sample median values, an average wiki in our full dataset contained 225 contributions to 217 pages by six unique contributors at the point of data collection in 2010. Most of these wikis are not produced through the type of mass collaboration that Benkler (2002, 2006) calls peer production. For this reason, and because both theories and measures of oligarchy assume large, active groups, we limit our analysis to a subset of large wiki communities from this population.

To build this subset, we first ranked all 76,473 wikis in terms of the number of unique contributors. Next, we selected the top 1% of wikis in the full dataset (732 wikis). We removed a number of wikis because their database dumps included corrupted or invalid data (e.g., edits marked as occurring before the wikis were founded) leaving a total of 683 wikis in our final dataset. Summary statistics for this subset of wikis are reported in Table 1. This sample is an inclusive sample of peer production Wikia wikis in that the smallest of these wikis includes contributions from only 69 accounts. Because a 1% cutoff is still arbitrary, we also carried out our analysis using less inclusive subsets of only the top 300 and top 100 wikis. In both cases, we found results that were substantively similar.

This dataset consists of rich longitudinal records, which include every contribution made to each wiki, recorded with timestamps accurate to within 1 second. It includes 33,278,993 distinct contributions by 469,524 different contributors to 6,167,797 different wiki pages—more than 264 gigabytes of raw data. To create the

measures in our analysis, we wrote custom software to parse the database dump files and to query the Wikia application programming interface (API) for additional metadata. Because statistical power is not a concern and because discrete longitudinal analysis is more easily interpreted, we "bin" our data into week-long periods for this analysis. As a result, our analytic unit is the *wiki week* and our longitudinal dataset includes 146,858 such observations.

Because Michels' iron law is stated in terms of membership growth or organizational complexity, our primary independent variable must capture the size of a project's membership. For all our hypotheses, our primary question predictor, $accounts_{total}$, measures the number of unique registered accounts that have made at least one contribution to a wiki. Because wikis can grow older as well as more complex—and because our detailed data can disentangle these effects—we include another variable, *week*, that is expressed as the age of each wiki in weeks from the time when the wiki recorded its first contribution. As introduced above, we use "administrator status" as an indication of structural and formal authority of individuals within wikis. We calculate each individual's administrator status on a weekly basis for every *wiki week* in the dataset. Because this data is longitudinal as well, activity of a contributor in 1 week may be classified as coming from an administrator, whereas activity from the same user in another week may not.

To test our three hypotheses, we construct three dependent variables. For our first hypothesis (H1) that the probability of a community adding a new administrator will decline as wikis' contributor bases grow, we create a dichotomous variable that is coded "1" in a week if a community adds a new administrator and "0" if it does not. The addition of a new administrator constitutes a rare event and only occurred in 2,911 wiki weeks (2% of the wiki weeks in our sample).

To test H2 that structural oligarchy will also be associated with increased administrator activity in governance, we first identify a set of edits to administrative pages in each wiki. We take advantage of a feature of MediaWiki software called "namespaces," which allows pages to be categorized into one of several groups. One namespace common to all wikis is the "project" namespace, which, according to documentation, is designed to be used for "meta-discussions related to the operation and development of the wiki." By convention, this namespace is used for administrative and policy activity including policy and guideline creation and decision-making (Kriplean et al., 2007). As a first step, we identify a subset of edits to pages in the project namespace and record a count of these edits ($proj\text{-}edits_{week}$) for use as a control. As a second step, we identify the subset of edits made by administrators ($proj\text{-}edits\text{-}admin_{week}$) for use as an outcome. Some wikis in our sample do not use the project namespace (129 wikis or 19% of our sample), thus we exclude these wikis from our analyses when testing H2. Our results are substantively unchanged if they are included.

To test H3 that the number of reverts by administrators of edits made by experienced editors will increase as wikis' contributor bases grow, we must first identify reverted edits. We identify reverts as revisions that return an article to its state prior to

the last editor's contribution.[6] As noted above, reverts that administrators make to the contributions of other community members provide a rough measure of the extent to which administrator interests diverge from those members' interests. Many reverts reflect work undoing vandalism and spam. Although it is not possible to automatically categorize contributions as having been made in either good or bad faith, we can consider a contributor's total edit history to identify experienced users with a history of accepted contributions. Toward this end, we construct a variable $admin\text{-}reverts_{week}$ that is a count of the number of edits of experienced editors that were reverted by administrators in each week on each wiki. We limit this count to reverts of edits made by experienced editors because we do not want the measures to reflect reverts of spam and vandalism. We label contributors "experienced" who have made at least five non-reverted edits and have not had more than 25% of their total edits reverted. We also construct a measure of the number of edits by these experienced editors in each week ($expr\text{-}edits_{week}$) for use as a control.

Finally, we construct a baseline set of control variables. To capture differences in activity that may affect the current amount of administrative work, we construct measures of the number of unique contributors ($editors_{week}$) and the number of distinct contributions ($edits_{week}$). We also include controls that capture the state of the wiki including the number of distinct pages in the wiki ($pages_{total}$) and the total number of administrators ($admins_{total}$).

Analytic strategy

To test the three hypotheses described above, we construct longitudinal models to estimate the extent to which the wikis in our sample become more oligarchic as their contributor bases grow. Because our dataset includes repeated observations for each wiki, we are concerned about autocorrelation of residuals over time. Following Singer and Willett (2003), we use hierarchical linear models as a multilevel model for change and fit random intercepts for each wiki to cluster within-wiki variance in a compound error covariance structure.

As is typical in datasets from online communities, many of the variables capturing activities in wikis are highly skewed. Our measures of accounts, editors, pages, administrators, edits, administrator-reverts, edits by experienced editors, and administrative edits all follow this pattern. We take the natural logarithm of each, resulting in approximately normal distributions. Preliminary exploration of our data revealed that the relationship between time and our dependent variables is curvilinear. As a result, we include a quadratic specification of *week* in our models.

Each of our models is fit with a measure of oligarchy as its dependent variable and each model corresponds to one of our hypotheses. In Model 1 (M1), we use a multilevel logistic regression to estimate the probability of a new administrator being added. Models 2 and 3 are hierarchical linear models on different dependent variables: M2 is the log-transformed number of edits to administrative pages by administrators and M3 is the log-transformed number of reverts of experienced users by administrators. We use a base model in which every variable is measured at the level of the

wiki week and which includes a set of controls as well as our compound error term:

$$Y = \beta \ln accounts + \beta \ln week + \beta \ln week^2 + \beta \ln pages + \beta \ln admins + \beta \ln \text{editors} + \beta \ln edits + [u + E]$$

M2 estimates log edits by administrators on administrative pages and adds a control for the total amount of such activity (ln *proj-edits*). M3 estimates the log number of administrator reverts of experienced contributors' edits and includes a control for the total number of these edits (ln *expr-edits*).

Results

We find that as the wikis in our sample attract contributors, they display characteristics consistent with increases in oligarchy. Leadership roles remain concentrated among an extremely small group of elites as organizational membership grows, affirming H1. This group of administrators is increasingly active in administrative activity, controlling for the total amount of such work, consistent with H2. Finally, we observe that as membership grows, the number of reverts of experienced users by administrators tends to rise, supporting H3. In all three measures, wikis in our sample conform to the patterns predicted by Michels' iron law.

Figure 1 provides an example of a single wiki from our dataset that reflects the basic pattern demonstrated in our models. The data shown are from *Seattle Wiki*, a website that contains information about "places to go, things to do, community, organizations, happenings and events, resources, politics, and anything else related to the greater Seattle area."[7] Figure 1 plots cumulative counts of several variables. The visualizations show that the number of contributions (*edits*), the number of registered users that have made at least one contribution (*accounts*), and the number of edits to "project" pages have each increased steadily from the time the community was created in late 2004 through the point of data collection in 2010. That said, it is clear in the third panel that the number of administrators has grown much more slowly and that no administrators have been added since late 2005. The bottom panel shows that this small group of administrators has, over time, continued to revert contributions from experienced users, but that this activity has slowed.

Fitted regression models suggest a substantively similar pattern of associations and provide a set of formal tests of our hypotheses. Results from these models are shown in Table 2. In our first model (M1), we find strong support for H1 that large communities are less likely to add new administrators as they add new members. Indeed, we find that, *ceteris paribus*, one log-unit increase in the number of registered contributors is associated with odds of adding a new administrator that is only 0.81 times as high as they would be with the smaller contributor pool. This estimate reflects a marginal negative effect controlling for the number of administrators, the amount of activity, and the number of individuals active in each week. As expected, a large administrator corps is associated with a lower probability of adding a new administrator whereas

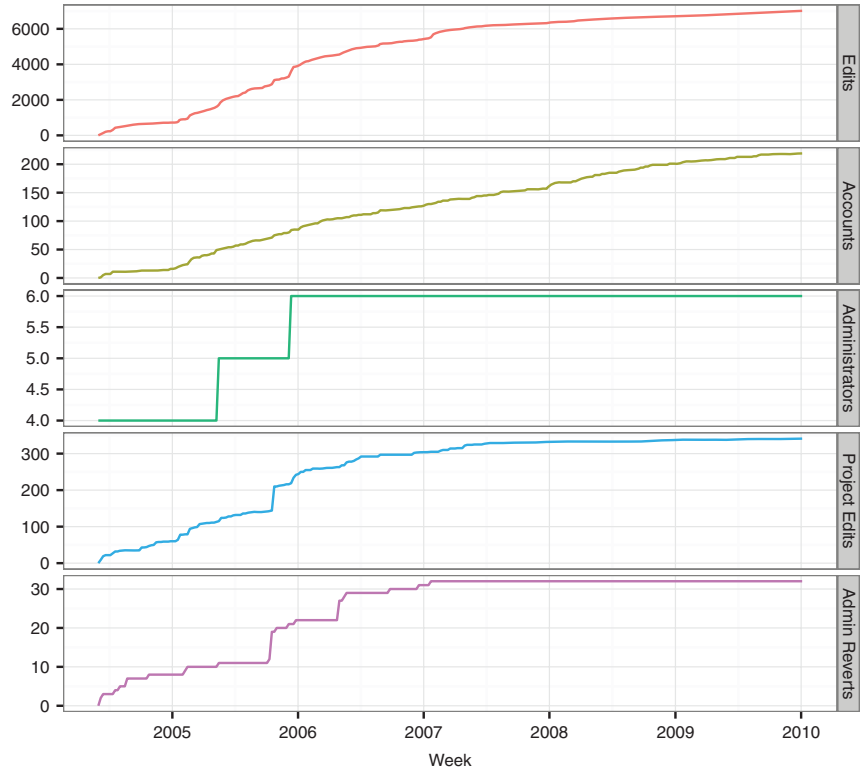


Figure 1 Cumulative plots of covariates for *Seattle Wiki*, a collaborative website for information about Seattle and one of the online communities in our dataset.

a large number of active editors in a given week are a positive predictor of a larger leadership body.

Our second model (M2) suggests support for H2 that as communities' contributor bases grow, the number of edits made by administrators to administrative "project" pages will increase. Because both our dependent and independent variables are expressed in natural log-transformed units, our parameters can be interpreted as elasticities. At the margin, we estimate that a 1% change in the total number of accounts on a wiki, controlling for the total edits to administrative pages, is associated with a 3% increase in the number of contributions made by administrators to the administrative parts of the wiki.

Our final model (M3) suggests support for H3 that as communities' contributor bases grow, the number of edits made by experienced contributors that are reverted by administrators also tends to grow. Again, the effect can be interpreted as an elasticity. At the margin, we estimate that a 1% change in the total number of accounts on a wiki, controlling for the total number of edits by experienced users in the wiki that week, is associated with a 5% increase in the number of contributions made by experienced editors being reverted by administrators.

Table 2 Table of Fitted Multilevel Regression Models. The unit of analysis in each case is the wiki week. M1 is a logistic regression model of the probability that a wiki will add a new administrator during a week. M2 is a linear model predicting the logged number of edits made by administrators on administrative "project" pages controlling for total edits to these pages. M3 is a linear model predicting the logged number of reverts of edits by experienced editors by administrators controlling for the number of edits by experienced editors. ($n = 146,858$ wiki weeks from 683 wikis.)

	M1	M2	M3
(Intercept)	-4.108^{***}	-0.138^{***}	-0.029^{*}
	(0.081)	(0.011)	(0.015)
week	-0.006^{***}	0.000^{***}	-0.002^{***}
	(0.001)	(0.000)	(0.000)
$week^2$	0.000^{***}	0.000^{***}	0.000^{***}
	(0.000)	(0.000)	(0.000)
ln $accounts_{total}$	-0.210^{***}	0.025^{***}	0.045^{***}
	(0.031)	(0.002)	(0.002)
ln $pages_{total}$	-0.763^{***}	-0.009^{***}	-0.021^{***}
	(0.023)	(0.001)	(0.002)
ln $admins_{total}$	0.666^{***}	0.070^{***}	-0.010
	(0.035)	(0.005)	(0.006)
ln $editors_{week}$	0.116^{**}	-0.036^{***}	0.125^{***}
	(0.038)	(0.002)	(0.003)
ln $edits_{week}$	0.996^{***}	0.010^{***}	-0.111^{***}
	(0.022)	(0.001)	(0.005)
ln $proj\text{-}edits_{week}$		0.608^{***}	
		(0.002)	
ln $expr\text{-}edits_{week}$			0.178^{***}
			(0.004)
Log likelihood	−11034.750	−38022.642	−103351.946
Number of observation	146,858	118,994	146,858
Number of groups: wiki	683	554	683
Variance: 1|wiki	0.325	0.034	0.080
Variance: residual	1.000	0.109	0.234

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

To assist in further interpretation and comparison, plots of predicted values for prototypical wikis for each of our three models are shown in Figure 2 along with 95% confidence intervals for the marginal effects. These model-derived plots emphasize what happens, on average, in our sample. The values along the *y*-axis correspond to estimated values of each of our dependent variables in our three models. Along the *x*-axis of each plot are a range of likely values of total registered accounts from 0 to 793 (the 95th percentile of observations in our dataset). These plots show the predicted values for wikis that differ in terms of the number of registered users with

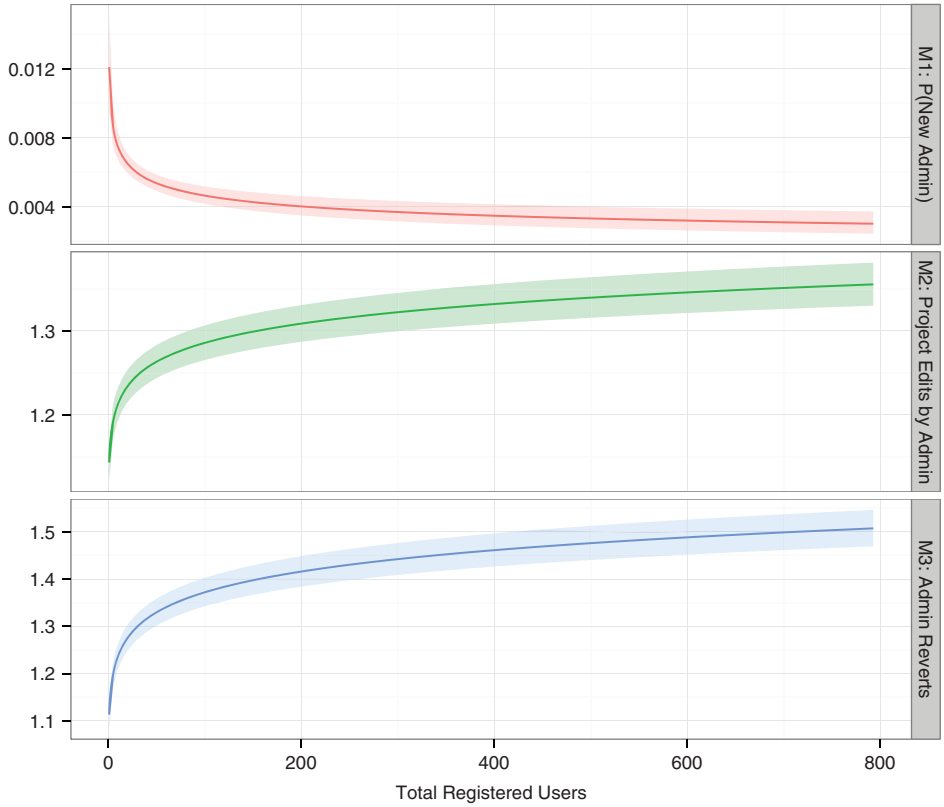


Figure 2 Plots showing predicted values from our models for wikis with varying number of accounts holding all other variables at sample medians. The graph also includes 95% confidence intervals for the marginal effects using the methods and tools described in Fox (2003). All outcome variables are measured in "per week" units.

at least one edit but are identical in every other respect. We have held each of our control variables constant at the sample median. As is the case in our models, each of these prototypical values should be interpreted in the context of a single week-long period. One important limitation is that each of these plots controls for time. As organizations grow larger and more complex, they also grow older. Although corresponding plots over time are not shown here, we find that the effect of age will, in most cases, compound the effects of organizational size. For example, older wikis are usually less likely to add administrators even if we hold the membership size constant.

Establishing the substantive significance of these results presents challenges for several reasons. Although M1 suggests new administrators become much less common in wikis with more contributors, the size of the estimated marginal effects in M2 and M3 for the prototypical wiki in Figure 2 is at the level of fractions of edits per week, and is estimated to be effectively flat over much of the variation

in community size in our sample. However, small movements in our indicators of oligarchy may have large effects. For example, Wikipedia researchers have shown that even a single revert can have powerful demotivating effects (Halfaker, Kittur, & Riedl, 2011) with broad implications (Halfaker, Geiger, Morgan, & Riedl, 2013). Then again, an increase of a single reverted edit every several weeks, or of a slightly larger share of administrative editing falling on administrators, might not be felt as a stark shift into oligarchy. Because of these challenges, and because Michel's iron law makes no attempt to quantify oligarchy, our hypotheses are framed in terms of directionality. Our results reflect consistent movement toward oligarchy in terms of our measures.

Discussion

The pattern of results in Table 2 and Figure 2 provides strong evidence that, on average, as wikis' contributor bases grow, a small group — present at the beginning — monopolizes positions of formal authority in the community and accounts for more administrative activity while also using their authority to restrict contributions from experienced community members. These findings contradict much theoretical and empirical scholarship suggesting that peer production communities will resist oligarchy and embrace participatory organizational practices. In the large sample of peer production communities we examined, wikis tend to reproduce undemocratic, noninclusive, organizational hierarchies. The wikis in our sample are not indicative of robustly democratic, participatory institutions. This is true despite the relative lack of formal bureaucratic structure or clearly defined roles within many wikis. These results are consistent with Michels' iron law of oligarchy and contradict prevailing notions regarding organizational democracy in peer production. Of course, our work is also subject to several threats and limitations.

Limitations

One important limitation concerns the effectiveness of the controls for the number of administrative edits (*proj-edits*) in M2 and contributions by experienced editors (*expr-edits*) in M3. Without these controls, more administrative edits or reverts by administrators might simply reflect an increase in administrative work and editing associated with a more active community. Although our controls for baseline activity in these two variables are among the strongest estimated effects in our models, they may still be insufficient. We gain confidence from several alternative specifications used as robustness checks included as an appendix in Table A1. In Model 4, we alter M2 to estimate the logged count of the number of edits to project pages by nonadministrators and find no statistically significant relationship to membership size.

Controlling for baseline project page activity, administrators increase their editing as wikis grow, whereas nonadministrators do not. An alternative approach to controls is to express our dependent variables as proportions. Models 5 and 6 in Table A1 are

specifications of M2 and M3, respectively, using proportions as a dependent variable instead of our activity-level controls.

Because proportions require a nonzero denominator, these datasets omit wiki weeks with no administrative edits or edits by experienced editors. These models show substantively similar results to M2 and M3: Membership size is positively associated with the proportion of contributions to administrative pages made by administrators and with the proportion of edits by experienced editors that are reverted by administrators.

Another potential threat to the internal validity of our findings concerns a key measure of oligarchic structure. In all three of our tests, our analysis relies on the administrator status as a way of identifying elites. However, as explained in our background section, the presence or absence of administrator rights is only one way of measuring elites in wikis. As wikis grow, contributors may take on the activity of much of the governance (i.e., creating, invoking, and enforcing rules) while merely relying on the small pool of administrators to carry out the final act of banning or deleting content. An alternate specification of M2 that predicts the logged number of editors contributing to project pages — ignoring whether they are administrators — can begin to address this concern. Estimates from this fitted model are shown in Model 7 in Table A1. In support of our finding in H2, we estimate a negative relationship between the number of editors contributing to project pages and membership size, controlling for activity in administrative pages. *Ceteris paribus* and administrative rights aside, fewer people are involved in governance activities as wikis attract contributors. Future work might use network measures or heuristics to identify a group of elite users without relying on administrator status.

Our analysis is also limited in ways that are more difficult to address. We employ "Big Data" methods from software engineering to construct our dataset, we test our hypotheses using more familiar regression-based tools. Although these methods are common in the social sciences, they import parametric assumptions about log-transformed data that can lead from subtle model misspecification to misleading findings. Our choice of methods was driven by the nature of our research questions, and we have built confidence in our results through alternative specifications and robustness checks, but we cannot eliminate these threats.

The breadth of our analysis and the size of our dataset signify that our measures do not include some of the nuance possible through other methodologies. Although we attempt to address qualities of individual communities through wiki-level random effects, our analysis is one of averages and cannot speak about differences that exist across the genres, languages, and demographic characteristics of the communities in our sample. By using data intensive techniques to look across millions of contributions and hundreds of communities, we lose the ability to speak in depth about some differences between and within those groups. We have spent time reading and contributing to some of the wikis in our dataset to build an intuition and familiarity with the users and communities in our data. That said, our measures necessarily provide an incomplete picture of a broad and rich dataset. The question of the extent to which

the patterns we observe constitute meaningful movement toward oligarchy requires further in-depth analysis of individual projects.

Generalizability

As is always the case with studies of organizations, movements, and communities, important questions about generalizability remain as well. Wikis are a particular type of technological platform and Wikia wikis tend to be used for particular types of projects. As a result, we cannot be certain about the implications of our findings for the broader universe of peer production. There are also important questions about generalizability of our results beyond peer production. Many of the theories of communal information goods and collective action were formulated without peer production or the Internet in mind. Peer production projects are also unusual kinds of organizations that share characteristics with firms, cooperatives, political parties, unions, social movements, fan clubs, self-help groups, and more. As a result, a comparative analysis of peer production employing organizational theory is, ideally, an exercise in careful analogy. At the same time, peer production projects are organizations in the fundamental sense that they are collectivities with goals and boundaries. Moreover, as voluntary organizations, peer production projects such as Wikipedia and GNU/Linux have had unprecedented success at mobilizing collective action and contributions to communal public goods. For this reason, we join communication and social movement scholars, as well as proponents of peer production, in the belief that theories of public goods production and voluntary organizations are a salient reference point for research in organizational democracy. The question of whether or not our findings regarding peer-produced communal information goods apply to other sorts of networked movement organizations remains for future research.

For scholars of collective action who are accustomed to environments where the objectives may include subjects such as information shared within or between firms, political rights, fair wages, or representation from the state, it may be tempting to dismiss the forms of leadership and activity on wikis as trivial or insubstantial. We believe that such dismissal is shortsighted, given the contemporary influence of peer production technologies and organizational strategies. For example, in Mexico, a newly created political party constructed around a call for participatory democratic political structures calls itself the *Wikipartido*.[8] Even the most playful peer production communities face many of the same obstacles to collective action and public goods production as other kinds of volunteer-based membership organizations. As such, peer production projects provide an ideal environment within which to test and elaborate existing theories and laboratory-based findings using observational data.

Conclusion

This article offers several contributions to existing research on peer production and collective action in participatory organizations. First, it joins a small number of studies

in expanding the existing domain of organizational research beyond political parties, firms, nonprofit organizations, and social movements in order to test influential organizational theories in the context of peer production. Second, it contributes to our understanding of peer production through the application and evaluation of an established domain of communication scholarship around communal public goods in interactive communication systems. Third, by conducting a comparative analysis across peer production communities in order to understand better their social and organizational dynamics, we also contribute an empirical test of the prevailing assumption that peer production projects inherently advance "small-d" democracy. Finally, our results provide evidence that the widely observed hierarchies and inequalities within peer production projects are associated with enhanced oligarchic organizational structures and behavior.

Our findings show that peer production communities tend to follow Michels' iron law of oligarchy, extending previous findings on the role of elites and powerful leaders in peer production and contradicting the notion that peer production and other forms of communal public goods production entail democratic organizational forms. This study therefore recalls the paradoxes of participatory processes observed by scholars of organizational democracy: There may be trade-offs and contradictions that accompany any effort to organize effective collaboration and incorporate meaningful participation at the same time (Stohl & Cheney, 2001).

We note that some wikis in our dataset appear more robustly democratic than others. Although these cases are exceptional, they signal that digital technologies, like their offline counterparts, might—or might not—be used to create participatory democratic organizations. The opportunities to do so are neither foretold nor foreclosed by the technologies themselves. Understanding why *some* peer production projects create robust democratic organizations is a promising area for future research. But this research is precluded when scholars select more democratic organizations a priori or simply take the participatory and democratic nature of peer production organizations for granted.

What sort of organizational culture and institutions will contemporary online associations, movements, and organizations create? The impact of peer production has grown as more organizations (firms, political parties, movements, and voluntary groups) adopt the platforms and practices of peer production to pursue their objectives. Some, such as the Occupy movement and the *Wikipartido*, have explicitly modeled their organization on wikis and other peer production communities in an effort to facilitate deeper participation and democracy. Our findings here suggest that the adoption of peer production's organizational forms may inhibit the achievement of enhanced organizational democracy.

The fact that peer production may bring with it oligarchic organizations recalls previous work on associational culture and its influence on broader social institutions. For example, when Alexis de Tocqueville visited America in the early 19th century, he saw a flowering of civic associations, the organizational structures of which he believed contributed to a democratic national culture (de Tocqueville,

2004). Similarly, when Seymour Martin Lipset and his colleagues looked to explain the remarkable democratic institutions within the International Typographical Union in the 1950s, they pointed to the robust culture of participation and engagement that cut across many levels of authority and experience within the union as a foundational factor. A few decades later, Robert Putnam bemoaned the apparent collapse of civic associationism in America, a pattern that many have subsequently sought to connect with changes in political culture. Consistent with Michels' original concerns, the rise of labor parties in various countries of the world has also brought about the nationalization of labor politics as the organization of unions becomes part and parcel of the organization of national political culture.

Although, invoking U.S. Supreme Court Justice Louis Brandeis, online collectives have been hailed as contemporary "laboratories of democracy," our findings suggest that they may not necessarily facilitate enhanced practices of democratic engagement and organization. Indeed, our results imply that widespread efforts to appropriate online organizational tactics from peer production may facilitate the creation of entrenched oligarchies in which the self-selecting and early-adopting few assert their authority to lead in the context of movements without clearly defined institutions or boundaries.

Acknowledgments

The authors gratefully acknowledge the input, support, and feedback of Yochai Benkler, Peter Evans, Eric von Hippel, Tom Malone, Mitch Resnick, Coye Cheshire, Lance Bennett, Helen Margetts, Abigail Andrews, Dave Karpf, Per Selle, the MIT Economic Sociology Working Group, the Cooperation Group at the Berkman Center for Internet and Society, the Center for Information Technology and Policy, the Online Collective Action Working Group at the 2013 ECPR Joint Sessions, the Northwestern Communication Studies Faculty Research in Progress Workshop, the Buffett Center at Northwestern University, the Computational Social Science Workshop at the University of Chicago, the City University of Hong Kong, Wikimania, two anonymous reviewers, and the editor and staff of the *Journal of Communication*.

Notes

1 As is common in peer production, ownership of the copyright on wiki content remains with the contributors but all material is licensed freely to the public as a condition of contribution. All Wikia content is released under the Creative Commons Attribution-ShareAlike license (the same license that Wikipedia uses) and is made publicly available for download.

2 More recent studies have turned to laboratory settings as well as firms and non-profit organizations, and have bypassed the question of whether specific modes of organizational governance more effectively support communal information resource production (Child & Shumate, 2007; Cho & Lee, 2008; Raban & Rafaeli, 2007; Yuan, Fulk, & Monge, 2007).

3 We do not attempt, in this article, to give a comprehensive overview of research into the iron law and its application. In addition to Michels' original text, we refer readers to work by Jenkins (1977), Leach (2005), and Voss and Sherman (2000).
4 It is worth noting that the conceptual elaboration of the iron law that we present here follows Jenkins (1977) and Voss and Sherman (2000) more closely than it does Leach (2005).
5 Research on Wikipedia has shown that reverts by administrators tend to drive away newcomers (Halfaker et al., 2011). Assuming a similar dynamic exists in wikis in our dataset, the presence of reverts by administrators might provide an indicator of the extent to which administrators prevent less experienced contributors from acquiring authority or control of the organization.
6 Specifically, we treat a contribution as reverted if the previous revision is identical to the subsequent revision. In other words, we treat a contribution as reverted if, and only if, a user makes a contribution and the next contribution simply undoes his/her work. Because a contributor's edits can be undone in ways that do not return the page to its previous state, this reflects a conservative measure of reversion.
7 http://seattle.wikia.com/.
8 http://wikipartido.mx/.

References

Andrews, K. T., Ganz, M., Baggetta, M., Han, H., & Lim, C. (2010). Leadership, membership, and voice: Civic associations that work. *American Journal of Sociology*, **115**, 1191–1242. doi:10.1086/649060.

Barakso, M., & Schaffner, B. F. (2008). Exit, voice, and interest group governance. *American Politics Research*, **36**, 186–209. doi:10.1177/1532673X07306545.

Benkler, Y. (2002). Coase's penguin, or, Linux and the nature of the firm. *Yale Law Journal*, **112**, 369.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Bennett, W. L., & Segerberg, A. (2012). The logic of connective action. *Information, Communication & Society*, **15**, 739–768. doi:10.1080/1369118X.2012.670661.

Bennett, W. L., & Segerberg, A. (2013). *The logic of connective action: Digital media and the personalization of contentious politics*. New York, NY: Cambridge University Press.

Bimber, B. A., Flanagin, A. J., & Stohl, C. (2012). *Collective action in organizations: Interaction and engagement in an era of technological change*. New York, NY: Cambridge University Press.

Black, L. W., Welser, H. T., Cosley, D., & DeGroot, J. M. (2011). Self-governance through group discussion in Wikipedia measuring deliberation in online groups. *Small Group Research*, **42**, 595–634. doi:10.1177/1046496411406137.

Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, **12**, 346–362. doi:10.1287/isre.12.4.346.9703.

Butler, B. S., Joyce, E., & Pike, J. (2008). Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in Wikipedia. In *Proceedings of the SIGCHI*

conference on human factors in computing systems (pp. 1101–1110). New York, NY: ACM. doi:10.1145/1357054.1357227.

Castells, M. (2012). *Networks of outrage and hope: Social movements in the internet age*. Malden, MA: Wiley.

Child, J. T., & Shumate, M. (2007). The impact of communal knowledge repositories and people-based knowledge management on perceptions of team effectiveness. *Management Communication Quarterly*, **21**, 29–54. doi:10.1177/0893318907301987.

Cho, H., & Lee, J.-S. (2008). Collaborative information seeking in intercultural computer-mediated communication groups testing the influence of social context using social network analysis. *Communication Research*, **35**, 548–573. doi:10.1177/0093650208315982.

Connolly, T., & Thorn, B. K. (1990). Discretionary databases: Theory, data, and implications. In J. Fulk & C. Steinfield (Eds.), *Organizations and communication technology* (pp. 219–233). Newbury Park, CA: Sage.

de Tocqueville, A. (2004). *Democracy in America*. (A. Goldhammer, trans.) New York, NY: Library of America.

Earl, J., & Kimport, K. (2011). *Digitally enabled social change: Activism in the internet age*. Cambridge, MA: MIT Press.

Forte, A., Larco, V., & Bruckman, A. (2009). Decentralization in Wikipedia governance. *Journal of Management Information Systems*, **26**, 49–72. doi:10.2753/MIS0742-1222260103.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, **8**, 1–27.

Fulk, J., Flanagin, A. J., Kalman, M. E., Monge, P. R., & Ryan, T. (1996). Connective and communal public goods in interactive communication systems. *Communication Theory*, **6**, 60–87. doi:10.1111/j.1468-2885.1996.tb00120.x.

Fuster Morell, M. (2010). *Governance of online creation communities: Provision of infrastructure for the building of digital commons* [Ph.D. Thesis]. European University Institute. (Defense date: 20/09/2010)

Fuster Morell, M. (2012). The free culture and 15 M movements in Spain: Composition, social networks and synergies. *Social Movement Studies*, **11**, 386–392. doi:10.1080/14742837.2012.710323.

Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, **57**, 664–688. doi:10.1177/0002764212469365.

Halfaker, A., Kittur, A., & Riedl, J. (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* (pp. 163–172). New York, NY: ACM. doi:10.1145/2038558.2038585.

Hardin, G. (1968). The tragedy of the commons. *Science*, **162**, 1243–1248 PMID: 17756331. doi:10.1126/science.162.3859.1243.

Healy, K., & Schussman, A. (2003). *The ecology of open-source software development*. Retrieved from http://kieranhealy.org/files/drafts/oss-activity.pdf.

Hess, C., & Ostrom, E. (Eds.) (2011). *Understanding knowledge as a commons: From theory to practice*. Cambridge, MA: MIT Press.

Hill, B. M., Burger, C., Jesse, J., & Bacon, J. (2008). *Official Ubuntu book* (Vol., 3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Ingo, H. (2006). *Open life: The philosophy of open source*. Retrieved from http://openlife.cc/online/.

Jenkins, J. C. (1977). Radical transformation of organizational goals. *Administrative Science Quarterly*, **22**, 568–586. doi:10.2307/2392401.

Joyce, E., Pike, J. C., & Butler, B. S. (2012). Rules and roles vs. consensus: Self-governed deliberative mass collaboration bureaucracies. *American Behavioral Scientist*. doi:10.1177/0002764212469366.

Karpf, D. (2012). *The MoveOn effect: The unexpected transformation of American political advocacy*. New York, NY: Oxford University Press.

Kittur, A., Chi, E., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Proceedings of Alt. CHI*. New York, NY: ACM.

Kollock, P. (1999). The economies of online cooperation: Gifts and public goods in cyberspace. In M. Smith & P. Kollock (Eds.), *Communities in cyberspace* (pp. 220–239). London, England: Routledge.

Konieczny, P. (2009). Governance, organization, and democracy on the internet: The iron law and the evolution of Wikipedia. *Sociological Forum*, **24**, 162–192. doi:10.1111/j.1573-7861.2008.01090.x.

Kriplean, T., Beschastnikh, I., McDonald, D. W., & Golder, S. A. (2007). Community, consensus, coercion, control: cs*w or how policy mediates mass participation. In *Proceedings of the 2007 international ACM conference on supporting group work* (pp. 167–176). New York, NY: ACM. doi:10.1145/1316624.1316648.

Leach, D. K. (2005). The iron law of what again? Conceptualizing oligarchy across organizational forms. *Sociological Theory*, **23**, 312–337. doi:10.1111/j.0735-2751.2005.00256.x.

Leuf, B., & Cunningham, W. (2001). *The wiki way: Quick collaboration on the web*. Boston, MA: Addison-Wesley Longman.

Lipset, S. M., Trow, M. A., & Coleman, J. S. (1956). *Union democracy: The internal politics of the international typographical union*. Glencoe, IL: Free Press.

Loubser, M. (2010). *Organisational mechanisms in peer production: The case of Wikipedia (DPhil)*. Oxford, UK: Oxford Internet Institute, Oxford University.

Michels, R. (1915). *Political parties: A sociological study of the oligarchical tendencies of modern democracy* (E. Paul & C. Paul, trans.). New York, NY: Hearst.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups* (Vol. ,1st ed.). Cambridge, MA: Harvard University Press.

O'Mahony, S., & Ferraro, F. (2007). The emergence of governance in an open source community. *Academy of Management Journal*, **50**, 1079–1106. doi:10.5465/AMJ.2007.27169153.

Ortega, F. (2009). Wikipedia: A quantitative analysis [PhD thesis]. Universidad Rey Juan Carlos.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. New York, NY: Cambridge University Press.

Panciera, K., Halfaker, A., & Terveen, L. (2009). Wikipedians are born, not made: A study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on supporting group work* (pp. 51–60). New York, NY: ACM. doi:10.1145/1531674.1531682.

Piven, F. F., & Cloward, R. A. (1977). *Poor people's movements: Why they succeed, how they fail*. New York, NY: Pantheon Books.

Polletta, F. (2013). Participatory democracy in the new millennium. *Contemporary Sociology: A Journal of Reviews*, **42**, 40 – 50. doi:10.1177/0094306112468718b.

Priedhorsky, R., Chen, J., Lam, S. T. K., Panciera, K., Terveen, L., & Riedl, J. (2007). Creating, destroying, and restoring value in Wikipedia. In *Proceedings of the 2007 international ACM conference on supporting group work* (pp. 259 – 268). New York, NY: ACM. doi:10.1145/1316624.1316663.

Raban, D. R., & Rafaeli, S. (2007). Investigating ownership and the willingness to share information online. *Computers in Human Behavior*, **23**, 2367 – 2382. doi:10.1016/j.chb.2006.03.013.

Rafaeli, S., & LaRose, R. J. (1993). Electronic bulletin boards and "Public goods" explanations of collaborative mass media. *Communication Research*, **20**, 277 – 297. doi:10.1177/009365093020002005.

Rucht, D. (1999). Linking organization and mobilization: Michels's iron law of oligarchy reconsidered. *Mobilization: An International Quarterly*, **4**, 151 – 169.

Schweik, C. M., & English, R. C. (2012). *Internet success: A study of open-source software commons*. Cambridge, MA: MIT Press.

Shaw, A. (2012). Centralized and decentralized gatekeeping in an open online collective. *Politics & Society*, **40**, 349 – 388. doi:10.1177/0032329212449009.

Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. New York, NY: Penguin Press HC.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence* (Vol. ,1st ed.). New York, NY: Oxford University Press.

Stohl, C., & Cheney, G. (2001). Participatory processes/paradoxical practices communication and the dilemmas of organizational democracy. *Management Communication Quarterly*, **14**, 349 – 407. doi:10.1177/0893318901143001.

Tufekci, Z., & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of Communication*, **62**, 363 – 379. doi:10.1111/j.1460-2466.2012.01629.x.

Viégas, F. B., Wattenberg, M., McKeon, M. M., & Schuler, D. (2007). The hidden order of Wikipedia. In *Online communities and social computing* (pp. 445 – 454). Berlin, Heidelberg, Germany: Springer-Verlag.

Voss, K., & Sherman, R. (2000). Breaking the iron law of oligarchy: Union revitalization in the American labor movement. *American Journal of Sociology*, **106**, 303 – 349. doi:10.1086/ajs.2000.106.issue-2.

Weber, S. (2004). *The success of open source*. Cambridge, MA: Harvard University Press.

Wu, F., Wilkinson, D. M., & Huberman, B. A. (2009). Feedback loops of attention in peer production. In *Proceedings of the 2009 international conference on computational science and engineering – Volume 04* (pp. 409 – 415). Washington, DC: IEEE Computer Society. doi:10.1109/CSE.2009.430.

Yuan, Y. C., Fulk, J., & Monge, P. R. (2007). Access to information in connective and communal transactive memory systems. *Communication Research*, **34**, 131 – 155. doi:10.1177/0093650206298067.

Zhu, H., Kraut, R. E., Wang, Y. – C., & Kittur, A. (2011). Identifying shared leadership in Wikipedia. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (pp. 3431 – 3434). New York, NY: ACM. doi:10.1145/1978942.1979453.

Appendix: Robustness checks

Table A1. Table of Robustness Checks. Each fitted multilevel regression model uses wiki week as its unit of analysis. M4 is a linear model predicting the logged number of edits made by non-administrators on administrative "project" pages controlling for total edits to these pages. M5 is a linear model predicting the proportion of edits of administrative "project" pages made by administrators. M6 is a linear model predicting the proportion of edits by experienced editors that are reverted by administrators. M7 is a linear model predicting the number of different editors of administrative "project" pages in a wiki week. ($n = 146{,}858$ wiki weeks from 683 wikis.)

	M4	M5	M6	M7
(Intercept)	0.135^{***}	-0.259^{***}	-0.002^{***}	0.012^{**}
	(0.010)	(0.021)	(0.000)	(0.004)
week	0.000	-0.001^{***}	0.000^{***}	0.000^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
$week^2$	0.000^{**}	0.000^{***}	0.000^{***}	0.000^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
ln $accounts_{total}$	−0.002	0.021^{***}	0.001^{***}	-0.002^{***}
	(0.002)	(0.005)	(0.000)	(0.001)
ln $pages_{total}$	0.001	0.021^{***}	0.000	0.007^{***}
	(0.001)	(0.004)	(0.000)	(0.001)
ln $admins_{total}$	-0.081^{***}	0.142^{***}	0.002^{***}	-0.013^{***}
	(0.004)	(0.009)	(0.000)	(0.002)
ln $editors_{week}$	0.086^{***}	-0.102^{***}	-0.001^{***}	0.062^{***}
	(0.002)	(0.005)	(0.000)	(0.001)
ln $edits_{week}$	-0.049^{***}	0.094^{***}	0.001^{***}	-0.027^{***}
	(0.001)	(0.003)	(0.000)	(0.001)
ln $proj\text{-}edits_{week}$	0.710^{***}			0.581^{***}
	(0.001)			(0.001)
Number of observations	11,8994	26,138	112,938	146,858
Number of groups: wiki	554	554	683	683
Variance: 1|wiki	0.027	0.045	0.000	0.004
Variance: Residual	0.094	0.114	0.000	0.022

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media

WeDo: End-To-End Computer Supported Collective Action

Haoqi Zhang[1], **Andrés Monroy-Hernández**[2,3], **Aaron Shaw**[1], **Sean A. Munson**[3], **Elizabeth Gerber**[1]
Benjamin Mako Hill[3], **Peter Kinnaird**[4], **Shelly D. Farnham**[2], **Patrick Minder**[5]
[1]Northwestern University, [2]Microsoft Research, [3]University of Washington
[4]Carnegie Mellon University, [5]University of Zurich

Abstract

Many celebrate the Internet's ability to connect individuals and facilitate collective action toward a common goal. While numerous systems have been designed to support particular aspects of collective action, few systems support *participatory, end-to-end collective action* in which a crowd or community identifies opportunities, formulates goals, brainstorms ideas and develops plans, mobilizes, and takes action. To explore the possibilities and barriers in supporting such interactions, we introduce *WeDo*, a system aimed at promoting simple forms of participatory, end-to-end collective action. Pilot deployments of WeDo illustrate that sociotechnical systems can support automated transitions through different phases of end-to-end collective action, but that challenges, such as the elicitation of leadership and the accommodation of existing group norms, remain.

Introduction

Social computing systems play an increasingly important role in collective action—that is, "actions taken by two or more people in pursuit of the same collective good" (Marwell & Oliver 1993). *Computer supported collective action* (CSCA) systems encompass the use and design of social media technology for collective action (Shaw et al. 2014). Successful instances of CSCA connect crowds and communities of participants, lower the cost of communication, facilitate deliberation, and help to coordinate action, thereby enabling new forms of collaboration.

Today, numerous platforms facilitate distinct pieces of CSCA or context-specific collaborations. Examples include *IfWeRanTheWorld*, *Stanford Catalyst*, *Twitter*, *Facebook*, *FixMyStreet*. While some of these systems help connect communities of interest, others enable identification what actions should be taken, discussion of what issues exist, or coordination of synchronous action (King & Brown 2007; Starbird 2013). In this work, we take inspiration from these systems to design a system to support *participatory* and *end-to-end* collective action, in which a crowd or community comes together to surface opportunities, formulate goals, brainstorm ideas, make plans, and mobilize a critical mass of participants to take action.

Copyright © 2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Participatory, end-to-end CSCA represents an important opportunity for social computing research for several reasons. First, CSCA that is both participatory and end-to-end may take advantage of collective intelligence and cooperation more effectively than systems that are top-down or hierarchically organized (Benkler 2006; Woolley et al. 2010). By involving participants in the entire process, we can explore opportunities for effective mobilization in the pursuit of important, but perhaps unanticipated, individual, social, and organizational goals. Second, ideas that emerge from a participatory process collect support along the way, which may increase the likelihood of collective action actually taking place. Finally, a single end-to-end CSCA system avoids the need for switching between tools to accomplish discrete tasks (e.g. voting), which can lead to inefficiency and frustration (González & Mark 2004). End-to-end CSCA systems can better nudge participants to follow through on promising ideas, and conversely, not to push particular courses of action without deliberation.

In this note, we present *WeDo*, a system aimed at promoting simple, participatory, and end-to-end collective action. We created and deployed WeDo as a technology probe (Hutchinson et al. 2003) to help identify possibilities for, and barriers to, building and deploying participatory, end-to-end CSCA systems. Built as a Twitter app with a complementary web interface (see Figure 1), WeDo supports: the creation of high-level "missions;" collecting ideas from a community for accomplishing the mission; browsing and voting on ideas; and following through on actions. The system automates the transition through these stages of collective action, so as to promote progress and follow-through. Results from our pilot deployments highlight opportunities and challenges for designers of end-to-end CSCA systems.

Related Work

Existing systems supporting collective action tend to fall into one or more of the following four categories:

Systems Supporting Few Tasks. Many systems focus exclusively on a few tasks within an individual stage of collective action. For example, *FixMyStreet* allows people to submit reports of issues such as potholes, broken lights, or graffiti (King & Brown 2007). *FixMyStreet* and similar tools support very specific actions (reporting the issue and location) within a very specific stage of collective action (identi-

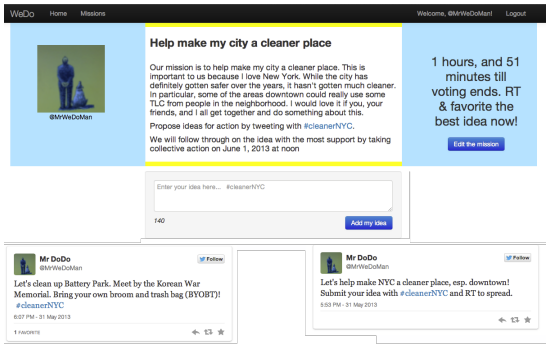


Figure 1: WeDo web interface for viewing and contributing to missions. Users can also contribute directly via Twitter.

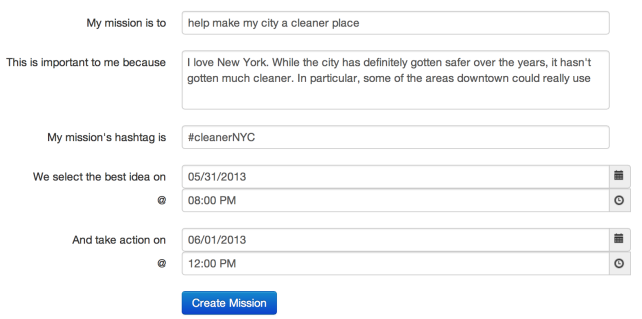


Figure 2: Creating a mission in WeDo

fying problems). *FixMyStreet* need not address other stages (such as deliberation), and has not been designed to support other challenges.

Systems Supporting Few Stages. Systems such as *Reddit* promote sharing of ideas, deliberating about their merits, and voting up popular ones, but do not provide support for following through on ideas to taking collective action. Systems such as *Stanford Catalyst* help community members share activities and sign up for activities they wish to participate in, but do not allow users to plan the details of another's suggested activity nor to brainstorm about possible activities collaboratively. In supporting their respective use cases, such systems give up on certain opportunities for collaboration that may help to stimulate collective action on ideas existing and new.

Systems Supporting Action in Few Contexts. Another class of systems support end-to-end CSCA by relying on a narrowly defined context or domain to structure group activities. Consider, for example, *GitHub* and similar tools that combine code repositories, bug trackers, and discussion boards. These tools allow people to surface issues, deliberate solutions, submit resolutions, and adopt them (Schweik & English 2012). However, this end-to-end action is enabled by the constraints of the domain: most issues with software programs can be solved by new code or images; a patch fixes a bug or it does not. Such systems rarely translate well to other contexts or solutions.

Systems Supporting Less Participatory Actions. Finally, there are systems that support less participatory forms of collective action. For example, *Kickstarter* helps people post projects, market their proposal, recruit and engage supporters, and receive funding (or not). Donors provide funds but do not participate in projects directly. Gallop's *IfWeRanTheWorld* breaks down broad goals into specific actions, invites individuals to participate, and motivates them through elements of gamification. Users participate by suggesting and taking actions, but the actions are mostly designed to be carried out by individuals on their own. *Pipeline* (Luther, Fiesler, and Bruckman 2013) helps leaders in online creative collaborations distribute responsibility by assigning tasks to trusted team members, while still leaving the leaders free to choose when and how to move forward. While team members all participate, the projects are not participatory insofar as they are conceptualized, designed, and often executed by members of a closed, pre-defined group.

These existing systems leave open the question of whether it is possible to create systems that support participatory, end-to-end CSCA without relying on features designed for specific domain- or context-specific constraints. Such systems promise potential enhancements over existing tools by (1) distributing tasks more effectively throughout the process of collective action; (2) facilitating the progression toward eventual action, partly by managing transitions across discrete types of tasks more smoothly; and (3) meeting needs of communities of end users who may benefit from participatory, end-to-end processes of collective action.

The WeDo Prototype

WeDo is a lightweight system for promoting participatory, end-to-end collective action. In the initial design of WeDo, we sought to use the Internet to reduce barriers to action by providing technical support for moving between phases of collective action and automating some elements of the leadership role (Marwell & Oliver 1993). Thus, WeDo supports collective action in four phases: start a "mission," collect ideas, vote on a plan for action, and notify people to coordinate action. The prototype system consists of a web interface for submitting, tracking, and participating in missions, as well as a Twitter bot that automates the distribution of messages for ideation, voting, selection, and action notification (see Figures 1 and 2).

For example, a user could create a mission to cleanup a local park. WeDo announces the mission via posts to the user's Twitter account as well as the WeDo account, soliciting suggestions. After a few hours, the system prompts interested participants to vote on promising ideas by retweeting or favoriting those they like. After the voting deadline, the system sends messages announcing the leading choice and prompting the submission of additional details. Finally, the system reminds participants immediately before the event.

Building on top of Twitter has many advantages. First, in any phase of the process, WeDo allows anyone to discuss ideas and refine plans directly on Twitter. Second, as a messaging system, Twitter makes it easy for people to contribute and to receive notifications anywhere, which facilitates transitioning through phases of collective action. Third, functionalities for expressing support or voting are already built in via retweeting and favoriting. Finally, the social network in Twitter can kickstart missions by leveraging a user's ex-

Mission	Who	Ideas	Votes	Top idea votes	Action resulted?	Winning idea	# of people acting
Celebrate end of workshop	Workshop attendees	8	26	9	yes	crowd scream	15
Conference banquet event	Conference attendees	10	18	5	yes	group hug	30
Twitter book club	Book club members	14	29	9	N/A	*Ocean at the End of the Lane*	N/A
Red Cross donations	Twitter followers	0	0	0	no	N/A	N/A

Table 1: Summary of WeDo deployments, participation, and outcomes

isting followers, conversations, and reputation.

To supplement Twitter's affordances, the WeDo prototype also includes a web interface for visualizing ideas and coordinating actions. The interface provides more context for a mission than Twitter and supports more structured interactions. In the current version, a user creates a mission (Figure 2) by filling out a form with fields for the name of the mission, a description prompted by the text, "this mission is important to me because," a hashtag for the mission, a date and time by which the idea is to be selected, and a date and time by which the mission is to be executed. When the form is submitted, the WeDo presents the user with a suggested tweet to start the mission. The user can edit the message but is constrained to the 140 character limit of Twitter. We found this addition was necessary to provide clarity of mission (Marwell & Oliver 1993).

Once the mission is created, interested participants can submit ideas either via Twitter or the WeDo web interface. Automated messages are sent out at their scheduled time. The web interface provides light support for browsing submitted ideas by placing the ideas with the most votes at the top. It also helps users keep track of different stages of the process by grouping messages based on the stages, and showing the amount of time left till the next stage (see Figure 1). Future iterations of the prototype may provide additional support for discussion and planning.

Deployments

We conducted four initial deployments (Table 1). The first two deployments were conducted using a wizard-of-oz version of WeDo that featured only the Twitter interface. Votes were counted manually, and scheduled tweets were sent manually. The latter two deployments were conducted using the current WeDo prototype. With each deployment, we also conducted informal observations of and discussions with system users to learn about their experiences and to inform subsequent iterations.

Wizard of Oz Deployments. We conducted our Wizard-of-Oz deployments at an academic conference. The conference had a few hundred attendees, many of whom had pre-existing friendship ties and familiarity with Twitter. First, attendees of a conference workshop were asked to suggest ideas and decide on a group activity to celebrate the end of the workshop. The entire process of ideation to collective action took about 10-15 minutes. We received a total of 8 ideas, which included taking a silly group photo, laughing, screaming to scare other workshops, paper lightsaber fights, writing code, singing, and drinking beer.

For our second deployment, we created an open-ended mission asking all conference attendees to suggest a group activity for the conference banquet the following day. We used our personal Twitter accounts to spread the word, reply to ideas, and provide examples. We also reached out to highly visible individuals and asked them to participate. We received a total of 11 ideas, which included suggestions such as a flash freeze, a group hug, gambling for charity, and Harlem Shake. Both deployments resulted in collective action, which we recorded via video and shared on YouTube.

Deployments using WeDo. After the initial wizard-of-oz deployments, a book club that "gathers" on Twitter every month to collectively choose a book to read and discuss contacted us about using WeDo to automatically facilitate the process of ideation, voting, and selection. As a test, the facilitator created a mission using his own Twitter account and the group's account (which had more than 90,000 followers). The mission invited people to nominate and vote for the book of the month using WeDo. 30 people contributed to the discussion.

In another deployment, we attempted to create a mission to gather support for the victims of a large tornado. We used one of the authors' account to post the mission, and even recruited a person with more than 18,000 Twitter followers to retweet that first tweet. However, this mission did not take off and no ideas were proposed.

Lessons Learned

Our deployment experiences demonstrate several ways in which WeDo facilitates aspects of end-to-end collective action. First, many users expressed that Twitter made adding ideas easy. They found voting on existing ideas through retweets and favorites intuitive, and liked being able to participate via mobile devices. Second, users mentioned that the system helped to move the process along, steering them toward subsequent action. Third, users appreciated how the system facilitates idea generation and decision-making. They understood that the decision was based on their votes, and that they generated the ideas that are voted on. Finally, even though most missions were open-ended and broad, a number of deployments were successful in that they resulted in participatory, end-to-end collective action.

The WeDo deployments also surfaced many challenges, some specific to WeDo's system design and others more general:

Fluidity and barriers between phases of collective action: While WeDo helped to automatically progress a mission through ideation, voting, selection, and action, barriers still remained at the intersections of these stages. For example, from the conference banquet mission we learned that once an action was planned it was not always clear how it would be "triggered," or whose responsibility it was to do so. To address this issue, future WeDo versions may require a user to agree to serve as a trigger and to identify a triggering action

as a prerequisite. As another example, the fluidity between stages confused some users, especially as the system relied on a single hashtag for many different actions throughout the missions. WeDo's web interface alleviates some of these problems, but users participating directly through Twitter may need additional support.

Confounding user expectations and existing norms: In the book club deployment, users had to adapt to WeDo's process. While they gained the ability to suggest their own ideas and can rely on WeDo to count their votes, they expected other features from the system such as being able to vote on a book suggested by multiple users and having those votes merged (even when retweets were modified). While software improvements can address some of this, human computation may also be necessary to help fill the gaps in data processing and for handling exceptions.

Clarity of high-level mission and task design: WeDo elicits a high-level, abstract mission, and leaves it to users to generate the specific ideas and steps to implement them. The early abstraction decentralizes control over decision-making, but sometimes, as in the "Red Cross" mission, this presented a barrier to entry. Newcomers did not know how to contribute when goals were too broad or vague. The system may need to do more to prompt clear, concrete missions, as well as to help users understand how to make small contributions that help achieve broad goals.

Identifying and mobilizing leadership: In the wizard-of-oz deployments, the authors helped to trigger actions and make sure things moved along. Subsequent deployments drew on existing socio-technical infrastructure to spread the word via existing communities and influential individuals. But without explicit support for identifying and mobilizing leaders, missions can fail to gather the necessary momentum to move forward despite collective interest.

Opportunities for peripheral participation: WeDo supports relatively short-term, lightweight activities, but the design currently assumes active participation throughout different phases of a collective mobilization process. This expectation of follow-up is prohibitive for some users. Participants who wish to join only the voting phases of a project may find it difficult to do so as it is difficult to discover existing missions or identify the current phase of a mission. Future versions can support peripheral participation more effectively by using the interface to make the social process more translucent.

Platform trust and security concerns: The submission process for new missions requires users to give WeDo permission to tweet on his or her behalf. This could prove worrisome for users fearful that a system glitch might spam their followers or choose an inappropriate idea by fraudulent votes. Although the problem is particular to Twitter in this case, questions about how to establish trust and transparency for participants who may not be completely familiar with the tools or procedures involved is a more general issue.

Acknowledgements

This work developed from the Crowdcamp Workshop at CSCW 2013. Thank you to workshop participants and all participants in our deployments.

Conclusion

The design of a system capable of comprehensively supporting participatory, end-to-end collective action across many contexts represents an ambitious, potentially unreachable goal. Our experiences with the WeDo prototype therefore recall the socio-technical gap (Ackerman 2000). Existing tools serve as building blocks which may gradually be assembled into new, more effective and comprehensive systems. In this way, our deployments of the WeDo system highlight potential opportunities for CSCA and provide insights that inform the design of future systems. We intend to pursue follow-up iterations and experimental deployments of the WeDo prototype in order to build on the lessons learned described above.

The domain of collective action presents numerous contexts in which social computing systems are already playing a variety of roles. One contribution of this paper is the elaboration of specific points of contact at which HCI researchers and designers can engage with these systems, and the class of problems they attempt to solve. We hope that others will join these efforts.

References

Ackerman MS. 2000. The intellectual challenge of cscw: the gap between social requirements and technical feasibility. *Human–Computer Interaction* 15(2-3):179–203.

Benkler Y. 2006. *The Wealth of Networks*. Yale.

González VM., Mark G. 2004. "constant, constant, multi-tasking craziness": managing multiple working spheres. In *CHI '04*, 113–120. ACM.

Hutchinson H; Mackay W; Westerlund B; Bederson BB; Druin A; Plaisant C; Beaudouin-Lafon M; Conversy S; Evans H; Hansen H; Roussel N; Eiderbäck B. 2003. Technology probes: inspiring design for and with families. In *CHI 2003*, 17–24. ACM.

King SF, Brown P. 2007. Fix my street or else: using the internet to voice local public service concerns. In *ICEGOV 2007*, 72–80. ACM.

Luther K; Fiesler C; Bruckman A. 2013. Redistributing leadership in online creative collaboration. In *CSCW 2013*, 1007–1022. ACM.

Marwell G, Oliver P. 1993. *The critical mass in collective action : a micro-social theory*. Cambridge.

Schweik C, English R. 2012. *Internet Success: A Study of Open Source Software Commons*. MIT Press.

Shaw A; Zhang H; Monroy-Hernández A; Munson SA; Hill BM; Gerber E; Kinnaird P; Minder P. 2014. Computer supported collective action. *interactions* 21(2):74–77.

Starbird K. 2013. Delivering patients to Sacré; Coeur: collective intelligence in digital volunteer communities. In *CHI 2013*, 801–810. ACM.

Woolley AW; Chabris CF; Pentland A; Hashmi N; Malone T. 2010. Evidence for a collective intelligence factor in the performance of human groups. *Science* 330(6004):686–688.

THIS IS A PRE-PRINT VERSION OF PUBLISHED BOOK CHAPTER. THE FINAL VERSION AND THE REFERENCE TO USE WHEN CITING THIS WORK IS: *Benkler, Yochai, Aaron Shaw, and Benjamin Mako Hill. (2015) "Peer Production: A Form of Collective Intelligence." In Handbook of Collective Intelligence, edited by Thomas Malone and Michael Bernstein. MIT Press, Cambridge, Massachusetts.*

Peer Production: A Form of Collective Intelligence

Yochai Benkler (ybenkler@law.harvard.edu)
HARVARD UNIVERSITY

Aaron Shaw (aaronshaw@northwestern.edu)
NORTHWESTERN UNIVERSITY

Benjamin Mako Hill (makohill@uw.edu)
UNIVERSITY OF WASHINGTON

INTRODUCTION

Wikipedia has mobilized a collective of millions to produce an enormous, high quality, encyclopedia without traditional forms of hierarchical organization or financial incentives. More than any other twenty-first century collaborative endeavor, Wikipedia has attracted the attention of scholars in the social sciences and law both as an example of what collective intelligence makes possible and as an empirical puzzle. Conventional thinking suggests that effective and successful organizations succeed through hierarchical control and management, but Wikipedia seems to organize collectively without either. Legal and economic common sense dictates that compensation and contracts are necessary for individuals to share the valuable products of their work, yet Wikipedia elicits millions of contributions without payment or ownership. Intuition suggests that hobbyists, volunteers, and rag-tag groups will not

Parts of this chapter build upon work published in: Y. Benkler, Peer. "Production and Cooperation" forthcoming in J. M. Bauer & M. Latzer (eds.), *Handbook on the Economics of the Internet*, Cheltenham and Northampton, Edward Elgar.

be able to create information goods of sufficient quality to undermine professional production, but contributors to Wikipedia have done exactly this. Wikipedia should not work, but it does.

The gaps between conventional wisdom about the organization of knowledge production and the empirical reality of collective intelligence produced in "peer production" projects like Wikipedia have motivated research on fundamental social scientific questions of organization and motivation of collective action, cooperation, and distributed knowledge production. Historically, researchers in diverse fields such as communication, sociology, law, and economics have argued that effective human systems organize people through a combination of hierarchical structures (e.g., bureaucracies), completely distributed coordination mechanisms (e.g., markets), and social institutions of various kinds (e.g., cultural norms). However, the rise of networked systems and online platforms for collective intelligence has upended many of the assumptions and findings from this earlier research. In the process, online collective intelligence systems have generated data sources for the study of social organization and behavior at unprecedented scale and granularity. As a result, social scientific research on collective intelligence continues to grow rapidly.

Wikipedia demonstrates how participants in many collective intelligence systems contribute valuable resources without the hierarchical bureaucracies or strong leadership structures common to state agencies or firms and in the absence of clear financial incentives or rewards. Consistent with this example, foundational social scientific research relevant to understanding collective intelligence has focused on three central concerns: (1) explaining the organization and governance of decentralized projects, (2) understanding the motivation of contributors in the absence of financial incentives or coercive obligations, and (3) evaluating the quality of the products generated through collective intelligence systems.

Much collective intelligence research in the social sciences has focused on peer production projects. Following Benkler (2013), we define peer production as a form of open creation and sharing performed by groups online that: set and execute goals in a decentralized manner; harness a diverse range of participant motivations, particularly non-monetary motivations; and separate governance and management relations from exclusive forms of property and relational contracts (i.e., projects are governed as open commons or common property regimes and organizational governance utilizes combinations of participatory, meritocratic and charismatic, rather than proprietary or con-

tractual, models). Peer production is the most significant organizational innovation that has emerged from Internet-mediated social practice, among the most visible and important examples of collective intelligence, and a central theoretical frame used by social scientists and legal scholars of collective intelligence.

Peer production includes many of the largest and most important collaborative communities on the Internet. The best known instances of the phenomena include collaborative free/libre and open source software (FLOSS), like the GNU/Linux operating system, and free culture projects, like Wikipedia. GNU/Linux is widely used in servers, mobile phones, and embedded systems like televisions. Other FLOSS systems, such as the Apache web server, are core Internet utilities. Through participation in FLOSS development, many of the largest global technology companies have adopted peer production as a core business strategy. Explicitly inspired by FLOSS, Wikipedia has been the subject of several books and over 6,200 academic articles.[1] Wikipedia is among the top six websites in the world with more than half a billion unique viewers each month.[2] The success of FLOSS and Wikipedia has coincided with a variety of peer production projects in an array of industries that includes stock photography, videos, and travel guides. Organizations, both for-profit and non-profit, have found ways to organize their products through peer production to overcome competition from more traditional, market- and firm-based approaches.

Although peer production is central to social scientific and legal research on collective intelligence, not all examples of collective intelligence created in online systems are peer production. First, (1) collective intelligence can involve centralized control over goal-setting and execution of tasks. For example, InnoCentive provides a platform for companies to distribute the search for solutions to difficult problems, but the problems themselves are defined within the boundaries of InnoCentive's client firms ("seekers") and solutions are returned to seekers as a complete product created by a single individual or team. Peer production, in contrast, decentralizes both goal setting and execution to networks of individuals or more structured communities (Brabham, 2013). Second, (2) many collective intelligence platforms, such as the ESP Game or Amazon's Mechanical Turk labor market, focus on optimizing systems around a relatively narrow set of motivations and incentives – ludic and

[1] http://scholar.google.com/scholar?q=intitle%3Awikipedia (Accessed July 10, 2013)

[2] http://www.alexa.com/siteinfo/wikipedia.org (Accessed June 15, 2013)

financial in ESP and Mechanical Turk respectively. Peer production projects involve a broad range of incentives. Finally, (3) collective intelligence often occurs within firms where participants are bound by the obligation of contracts and in contexts where the resources used or products of collective effort are managed through exclusive property rights. Peer production exists without these structures.

Since the late 1990s when FLOSS first acquired widespread visibility and broad economic relevance, peer production has attracted scholarly attention. The earliest work sought to both explain the surprising success of peer production and to draw distinctions between peer production and traditional models of firm-based production. While many of these first analyses focused on software, the concept of peer production emerged as the most influential attempt to situate FLOSS as an instance of a broader phenomenon of online cooperation (Benkler, 2001, 2002). A series of scholars then adopted the concept of peer production to emphasize the comparative advantages of Internet-based cooperation as a powerful new mode of innovation and information production (e.g., von Hippel and von Krogh, 2003; Weber, 2004).

In the rest of this chapter, we describe the development of the academic literature on peer production and collective intelligence in three areas – organization, motivation, and quality. In each area, we introduce *foundational work* consisting primarily of earlier scholarship that sought to describe peer production and establish its legitimacy. Subsequently, we characterize work, usually more recent, that seeks to pursue *new directions* and to derive more nuanced analytical insights. Because FLOSS and Wikipedia have generated the majority of the peer production research to date, we focus on research analyzing those efforts. For each theoretical area, we briefly synthesize both foundational work and new directions and describe some of the challenges for future scholarship. Neither our themes nor our periodization are intended to encompass the complete literature on peer production. Instead, they reflect core areas of research that speak most directly to the literature on collective intelligence and locate peer production within that broader phenomenon. We conclude with a discussion of several issues that traverse our themes and implications of peer production scholarship for research on collective intelligence more broadly.

ORGANIZATION

Early scholars of peer production were struck by the apparent absence of formal hierarchies and leadership structures. Indeed, peer production communities perform all of the "classical" organizational functions like coordination, division of labor, recruitment, training, norm creation and enforcement, conflict resolution, and boundary maintenance – but do so in the absence of many of the institutions associated with more traditional organizations. Put in Williamson's (1985) terms, peer production communities thrive despite a relative absence of bureaucratic structure, exclusive property rights, and relational contracts. However, as peer production communities have aged, some have acquired increasingly formal organizational attributes, including bureaucratic rules and routines for interaction and control. The challenge for early peer production research was to elaborate an account of the characteristics that differentiate peer production from traditional organizations. As literature on peer production has matured, more recent work has sought to understand when the organizational mechanisms associated with peer production are more and less effective and how organizational features of communities change over time.

Foundational Work

A diverse and interdisciplinary group of scholars provided initial explanations for how and why peer production communities function. Early organizational analyses of peer production focused on discussions of how peer production compared to modes of organization in firms, states, and markets. Descriptive work also characterized patterns and processes of governance and leadership within some of the most prominent peer production communities.

Initial scholarship on the organization of peer production emphasized theoretical distinctions between peer production, bureaucracy, and transaction cost explanations of firms and markets. Specifically, Benkler (2002; 2004; 2006) focused on the role of non-exclusive property regimes and more permeable organizational boundaries (i.e., organization without relational contracts) to explain how peer production arrived at more efficient equilibria for the production of particular classes of informational goods. Peer production could, Benkler argued, outperform traditional organizational forms under conditions of widespread access to networked communications technologies, a multitude of motivations driving contributions, and non-rival information capable of being broken down into granular, modular, and easy-to-integrate

pieces. Other foundational accounts, like Moglen (1999) and Weber (2004), attended to the emergence of informal hierarchies and governance arrangements within communities. This earliest work relied on case study examples and descriptive characterizations of how these arrangements were lightweight compared to traditional bureaucratic practices within organizations historically tasked with the production of information goods. At the extreme, peer production communities were treated as a new category of "leaderless" social systems relying on informal norms such as reciprocity and fairness to solve complex problems *without* organizations.

Much of the earliest empirical research inductively sought to describe the structure and organization of FLOSS communities (e.g., von Krogh et al., 2003; Kelty, 2008) by, for example, characterizing contributions across FLOSS projects through reference to a core and a periphery of participants (e.g., Lakhani, 2006). Other work focused on describing the interactions between formal organizations and FLOSS communities in a variety of contexts (e.g., O'Mahony and Bechky, 2008; West and O'Mahony, 2005; Spaeth et al., 2010). In their review article, Crowston et al. (2010) provide a detailed overview of the scholarly literature on FLOSS organization and practice focusing on work that we characterize as foundational.

A large body of descriptive work has also sought to characterize organizational dimensions of Wikipedia and the practices of its contributors. A series of qualitative accounts examined the norms of participation and consensus among Wikipedians (e.g., Bryant et al., 2005; Forte and Bruckman, 2008; Pentzold, 2011). Reagle (2010) and Jemielniak (2014) have provided the most influential book-length scholarly descriptions of Wikipedia and its processes. Another influential approach employed data-mining to build inductive, quantitative descriptions of Wikipedia's organization (Ortega, 2009; Priedhorsky et al., 2007; Viegas et al., 2007; Yasseri et al., 2012). A growing body of related research has characterized Wikipedia's governance systems (e.g., Butler et al., 2008; Konieczny, 2009); the emergence of leadership within the Wikipedia community (Forte et al., 2009; Zhu et al., 2011, 2012); mechanisms through which new community members are incorporated and socialized (Antin et al., 2012; Antin and Cheshire, 2010; Halfaker et al., 2011); as well as the ways in which bureaucratic decisions like promotions are made (Burke and Kraut, 2008; Leskovec et al., 2010).

Although they are a relative minority, foundational scholarly descriptions of organizational elements of peer production communities other than FLOSS

and Wikipedia have emphasized similar concerns. For example, analyses of Slashdot's system of distributed moderation provided an early illustration of how recommendation and filtering constituted an informal mode of governance and supported a division of labor on the site (Lampe and Resnick, 2004; Lampe et al., 2007). This research built on earlier work showing that contributors to bulletin boards, newsgroups, forums, and related systems adopted durable "social roles" through their patterns of contribution (e.g., Fisher et al., 2006) – an approach that was subsequently applied to Wikipedia by Welser et al. (2011) and McDonald et al. (2011).

New Directions

More recent work on organizational aspects of peer production has begun to question the "stylized facts" that prevailed in earlier research. Examples of this work compare across projects and routines to understand the relationship between organizational attributes of communities and their outputs. Other work engages in interventions and field experiments to test and improve organizational operations in peer production communities. Many of the best examples of this newer wave of work have incorporated collections of peer production communities in comparative analyses. Some of these studies also articulate connections between peer production research and theories from a range of other organizational fields including academic studies of firms, states, political parties, social movement organizations, and civic associations.

Although some of the earliest theories of the organization of peer production celebrated the phenomena as non-hierarchical, more recent work has questioned both the putative lack of hierarchy and its purported benefits (e.g., Kreiss et al., 2011). In empirical work, several studies have underscored how feedback loops of attention and cumulative advantage can perpetuate starkly unequal distributions of influence and hierarchy (Hindman, 2008; Wu et al., 2009). Healy and Schussman (2003) have gone as far as to propose that hierarchy may even contribute to peer production success and a series of empirical studies by Keegan and Gergle (2010) and Shaw (2012) have documented gate-keeping behavior in peer production and suggested that it may benefit projects. Several studies have looked at wikis and shown that governance and hierarchies tend to become more pronounced as peer production projects mature (Loubser, 2010; Shaw and Hill, 2014). O'Mahony and Ferraro (2007) argued that FLOSS projects can depend and reflect existing firm-based hierarchies. This body of scholarship does not refute the idea that peer production

is less hierarchical than alternative forms of organization or that it uses novel governance structures; however, the authors suggest a different model of peer production than the stylized anti-hierarchies depicted in earlier work.

While enormous variation in organizational form exists among peer production projects, until recently, very few studies had systematically compared across communities. An important step toward such cross-organizational comparisons has been the creation of large data sets drawn from many communities. The FLOSSmole project has helped facilitate recent comparative scholarship by assembling and publishing a series of cross-project datasets in the area of software, including a widely used dataset from the FLOSS hosting website SourceForge published by Howison et al. (2006). The most in-depth comparative study of FLOSS is Schweik and English (2012), who build on the work of Ostrom (1990) to compare characteristics of FLOSS projects' organization, resources, governance, and context, using the FLOSSmole dataset of SourceForge projects. These comparative studies have not only revealed that large, collaborative projects are extraordinarily rare, but also that many projects generating widely downloaded FLOSS resources are, in fact, not peer production at all but rather the products of firms or individuals working alone.

Similar cross-organization studies in other areas of peer production, or studies comparing across different *types* of peer production, have remained challenging and rare. One difficulty with comparative work across organizations, in general, is designing research capable of supporting inference into the causes of organizational success and failure. Historically, failed attempts to build organizations never enter into research datasets or attract scholarly attention. Two studies of peer production capturing these failures are Kittur and Kraut (2010) who draw from a dataset of 7,000 peer production wikis from the hosting firm Wikia, and Shaw and Hill (2014), who use an updated version of this dataset with ten times as many wikis. Others including Ransbotham and Kane (2011), Wang et al. (2012), and Zhu et al. (2011) have tried to look across wikis by considering variation in sub-organizations within Wikipedia. Fuster Morell (2010) describes a comparative case study of several hundred peer production projects active in Catalonia. Despite these exceptions, there exist very few publicly available large-scale comparative datasets for types of peer production projects outside of FLOSS.

Another exciting path forward in organizational research in peer production lies in the use of field experiments and organizational interventions. Using a series of examples from peer production, Kraut and Resnick (2012) draw

heavily on experimental results to form principles of community design. Other examples include Halfaker et al. (2013b), who report experiments by the Wikimedia Foundation to alter the organizational structure around participation and work by Luther et al. (2013), who analyze the impact of a new system intended to support peer produced animations through more effective allocation of leaders' time and effort. By structuring design changes as experiments, these studies make credible causal claims about the relationship of organizational structure and project outcomes that previous work struggled to establish. By intervening in real communities, these efforts achieve a level of external validity that lab-based experiments cannot.

Organizational analysis of peer production remains a critical area for future work as an increasing number of organizations perform their activities in computer-mediated environments using the tools, techniques, and resources of peer production. As a result, a science of the mechanisms, procedures, and techniques necessary for the effective production of networked informational resources is increasingly important in various spheres of activity. Through challenging stylized conceptions of peer production organization, through comparative work, and through causal inference, future peer production research can continue to deepen and extend foundational insights. Future work can also begin to address questions of whether, how, and why peer production systems have transformed some existing organizational fields more profoundly than others. An empirically-informed understanding of when and where organizational practices drawn from peer production provide efficient and equitable means to produce, disseminate, and access information can provide social impact beyond the insights available through the study of any individual community.

MOTIVATION

A second quality of peer production that challenged conventional economic theories of motivation and cooperation was the absence of clear extrinsic incentives like monetary rewards. Traditional economic explanations of behavior rely on the assumption of a fundamentally self-interested actor mobilized through financial or other incentives. In seeking to explain how peer production projects attract highly skilled contributors without money, much of the literature on peer production has focused on questions of participant motivation. In the earliest work on motivation in the context of FLOSS, an important goal was simply explaining how a system abandoning financial incen-

tives could mobilize participants to freely share their code and effort. Even as it became apparent that peer production attracted contributors motivated by diverse, non-monetary factors, work in economics was focused on whether it was feasible to collapse these motivations into relatively well-defined models of self-interest. A newer wave of work has stepped back from this approach and sought to explain how multiple motivational "vectors" figure in the creation of common pool resources online – an approach that underscores a core advantage of peer production in its capacity to enable action without requiring translation into a system of formalized, extrinsic, carrots and sticks. Systems better able to engage self-motivated action will be better able to attract the kind of decentralized discovery of projects, resources, and solutions that are neither mediated by prices nor command-and-control hierarchies of obligation (Osterloh et al., 2002).

Foundational Work

The earliest research on motivation in peer production by Ghosh (1998) and Lerner and Tirole (2002) used case studies of FLOSS projects to assert that developers' motivations which could be easily assimilated into standard economic models. Frequently cited motivations in foundational work by these authors, von Hippel and von Krogh (2003), von Krogh (2003), and others included: the use value of the software to the contributing developer; the hedonic pleasure of building software; the increased human capital, reputation, or employment prospects; and social status within a community of peers. Other early accounts analyzing examples of peer production beyond FLOSS suggested additional motivations. For example, Kollock (1999) emphasized reciprocity, reputation, a sense of efficacy, and collective identity as salient social psychological drivers of contribution to online communities and forums. Early studies of Wikipedia contributors suggested that Wikipedians edited in response to a variety of motives, many of which were social psychological in character (Beenen et al., 2004; Forte and Bruckman, 2008; Nov, 2007; Rafaeli and Ariel, 2008; Panciera et al., 2009). In studies of peer-to-peer file sharing networks and online information sharing communities, Cheshire (2007) and Cheshire and Antin (2008) emphasized social psychological motivation and demonstrated that peer feedback could provide a mechanism of activating reputational concerns and encouraging additional contributions. Benkler (2002, 2006) argued that the combination of many incentives created a potentially fragile interdependence of motivations in peer production. In anthropological accounts of FLOSS, Coleman suggested that political ideolo-

gies of freedom play an important role in FLOSS production (Coleman, 2012; Coleman and Hill, 2004). von Krogh et al. (2012) provide a comprehensive recent survey of work characterizing motivations in open source software.

Much of the scholarship on individual motivation in FLOSS, Wikipedia, and other communities has relied on surveys. Some of the most influential survey research on FLOSS are studies by Rishab Ghosh and colleagues (e.g., Ghosh and Prakash, 2000; Ghosh et al., 2002; Ghosh, 2005), the Boston Consulting Group Hacker Survey (Lakhani and Wolf, 2005), and the US FLOSS study (David and Shapiro, 2008). Lakhani and Wolf (2005) emphasized the self-reported motivations of intellectual stimulation, hedonic gain, and skills building, while Ghosh et al. (2002) found reciprocity and skills development as the core motivation – a finding that Hars and Ou (2002) found was supported even among contributors who earned a living from their work on FLOSS. Ghosh and colleagues at the UN University in Maastricht also collaborated with the Wikimedia Foundation to conduct an early survey of Wikipedia editors and readers (Glott et al., 2010) – a study repeated several times – finding that Wikipedia contributors likewise identified many different reasons for participation. Research in remixing communities have suggested a similarly diverse range of motivation. Despite their differences in emphasis, scope, and genre, all of these surveys support the claim that motivations in peer production are diverse and heterogeneous.

A related body of early observational research into motivations of peer production participants underscored their self-selection into particular social roles within projects and communities (Fisher et al., 2006). For example, Shah (2006) found that developers motivated by use value contributed more to what she called "gated" source communities, while developers who were primarily motivated by fun or pleasure contributed primarily to more purely peer production-based projects. Using contributions to differently-licensed FLOSS projects, Belenzon and Schankerman (2008) found that contributors are heterogeneous in their motivational profiles, but also argued that contributors self-sorted among projects. For example, contributors who were more responsive to extrinsic motivations like reputation and employment tended to contribute more to larger corporate-sponsored projects. Studies of free culture communities including Slashdot and Wikipedia suggest comparable findings: participant behavior patterns reflect the sorts of editorial and coordination tasks necessary to produce relevant news and encyclopedic content (Lampe et al., 2007, 2010; Welser et al., 2011; Viegas et al., 2007; Kriplean et al., 2008; McDonald et al., 2011).

Other foundational research on motivation in peer production by Lerner and Schankerman (2010) and others has explored why organizations, firms and governments, rather than individual users, choose to participate in open source software. In this vein, Schweik and English (2012) discuss firm-level motivations including the rate of innovation, the capacity to collaborate with other firms, and avoiding dependence on sole-source providers. Another major motivation for organizations is that participating in peer production could be to permit firms to develop in-house expertise in a tacit-knowledge rich innovation system and to increase the absorptive capacity of the firm (King and Lakhani, 2011).

New Directions

Recent empirical studies of motivation in peer production communities have drawn increasingly on observational data and field experiments. Much of this work has supported the finding that a mixture of motivations attract participants to contribute in peer production. That said, a growing number of studies also suggest that these motives interact with each other in unpredictable ways and, as a result, are vulnerable to "crowding out" when the introduction of extrinsic incentives undermines intrinsic motivation (Frey and Jegen, 2001). Robust evidence of varied participation patterns and motivations of contributors to peer production communities has given rise to other explanations of why peer producers do what they do. In particular, these newer accounts have focused on social status, peer effects, prosocial altruism, group identification, and related social psychological dimensions of group behavior.

Evidence from recent work on Wikipedia and related peer production projects has used observational data as well as field and laboratory experiments. The most important insight provided by some of this newer work is that contributors act for different reasons, and that theories based on a single uniform motivational model are likely to mischaracterize the motivational dynamics. For example, Restivo and van de Rijt (2012) randomly distributed "barnstars" (an informal award that any Wikipedian can give to another in recognition of contributions) to a sample of editors with high numbers of edits and found that the awards caused an increase of subsequent contributions. In a followup study, Restivo and van de Rijt (2014) use an additional experiment to show that the positive effects of the awards is only true among the most active Wikipedians. Similarly, Hill et al. (2012) use observational data to compare award recipients who displayed their barnstars as a public signal of

accomplishment with those who did not and found that the "signalers" edited more than the "non-signalers." Algan et al. (2013) combined observational data on the contribution history of 850 Wikipedians with the performance of the same individuals in a battery of laboratory social dilemmas. They show that that a preference for reciprocity (measured by contributions to public goods games), and trustworthy behavior (measured in a trust game) predict an increase in contributions up to the median editor, but that a preference for reciprocity no longer predicts contributions for editors above the median level. Instead, a preference for social signaling (identified partly using the Hill et al. dataset) predicts above-median contributions, but these high contributors who are signalers do not exhibit particularly prosocial behavior in a laboratory setting. This study also shows that a lab-measured taste for altruism is not associated with increased contributions to Wikipedia. In field experiments conducted by affiliates of the Wikimedia Foundation, Halfaker et al. (2013b) demonstrated that a feedback tool could elicit increased contributions among new editors, so long as the cost (in terms of effort) of contribution remained low. Another study by Antin et al. (2012) combined longitudinal and survey data from a cohort of new Wikipedia contributors to show that differential patterns of contribution – as well as different reasons for contributing – emerge early in individual editors' contribution history.

Evidence from this newer body of research shows that motivations are diverse *within* contributors and that different contributors have different mixes of motivations. This depth and variety makes the problem of designing incentives for peer production systems more complex than foundational work suggested. Such complexity is further compounded by the fact that the impact of any given design intervention (e.g. the introduction of a system of reward or punishment) will necessarily focus on harnessing a particular motivational driver and will necessarily be non-separable from its effects on other motivational drivers. Experimental and observational data have exhaustively documented that the effects of the standard economic incentives like payment or punishment are not separable from their effects on social motivational vectors (Bowles and Hwang, 2008; Bowles and Polania-Reyes, 2012; Frey and Jegen, 2001; Frey, 1997). Similarly, social psychological motivations do not offer a panacea: although evidence from Willer (2009) and Restivo and van de Rijt (2012) and others indicates that the conferral of social status can provide an effective mechanism for eliciting prosocial or altruistic behavior under many conditions, more recent work by Restivo and van de Rijt (2014) suggests that the very same status-based incentives can generate diminishing returns be-

yond a certain point and may even hurt in some situations.

Resolving the tensions between different motivations and incentives presents a design challenge for peer production systems and other collective intelligence platforms. The complex interdependence of motivations, incentive systems, and the social behaviors that distinct system designs elicit has led Kraut and Resnick (2012) to call for evidence-based social design and Benkler (2009, 2011) for cooperative human system design. Research on FLOSS by Alexy and Leitner (2011) has demonstrated that, with the appropriate normative framing, it is sometimes possible to combine paid and unpaid contributions without crowding out intrinsic motivation. That said, we are not aware of any comparable studies of the successful integration of material and prosocial rewards in other areas of peer production. Indeed, websites like Weblogs Inc. have tried, and failed, to augment pure peer production systems by offering material rewards to top contributors. With habituation and practice, internalized prosocial behaviors may lead people to adopt a more, or less, cooperative stance in specific contexts based on their interpretation of the appropriate social practice and its coherence with their self-understanding of how to live well (von Krogh et al., 2012; Benkler and Nissenbaum, 2006). Similarly, practice and experience may shift individuals' proclivity to cooperate and make the work of designers or organizational leaders easier. Antin and Cheshire (2010) have argued that motivations can change and shift over time as certain forms of activity not visible in contribution logs, like reading Wikipedia, can both incentivize others to contribute and constitute an essential step toward participation.

Observational work that estimates the effects of motivational drivers across real communities offers an especially promising avenue for future work. Within this vein, planned and unplanned design changes to communities can provide "natural" experiments and may present particularly useful opportunities for researchers and community managers to evaluate changes to existing incentive systems. For example, Zhang and Zhu (2010) treat the Chinese government's decision to block Wikipedia as a natural experiment and estimate the effect of changes in "group size" on motivation to contribute among non-blocked users. Hill (2011) uses a change to the remixing site Scratch to estimate the causal effect of a newly introduced reputational incentive. Because most peer production systems are facilitated by software-defined web applications, changing the fundamental rules of engagement within peer production projects is simple and common. The diffusion of techniques like A/B testing and quasi-experimental data analysis thus provide an enormous untapped

pool of opportunities to study motivation.

Peer production successfully elicits contributions from diverse individuals with diverse motivations – a quality that continues to distinguish it from similar forms of collective intelligence. Although an early body of literature helped establish and document a complex web of motivations, more recent studies have begun to disentangle the distinct motivational profiles among contributors across cultures and communities; the interactions between motivational drivers and motivational profiles; as well as the design choices that can elicit and shape contribution.

QUALITY

A third focus of research into peer production focuses on the quality of peer production's outputs. The capacity of peer production communities like FLOSS and Wikipedia to create complex, technical products whose quality and scale rivals, and frequently outcompetes, the products of professional workforces in resource-rich firms has driven both popular and academic interest. Indeed, the presence of novel organizational structures and unusual approaches to incentivizing contributors may have been uninteresting to many scholars of peer production if the products had been consistently of low quality. However, by the late 1990s, FLOSS projects including GNU/Linux and Apache were widely considered to be of higher quality than proprietary alternatives (i.e., proprietary UNIX operating systems and proprietary web servers like Microsoft's IIS, respectively). Over time, other peer produced goods have also equaled or overtaken their direct proprietary competitors. Recent work has begun to probe more deeply into different dimensions along which quality can be conceptualized and measured. This new scholarship has given rise to a more nuanced understanding of the different mechanisms through which high quality resources arise, and founder, in peer production.

Foundational Work

FLOSS has been described as built upon a process that will inherently lead to high quality outputs (e.g., Weber, 2004). A prominent FLOSS practitioner and advocate, Raymond (1999) used the example of the Linux kernel to coin what he called Linus' Law – "with enough eyeballs, all bugs are shallow" – suggesting that FLOSS projects, by incorporating contributions from a large number of participants, would have fewer bugs than software developed through

closed models. The mission statement of Open Source Initiative, the non-profit advocacy organization founded by Raymond in 1998, argues that, "the promise of open source is better quality, higher reliability, more flexibility, lower cost, and an end to predatory vendor lock-in"[3] and is the cited inspiration for many early academic studies of FLOSS (e.g., von Krogh and von Hippel, 2006). Practitioner-advocates, FLOSS entrepreneurs, and scholars have each assumed that the peer produced nature of FLOSS carried inherent quality benefits.

Raymond's argument for the superiority of FLOSS' development methodology gained currency as FLOSS emerged from relative obscurity in the late 1990s and began to be adopted by individuals and firms. The GNU/Linux operating system became peer production's most widely discussed success, and remains widely used today. Recognizing the value produced through peer production, many established companies, including Google, HP, and Oracle, followed IBM's early lead and adopted FLOSS in major parts of their technology business. For example, Google's strategic decision to develop Android as FLOSS allowed its product to catch up to and overtake Apple's iOS as the dominant smart phone operating system. FLOSS-focused start-ups, like Red Hat, have ranked among the fastest growing technology companies over the last decade (Savitz, 2012). FLOSS has also been particularly important in the creation of Internet infrastructure. This includes both web servers and server-side scripting languages where, in both cases, a series of FLOSS systems have held a majority of market share for as long as statistics have been kept (Netcraft, 2013; Q-Success, 2013). Although Microsoft's Internet Explorer once held over 95% of the market after it squeezed Netscape Navigator out (illegally, according to antitrust adjudications in both the US and EU), FLOSS browsers Chrome and Firefox now hold nearly half of the market (Vaughan-Nichols, 2013). Recent industry surveys suggest that almost 40% of firms engaged in software development develop and contribute to FLOSS (Lerner and Schankerman, 2010). Bonaccorsi and Rossi (2003), von Krogh (2003), and Lorenzi and Rossi (2008) have each suggested that FLOSS is more innovative in at least particular contexts.

Like FLOSS, Wikipedia's success has been connected to its quality which, in turn, has also been associated with its basis in collaboration from peer production. In a widely cited study, writers at the journal *Nature* presented evidence that a series of Wikipedia articles were of comparable quality to articles

[3] http://opensource.org/about (Accessed July 6, 2013)

on the same subjects in the Encyclopedia Britannica after experts recruited by the journal found approximately the same number of errors in each (Giles, 2005). Although there is no statistically significant correlation between the number of edits to a Wikipedia article and the number of errors in each article within the *Nature* sample,[4] other studies of Wikipedia by Wilkinson and Huberman (2007) and Kittur et al. (2009) have shown that articles with more intense collaboration tend to be rated as being of higher quality by other Wikipedia contributors. Similarly, Greenstein and Zhu (2012) have shown that articles on US politics tend to become less biased as the number and intensity of contributors grows. Viegas et al. (2004) and many others have shown that Wikipedia's open editing model leads to vandalism being detected and removed within minutes or seconds.

The success of FLOSS and Wikipedia has inspired the adoption of peer production in a wide array of industries. Maurer (2010) describes instances where distributed, non-state, non-market action was able to deliver public goods, ranging from nanotechnology safety standards to a synthetic DNA anti-terrorism code. Online business models that depend on peer production (and related commons-based approaches) have out-competed those that depend on more traditional, price-cleared or firm-centric models of production. For example, Flickr, Photobucket, and Google Images, all of which incorporate peer production as a core driver of their platform, have produced vast repositories of stock photography and overshadowed Corbis, the largest stock photography firm using a purely proprietary business model. YouTube, Google video, and Vimeo all rely heavily on peer production for their video content and each is more popular than the studio-produced models of Hulu, Vevo, or even Netflix (though Netflix, the most widely used among these, is roughly as popular as Vimeo). The peer produced TripAdvisor is more popular than Lonely Planet, Fodor's, or Frommers for travel guides. In restaurant reviews, Yelp is more widely used than alternatives. Both for-profit and non-profit organizations that have incorporated peer production models have thrived in the networked environment, often overcoming competition from more traditional, market- and firm-based models.

[4] Analysis of data by the authors of this article using public data from nature study published line. http://www.nature.com/nature/journal/v438/n7070/extref/438900a-s1.doc (Accessed July 6, 2013)

Recent research has stepped back from the simple association of peer production with high quality and has problematized earlier celebratory accounts. This has entailed moving beyond the rhetoric of Raymond (1999) to cease taking the high quality of peer produced resources for granted. Toward that end, this newer wave of scholarship has looked at large datasets of attempts of peer production and documented that the vast majority of would-be peer production projects fail to attract communities – an obvious prerequisite to high quality through collaboration. Additional scholarship has suggested that quality cannot be taken for granted even in situations where large communities of collaborators are mobilized. Finally, recent work has evaluated peer production in terms of a series of different types and dimensions of quality. In following these three paths, scholars have begun to consider variation within peer production projects to understand when and why peer production leads to different kinds of high quality outputs.

Contrary to the claims of foundational accounts, studies have suggested that the vast majority of attempts at peer production fail to attract the large crowd whose aggregate contributions might invoke Linus' law. Lacking any substantive collaboration, these projects do not constitute examples of peer production. Healy and Schussman (2003) first showed that the median number of contributors to a FLOSS project hosted on SourceForge is one. Schweik and English (2012) demonstrated that even among FLOSS projects that have produced successful and sustainable information commons, the median number of contributors is a single individual. A series of studies of remixing suggest that more than 90% of projects uploaded to remixing sites are never remixed at all (Luther et al., 2010; Hill and Monroy-Hernandez, 2013b). Reich et al. (2012) have shown that most attempts to create wikis within classroom environments fail to attract contributions. Hill (2013) has shown that even Wikipedia was preceded by seven other attempts to create online, freely-available, collaborative encyclopedias – all of which failed to created communities on anything approaching the scale of Wikipedia.

The relative infrequency of peer production within domains like FLOSS does not necessarily pose a theoretical problem for peer production since the success of the phenomenon is not defined in terms of the proportion of viable projects. That said, the fact that peer production seems so difficult to achieve, and that the vast majority of attempts at peer production never attract a community, is worrisome for those who value the high quality public goods made

available freely on the Internet through peer production. Although this has constituted an inconvenient fact in peer production practice, it also reflects an important opportunity for future research. By focusing only on the projects that successfully mobilize contributors, researchers interested in *when* peer production occurs or the reasons *why* it succeeds at producing high quality outputs have systematically selected on their dependent variables. An important direction for peer production research will be to study these failures.

Several critics of peer production have questioned whether large scale collaborations can ever lead to high quality outputs. Although relying primarily on anecdotal evidence, Keen (2007) argued that the openness of peer production to participation leads to products that are amateurish and of inferior quality. Similarly, Lanier (2010) suggested that Wikipedia's reliance on mass collaboration results in articles that are sterile and lack a single voice. These dismissive treatments are problematic because peer production obviously does lead to high quality outputs in some situations. A more fruitful approach considers variation in peer production success to understand when and where it works better and worse. For example, Duguid (2006) raised a series of theoretical challenges for peer production and suggests that it may work less well outside of software development. Benkler (2006) speculated that peer production may be better at producing functional works like operating system software and encyclopedias than creative works like code or art. In support of this idea, research by Hill and Monroy-Hernandez (2013a) from the youth-based remixing community *Scratch* suggested that collaborative works tend to be rated lower than *de novo* projects and that the gap in quality between remixes and non-remixes narrows for code-heavy projects, but expands enormously for media-intense works.

Just as peer produced goods vary in their nature and form, there is also enormous variation between and within projects in terms of the dimensions along which quality might be evaluated. In the case of Wikipedia, scholars have assessed the encyclopedia in terms of factual accuracy, scope of coverage, political bias, expert evaluation, and peer evaluation – often drawing different conclusions about the quality of Wikipedia or particular articles. Many studies have suggested that Wikipedia has uneven topic coverage. One group of studies has found uneven cross-cultural and cross-language coverage between Wikipedias and shown that the unevenness corresponds to other forms of socioeconomic and cultural inequality (Arazy et al., 2015; Royal and Kapila, 2009; Pfeil et al., 2006; Bao et al., 2012; Hecht and Gergle, 2009). Inequality in gender participation (see Glott et al., 2010; Hill and Shaw, 2013) has also

been connected to underproduction of Wikipedia articles on topics related to women by Antin et al. (2011) and Reagle and Rhue (2011). Many explanations for this variation exist, but Halavais and Lackaff (2008) offer a succinct summary: "Wikipedia's topical coverage is driven by the interests of its users, and as a result, the reliability and completeness of Wikipedia is likely to be different depending on the subject area of the article."

A related series of studies have sought to unpack the dynamics of collaboration and to understand which features of peer productions support the creation of higher quality content. This topic has been studied especially closely in the case of Wikipedia, where particular organizational attributes, routines, norms, and technical features impact the quality of individual contributions as well as the final, collaborative product. For example, Priedhorsky et al. (2007) and Halfaker et al. (2009) have provided evidence to support the claim that experienced contributors possess a deep sense of ownership and commitment leading them to make more durable contributions that, in Wikipedia's case, have driven the growth of the encyclopedia. Other work has shown that persistent subcommunities of contributors can develop stable dynamics of coordination, leading to more productive collaboration with higher quality products (Kittur and Kraut, 2008). At the same time, Halfaker et al. (2011, 2013a) have shown that experienced editors can wield their expertise against new contributors, removing their contributions and undermining their motivation to edit in the future. Loubser (2010) has suggested that this process can shrink the pool from which future experienced contributors and leaders can be drawn. Gorbatai (2012) has demonstrated that harsh treatment of newcomers can also undermine quality in other, less obvious ways. For example, while experienced editors may perceive low quality contributions from inexperienced editors as a nuisance, "newbie" edits also attract the attention of experienced community members to improve popular pages they might otherwise have ignored.

An important remaining question is whether, and to what extent, the dynamics, routines, infrastructure, and mechanisms that support peer production in intensively studied projects like the English language version of Wikipedia generalize to other communities engaged in other kinds of collaborative activity. Here, peer production researchers may do well to draw on approaches applied to other areas of collective intelligence, such as crowdsourcing, where more precise analyses of the relationship between tasks, system design, and the quality of collaborative outputs have already been performed (e.g., Yu and Nickerson, 2011). Ultimately, this is an area where fur-

ther comparative study across platforms, communities, contexts, and tasks – research like Roth et al. (2008), Kittur and Kraut (2010), and Schweik and English (2012) – will be imperative.

CONCLUSION

For scholars of collective intelligence in the social sciences and legal scholarship, three of the most important and confounding challenges have been in regards to organization, motivation, and quality in the context of peer production. As we have suggested, these themes do not reflect an attempt to be exhaustive, proscriptive, or predictive of future research. In addition to our incomplete reviews of the literature addressing each of these themes, there are many bodies of research that build on, and are related to, peer production that we have not addressed. Nevertheless, our approach illustrates how studies of peer production and collective intelligence have already engaged with sustained debates that range across the social sciences and suggests how they might productively continue to do so. Interested readers seeking a deeper understanding of collective intelligence in the social sciences and legal scholarship should recognize that this chapter provides an entry point into these topics, but remains far from a comprehensive summary.

The limitations of prior peer production scholarship underscores the opportunities for continued social scientific study of peer production and collective intelligence from a variety of theoretical and methodological perspectives. We have highlighted several of what we feel are the most promising areas for future work. As we have suggested, studies that engage in comparative analysis, consider failed as well as successful efforts, and identify causal mechanisms can contribute a great deal. Likewise, work that challenges, extends or elaborates the major theoretical puzzles encompassed by any of the three areas will speak to a wide variety of peer produced phenomena.

Finally, there is an important possibility to use evidence from peer production as a foundation to challenge and extend the literature on collective intelligence and, perhaps, to contribute to more general theories of human behavior and cooperation. Peer production systems combine novel and creative social practices and forms of organization with the creation of truly unprecedented sources of behavioral data. Never, before peer production, have scholars possessed full transcripts of every interaction, communication, and contribution to a collective endeavor or to a particular public good. With the widespread availability of massive datasets of peer production communities

– many of them publicly released under free or open licenses that permit research and reuse – it has become possible to harness the tactics and tools of computational science for the comparative study of social action. Substantively, this means that there are many opportunities to conduct research that cuts across traditional divisions between micro-, meso-, and macro-level social analysis. There are also opportunities to revisit classical concerns of social scientific theory through a lens of evidence that is granular and exhaustive.

REFERENCES

Alexy, O. and Leitner, M. (2011). A fistful of dollars: Are financial rewards a suitable management practice for distributed models of innovation? *European Management Review*, 8(3):165–185.

Algan, Y., Benkler, Y., Fuster Morell, M., and Hergueux, J. (2013). Cooperation in a peer production economy: Experimental evidence from Wikipedia.

Antin, J. and Cheshire, C. (2010). Readers are not free-riders. *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW '10)*, pages 127–130.

Antin, J., Cheshire, C., and Nov, O. (2012). Technology-mediated contributions: editing behaviors among new wikipedians. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 373–382, New York, NY, USA. ACM.

Antin, J., Yee, R., Cheshire, C., and Nov, O. (2011). Gender differences in Wikipedia editing. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, pages 11–14, Mountain View, California. ACM.

Arazy, O., Ortega, F., Nov, O., Yeo, L., and Balila, A. (2015). Determinants of wikipedia quality: the roles of global and local contribution inequality. In *Proceedings of the 2015 ACM conference on Computer supported cooperative work*, CSCW '15, pages 233–236, New York, NY, USA. ACM.

Bao, P., Hecht, B., Carton, S., Quaderi, M., Horn, M., and Gergle, D. (2012). Omnipedia: bridging the wikipedia language gap. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 1075–1084, New York, NY, USA. ACM.

Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., and Kraut, R. E. (2004). Using social psychology to motivate contributions to

online communities. In *Proceedings of the 2004 ACM conference on Computer Supported Cooperative Work*, pages 212–221, Chicago, Illinois, USA. ACM.

Belenzon, S. and Schankerman, M. A. (2008). Motivation and sorting in open source software innovation. SSRN Scholarly Paper ID 1401776, Social Science Research Network, Rochester, NY.

Benkler, Y. (2001). The battle over the institutional ecosystem in the digital environment. *Commun. ACM*, 44(2):84–90.

Benkler, Y. (2002). Coase's penguin, or, linux and the nature of the firm. *Yale Law Journal*, 112(3):369–446.

Benkler, Y. (2004). Sharing nicely: On shareable goods and the emergence of sharing as a modality of economic production. *The Yale Law Journal*, 114(2):273–358.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.

Benkler, Y. (2009). Law, policy, and cooperation. In Balleisen, E. and Moss, D., editors, *Government and Markets: Toward A New Theory of Regulation*. Cambridge University Press.

Benkler, Y. (2011). *The Penguin and the Leviathan: How Cooperation Triumphs over Self-Interest*. Crown Business.

Benkler, Y. (2013). Peer production and cooperation. In Bauer, J. M. and Latzer, M., editors, *Handbook on the Economics of the Internet*. Edward Elgar.

Benkler, Y. and Nissenbaum, H. (2006). Commons-based peer production and virtue. *Journal of Political Philosophy*, 14(4):394–419.

Bonaccorsi, A. and Rossi, C. (2003). Why open source software can succeed. *Research Policy*, 32(7):1243–1258.

Bowles, S. and Hwang, S.-H. (2008). Social preferences and public economics: Mechanism design when social preferences depend on incentives. *Journal of Public Economics*, 92(8-9):1811–1820.

Bowles, S. and Polania-Reyes, S. (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature*, 50(2):368–425.

Brabham, D. C. (2013). *Crowdsourcing*. The MIT Press, Cambridge, Massachusetts ; London, England.

Bryant, S. L., Forte, A., and Bruckman, A. (2005). Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, GROUP '05, pages 1–10, New York, NY, USA. ACM.

Burke, M. and Kraut, R. (2008). Mopping up: modeling wikipedia promotion decisions. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, pages 27–36, San Diego, CA, USA. ACM.

Butler, B. S., Joyce, E., and Pike, J. (2008). Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1101–1110, New York, NY, USA. ACM.

Cheshire, C. (2007). Selective incentives and generalized information exchange. *Social Psychology Quarterly*, 70:82–100.

Cheshire, C. and Antin, J. (2008). The social psychological effects of feedback on the production of internet information pools. *Journal of Computer-Mediated Communication*, 13(3):705–727.

Coleman, E. G. (2012). *Coding Freedom: The Ethics and Aesthetics of Hacking*. Princeton University Press.

Coleman, G. and Hill, B. M. (2004). The social production of ethics in debian and free software communities: Anthropological lessons for vocational ethics. In Koch, S., editor, *Free/Open Source Software Development*.

Crowston, K., Wei, K., Howison, J., and Wiggins, A. (2010). Free/libre open source software: What we know and what we do not know. *ACM Computing Surveys*, 44(2):2012.

David, P. A. and Shapiro, J. S. (2008). Community-based production of open-source software: What do we know about the developers who participate? *Information Economics and Policy*, 20(4):364–398.

Duguid, P. (2006). Limits of self-organization: Peer production and "laws of quality". *First Monday*, 11(10).

Fisher, D., Smith, M., and Welser, H. T. (2006). You are who you talk to: Detecting roles in usenet newsgroups. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences - Volume 03*, HICSS '06, pages 59.2–, Washington, DC, USA. IEEE Computer Society.

Forte, A. and Bruckman, A. (2008). Scaling consensus: Increasing decentralization in Wikipedia governance. In *Proceedings of the 41st Annual Hawaii International Conference on System Science*, page 157. IEEE.

Forte, A., Larco, V., and Bruckman, A. (2009). Decentralization in Wikipedia governance. *J. Manage. Inf. Syst.*, 26(1):49–72.

Frey, B. S. (1997). A constitution for knaves crowds out civic virtues. *The Economic Journal*, 107(443):1043–1053.

Frey, B. S. and Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, 15(5):589–611.

Fuster Morell, M. (2010). *Governance of Online Creation Communities: Provision of infrastructure for the building of digital commons*. Ph.d. dissertation, European University Institute, Florence, Italy.

Ghosh, R. A. (1998). Cooking pot markets: An economic model for the trade in free goods and services on the internet. *First Monday*, 3(3).

Ghosh, R. A. (2005). Understanding free software developers: Understandings from the FLOSS study. In Feller, J., Fitzgerald, B., Hissam, S. A., and Lakhani, K. R., editors, *Perspectives on Free and Open Source Software*. MIT Press.

Ghosh, R. A., Glott, R., Krieger, B., and Robles, G. (2002). Free/libre and open source software: Survey and study. Technical report, Maastricht Economic Research Institute on Innovation and Technology, University of Maastricht, The Netherlands.

Ghosh, R. A. and Prakash, V. V. (2000). The Orbiten free software survey. *First Monday*, 5(7).

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901.

Glott, R., Ghosh, R., and Schmidt, P. (2010). Wikipedia survey. Technical report, UNU-MERIT, Maastricht, Netherlands. Accessed April 4, 2011.

Gorbatai, A. (2012). Aligning collective production with demand : Evidence from Wikipedia.

Greenstein, S. and Zhu, F. (2012). Is Wikipedia biased? *American Economic Review*, 102(3):343–348.

Halavais, A. and Lackaff, D. (2008). An analysis of topical coverage of Wikipedia. *Journal of Computer-Mediated Communication*, 13(2):429–440.

Halfaker, A., Geiger, R. S., Morgan, J. T., and Riedl, J. (2013a). The rise and decline of an open collaboration system how Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 57(5):664–688.

Halfaker, A., Keyes, O., and Taraborelli, D. (2013b). Making peripheral participation legitimate: reader engagement experiments in wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 849–860, New York, NY, USA. ACM.

Halfaker, A., Kittur, A., Kraut, R., and Riedl, J. (2009). A jury of your peers: quality, experience and ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, WikiSym '09, pages 15:1–15:10, New York, NY, USA. ACM.

Halfaker, A., Kittur, A., and Riedl, J. (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, pages 163–172, New York, NY, USA. ACM.

Hars, A. and Ou, S. (2002). Working for free? motivations for participating in open-source projects. *International Journal of Electronic Commerce*, 6(3):25–39.

Healy, K. and Schussman, A. (2003). The ecology of open-source software development.

Hecht, B. and Gergle, D. (2009). Measuring self-focus bias in community-maintained knowledge repositories. In *Proceedings of the fourth international conference on Communities and technologies*, C&T '09, pages 11–20, New York, NY, USA. ACM.

Hill, B. M. (2011). Causal effects of a reputation-based incentive in an peer production community.

Hill, B. M. (2013). *Essays on volunteer mobilization in peer production*. Ph.d. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Hill, B. M. and Monroy-Hernandez, A. (2013a). The cost of collaboration for code and art: evidence from a remixing community. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 1035–1046, New York, NY, USA. ACM.

Hill, B. M. and Monroy-Hernandez, A. (2013b). The remixing dilemma the trade-off between generativity and originality. *American Behavioral Scientist*, 57(5):643–663.

Hill, B. M. and Shaw, A. (2013). The Wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PLoS ONE*, 8(6):e65782.

Hill, B. M., Shaw, A., and Benkler, Y. (2012). Status, social signalling and collective action in a peer production community. Working Paper.

Hindman, M. (2008). *The Myth of Digital Democracy*. Princeton University Press, Princeton, NJ.

Howison, J., Conklin, M., and Crowston, K. (2006). FLOSSmole. *International Journal of Information Technology and Web Engineering*, 1(3):17–26.

Jemielniak, D. (2014). *Common Knowledge?: An Ethnography of Wikipedia*. Stanford University Press, Stanford, California.

Keegan, B. and Gergle, D. (2010). Egalitarians at the gate. *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW '10)*, (Savannah, Georgia, ACM Press):131–134.

Keen, A. (2007). *The Cult of the Amateur: How Today's Internet is Killing Our Culture*. Crown Business, 3rd printing edition.

Kelty, C. (2008). *Two Bits: The Cultural Significance of Free Software*. Duke University Press, Durham.

King, A. A. and Lakhani, K. R. (2011). The contingent effect of absorptive capacity: An open innovation analysis.

Kittur, A. and Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW)*, pages 37–46, San Diego, CA, USA. ACM.

Kittur, A. and Kraut, R. E. (2010). Beyond Wikipedia: coordination and conflict in online production groups. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW)*, pages 215–224, Savannah, Georgia, USA. ACM.

Kittur, A., Lee, B., and Kraut, R. E. (2009). Coordination in collective intelligence: the role of team structure and task interdependence. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 1495–1504.

Kollock, P. (1999). The economies of online cooperation: Gifts and public goods in cyberspace. In Smith, M. and Kollock, P., editors, *Communities in Cyberspace*, pages 220–239. Routledge, London.

Konieczny, P. (2009). Governance, organization, and democracy on the internet: The iron law and the evolution of Wikipedia. *Sociological Forum*, 24(1):162–192.

Kraut, R. E. and Resnick, P. (2012). *Building Successful Online Communities: Evidence-Based Social Design*. The MIT Press.

Kreiss, D., Finn, M., and Turner, F. (2011). The limits of peer production: Some reminders from max weber for the network society. *New Media & Society*, 13(2):243–259.

Kriplean, T., Beschastnikh, I., and McDonald, D. W. (2008). Articulations of wikiwork: uncovering valued work in wikipedia through barnstars. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, pages 47–56, New York, NY, USA. ACM.

Lakhani, K. (2006). *The core and the periphery in distributed and self-organizing innovation systems*. Ph.d. dissertation, Massachusetts Institute of Technology, Sloan School of Management.

Lakhani, K. and Wolf, B. (2005). Why hackers do what they do: Understanding motivation and effort in free/open source software projects. In Feller, J., Fitzgerald, B., Hissam, S. A., and Lakhani, K. R., editors, *Perspectives on Free and Open Source Software*, pages 3–22. MIT Press.

Lampe, C. and Resnick, P. (2004). Slash(dot) and burn: Distributed moderation in a large online conversation space. In *Conference on Human Factors in Computing Systems (CHI)*, pages 543–550, Vienna, Austria. ACM Press.

Lampe, C., Wash, R., Velasquez, A., and Ozkaya, E. (2010). Motivations to participate in online communities. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1927–1936, Atlanta, Georgia, USA. ACM.

Lampe, C. A., Johnston, E., and Resnick, P. (2007). Follow the reader: filtering comments on slashdot. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, pages 1253–1262, New York, NY, USA. ACM.

Lanier, J. (2010). *You Are Not a Gadget: A Manifesto*. Knopf, 1 edition.

Lerner, J. and Schankerman, M. (2010). *The comingled code: open source and economic development*. MIT Press, Cambridge, Massachusetts.

Lerner, J. and Tirole, J. (2002). Some simple economics of open source. *Journal of Industrial Economics*, 50(2):197–234.

Leskovec, J., Huttenlocher, D., and Kleinberg, J. (2010). Governance in social media: A case study of the Wikipedia promotion process. In *International AAAI Conference on Weblogs and Social Media (ICWSM)*, Washington D.C. ICWSM 2010 - International AAAI Conference on Weblogs and Social Media.

Lorenzi, D. and Rossi, C. (2008). Assessing innovation in the software sector: Proprietary vs. FOSS production mode. preliminary evidence from the Italian case. In Russo, B., Damiani, E., Hissam, S., Lundell, B., and Succi, G., editors, *Open Source Development, Communities and Quality*, number 275 in IFIP - The International Federation for Information Processing, pages 325–331. Springer US.

Loubser, M. (2010). *Organisational Mechanisms in Peer Production: the Case of Wikipedia*. DPhil, Oxford Internet Institute, Oxford University, Oxford, UK.

Luther, K., Caine, K., Zeigler, K., and Bruckman, A. (2010). Why it works (when it works): Success factors in online creative collaboration. In *Proceedings of the ACM Conference on Supporting Group Work.*, Sanibel Island, Florida, USA. ACM.

Luther, K., Fiesler, C., and Bruckman, A. (2013). Redistributing leadership in online creative collaboration. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 1007–1022, New York, NY, USA. ACM.

Maurer, S. (2010). Five easy pieces: Case studies of entrepreneurs who organized private communities for a public purpose. SSRN Scholarly Paper ID 1713329, Social Science Research Network, Rochester, NY.

McDonald, D. W., Javanmardi, S., and Zachry, M. (2011). Finding patterns in behavioral observations by automatically labeling forms of wikiwork in barnstars. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, pages 15–24, New York, NY, USA. ACM.

Moglen, E. (1999). Anarchism triumphant: Free software and the death of copyright. *First Monday*, 4(8).

Netcraft (2013). January 2013 web server survey. http://news.netcraft.com/archives/2013/01/07/january-2013-web-server-survey-2.html.

Nov, O. (2007). What motivates Wikipedians? *Commun. ACM*, 50(11):60–64.

O'Mahony, S. and Bechky, B. A. (2008). Boundary organizations: Enabling collaboration among unexpected allies. *Administrative Science Quarterly*, 53(3):422–459.

O'Mahony, S. and Ferraro, F. (2007). The emergence of governance in an open source community. *Academy of Management Journal*, 50(5):1079–1106.

Ortega, F. (2009). *Wikipedia: A Quantitative Analysis*. Ph.d. dissertation, Universidad Rey Juan Carlos, Madrid, Spain.

Osterloh, M., Frost, J., and Frey, B. S. (2002). The dynamics of motivation in new organizational forms. *International Journal of the Economics of Business*, 9(1):61–77.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, New York, NY.

Panciera, K., Halfaker, A., and Terveen, L. (2009). Wikipedians are born, not made: a study of power editors on wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work*, GROUP '09, pages 51–60, New York, NY, USA. ACM.

Pentzold, C. (2011). Imagining the Wikipedia community: What do wikipedia authors mean when they write about their 'community'? *New Media & Society*, 13(5):704–721.

Pfeil, U., Zaphiris, P., and Ang, C. S. (2006). Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer-Mediated Communication*, 12(1):88–113.

Priedhorsky, R., Chen, J., Lam, S. T. K., Panciera, K., Terveen, L., and Riedl, J. (2007). Creating, destroying, and restoring value in Wikipedia. In *Proceedings of the 2007 international ACM conference on Supporting group work*, GROUP '07, pages 259–268, New York, NY, USA. ACM.

Q-Success (2013). Usage statistics and market share of server-side programming languages for websites, July 2013.

Rafaeli, S. and Ariel, Y. (2008). Online motivational factors: Incentives for participation and contribution in Wikipedia. In Barak, A., editor, *Psychological Aspects of Cyberspace: Theory, Research, Applications*. Cambridge University Press.

Ransbotham, S. and Kane, G. C. (2011). Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quarterly-Management Information Systems*, 35(3):613.

Raymond, E. S. (1999). *The cathedral and the bazaar: Musings on Linux and open source by an accidental revolutionary*. O'Reilly and Associates, Sebastopol, CA.

Reagle, J. (2010). *Good Faith Collaboration: The Culture of Wikipedia*. MIT Press, Cambridge Mass.

Reagle, J. and Rhue, L. (2011). Gender bias in Wikipedia and britannica. *International Journal of Communication*, 5:21.

Reich, J., Murnane, R., and Willett, J. (2012). The state of wiki usage in u.s.: K-12 schools leveraging web 2.0 data warehouses to assess quality and equity in online learning environments. *Educational Researcher*, 41(1):7–15.

Restivo, M. and van de Rijt, A. (2012). Experimental study of informal rewards in peer production. *PLoS ONE*, 7(3):e34358.

Restivo, M. and van de Rijt, A. (2014). No praise without effort: experimental evidence on how rewards affect Wikipedia's contributor community. *Information, Communication & Society*, pages 1–12.

Roth, C., Taraborelli, D., and Gilbert, N. (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*, pages 1–5, Porto, Portugal. ACM.

Royal, C. and Kapila, D. (2009). What's on Wikipedia, and what's not ... ?: Assessing completeness of information. *Social Science Computer Review*, 27(1):138–148.

Savitz, E. (2012). The forbes fast tech 25: Our annual list of growth kings. *Forbes*.

Schweik, C. M. and English, R. C. (2012). *Internet success: a study of open-source software commons*. MIT Press, Cambridge, Mass.

Shah, S. K. (2006). Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science*, 52(7):1000–1014.

Shaw, A. (2012). Centralized and decentralized gatekeeping in an open online collective. *Politics & Society*, 40(3):349–388.

Shaw, A. and Hill, B. M. (2014). Laboratories of oligarchy? how the iron law applies to peer production. *Journal of Communication*, 64(2):215–238.

Spaeth, S., Stuermer, M., and Krogh, G. V. (2010). Enabling knowledge creation through outsiders: towards a push model of open innovation. *International Journal of Technology Management*, 52(3/4):411 – 431.

Vaughan-Nichols, S. J. (2013). The web browser wars continue, and #1 is... well, that depends on whom you ask. *ZDNET*.

Viegas, F. B., Wattenberg, M., and Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 575–582, New York, NY, USA. ACM Press.

Viegas, F. B., Wattenberg, M., McKeon, M. M., and Schuler, D. (2007). The hidden order of Wikipedia. In *Online Communities and Social Computing*, HCII, pages 445–454. Springer-Verlag, Berlin, Heidelberg.

von Hippel, E. and von Krogh, G. (2003). Open source software and the 'private-collective' innovation model: Issues for organization science. *Organization Science*, 14(2):209–223.

von Krogh, G. (2003). Open-source software development. *MIT Sloan Management Review*, Spring.

von Krogh, G., Haefliger, S., Spaeth, S., and Wallin, M. W. (2012). Carrots and rainbows: Motivation and social practice in open source software development. *MIS Quarterly*, 36(2):649–676.

von Krogh, G., Spaeth, S., and Lakhani, K. R. (2003). Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, 32(7):1217–1241.

von Krogh, G. and von Hippel, E. (2006). The promise of research on open source software. *Management Science*, 52(7):975–983.

Wang, L. S., Chen, J., Ren, Y., and Riedl, J. (2012). Searching for the goldilocks zone: trade-offs in managing online volunteer groups. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 989–998, New York, NY, USA. ACM.

Weber, S. (2004). *The Success of Open Source*. Harvard University Press.

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., and Smith, M. (2011). Finding social roles in Wikipedia. In *Proceedings of the 2011 iConference*, iConference '11, pages 122–129, New York, NY, USA. ACM.

West, J. and O'Mahony, S. (2005). Contrasting community building in sponsored and community founded open source projects. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 7 - Volume 07*, page 196.3. IEEE Computer Society.

Wilkinson, D. M. and Huberman, B. A. (2007). Cooperation and quality in wikipedia. In *Proceedings of the 2007 international symposium on Wikis*, WikiSym '07, pages 157–164, New York, NY, USA. ACM.

Willer, R. (2009). Groups reward individual sacrifice: The status solution to the collective action problem. *American Sociological Review*, 74(1):23 –43.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. Free Press, New York, New York.

Wu, F., Wilkinson, D. M., and Huberman, B. A. (2009). Feedback loops of attention in peer production. In *Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 04*, CSE '09, pages 409–415, Washington, DC, USA. IEEE Computer Society.

Yasseri, T., Sumi, R., Rung, A., Kornai, A., and Kertesz, J. (2012). Dynamics of conflicts in Wikipedia. *PLoS ONE*, 7(6):e38869.

Yu, L. and Nickerson, J. V. (2011). Cooks or cobblers? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 1393. ACM Press.

Zhang, X. M. and Zhu, F. (2010). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review*.

Zhu, H., Kraut, R., and Kittur, A. (2012). Organizing without formal organization: group identification, goal setting and social modeling in directing online production. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 935–944, New York, NY, USA. ACM.

Zhu, H., Kraut, R. E., Wang, Y.-C., and Kittur, A. (2011). Identifying shared leadership in Wikipedia. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 3431–3434, New York, NY, USA. ACM.



Routledge
Taylor & Francis Group

Information, Communication & Society

ISSN: 1369-118X (Print) 1468-4462 (Online) Journal homepage: http://www.tandfonline.com/loi/rics20

Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia

Eszter Hargittai & Aaron Shaw

To cite this article: Eszter Hargittai & Aaron Shaw (2015) Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia, Information, Communication & Society, 18:4, 424-442, DOI: 10.1080/1369118X.2014.957711

To link to this article: http://dx.doi.org/10.1080/1369118X.2014.957711

Published online: 04 Nov 2014.

Submit your article to this journal

Article views: 693

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=rics20

Download by: [8.28.179.180] **Date:** 06 October 2015, At: 14:09

Information, Communication & Society, 2015
Vol. 18, No. 4, 424–442, http://dx.doi.org/10.1080/1369118X.2014.957711

Routledge
Taylor & Francis Group

Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia

Eszter Hargittai* and Aaron Shaw

Communication Studies, Northwestern University, Evanston, IL 60208, USA

(*Received 10 June 2014; accepted 20 August 2014*)

Despite the egalitarian rhetoric surrounding online cultural production, profound gender inequalities remain in who contributes to one of the most visited Web sites worldwide, Wikipedia. In analyzing this persistent disparity, previous research has focused on aspects of current contributors and the existing Wikipedia community. We draw on unique panel survey data of young adults with information about both Wikipedia contributors and non-contributors. We examine the role of people's background attributes and Internet skills in participation on the site. We find that the gender gap in editing is exacerbated by a similarly significant Internet skills gap. Our results show that the most likely contributors are high-skilled males and that among low-skilled Internet users no gender gap in Wikipedia contributions exists. Our findings suggest that efforts to understand the gender gap must also take Internet skills into account.

Keywords: digital inequality; gender; Internet skills; participation; peer production; Wikipedia

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Introduction

The growth of online communities and social media has inspired enthusiasm about new opportunities for people to express themselves and participate in the production of freely available cultural goods (Benkler, 2006; Jenkins, 2006). Indeed, millions of people worldwide have contributed their time and effort toward the creation of public resources such as Wikipedia, Free/Libre Open Source Software, social question-and-answer sites like Yahoo! Answers, StackOverflow, and more. Yet, research has consistently found that women are less likely to contribute to many such endeavors than men despite being online at similar rates (Zickuhr & Smith, 2012), with especially robust results found among Wikipedia editors (Antin, Yee, Cheshire, & Nov, 2011; Benkler, 2002; Cohen, 2011; Collier & Bear, 2012; Glott, Ghosh, & Schmidt, 2010; Hill & Shaw, 2013). Given that large numbers – over 87 million unique monthly visitors in 2013 – of Internet users consult Wikipedia regularly its content reaches a vast audience (ComScore, 2013). Accordingly, gender disparities in who authors the material on the site are increasingly a matter of public concern as the content disproportionately reflects the interests and perspectives of its most active contributors who are mainly men. This 'gender gap' in contributions has persisted despite public outcry (Eckert & Steiner, 2013; Filipacchi, 2013; Gleick, 2013). Campaigns by the Wikimedia Foundation (WMF) intended to encourage the retention of new female editors (Morgan, Bouterse, Walls, & Stierch, 2013) have not resulted in vast changes.

*Corresponding author. Email: eszter@northwestern.edu

© 2014 Taylor & Francis

Downloaded by [8.28.179.180] at 14:09 06 October 2015

The causes and mechanisms that sustain the Wikipedia gender gap – as with other forms of gender inequity in society – are likely complex. However, most previous research on the Wikipedia gender gap has looked for answers *within* the community of Wikipedia editors, analyzing the community culture, the dynamics of contribution, and testing design solutions intended to mitigate the problem among current participants (Antin et al., 2011; Halfaker, Kittur, & Riedl, 2011; Morgan et al., 2013). We contend that these approaches are an important step, but are necessarily limited because they only consider existing Wikipedia contributors when seeking to understand or resolve participation inequalities. Some of the strongest mechanisms of exclusion likely concern barriers that prevent initial contribution, leading to differentiation in who becomes an editor in the first place (Collier & Bear, 2012). Accordingly, by limiting studies to those who have at one point or another already contributed, research misses an important part of the puzzle: What factors explain who selects into contributing to Wikipedia in the first place?

Differential rates of cultural participation are an important line of sociological inquiry, as they have been linked to social inequality both in terms of who participates and the resulting benefits of participation (Alderson, Junisbai, & Heacock, 2007; DiMaggio & Ostrower, 1990; DiMaggio & Useem, 1978). Traditionally, participation has mainly consisted of audience membership, for example museum or concert visits, since contributing content to public discourse, while possible (Perrin & Vaisey, 2008) had much more considerable barriers to entry. In the digital age, new modes of content production are available with lower barriers to entry and larger potential audiences than was previously within reach (Hargittai, 2000), but inequalities have nonetheless persisted in who contributes written, audio and visual content to the Web (Blank, 2013; Correa, 2010; Hargittai & Walejko, 2008; Schradie, 2011).

Among the many potential barriers to initial participation, we focus our analysis on the role of Internet experiences and skills. In doing so, we build on an extensive body of research considering the role of Internet experiences and skills in conjunction with various demographic and socioeconomic factors, in shaping what people do online (Hargittai & Hsieh, 2013; Howard, Rainie, & Jones, 2001; Schradie, 2011). This previous work has shown that comparable access to information resources and online tools does not necessarily lead to similar online behavior (Zillien & Hargittai, 2009). A persistent 'skills divide' separates those who engage in many online behaviors from those who do not (Hargittai & Hinnant, 2008). Of particular relevance to the case of the Wikipedia gender gap, research on differentiated Internet experiences and skills has shown that socioeconomic inequalities and demographic differences tend to translate into differentiated patterns of Internet use across cultural, civic, and relational spheres of activity (see Hargittai & Hsieh, 2013 for a review of related literature). We have every reason to believe that this might also be the case with Wikipedia and the gender gap.

Based on earlier work highlighting the relationship between gender and Internet experiences and skills (Correa, 2010; Hargittai & Shafer, 2006), we hypothesize that gender inequalities in cultural production in Wikipedia contribution are likely intertwined with Internet experiences and skills. In other words, we ask whether the gender gap in contributions is more, less, or equally present among low- and high-skill Internet users.[1] More formally, we test (a) whether an Internet skills divide separates those who do and do not contribute to Wikipedia; and, if so, (b) how this Internet skills divide relates to the gender gap in participation. We conduct these tests using a unique panel survey data set of 547 diverse young adults (ages 21–22 years at the time of the second wave) to examine what factors are related to Wikipedia contributions.

We operationalize participation on Wikipedia as any experience making edits to existing entries (including the addition of an image) or starting a new entry. For the purposes of this paper, we do not discriminate between the value of fixing a mistake versus adding new material, editing an existing entry or starting a new one. This is undoubtedly a low threshold for

participation, but as we will show, it is meaningful given the low proportion of respondents who engage with Wikipedia in any of those ways.

We use the terms 'editor' and 'contributor' somewhat interchangeably as the former reflects the common term used within the Wikipedia community to describe anyone who has made any contribution of any kind to the encyclopedia. Unlike previous studies of the Wikipedia-editing gender gap, our data set includes both individuals who have edited Wikipedia and many (indeed, the majority) who have never done so. This diversity makes it possible for us to construct a meaningful test of whether or not Internet experiences and skills may predict whether or not male and female respondents are likely to have ever contributed to Wikipedia.

Our central contribution in this paper consists of an empirical analysis of whether and how Internet experiences and skills may or may not interact with the gender gap in online participation regarding Wikipedia editing. As previous literature on Wikipedia contributions has largely overlooked formal analyses of Internet experiences and skills, the tests we conduct here constitute a novel contribution. Similarly, none of the past work examining Internet skills has focused on contributions to Wikipedia, one of the most widely used online resources and instances of participatory cultural production (ComScore, 2013). In addition, previous work examining gender as well as Internet experiences and skills has not specified the relationship between these variables with much precision. Applying nuanced techniques, we conclude that gender and Internet skills likely have a relatively mild interaction with each other, reinforcing the gender gap at the high end of the Internet skills spectrum.

The rest of the paper is structured as follows: in the next section, we frame our approach and research questions in relation to previous findings on the factors that shape patterns of online participation, and in particular previous work considering the gender gap in Wikipedia contributions. We also briefly review prior work on the role of Internet experiences and skills in explaining variations in a range of online behaviors, including social media participation and other sorts of online cultural production. In the following section, we describe the data set and methods we use to address our research questions, emphasizing the unique characteristics of our sampling frame and longitudinal data collection that make for a particularly compelling test of our research questions. Then we present the results of our analyses, in which we model the relationship between Wikipedia contributions, gender, and Internet experiences and skills as well as a number of other variables. We end with a discussion of our findings and their limitations, as well as their significance and implications for future work.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

What determines who participates in online communities and Wikipedia?

This study contributes to the extensive literature on inequalities in cultural participation specifically focusing on contributions to Wikipedia. Previous studies have emphasized several possible determinants of whether or not individuals contribute to online communities more generally, and to this online encyclopedia more specifically, including background characteristics such as gender as well as Internet experiences and skills. Below, we briefly review related literature.

Background characteristics including gender

Previous work has focused on how certain background characteristics such as gender may shape who participates in online communities. The Wikipedia gender gap represents arguably the most prominent example of this line of inquiry (Cohen, 2011; Glott et al., 2010). Research into the Wikipedia gender gap originated with the results of a 2008 survey conducted by the WMF and the United Nations University Maastricht that found that less than 13% of all editors at the time were women, 18% among US adult editors (Glott et al., 2010). Recent analyses taking

into account sampling bias introduced by the fact that said survey allowed opt-in participation suggest that these estimates likely overstated the extent of the gender gap whereby 16% of editors globally and 23% in the United States are female (Hill & Shaw, 2013). Even the updated figures suggest a considerable gender gap, nonetheless. The presence of such extreme inequality has motivated a growing body of work investigating the relationship between participation in Wikipedia and gender. This work has shown that gender differences extend to editors' earliest patterns of contribution (Antin et al., 2011) and also manifest in how contributors interact with other editors (Collier & Bear, 2012; Halfaker, Geiger, Morgan, & Riedl, 2012). Related research considers the effects such diversity and differential experiences may have on the quality and quantity of contributions (Chen, Ren, & Riedl, 2010; Lam et al., 2011) and describes the effects of design interventions intended to reverse the trend (Morgan et al., 2013).

While these previous studies offer insights into gender differences among Wikipedia contributors, the majority of them have sampled exclusively from within the population of Wikipedia editors or used recruitment materials posted on Wikipedia itself. As a result, the respondent pools are biased to include those who already participate or those willing to participate in research announced *through* Wikipedia itself. The results therefore cannot speak to factors that may prevent the majority of females from ever contributing to the encyclopedia in the first place.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Differentiated Internet experiences and skills

Digital inequality research emphasizes that beyond mere physical access to the Internet, people's online experiences also vary based on their autonomy of use, how long they have been Internet users, the amount of time they spend online, their self-efficacy, and their Internet skills (van Dijk, 2005; DiMaggio, Hargittai, Celeste, & Shafer, 2004). People with more autonomy of use, that is, freedom to use the Internet when and where they wish, engage in more capital-enhancing activities (Hassani, 2006) as do people who spend more time online (Hargittai & Hinnant, 2008). Self-efficacy (Bandura, 1977) or the belief in one's ability to achieve certain outcomes has also been linked to types of Internet uses (Eastin & LaRose, 2000; Zhao, Lu, Huang, & Wang, 2010). Research has also found that Web-use skills, that is, the ability to use the Internet effectively and efficiently, is related to types of online behavior from the likelihood of joining certain communities like Twitter to online content creation and engagement (Correa, 2010; Hargittai & Hinnant, 2008; Hargittai & Litt, 2011; Wasserman & Richmond-Abbott, 2005). Accordingly, it is important to consider such nuanced measures of Internet experiences when looking at predictors of online engagement, in this case Wikipedia contributions. While studies of online behavior often include some measure of frequency of use, the other concepts are less likely to show up in analyses, likely in part due to the lack of appropriate measures in data sets. Our study overcomes this challenge by drawing on a data set with measures of all of these factors.

Research questions

Our study seeks to understand the relationship of Internet experiences and skills to the act of participating in online cultural production, specifically, contributing to Wikipedia. We ask whether and to what extent Internet experiences and skills – in addition to gender, socioeconomic status, and other background attributes – may help explain who participates in editing Wikipedia. Drawing on the growing literature examining the Wikipedia gender gap as well as past work demonstrating a relationship between online behavior, gender, and skill, we also focus closely on how gender and skill together may account for variation in Wikipedia contribution.

Data and methods

We analyze panel survey data from a diverse group of young adults to look at who contributes to Wikipedia. Selection into the study was completely independent of Wikipedia use thereby avoiding concerns of sampling on the dependent variable. The data set includes a rich set of covariates about both contributors and non-contributors making it possible to examine what factors predict engagement in Wikipedia editing.

We know of no nationally representative panel data set that includes detailed information about people's Internet uses and skills, making it impossible to explore these questions on such a sample. Our respondents are diverse on several factors, but homogenous regarding their age and level of education. These two variables have been shown as important correlates of Internet usage (Dutton & Blank, 2011). By focusing on a generation that was 'born digital' (Palfrey & Gasser, 2008) we are excluding people with little-to-no online experiences and thus ensuring some level of online know-how for the majority of respondents. Accordingly, findings about differences in online participation are likely to be conservative. That is, were we to find effects of Internet experiences and skills on Wikipedia contributions, those are likely to be conservative compared to a more diverse – both demographically and in the related realm of online experiences – sample. We address the implications of the sample's particular make-up in the discussion of our findings.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Data collection

The results reported here draw on two waves of survey data collected about the same people over four years: the first administered in spring 2009; the second in summer 2012 on a diverse group of 547 young adults. In spring 2009, we surveyed a representative sample of first-year students at an urban public university, a school with which neither author has ever had any affiliation beyond this study. Its selection for the project was due to (1) it being one of the most racially and ethnically diverse research universities in the United States (US News & World Report, 2009) relevant for the overall research project's focus on questions of social inequality and (2) its curriculum requiring that all students take a particular class: the First-Year Writing Program. Working with this course allowed reaching an unbiased sample of the university's student body. It also led to a high response rate since surveys were administered on paper during class sessions.

We used paper-pencil methods as an online questionnaire would have biased against people who are less likely to spend time on the Web, have less private access to the Internet, and who feel less comfortable filling out online forms. For similar reasons, and to control for mode of data collection, the 2012 follow-up study also relied on a paper-pencil instrument.

All 2009 respondents who had consented in 2009 to being re-contacted (the majority, 98%, had done so) received the 2012 survey in postal mail with a self-addressed stamped envelope. In the first wave of the study, respondents did not receive any incentives. In the 2012 wave, they were offered a $25 gift certificate to Amazon and were entered into a drawing for one of two iPads.

The initial data set in 2009 included valid responses from 1115 first-year students. The questionnaire included an item to verify respondents' attentiveness to the survey. A small portion, 4.5%, answered incorrectly to this verification question, suggesting that they were checking off responses randomly. These students were excluded from the study so as to minimize error. The 1115 students all answered the verification question correctly. Of the 92 course sections in the program, 86 participated in the study for a 93.5% participation rate on the part of sections. Of all the students enrolled in the class, 80.5% filled out the questionnaire.

In 2012, we received 547 valid responses. We had permission to follow up with and contact information for 1066 people for a 4% refusal from the original sample of 1115. For 17 people, constituting one and a half percent of the original group, both email and postal mail addresses bounced meaning no possibility for re-contact.[2] The 2012 questionnaire included two items to verify respondents' attentiveness to question wording. We received 567 surveys, 20 of which – accounting for three and a half percent of the received surveys – had either or both of these questions marked incorrectly or left blank. We excluded these responses from the analyses resulting in 547 valid surveys with data from both years. The 2012 sample is representative of the 2009 sample on most measures regarding background characteristics, Internet experiences and skills. The one statistically significant difference is that a disproportionately smaller number of African Americans participated in the second wave (11% in 2009 versus 8% in 2012).

By 2012, about half of the initial sample was no longer enrolled in college either having graduated (48% had done so) or having left school altogether (3.5% had done so). For this reason, we refer to respondents as young adults rather than college students.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Measures: independent variables

The survey asked respondents about their demographic and socioeconomic background. We use year of birth to calculate age. We use parental education as a proxy for socioeconomic status, an established practice in the literature (e.g., Hargittai & Litt, 2011). The survey asked about the level of education of both parents using the following categories: (a) less than high school degree; (b) high school degree; (c) some college; (d) college degree (e.g.: B.A., B.S., B.S.E); (e) advanced graduate (for example: master's, professional, Ph.D., M.D., Ed.D.). Relying on information from these two questions, we created a parental education variable that is assigned the value of the highest education by either parent, for example, if a student has a mother with a high school degree and a father with a college degree, the parental education variable for that student is coded as 'college degree'. We then recoded the parental education variable into three mutually exclusive categories: (1) high school degree or less; (2) some college; and (3) college degree or more.

To measure race and ethnicity, the survey first asked participants if they were Hispanic or of Latino origin. Then they were asked their race using the following categories: (a) White/Anglo/Caucasian/Middle Eastern; (b) Black/African American; (c) Asian; (d) American Indian or Alaskan Native; and (e) Other. Most responses in the 'Other' category indicated Hispanic origin and were recoded accordingly. The final categories are: Hispanic, non-Hispanic African American, non-Hispanic Asian American, non-Hispanic Native American, and non-Hispanic White.

To assess how much discretionary time respondents may have, we asked them about their current employment activity. Answer options included 'not working' and then several options specifying amount of weekly work hours. We recoded the responses into a binary variable indicating any time spent working.

In order to gauge the importance of Internet-related experiences for Wikipedia contributions, we collected data on how long people have been Internet users, autonomy of use, and frequency of use. We calculated years of use using information from survey questions that asked about the stage in a student's academic career when he or she first became an Internet user (i.e. during elementary school, middle school, or a particular year in high school). We capped user years at 10 (a response given by just under 20% of the sample). To calculate autonomy of use, we relied on questions asking about different locations where participants have access to the Internet (as opposed to actually using it regularly at various locations) by having the choice to check off all applicable locations from a list of 11 options. We created a summary variable from these; the final

measure ranges from zero to 11 locations. We logged this measure in the analyses due to the idea that there are diminishing returns to additional locations of access. We derived time spent on the Web weekly (excluding email, chat and voice services) from answers to two questions asking about hours spent on the Web on an average day; one inquiring about weekdays, the other about an average Saturday or Sunday. This measure ranges from 0 to 42 hours and is also logged in the analyses for reasons similar to logging number of access locations.

To measure Web-use skill, the data set includes a previously developed and validated instrument included in the Internet Society Module of the General Social Survey 2000 (National Opinion Research Center, 2000; Wasserman & Richmond-Abbott, 2005) although more elaborate and used by others in the literature (Correa, 2010; Park, 2013). Respondents were asked to rate their level of understanding of 27 Internet-related terms on a five-point scale (Hargittai & Hsieh, 2012). Although a self-report of skill can be prone to gender bias – women tend to rate their skill level lower than men even when controlling for their actual online abilities (Hargittai & Shafer, 2006) – research has shown that self-perceived skill can itself have implications for what people do online (Hargittai & Walejko, 2008) so it is nonetheless a relevant measure. We averaged the 27 items to create the skill index (Cronbach's alpha = 0.94) with a possible range of 0–4 where the lowest possible score was set to zero.

To get a sense of people's confidence with editing Wikipedia, that is, self-efficacy (Bandura, 1977), the questionnaire included an item asking respondents to rate how confident they felt in the following statement on a 1–5 point scale (1 referring to 'strongly disagree' and 5 meaning 'strongly agree'): 'I feel confident changing information on a Wikipedia entry'.

To account for whether students may have contributed to Wikipedia due to a class assignment, the survey asked the following: 'As part of school work, have you ever been assigned to do any of the following?' with 'edit an entry on Wikipedia' and 'create a new entry on Wikipedia' as answer options. We created a binary variable to signal experience with receiving such an assignment (i.e. a 'yes' if the person checked either or both of the above two options).

Measures: dependent variable

To measure contributions to Wikipedia, the instrument asked respondents the following: 'How often, if ever, have you done each of the following on Wikipedia?' The activities of interest here are: (1) edited an entry by fixing a mistake; (2) edited an entry by adding new material; (3) started a new entry; and (4) added an image/graphic to an entry. Answer options included 'never', 'have done it once', 'have done it 2–3 times', and 'have done it 4 or more times'. For the variable measuring Wikipedia contributions, we recoded responses to the above question resulting in a binary measure indicating those who had engaged in any of the four activities ever. The survey also included 'read an entry' among the Wikipedia activities, which we use to establish popularity of the site among respondents.

Analyses

First, we present basic descriptive statistics about the popularity of the individual Wikipedia editing activities listed above, including details about what proportion of respondents engages more actively on the site. Next, we discuss bivariate descriptive comparisons of how likely people are to contribute to Wikipedia depending on their background characteristics, Internet experiences and skills. Then we run logistic regression analyses to examine what explains differences in likelihood of contributing when controlling for other factors, presenting multiple specifications with and without interaction terms. We also create a graphical representation of the relationship between gender and Internet skills in predicting Wikipedia contribution to highlight

Downloaded by [8.28.179.180] at 14:09 06 October 2015

the main findings of the regression analyses. For most of the independent variables (background characteristics, Internet experiences and skills, as well as confidence with Wikipedia editing), we use lagged data (data from 2009) to predict experiences with contributing to Wikipedia by 2012. The two independent variables we do not lag are current employment activity and whether the person had ever been assigned to edit Wikipedia for a class.

As previous research on gender and Internet skills suggests either an additive (linear) or multiplicative (interactive) relationship with our outcome variable (editing Wikipedia), we incorporate an interaction term and corresponding hypothesis tests into our regression model (Braumoeller, 2004; Friedrich, 1982). We did not have any basis upon which to anticipate whether the relationship would involve interaction or not (Berry, DeMeritt, & Esarey, 2010; Berry, Rubin, & Esarey, in press), so we report the marginal effects on the dependent variable using simulation-based estimates of the 95% confidence interval (King, Tomz, & Wittenberg, 2000).[3]

The resulting analysis is similar to statistical tests for 'moderation', although we avoid the use of this concept and related terminology of mediation/moderation. The framework of mediation and moderation presumes (1) the presence of a causal process and (2) that the accompanying statistical techniques can credibly estimate the effects of a causal relationship (Baron & Kenny, 1986; Hayes, 2013). However, following the growing literature on counterfactuals and the potential outcomes framework for causal inference (Holland, 1986; Morgan & Winship, 2007), we contend that such counterfactual outcomes are simply implausible for most analyses involving gender. In the language used by Holland (1986), gender is a stable attribute and therefore cannot provide a suitable basis for credible causal inference using any technique of statistical estimation. Furthermore, the idea that we could test for a causal relationship between gender, Internet skills, and Wikipedia editing is all the more implausible given the nature of our data. Even our detailed, longitudinal survey measures do not provide a credible basis for testing causal claims in the absence of some exogenous intervention or carefully identified instrument (Morgan & Winship, 2007). As a result of these concerns, we adopt a more conservative, careful approach to our analysis and the presentation of our findings. We report the results for multiple model specifications, consider their implications, and use the most compelling specification to build a more intuitive, graphical explanation of the model that also incorporates simulation-based measures of the uncertainty surrounding our estimates (King et al., 2000).

Downloaded by [8.28.179.180] at 14:09 06 October 2015

The sample

Everybody in the sample was either 21 or 22 years old so we exclude this variable from the analyses. Women were somewhat more represented than men (see Table 1 for details). Less than half of the sample was White, one-quarter were Asian or Asian American, more than one-fifth were Hispanic, and a little less than a tenth were African American. Just under one-fifth came from families in which neither parent had more than a high school education, just over a fifth had at least one parent with some college experience, and half had at least one parent with a college degree. Overall, these figures indicate that the sample is quite diverse from a socioeconomic perspective.

In terms of online experiences, on average, these young adults had been using the Internet for almost six years at the time of the first wave of data collection (2009), had numerous locations to access the Web, and spent approximately 20 hours per week online (not counting email, chat, or voice services). The average respondent's confidence in editing Wikipedia is relatively low, falling between the 'disagree' and 'neutral' sentiments regarding the confidence statement, but tending toward the former. About one in eight students had been given an assignment in class at some point either to edit or create a new entry on Wikipedia. While the sample consists of a

Table 1. Sample descriptives.

User attributes	Percent
Gender	
Men	39
Women	61
Race/ethnicity	
African American, non-Hispanic	8
Asian American, non-Hispanic	24
Hispanic	23
Native American, non-Hispanic	1
White, non-Hispanic	44
Parental education	
High school or less	23
Some college	26
College or more	50
Employment activity	
Currently working	76
Currently not working	24

Table 2. Internet experiences (2009 data).

	Mean	Standard deviation
Number of use years	5.71	2.25
Number of Internet access locations	6.48	2.18
Hours spent on the Web weekly	17.16	9.76
Confidence in editing Wikipedia (1–5)	2.17	1.14
School assignment to edit Wikipedia (2012 data)	0.12	0.33
Internet skills (0–4)	2.2	0.82

highly wired group of young adults, we observe considerable variation in Internet skills, ranging from 0.33 to 4 on a 0–4 scale. See Table 2 for details.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Popularity of contributing to Wikipedia

The vast majority (99%) of respondents reported having read an entry on Wikipedia, signaling that lack of contributions cannot be attributed to not knowing about or not visiting the site. Looking at aggregate statistics of how many respondents made edits to existing entries (including the addition of an image) or started a new entry, we find that just over a quarter (28%) of the sample reports such experiences (many having done this just once ever). This figure drops to just a fifth of respondents if we exclude those who had been assigned to make edits to Wikipedia as part of a class assignment.

As Table 3 presents, the most popular engagement is editing an entry by fixing a mistake, which just under a quarter of the sample reported having done at least once. Less than a fifth claimed experiences with adding new material while less than a tenth had either started a new entry or had added an image or a graphic to an entry, respectively. Overall, these figures suggest that contributing to Wikipedia is not a widespread activity. On the whole, 16% of the sample reported having ever engaged in more than one of the four activities and only 12% had engaged in any of the activities more than once (less than 5% had engaged in any of the activities 4 or more times).

Table 3. Level of engagement with Wikipedia editing activities.

	Never	Has done it once	Has done it two to three times	Has done it four or more times
Read an entry	1.10	1.29	5.71	91.90
Edited an entry by fixing a mistake	75.87	12.89	7.18	4.05
Edited an entry by adding new material	82.69	10.50	3.68	3.13
Added an image/graphic to an entry	90.81	5.33	1.84	2.04
Started a new entry	91.18	5.51	1.10	2.21

Wikipedia contributions by user attributes

Table 4 presents what percentage of sample subgroups reported any Wikipedia contributions. Consistent with the literature, women are much less likely to have contributed to Wikipedia than men. Only one in five females (21%) ever made any Wikipedia edits compared to over a third of men (38%). The discrepancies are even larger when we only look at those who did not have a related class assignment: 13% of such women compared to almost 32% of such men

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Table 4. Wikipedia contributions by user attribute, prior Internet experiences and skills.

User attributes	Full sample	Excluding those with a related school assignment
Gender		
Men	38.21***	31.55***
Women	20.90***	12.97***
Race/ethnicity		
African American, non-Hispanic	20.93	15.38
Asian American, non-Hispanic	23.44	20.17
Hispanic	25.62	14.71
White, non-Hispanic	31.65*	23.90*
Parental education		
High school or less	25.20	14.81
Some college	28.37	18.85
College or more	28.57	23.36*
Employment activity		
Currently working	29.09	21.70
Currently not working	23.08	15.65
Internet experiences		
Number of use years LQ	14.29	16.67
Number of use years HQ	29.35	21.41
Autonomy of use LQ	24.04*	17.39
Autonomy of use HQ	33.33*	25.15
Frequency of use LQ	21.21*	11.76*
Frequency of use HQ	33.11*	26.52*
Confidence in editing Wikipedia Lowest Third	24.49***	15.29***
Confidence in editing Wikipedia Highest Third	37.56***	31.18***
Internet skills LQ	14.73***	7.76***
Internet skills HQ	39.86***	32.50***

Note: Asterisks signify statistically significant differences between the subgroups of the category.
*$p < 0.1$.
***$p < 0.001$.

made some edits to the site. The observed differences in any level of engagement also extend to intensity of engagement. Men are much more likely to have engaged more than once in Wikipedia edits than women, a result especially salient for fixing a mistake and adding new material to an entry. Generally speaking, very few participants had started new entries or added images to entries more than once regardless of their gender.

Additional relationships reported in Table 4 suggest that Whites are more likely to edit than others as are those from more educated families. These are the only background characteristics that exhibit statistically significant relationships with Wikipedia contributions. Various Internet-use variables do matter considerably, however. Because these measures are continuous, we created quartiles and present the percent of contributors for the lowest (LQ) and highest (HQ) quartile groups of each measure. Frequent Web users are more likely to contribute as are those who are more confident in their Wikipedia editing abilities, both concerning lagged measures. The biggest contrast between contributors and non-contributors concerns general Internet skills as a considerably larger segment of those in the highest skill quartile have made at least one edit on Wikipedia compared to those in the LQ. Next we look at whether these findings hold when considering people's background characteristics and Internet experiences simultaneously.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Predicting Wikipedia contributions

The results presented in Table 5 of logistic regression analyses show evidence that both gender and skill explain variation in Wikipedia contributions in several model specifications. Remarkably, these two variables – along with the measure of whether or not respondents had been assigned to edit Wikipedia for a class – are the only significant predictors of Wikipedia contributions when holding other factors constant.[4] None of the other variables in the models – including measures of race, ethnicity, socioeconomic status, availability, Internet experience, and confidence in editing Wikipedia – are significant.

Consistent with previous research on who edits Wikipedia, female respondents were less likely than males to have ever contributed to the encyclopedia across most specifications. Adding to this literature is our novel introduction of Internet skills in the model, a variable that is a significant predictor of editing as well, even while controlling for gender. Individuals with high Internet skills were considerably more likely to contribute than those with low Internet skills. We note that the addition of Internet skills attenuates the relationship between gender and Wikipedia contribution (Models 3 and 5).

Models 4 and 6 introduce the multiplicative term capturing the interaction of gender and Internet skills, and both show that the interaction term itself is not a significant predictor of editing Wikipedia. In Models 3 and 4, the Internet skills measure runs from 0 to 4, whereas in Models 5 and 6 we present results with Internet skills centered around its mean value. As expected, these models are identical except for the coefficients for the Intercept and Female. The differences between these specifications illustrate the nature of the relationship between gender and Internet skills and their joint association with Wikipedia editing. The addition of the interaction term in Model 4 attenuates the association between gender and Wikipedia contribution completely. However, centering Internet skills around the mean recovers the significant relationship between gender and the dependent variable.[5] The interaction term for gender and Internet skills is not significant, but this provides inconclusive evidence for the presence or absence of an interactive relationship (Berry et al. in press).

We investigate this relationship between gender and Internet skills further in a plot of marginal effects in Figure 1. All covariates other than gender and Internet skills are held at their sample

Table 5. Logistic regression of contribution.

	1	2	3	4	5	6
Background						
Female	−0.90*** (0.20)	−0.91*** (0.24)	−0.71** (0.26)	0.00 (0.69)	−0.71** (0.26)	−0.66* (0.27)
Hispanic	−0.13 (0.28)	−0.19 (0.34)	−0.20 (0.34)	−0.21 (0.35)	−0.20 (0.34)	−0.21 (0.35)
African American non-Hispanic	−0.37 (0.41)	−0.33 (0.48)	−0.44 (0.48)	−0.40 (0.48)	−0.44 (0.48)	−0.40 (0.48)
Asian American non-Hispanic	−0.35 (0.26)	−0.28 (0.29)	−0.39 (0.30)	−0.37 (0.30)	−0.39 (0.30)	−0.37 (0.30)
Parental Education – Some College	−0.07 (0.25)	−0.08 (0.28)	−0.04 (0.29)	−0.04 (0.29)	−0.04 (0.29)	−0.04 (0.29)
Parental Education – High School or less	−0.18 (0.28)	−0.23 (0.33)	−0.17 (0.34)	−0.17 (0.34)	−0.17 (0.34)	−0.17 (0.34)
Working (2012)	0.31 (0.25)	0.32 (0.29)	0.32 (0.29)	0.31 (0.29)	0.32 (0.29)	0.31 (0.29)
Internet experiences						
Internet use years (2009)		0.06 (0.05)	0.03 (0.06)	0.03 (0.06)	0.03 (0.06)	0.03 (0.06)
Internet autonomy (2009)		0.14 (0.34)	−0.04 (0.34)	−0.03 (0.34)	−0.04 (0.34)	−0.03 (0.34)
Internet use frequency (2009)		0.40 (0.20)	0.31 (0.21)	0.30 (0.21)	0.31 (0.21)	0.30 (0.21)
School assignment to edit Wikipedia (2012)		3.00*** (0.36)	2.97*** (0.36)	2.98*** (0.36)	2.97*** (0.36)	2.98*** (0.36)
Confidence editing Wikipedia (2009)		0.16 (0.10)	0.06 (0.11)	0.04 (0.11)	0.06 (0.11)	0.04 (0.11)
Internet skills (2009)			0.52** (0.17)	0.69** (0.23)	0.52** (0.17)	0.69** (0.23)
Female*Internet skills				−0.35 (0.31)		−0.35 (0.31)
(Intercept)	−0.51 (0.29)	−3.03** (0.96)	−3.16** (0.96)	−3.50*** (1.02)	−2.18* (0.99)	−2.20* (1.00)
AIC	612.20	510.84	503.00	503.75	503.00	503.75
BIC	646.38	566.14	562.53	567.53	562.53	567.53
Log-likelihood	−298.10	−242.42	−237.50	−236.88	−237.50	−236.88
Deviance	596.20	484.84	475.00	473.75	475.00	473.75
Num. obs.	530	520	519	519	519	519

Notes: Model 1 includes background variables only; Model 2 adds Internet experiences variables; Model 3 adds Internet skills; Model 4 adds an interaction of Internet skills and Female; Models 5 and 6 use Internet skills centered at the sample mean; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.
*$p < 0.05$
**$p < 0.01$.
***$p < 0.001$.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

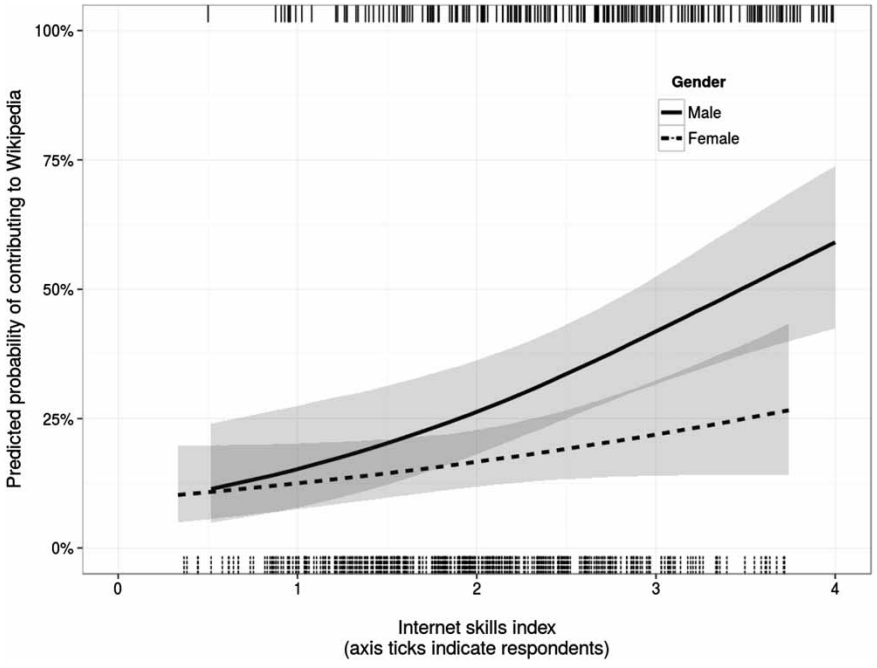


Figure 1. Marginal effects plot: predicted probability of contribution to Wikipedia by gender.

mean values (modal values for categorical predictors) and the coefficients from Model 6 (Table 5) are used to estimate predicted probabilities that a male or female respondent at each observed point on the Internet skills index would have edited Wikipedia.[6] In addition, 'rug plots' (ticks along the top and bottom of the horizontal axes of the figure) show the distribution of male and female respondents respectively along the Internet skills index. The grey 'ribbons' accompanying the two curves represent the 95% confidence interval around these predictions, which we generated by simulating 5000 randomly drawn parameter values with identical variance–covariance matrices to the observed model from a multivariate normal distribution (King et al., 2000). The divergence of the curves further illustrates the nature of the interaction: at low values for Internet skills, the predicted probabilities of Wikipedia contribution for males and females overlap almost perfectly. As the observed level of Internet skills increases, the curves drift apart and the accompanying confidence intervals nearly separate. We discuss these results in greater detail below.

Finally, Figure 2 visualizes the interactive relationship between gender, Internet skills, and Wikipedia editing estimated by Model 6 in greater detail.[7] Here the test statistics for the model coefficient for gender are visualized at discrete points along the distribution of Internet skills and the critical value for the test statistic ($|t| > 1.97$) is indicated by the red horizontal segment. The plot shows that for values of Internet skills roughly below the sample mean (2.21) the model cannot reject the null hypothesis of no effect for gender. In other words, below the sample mean for Internet skills we find no support for the hypothesis of a relationship between gender and Internet skills, but above it we do. These results are suggestive of an interaction between gender and Internet skills, but remain inconclusive for reasons addressed by Berry et al. (in press).

Downloaded by [8.28.179.180] at 14:09 06 October 2015

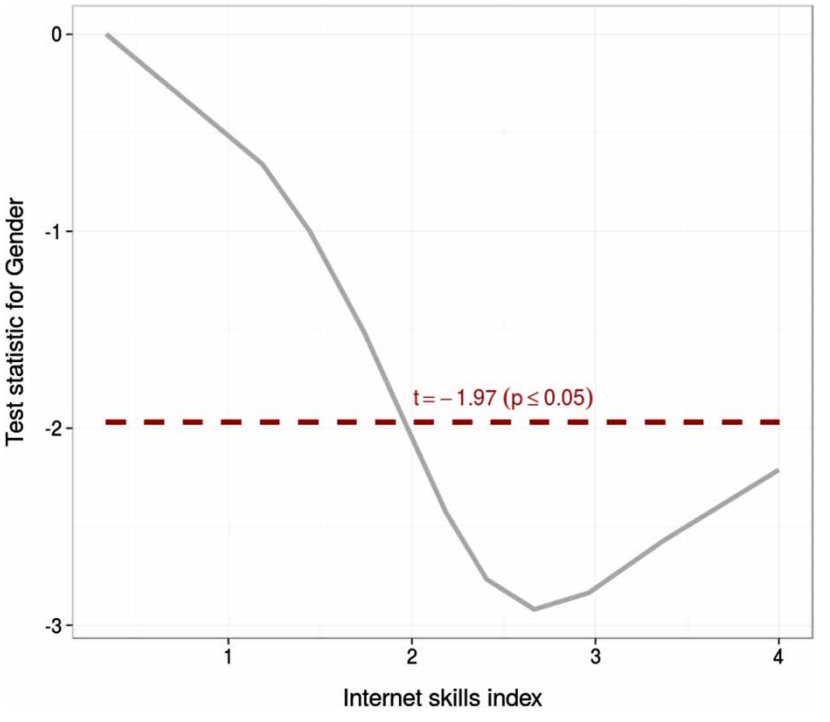


Figure 2. Test statistic for the relationship between gender and Wikipedia editing across the range of observed values for Internet skills.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Discussion and conclusion

Our findings suggest that Internet skills are an extraordinarily robust predictor of contributing to Wikipedia and also help explain a key dimension of the gender gap among Wikipedia editors. Using data that include detailed and reliable measures of demographic attributes as well as socio-economic status, we find that only gender and Internet skills help explain variations in Wikipedia contribution when excluding related school assignments. Most importantly, we find that the variation in Internet skills (measured in 2009) and gender create a consistent pattern in terms of who contributes to Wikipedia years later (by 2012). The models also suggest the presence of a substantively meaningful interaction between gender and Internet skills, despite failing to reject the null hypothesis of no association for the multiplicative term itself. As the marginal-effects plot presented in Figure 1 illustrates, we find that higher levels of Internet skills predict much greater probability of contribution for men than for women. Given that women tend, on average, to have lower Internet skills than men, this indicates a key reason why the gender gap may be so persistent.

The results also suggest that low-skilled men and low-skilled women are equally highly *unlikely* to contribute to Wikipedia. In other words, there is no gender gap in Wikipedia contribution among individuals with low Internet skills. The fact that we see these results in a panel data set where we surveyed the same individuals years apart suggests that differentiated Internet know-how has long-term consequences. Women in the study have lower skills than comparable men, helping to explain why they are so much less likely ever to have edited Wikipedia than their male counterparts.

Why women, on average, report lower level understanding of Internet-related terms remains a puzzle. Although studies with detailed data about actual skills based on performance tests suggest

no gender differences in the observed skills (van Deursen & van Dijk, 2010; Hargittai & Shafer, 2006), research that looks at self-rated know-how consistently finds gender variation (Correa, 2010; Wasserman & Richmond-Abbott, 2005) with real consequences for online behavior (e.g. differentiated rates of content sharing online [Hargittai & Walejko, 2008]). That is, even if it is just a perceived variation in know-how, it has real consequences for online behavior, as demonstrated in the present study as well. One possible source of this discrepancy could be confidence in using the Internet, but as noted above, we control for confidence with the specific activity of editing Wikipedia and that does not seem to make a difference for actual editing experiences. Another possible source of this variation concerns the level of encouragement, or lack thereof, that boys versus girls receive for tinkering with technology while growing up. If parents place more restrictions on girls' technology uses out of safety concerns, that may scare girls away from open exploration with potential long-term consequences for understanding the ins-and-outs of the medium. Future research should tackle such potential negative effects of restrictions on technology use during childhood.

This study offers novel contributions and also opens up new arenas of inquiry for research on social media, peer production, online behavior more broadly and the general question of cultural production in the digital age. Given the type of data we have, we are unable to differentiate among various types of 'lurkers' for example, people who visit Wikipedia but have not contributed in the ways we measure participation (i.e. specific content additions to entries). Scholars have debated how being a member of the audience is itself a type of participation (Crawford, 2009; van Dijck, 2009; Jenkins, Ford, & Green, 2013), an area of inquiry that offers interesting questions for future research that will require different data to address.

Having measured Wikipedia editing experiences through class assignments, we are able to show that educational interventions can matter to increasing participation. Our project did not focus on such interventions, but they do exist (e.g. FemTechNet) and are ripe areas for research to see whether they can help alleviate gender and skill differences in Wikipedia contributions.

We extend earlier work on the Wikipedia gender gap by demonstrating that variation in Internet skills helps explain a critical aspect of the variation in the contribution rates among men and women. Our novel finding that the gender gap in Wikipedia contributions does not exist for individuals with low Internet skills complicates prior explanations of these phenomena. While studies have demonstrated differential participation practices among Wikipedia editors (Antin et al., 2011; Kriplean, Beschastnick, & McDonald, 2008; McDonald, Javanmardi, & Zachry, 2011; Welser et al., 2011), we are not aware of work offering insights into specific factors that explain how people select into contributing to Wikipedia in the first place, beyond gender. We are able to make these unique contributions, because we have comparable data on both contributors and non-contributors thereby avoiding a common limitation of prior work: sampling on the dependent variable, that is, analyzing data only about contributors.

The fact that Internet skills are such a robust predictor – at least as robust as gender – of who participates in Wikipedia editing underscores the importance of incorporating skill as both an object of future research studies and as a central design consideration for social media and peer production systems. For example, much work has focused on the motivations driving participation in online communities and on empirically testing design interventions that effectively incentivize participants to contribute (Kraut & Resnick, 2011). These are important dimensions of building successful, dynamic online communities, but interventions must also address differentiated levels of Internet skills.

One limitation of this study, the fact that it examines a non-random and statistically non-representative sample of young adults, restricts our ability to generalize on the basis of these findings. Nevertheless, panel data with detailed measures of Internet skills and use remain extremely rare and are non-existent for nationally representative samples. Thus, data even on non-representative

Downloaded by [8.28.179.180] at 14:09 06 October 2015

samples can offer unique insights into the social dynamics and determinants of participation in online communities. In addition, the fact that this data set incorporates many non-contributors to Wikipedia allows us to explore the factors that predict contribution in the first place. Moreover, given that respondents represent a highly wired panel of young adults (i.e. not novices when it comes to digital media) the findings are likely conservative when it comes to the importance of skill among a more diverse group.

A great deal of work on these topics remains to be done. How do gender and skill differences in contribution manifest themselves among people heterogeneous in age and education? Why is it that women with high levels of Internet skills are so much less likely to contribute to Wikipedia than comparably skilled men? While research has shown that one of the reasons women use the Internet differently from men concerns their parental responsibilities due to less available time (Kennedy, Wellman, & Klement, 2003), given that only 3% of the sample reported having a child, this factor cannot explain the gender differences for this group. Self-selection into a project like Wikipedia also plays a critical role, but specific reasons why women might find encyclopedia writing (or the practice of doing so in Wikipedia) disproportionately unappealing requires further research.

Additionally, future research should seek to explain the pathways through which Internet skills relate to participation in Wikipedia and comparable online communities. More work is needed in understanding what particular Web-use skills – not necessarily specific to Wikipedia – are especially helpful for encouraging contributions and what types of interventions can improve them, especially among women. In addition, the relationship between skills, gender, and other predictors of participation should be examined in greater depth through multiple methods. Our findings partially contradict previous studies of the Wikipedia gender gap for participation and begin to unpack the mechanisms by which Internet skills (along with gender and other factors) contribute to differentiated participation in cultural production online.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Acknowledgements

The authors greatly appreciate the generous support of the John D. and Catherine T. MacArthur Foundation and the Robert and Kaye Hiatt Fund that made this project possible. The authors thank Benjamin Mako Hill and members of the audience at the Berkman Center for Internet & Society at Harvard University for their helpful input as well as the undergraduate and graduate research assistants in the Web Use Project group during the 2009–2012 academic years for their assistance with data collection and data entry. The authors are grateful for support from the Northwestern University Communication Studies Department as well as Nokia Research Center.

Notes

1. For reasons discussed in detail below, we do not test for the presence of a conditional causal process, such as mediation or moderation.
2. After two attempts at reaching respondents in postal mail, we sent an email to those who had not yet responded to request a verified postal address for mailing the survey in hard copy. No one filled out the survey online.
3. This approach follows the procedures recommended by Berry et al. (in press) for circumstances when theoretical support for interaction terms is ambiguous. They demonstrate that normal statistical tests for multiplicative interactions do not possess sufficient statistical power to detect the alternative hypotheses (or reject the null) with conventional levels of precision.
4. All results reported here are robust to the exclusion of the subset of respondents who had been assigned to edit Wikipedia for a class. In additional robustness checks, we included a control for whether or not respondents had ever contributed to Wikipedia by 2009, and our findings were unchanged substantively.

5. This is due to the fact that, in the interaction models, the coefficient for gender only captures the relationship with the dependent variable when the other component of the interaction term (Internet skills) is equal to zero.
6. The figure is substantively the same if we use the coefficients from Model 4.
7. Again, the figure is unchanged if we use Model 4.

Notes on contributors

Eszter Hargittai is Delaney Family Professor in the Communication Studies Department and Faculty Associate of the Institute for Policy Research at Northwestern University where she heads the Web Use Project. Her research looks at how people may benefit from their digital media uses with a particular focus on how differences in people's Web-use skills influence what they do online. [email: eszter@northwestern.edu]

Aaron Shaw is an Assistant Professor of Communication Studies at Northwestern University and a Faculty Associate of the Berkman Center for Internet and Society at Harvard University. He studies collective action, collaboration, and mobilization in online communities and peer production projects. [email: aaronshaw@northwestern.edu]

Downloaded by [8.28.179.180] at 14:09 06 October 2015

References

Alderson, A. S., Junisbai, A., & Heacock, A. (2007). Social status and cultural consumption in the United States. *Poetics*, *35*, 191–212.

Antin, J., Yee, R., Cheshire, C., & Nov, O. (2011). Gender differences in Wikipedia editing. *ACM WikiSym*, Mountain View, California, 11–14.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215.

Baron, R. M., & Kenny, D. A. D. (1986). The moderator-mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Benkler, Y. (2002). Coase's penguin, or, Linux and the nature of the firm. *The Yale Law Journal*, *112*, 369–446.

Benkler, Y. (2006). *The wealth of networks*. New Haven, CT: Yale University Press.

Berry, W. D., DeMeritt, J. H. R., & Esarey, J. (2010). Testing for interaction in binary logit and probit models: Is a product term essential? *American Journal of Political Science*, *54*, 248–266.

Berry, W. D., Rubin, J. H., & Esarey, J. (in press). Bias and overconfidence in parametric models of interactive processes. *American Journal of Political Science*.

Blank, G. (2013). Who creates content? Stratification and content creation on the Internet. *Information, Communication & Society*, *16*, 590–612. doi:10.1080/1369118x.2013.777758

Braumoeller, B. F. (2004). Hypothesis testing and multiplicative interaction terms. *International Organization*, *58*, 807–820.

Chen, J., Ren, Y., & Riedl, J. (2010). *The effects of diversity on group productivity and member withdrawal in online volunteer groups*. Proceedings of the 28th international conference on Human factors in computing systems, ACM, New York, NY, 821–830.

Cohen, N. (2011). Define gender gap? Look up Wikipedia's contributor list. *The New York Times*, New York.

Collier, B., & Bear, J. (2012). *Conflict, criticism, or confidence: An empirical examination of the gender gap in Wikipedia contributions*. Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, ACM, 383–392.

ComScore. (2013). *comScore Media Metrix Ranks Top 50 US Web Properties for May 2013*. Reston, VA: ComScore.

Correa, T. (2010). The participation divide among 'Online Experts': Experience, skills and psychological factors as predictors of college students' web content creation. *Journal of Computer-Mediated Communication*, *16*, 71–92. doi:10.1111/j.1083–6101.2010.01532.x

Crawford, K. (2009). Following you: Disciplines of listening in social media. *Continuum: Journal of Media & Cultural Studies*, *23*(4), 525–535.

van Deursen, A., & van Dijk, J. (2010). Measuring internet skills. *International Journal of Human-Computer Interaction*, *26*(10), 891–916.

van Dijck, J. (2009). Users like you? Theorizing agency in user-generated content. *Media Culture and Society*, *31*, 41–58.

van Dijk, J. A. G. M. (2005). *The deepening divide*. Thousand Oaks, CA: Sage.

DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). Digital inequality: From unequal access to differentiated use. In K. Neckerman (Ed.), *Social inequality* (pp. 355–400). New York, NY: Russell Sage Foundation.
DiMaggio, P., & Ostrower, F. (1990). Participation in the arts by Black and White Americans. *Social Forces*, *68*, 753–778.
DiMaggio, P., & Useem, M. (1978). Social class and arts consumption: The origins and consequences of class differences in exposure to the arts in America. *Theory and Society*, *5*, 141–161.
Dutton, W. H., & Blank, G. (2011). *Next generation users: The Internet in Britain*. Oxford Internet Survey 2011. Oxford Internet Institute, University of Oxford. Retrieved from http://dx.doi.org/10.2139/ssrn.1960655
Eastin, M. S., & LaRose, R. (2000). Internet self-efficacy and the psychology of the digital divide. *Journal of Computer Mediated Communication*, *6*. doi:10.1111/j.1083–6101.2000.tb00110.x
Eckert, S., & Steiner, L. (2013). (Re)triggering backlash: Responses to news about Wikipedia's gender gap. *Journal of Communication Inquiry*, *37*, 284–303. doi:10.1177/0196859913505618
Filipacchi, A. (2013). Wikipedia's sexism. *The New York Times*, p. SR3.
Friedrich, R. J. (1982). In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science*, *26*, 797–833.
Gleick, J. (2013, April 29). Wikipedia's women problem. *The New York Review of Books*. Retrieved from http://www.nybooks.com/blogs/nyrblog/2013/apr/29/wikipedia-women-problem/
Glott, R., Ghosh, R., & Schmidt, P. (2010). *Wikipedia survey*. Maastricht: UNU-MERIT.
Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2012). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, *57*, 664–688. doi:10.1177/0002764212469365
Halfaker, A., Kittur, A., & Riedl, J. (2011). *Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work*. Paper presented at the ACM WikiSym Mountain View, California.
Hargittai, E. (2000). Open portals or closed gates? Channeling content on the world wide web. *Poetics*, *27*(4), 233–254.
Hargittai, E., & Hinnant, A. (2008). Digital inequality: Differences in young adults' use of the Internet. *Communication Research*, *35*, 602–621.
Hargittai, E., & Hsieh, Y. P. (2012). Succinct survey measures of Web-use skills. *Social Science Computer Review*, *30*(1), 95–107.
Hargittai, E., & Hsieh, Y. P. (2013). Digital inequality. In W. H. Dutton (Ed.), *Oxford handbook for Internet research* (pp. 129–150). Oxford: Oxford University Press.
Hargittai, E., & Litt, E. (2011). The tweet smell of celebrity success: Explaining variation in Twitter adoption among a diverse group of young adults. *New Media & Society*, *13*(5), 824–842. doi:10.1177/1461444811405805
Hargittai, E., & Shafer, S. (2006). Differences in actual and perceived online skills: The role of gender. *Social Science Quarterly 87*(2), 432–448.
Hargittai, E., & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication and Society*, *11*, 239–256.
Hassani, S. N. (2006). Locating digital divides at home, work, and everywhere else. *Poetics*, *34*, 250–272.
Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: The Guilford Press.
Hill, B. M., & Shaw, A. (2013). The Wikipedia gender gap revisited: Characterizing survey response bias in peer production communities. *PLOS ONE*, *8*(6), e65782. doi:10.1371/journal.pone.0065782
Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960.
Howard, P. N., Rainie, L., & Jones, S. (2001). Days and nights on the Internet: The impact of a diffusing technology. *American Behavioral Scientist*, *45*, 383–404.
Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York: New York University Press.
Jenkins, H., Ford, S., & Greene, J. (2013). *Spreadable media: Creating value and meaning in a networked culture*. New York: New York University Press.
Kennedy, T., Wellman, B., & Klement, D. (2003). Gendering the digital divide. *IT & Society*, 1.
King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, *44*, 341–355.
Kraut, R. E., & Resnick, P. (2011). *Building successful online communities: Evidence-based social design*. Cambridge, MA: MIT Press.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Kriplean, T., Beschastnikh, I., & McDonald, D. W. (2008). *Articulations of Wikiwork: Uncovering valued work in Wikipedia through barnstars*. ACM CSCW '08, 47–56.

Lam, S. K., Uduwage, A., Dong, Z., Sen, S., Musicant, D. R., Terveen, L., & Riedl, J. (2011). WP: Clubhouse?: An Exploration of Wikipedia's Gender Imbalance. *ACM WikiSym '11*, pp. 1–10.

McDonald, D. W., Javanmardi, S., & Zachry, M. (2011). *Finding patterns in behavioral observations by automatically labeling forms of wikiwork in Barnstars*. ACM WikiSym '11, 15–24.

Morgan, J. T., Bouterse, S., Walls, H., & Stierch, S. (2013). *Tea and sympathy: Crafting positive new user experiences on Wikipedia*. ACM Computer Supported Cooperative Work, San Antonio, Texas, 839–848.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. New York, NY: Cambridge University Press.

National Opinion Research Center. (2000). *General social survey information society module*. Chicago, IL: NORC.

Palfrey, J., & Gasser, U. (2008). *Born digital: Understanding the first generation of digital natives*. New York, NY: Basic Books.

Park, Y. J. (2013). Digital literacy and privacy behavior online. *Communication Research*, *40*, 215–236. doi:10.1177/0093650211418338

Perrin, A., & Vaisey, S. (2008). Parallel public spheres: Distance and discourse in letters to the editor. *American Journal of Sociology*, *114*, 781–810. doi:10.1086/590647

Schradie, J. (2011). The digital production gap: The digital divide and Web 2.0 collide. *Poetics*, *39*, 145–168.

US News & World Report. (2009). America's Best College 2010. Washington, DC.

Wasserman, I. M., & Richmond-Abbott, M. (2005). Gender and the Internet: Causes of variation in access, level, and scope of use. *Social Science Quarterly*, *86*, 252–270.

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., & Smith, M. (2011). *Finding social roles in Wikipedia*. iConference Fort Worth, Texas, 122–129.

Zhao, L., Lu, Y., Huang, W., & Wang, Q. (2010). Internet inequality: The relationship between high school students' Internet use in different locations and their Internet self-efficacy. *Computers & Education*, *55*, 1405–1423. doi:10.1016/j.compedu.2010.05.010

Zickuhr, K., & Smith, A. (2012). *Digital differences*, Pew Internet & American Life Project, Washington, DC.

Zillien, N., & Hargittai, E. (2009). Digital distinction: Status-specific Internet uses. *Social Science Quarterly*, *90*(2), 274–291.

Downloaded by [8.28.179.180] at 14:09 06 October 2015

Page Protection: Another Missing Dimension of Wikipedia Research

Benjamin Mako Hill
University of Washington
Department of Communication
makohill@uw.edu

Aaron Shaw
Northwestern University
Department of Communication Studies
aaronshaw@northwestern.edu

ABSTRACT

Page protection is a feature of wiki software that allows administrators to restrict contributions to particular pages. For example, pages are frequently protected so that they can only be edited by administrators. Page protection affects tens of thousands of pages in English Wikipedia and renders many of Wikipedia's most visible pages uneditable by the vast majority of visitors. That said, page protection has attracted very little attention and is rarely taken into account by researchers. This note describes page protection and illustrates why it plays an important role in shaping user behavior on wikis. We also present a new longitudinal dataset of page protection events for English Wikipedia, the software used to produce it, and results from tests that support both the validity of the dataset and the impact of page protection on patterns of editing.

Categories and Subject Descriptors

H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work—*web-based interaction*

General Terms

Wikipedia, Peer Production, Page Protection, Wikis

1. INTRODUCTION

Page protection is a feature of wiki software that allows administrators to restrict contributions to particular pages. For example, a page can be protected so that only administrators can edit it. Protected pages can be distinguished from normal pages by the replacement of the "Edit" button with a "View Source" button and, in many cases, by a small lock icon that appears near the top-right corner of the page (see Figure 1). Protection might involve "full protection" where a page can only be edited by administrators (i.e., "sysops"), or "semi-protection" where a page can only be edited by accounts with a history of good edits (i.e., "autoconfirmed" users). Additionally, protection can prevent specific types of contributions such as editing, moving, creating, or uploading.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/authors.
OpenSym '15, August 19-21, 2015, San Francisco, CA, USA
ACM 978-1-4503-3666-6/15/08.
http://dx.doi.org/10.1145/2788993.2789846

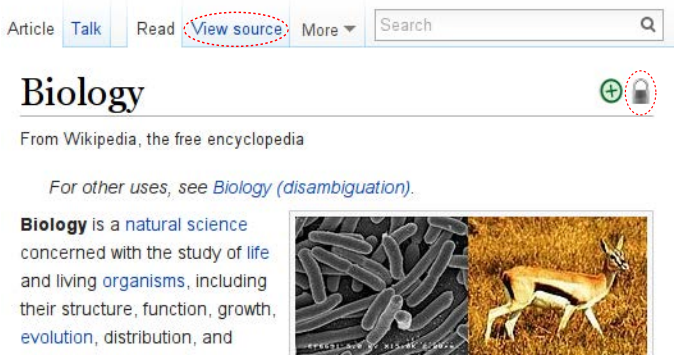


Figure 1: Example of the English Wikipedia article on Biology. Circled in red are the "View Source" button (instead of "Edit") and the small lock icon which each signal that the page is protected.

Page protection profoundly shapes activity on Wikipedia. For example, page protection is an important tool used to manage access and participation when vandalism or interpersonal conflict threatens to undermine content quality. In this way, page protection is a key aspect of the encyclopedia's "hidden order" [11] and contributes to wikis' usability and user experience. While it affects only a small portion of pages in English Wikipedia, many of the most highly viewed pages are protected. For example, the "Main Page" in English Wikipedia has been protected since February 2006. Articles are protected when they are featured on or promoted from the Main Page. Millions of viewers cannot edit protected pages because they never see an edit button.

Page protection has attracted scholarly attention indirectly. It is often invoked in the context of vandalism and anti-vandalism work, as well as related modes of behind-the-scenes, organizational "wikiwork" [5, 7]. Nevertheless, we are not familiar with any quantitative research on Wikipedia that takes page protection into account.

This note makes several contributions. First, we introduce a longitudinal dataset of nearly 300,000 page protection "spells" (time periods during which specific articles were protected) in English Wikipedia between September 2008 and January 2015. We also describe the software used to create the dataset. Second, we use this dataset to characterize page protection activity in English Wikipedia. Finally, we pro-

vide results from validation tests of this dataset and show that accounting for page protection reveals dynamics that researchers should address. We conclude with guidance for how future research might incorporate page protection.

2. DESCRIBING PAGE PROTECTION

Wikipedia administrators protect pages for many reasons. Pages about controversial topics are protected to prevent "edit wars" between contributors with divergent views. Prior work has shown that viewership is positively associated with vandalism and low quality contributions by new users [3]. As a result, popular articles are often protected. For example, consistently high traffic pages like *Biology* may be protected for extended periods. Articles related to current events may be protected while they are in the news (e.g., the article on the February 2014 sports event *Super Bowl XLVIII* was protected during a seven week window in early 2014). Articles "featured" on the front page of Wikipedia are protected as a matter of course. Additionally, protection is used to ensure the stability of pages the community has determined should provide consistent content, like policy pages. Protection is also used by the Wikipedia community and the Wikimedia Foundation (WMF) to enforce policies and the law.

Protection is highly configurable. On English Wikipedia, protection comes in at least ten varieties, including "full protection" and "semi-protection."[1] Protection status can be configured to expire at a predefined time or to last indefinitely. MediaWiki, the software that runs Wikipedia and many other wikis, enforces page protection by preventing actions prohibited by a page's protection status. Protection status can be modified and is subject to discussion and controversy. When the protection status of an article changes, the change is normally recorded by MediaWiki in a log.[2]

Page protection is an important example of how less visible aspects of wikis and social computing platforms shape users' participation and experience. Previous research has analyzed similarly unobtrusive and "hidden" elements of Wikipedia, including aspects of its dispute resolution procedures and coordination work [11, 8]. Studies of peripheral forms of participation have shown that the work of producing the encyclopedia consists of much more than editing the content of articles [1, 2]. Specific types of "wikiwork" have also attracted scholarly attention [7], including recent analysis of other dimensions of the encyclopedia hitherto ignored in most Wikipedia research [4]. None of these studies have examined or accounted for page protection. Page protection is both another form of hidden wikiwork as well as a critical part of the encyclopedia's socio-technical infrastructure.

3. PAGE PROTECTION DATASET

Adopting a strategy similar to one used in our previous work on redirects [4], we have created and published a dataset of *page protection spells* that represent the periods when pages were protected. Table 1 contains four example spells from the article *Biology*. Each spell includes two types of metadata: (1) *type* describing the actions restricted ("edit," "move," or "create"); and (2) *level* describing the class of users to whom that action is restricted ("sysop" means administrators; "autoconfirmed" means established editors; and "templateeditor" indicates a third class of users).[3]

Building a dataset of page protections is challenging because protection information is represented differently and incompletely across different sources of Wikipedia data. One source of page protection data is a "log" of events available through the MediaWiki API[4] and published as an XML database dump.[5] The log records page protection actions alongside other events including page moves, deletions, and editor blocks. In addition to log data, the WMF publishes SQL "page info" database dumps that provide a snapshot of page protection status at the time that the logs and dumps are created. We use both the complete XML log and SQL data to create and validate our dataset.

Unfortunately, the standard for recording page protection data in the log has shifted over time. As a result, log records for page protection events have been recorded incompletely and inconsistently. The most complete and reliable data on page protection events began in late September 2008, when MediaWiki introduced the format for describing page protection log events used today. The spells dataset described in this note is limited to the period between late September 2008, when the current standard was introduced, and the point of data collection in January 2015.

To construct our spells dataset for English Wikipedia, we parsed both the XML log data and SQL page information dumps. Creating spell data required that we track all available records of page protection and unprotection events, page deletions that end protection spells, and events when protected pages were moved. Protection settings often "follow" a page when it is moved. In our dataset, move events are coded as the end of one spell at the source of the move and the start of a new spell at the destination. We have released our full source code to document and reproduce our process.[6] Our code consists of more than 300 lines of Python used to parse XML and SQL dumps and more than 650 lines of R to create spells.

Our dataset remains necessarily limited, incomplete, and subject to missing and incorrect input data. In the case of 23,272 spells affecting 12,833 pages, we use SQL dumps to infer protection periods that are "left censored," (i.e., they began before our data collection window [10]). In addition, because protection expirations and page deletions are not noted as protection events in the log, our dataset is missing spells for pages protected before our data collection window if protection expired or if these previously protected pages were deleted within the window. If the protection status of a page protected prior to the window is changed, we report that the earlier protection status existed, but the level and type of prior status is missing. Our dataset does not include "cascading" protection that affects pages that are linked from special protected pages.[7]

[1] https://en.wikipedia.org/wiki/Wikipedia:Protection_policy

[2] https://en.wikipedia.org/wiki/Special:Log

[3] https://en.wikipedia.org/wiki/Wikipedia:User_access_levels

[4] http://www.mediawiki.org/wiki/API:Logevents

[5] http://dumps.wikimedia.org

[6] http://communitydata.cc/wiki-protection

[7] Cascading protection is rare and was applied to only 97 of

	Page Title	Type	Level	Start	End
1	Biology	edit	autoconfirmed	2008-09-29 15:14:44	2008-10-29 15:14:00
2	Biology	edit	autoconfirmed	2008-12-04 03:44:45	2008-12-25 03:44:00
3	Biology	edit	autoconfirmed	2010-03-01 16:20:48	<NA>
4	Biology	move	sysop	2010-03-01 16:20:48	<NA>

Table 1: Example protection spells for one article in our dataset. "<NA>" indicates that the data is censored because the spell was ongoing at the point of data collection.

In other cases, we must make decisions when our two input sources contain contradictory data. We include "bookmarks" in our code (e.g., "BK-X") to help readers to find and inspect these choices. For example, we exclude spells from 8 pages where log and SQL data suggest that a page was protected at contradictory levels (BK-A). We also exclude 4,148 spells whose final protection status in the SQL dumps is directly contradicted by our parsed log data (BK-B). We include 685 additional spells that final state data indicates have ended, but for which the logs do not indicate an end-point (BK-C).

While further development may improve data quality, hand-checking random subsets of contradictory spells has revealed some cases when the log and SQL dumps are simply ambiguous, missing, or incorrect in ways that additional analysis will not be able to fix. For example, at the time of writing, the log clearly records that the last event the disambiguation page "Moist" experienced was an indefinite semi-protection in March 2010,[8] while the page was verifiably editable by unregistered users. We include an "open" spell even though other evidence suggests it ended at some point between 2011 and 2015. Our source code can be modified to handle these cases differently.

Consistent with the WMF Open Access policy, we have published this page protection dataset freely for other researchers under the same license used for all Wikipedia content. Because WMF continues to publish new SQL snapshots of page info – and because other wikis using MediaWiki could be used to generate similar datasets – we have also released all of the software used to create the page protection dataset under the GNU GPLv3.

4. PAGE PROTECTION IN WIKIPEDIA

Our dataset contains 355,532 page protection spells applied to 125,951 unique pages or 0.36% of all pages in English Wikipedia. This counts all of the different "namespaces" in Wikipedia, including pages used for administrative work, discussions, file hosting, and more. Page protection occurs unevenly across these namespaces. For example, the article or "main" namespace accounts for less than 14% of the total pages in the encyclopedia, but 66% (235,458) of all spells.[9] Within the article namespace, 0.67% of pages have been protected during our data collection window.

The dataset suggests that the total number of pages protected at any given point has remained relatively stable increasing at about 1,000 new protected pages per year from 45,000 in 2008. Despite the relatively stable number of protected pages, there is a great deal of dynamism and variation within the population of spells. Among the page protections that start and end within the period covered by our dataset, protection spells are as short as a few seconds and as long as 6 years with a median of 14 days ($\bar{x}$: 123, σ: 292). At the point of data collection, 27% of the spells in the dataset were ongoing. Page protection shifts frequently and is distributed unevenly, with 96% of pages that have ever been protected undergoing multiple protection spells. Among pages that have experienced at least one protection, the median number of protections is 2 ($\bar{x}$: 2.8, σ: 3.5) with the talk page of one administrator, "NawlinWiki" – a target of persistent vandalism – experiencing a total of 797 protection spells during the window.

	Protected	Min	Median	Mean	Max
Unreg.	Yes	0	0	0.6	1,851
only	No	0	0.2	13	151,200
All	Yes	0	0.48	410	483,839
editors	No	0	1.1	110	302,400

Table 2: Edits per week by unregistered editors and all editors for pages in our dataset while protected and unprotected.

5. IMPACT OF PAGE PROTECTION

Page protection has dramatic effects on the dynamics of editing. Comparing editing activity by unregistered editors within pages that are protected and unprotected illustrates the impact of page protection and also provides a validity check for our dataset. The top section of Table 2 compares edits per week from unregistered editors for pages included in our dataset during periods when they are protected and unprotected. Unregistered editors make edits during only 6% of protection spells. By contrast, unregistered editors edit in 65% of the unprotected spells for the same pages. Edits from unregistered editors while pages are protected reflect the irregularities described above. The bottom section of Table 2 compares total edits per week for pages included in our dataset. Both a t-test for a difference of means ($t = 5.53$, p: 0) and a non-parametric Kolmogorov-Smirnov test for whether the two groups are drawn from the same underlying distribution (D: 0.15, p: 0) indicate that these differences are statistically significant.

Research analyzing the patterns and structure of editing should take page protection into account because protection systematically prevents some editors from contributing. For example, Keegan et al. published a series of influential papers comparing the structure of editing on breaking news articles about airplane crashes with articles written about historical accidents (e.g., [6]). Although the large majority of the contributions to the 249 articles studied by Kee-

106,103 ongoing spells at the point of data collection.
[8]https://en.wikipedia.org/wiki/Special:Log?page=Moist
[9]https://en.wikipedia.org/w/index.php?title=Wikipedia:Statistics&oldid=663786511

	M	$\bar{x}$	Range
Mean Views (All)	5.23	5.24	[3.89,6.06]
Mean Views (Protected)	207	195	[71.8,294.07]
% Views (Protected)	14.3	13.6	[6.66,21.46]
% Articles (Protected)	0.36	0.37	[0.33,0.57]
% in Top 1000 (Protected)	34.1	34.2	[27.3,42]

Table 3: Summary statistics for measures of views during 168 one-hour periods in December 2013. Columns are included for the median (M), mean ($\bar{x}$), and range.

gan et al. were made before the window of our dataset, we found protection spells for 12 articles within our window. Although 93 articles were classified as breaking news, 9 of the 12 protected articles were breaking news events, more than we would expect by chance (χ^2: 6.04, df: 1, p: 0.01).

Accounting for page protection should improve the precision of Keegan and colleagues' analysis of editing dynamics around breaking news. The fact that breaking news articles are more likely to be protected shapes the relationship between breaking news status and key variables like the distribution of editors across articles, the "tempo" of activity, and back-and-forth patterns of editing. Likewise, studies of the relationship between editing activity and article quality as measured by featured article status (e.g., [9]) should address the impact of page protection since featured articles are protected when they are posted on the Wikipedia front-page.

Protected pages are also disproportionately viewed by Wikipedia readers. Table 3 shows data on viewership during during the first week of December 2013 (168 hour-long periods) for pages that were protected for at least part of each hour and for all viewed pages. On average, protected articles were only 0.37% of the articles viewed, but received 14% of views. The median protected article received 26 views ($\bar{x}$: 195, σ: 6,363) while the median unprotected article received 1 view ($\bar{x}$: 5, σ: 391). Among the top 1,000 most viewed articles during each of these hours, an average of 342 were protected.

Research that exploits the role of page views in shaping contributions is at risk of systematically underestimating the strength of this relationship unless it accounts for page protection. For example, Gorbatai [3] shows how viewership drives edits by newcomers, which in turn drives engagement by established editors, leading to improved article quality. Because highly-viewed articles are more likely to be protected, the association between edits from newcomers and edits from established editors would likely be even stronger if page protection were considered. These findings also illustrate how page protection may undermine a key mechanism driving the improvement of highly-viewed articles.

6. CONCLUSIONS

Page protection shapes the structure and experience of collaborative work in wikis. We have demonstrated how page protection impacts the relationship between article views and edits. Studies of wiki work and Wikipedia governance might use this dataset to investigate how Wikipedians manage the boundaries of collaboration. Granular, precise data about page protection also facilitates studies of related phenomena. For example, sudden changes in the number of edits to a page brought about by a protection event create a discontinuity that could be used to estimate the impact of blocking unregistered contributors.

Future work can determine how best to incorporate a detailed understanding of page protection into existing knowledge about Wikipedia and other wikis. Page protection also merits more systematic analysis in its own right. Such analysis lies beyond the scope of this paper, but we hope that others will use our dataset and software toward this end. As researchers begin to consider protection, we invite contributions to our code that address the limitations we have described and other issues of which we are unaware.

7. REFERENCES

[1] J. Antin and C. Cheshire. Readers are not free-riders. In *Proc. CSCW '10*, pages 127–130. ACM Press, 2010.

[2] S. L. Bryant, A. Forte, and A. Bruckman. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proc. SIGGROUP '05*, pages 1–10. ACM Press, 2005.

[3] A. Gorbatai. Aligning Collective Production with Demand : Evidence from Wikipedia. Available at SSRN 1949327, 2011.

[4] B. M. Hill and A. Shaw. Consider the redirect: A missing dimension of Wikipedia research. In *Proc. OpenSym '14*, pages 28:1–28:4. ACM Press, 2014.

[5] D. Jemielniak. *Common Knowledge?: An Ethnography of Wikipedia*. Stanford University Press, Stanford, California, May 2014.

[6] B. Keegan, D. Gergle, and N. Contractor. Hot Off the Wiki Structures and Dynamics of Wikipedia's Coverage of Breaking News Events. *American Behavioral Scientist*, 57(5):595–622, May 2013.

[7] D. W. McDonald, S. Javanmardi, and M. Zachry. Finding patterns in behavioral observations by automatically labeling forms of wikiwork in Barnstars. In *Proc. WikiSym '11*, pages 15–24. ACM Press, 2011.

[8] R. Priedhorsky, J. Chen, S. T. K. Lam, K. Panciera, L. Terveen, and J. Riedl. Creating, destroying, and restoring value in Wikipedia. In *Proc. SIGGROUP '07*, pages 259–268. ACM Press, 2007.

[9] S. Ransbotham and G. C. Kane. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quarterly*, 35(3):613, 2011.

[10] J. D. Singer and J. B. Willett. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press, New York, 2003.

[11] F. B. Viegas, M. Wattenberg, M. M. McKeon, and D. Schuler. The hidden order of Wikipedia. In *Online Communities and Social Computing*, HCII, pages 445–454. Springer-Verlag, Berlin, 2007.

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

7

Hired Hands and Dubious Guesses: Adventures in Crowdsourced Data Collection

Aaron Shaw

"What if we used Mechanical Turk to do it?" It was one of those hare-brained, off-hand suggestions that bubbled up frequently enough in our brainstorming sessions that I no longer remember exactly who said it. There were three of us around the conference table that day, our diagrams and notes on the floor-to-ceiling dry-erase boards, our laptops, papers, and cell phones spread across the table. As soon as the sentence came out, smiles spread across each of our faces. The faculty principal investigator on the project started to chuckle out loud. The laughter spread, and in a few seconds we were mapping out the hypothetical steps and implications of the idea. We left the room still talking, the conversation light with excitement and anticipation.

Something about the idea of crowdsourcing our data collection seemed completely appropriate. We had struggled for weeks to arrive at a compelling research design and found ourselves confronting a very concrete data collection problem: how to do an extensive content analysis of several hundred websites at once. The questions we wanted to ask about each site were fairly simple, but far too numerous to make it practical to answer them all without the help of a very large team of research assistants. Knowing that we didn't really have the organizational or financial resources to undertake that effort, we had started to consider scaling back the study and other alternatives. Then came the Mechanical Turk idea: we thought we would distribute the work as a massive set of microtasks through an online labor market. As it turned out, we had no idea what that would entail.

Mechanical Turk is a Web service operated by Amazon.com and the most widely known commercial platform for distributed work online, which is also known as crowdsourcing, or human computation. A decent

9386_007.indd 155 7/6/2015 3:18:02 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

definition of crowdsourcing is *the disaggregation and distribution of large tasks across a large group of people over the Internet*.[1] To illustrate what this actually means, imagine that you have 100 million photos that you need to make sure do not include any child pornography, hate speech, copyright violations, or something similar.[2] The automated methods of filtering images for this kind of stuff turn out to be imperfect in a bunch of ways, so you cannot trust them 100% of the time. As a result, you need to verify that these automated methods work by at least reviewing a sample of the images by hand. Crowdsourcing the solution to this problem would entail that you use some system (probably also online) to enable as many people as possible to review as few or as many photos at a time as they want (probably in exchange for very tiny amounts of money for every photo reviewed).[3] Other examples of crowdsourcing and human computation involve very different kinds of problems and people, as well as more or less complicated procedures.

Over the past five years or so, crowdsourcing of many different kinds has become a growing area of investment and innovation in Silicon Valley as well as in several research communities. At the same time, crowdsourcing remains a relatively foreign concept among social scientists. The purpose of this chapter is to both introduce the idea of using crowdsourcing for social science research and to address some of the challenges I have encountered in trying to actually do so. Despite the fact that writing something like this makes it seem like I am setting myself up as an expert on the subject, I have really been more of a spectator sitting somewhere off to the side of the avant garde of the crowdsourcing industry and the research communities involved in advancing the development of new techniques and tools for managing "collective intelligence." In the process, I have developed multiple research projects that utilize crowdsourcing both as a field site where I have recruited research subjects as well as a source for distributed research assistance. In the material that follows, I discuss practical issues involved in this work that extend beyond such a novel, unusual field of study.

The biggest concerns with crowdsourcing any aspect of a research project boil down to the fact that the practice of crowdsourcing tends to be (1) organized like poorly compensated online piecework, and (2) very boring. In addition, customizing many of the interfaces through which online crowdsourcing takes place can require significant computer

9386_007.indd 156 7/6/2015 3:18:02 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

programming knowledge, and some of us who choose to use them enlist the help of a software engineer at some point. As a result, you, the researcher, wind up confronting digital-age variants of organizational problems that have plagued large-scale divisions of scientific, industrial, and bureaucratic labor for several centuries. How do you hire and manage employees to contribute to a project without sacrificing quality? How do you collaborate with colleagues who have completely different training and vocabulary from your own? How do you learn to use new tools without wasting resources? None of my graduate theory and methods seminars prepared me adequately to confront these challenges, despite the fact that I now believe they have been commonplace in academic research for a long time.

In a classic 1966 essay, Julius Roth described the problems with using what he called "hired hands" to conduct research work.[4] His analysis speaks to a general problem of any research project employing multiple people and technologies to perform data collection or analysis: "When the tasks of a research project are split up into small pieces to be assigned to hired hands, none of these data-collectors and processors will ever understand all the complexities and subtleties of the research issues in the same way as the person who conceived of the study. ... Since the [research] director often cannot be sure what conceptions of the issues the hired hands have as a result of his explanations and 'training,' he must make *dubious guesses* about the meaning of much of the data they return to him. … As he gains in quantity of data, he loses in validity and meaningfulness" (Roth 1966: 193, emphasis added).

Roth could just as easily be speaking of a more typical study with one or two paid research assistants as a crowdsourcing study employing thousands on Mechanical Turk or some other Web-based platform. Irrespective of the type or scale of work you do, the critique (as well as his humorous ethnographic examples of hired hands undermining researchers' objectives) should strike terror into your heart if you had assumed that such farming out your scholarship to people or machines was going to be a straightforward affair.

At the same time, I disagree with part of Roth's premise, as well as his conclusion. Problems of data quality and uncertainty apply even when research assistants do not perform your data collection for you. The difficulties of calibrating and coordinating diverse sets of people,

9386_007.indd 157 7/6/2015 3:18:02 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

sociotechnical systems, and techniques affect even those projects where there is only one researcher involved. Data collection methods and artifacts may or may not, in Langdon Winner's (1986) famous phrase, "have politics," but they are also far from neutral instruments that exist only to serve the scholarly will in a direct or unmediated way. It takes a lot of practical learning and work to apply any method effectively, and the process usually entails numerous failures and mistakes. Furthermore, the quality of data does not necessarily increase in inverse proportion to its quantity. The failures, mistakes, and "dubious guesses" that so often look like bad research work often feed into the construction of a more nuanced, intelligent understanding of a particular process, method, or person.

A narrower form of my argument may be more straightforward: *creating and managing an effective division of intellectual labor lies at the heart of any ambitious research project*. By "research project" and "intellectual labor" here I really mean the whole gamut—all the stuff that researchers, research assistants, colleagues, and supervisors do in addition to the methods and the tools involved in data collection and analysis. For a variety of reasons I explain below, crowdsourcing has illustrated the value of this broad perspective to me many times over. As social scientists, many of us may cling to the myth of the heroic scholar, laboring away in monastic, dusty solitude and publishing brilliant (sole-authored!) articles or books. However, the reality of scholarly labor is almost always more interesting and complicated, involving teams of assistants, administrative staff, librarians, statisticians, or—increasingly for those of us doing work involving digital media—software, engineers, websites, and Internet users about whom we may know shockingly little. The names or accomplishments of these individuals or tools may appear in footnotes and acknowledgments, but the labor of their contributions often remains invisible, with the heavy brush strokes of collaboration, paid work, and research administration obscured behind a polite veneer of attribution.

Scrutinizing the division of academic labor from this point of view can get a little uncomfortable. What began in my research group meeting that day as a starry-eyed foray into crowdsourcing research has forced me to recognize some of my shortcomings as a scholar as well as the extent to which I have capitalized on the contributions of others to my work in a way that feels unfair and even a little exploitative. At the same time, I still

9386_007.indd 158 7/6/2015 3:18:02 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

hope that these realizations may make me a better researcher in the end. Crowdsourcing has shown me that in the course of a collective endeavor, almost everyone and everything involved will, at some point, make "contributions" that are either unproductive or an active hindrance to the progress of the project. The interesting questions arise when you try to figure out how to learn from all these failures and "dubious guesses."

Crowdsourcing and the Division of Scientific Labor

As I suggested above, the need to divide an overwhelming amount of academic labor drove my collaborators and I to use crowdsourcing for our data collection. In this regard, our motives resemble the reasons human computation and crowdsourcing were developed in the first place. Historical examples of the division of academic labor overlap with the earliest cases of human computation, which in turn gave rise to contemporary crowdsourcing techniques.[5] In 1759, a team of three French astronomers used calculus to work out an accurate estimate for the date on which Halley's Comet would pass closest to Sun. In doing so, the three, Alexis Clairaut, Joseph Lalande, and Nicole-Reine Lepaute, had to perform literally thousands of arithmetic calculations *by hand*, a task that they completed in part by dividing the labor and devising systematic procedures to check for errors. The techniques they developed became the foundation for subsequent practices of scientific calculation and computation, which grew in scope and ambition as calculus and statistics became prominent fields of research and innovation. Following the two world wars, during which human computers even calculated the ballistics tables used in the field to aim various weapons, human computation waned in prominence as mainframe computing became feasible and outpaced the processing power of people alone.[6]

Ironically, the diffusion of digital computing networks that followed the mainframe era has also brought with it a renaissance of human computing, more commonly referred to now as crowdsourcing. As was the case with human computation, much of the innovation in crowdsourcing has happened through scientific research. The 2001 NASA clickworkers project, in which visitors to a NASA website could volunteer to label craters in photos of the surface of Mars, represents the foundational example of contemporary academic crowdsourcing (Benkler, 2006: 136–138;

9386_007.indd 159 7/6/2015 3:18:02 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

Kanefsky, Barlow, and Gulick, 2001).[7] Since then, a number of scientific projects have adopted similar techniques to engage amateur citizen scientists in data collection and classification.[8] In addition, a growing number of commercial crowdsourcing platforms have made paid, distributed work more accessible. Using these platforms, computer scientists and computational linguists have begun to apply more refined filtering techniques to some of the data collected by crowds in order to improve the precision of results (e.g., Quinn and Bederson, 2011; Sheng, Provost, and Ipeirotis, 2008; Snow et al. 2008). The result has been a concomitant growth in research on the dynamics and applications of crowdsourcing, including a growing body of work that uses crowdsourcing for social scientific inquiry and to address topics of relevance to social scientists (Horton, Rand, and Zeckhauser 2011).

There are now several websites that make it feasible to recruit a crowd of your own (in exchange for payment or not) and, indeed, several companies have received venture funding for business models based on commercial crowdsourcing of various kinds. In terms of research, the diffusion of these tools and techniques has made it so that you no longer need to be an engineer or a physicist at NASA to experiment with crowdsourcing. Indeed, a few dollars and an hour or so of tinkering can get you access to a crowd of your own through Web services like Amazon's Mechanical Turk, CrowdFlower, or Clickworker (a private company founded in Germany that shares the name of the original NASA project).[9] Likewise, the widespread availability of tools for building Web applications make it so that a few hours of programming from a skilled developer are all that stand between you and your own custom-built crowdsourcing platform … in theory.

User Error

Thinking back to the conversation in the conference room, I can see that my colleagues and I had bought into a number of prevailing theories about crowdsourcing, most of which were wrong. To start with, consider the basic insight behind using crowds for any kind of work: that it can be faster and more effective to recruit a large group of nonspecialist people to perform a research task in a distributed fashion over the Internet than to employ either a small number of highly trained research assistants or

9386_007.indd 160 7/6/2015 3:18:02 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

computer algorithms. I do not consider this statement untrue, but it leaves out so many important details as to be totally useless when you actually try to crowdsource something. I learned just how useless as soon as I began trying to execute our research ideas on Mechanical Turk.

Once we had secured human subjects approval for our project, I started running pilot jobs on Mechanical Turk. Immediately, I found numerous ways to mess them up. My problems started with simple interface design mistakes, such as asking the workers to enter a date in response to a question, but not requiring that their answer be in a particular format. The resulting data was a mess of words, numbers, and variations of month-day-year formats I had never seen. It also took me a while to internalize some of the terminology Amazon uses to distinguish between the individual microtasks that make up a larger job (they call these "human intelligence tasks," or "HITs," and the larger job itself (these used to be known as tasks, but are now called groups). As a result, I ordered several jobs at once in which each of 10 questions were answered 100 times by a single Mturk worker, instead of 100 workers answering 10 questions one time each (feel free to read that over if you need to, or just trust me when I say it's a totally bone-headed move). Luckily, within a few minutes, one of the workers emailed through the Mturk interface to alert me to the situation and I stopped the job after only 200 or so of the HITs had been completed. Such mistakes were (and still are!) embarrassing as well as costly. I had not accounted for my own ignorance in the original research budget, and after accumulating a handful of similar screw-ups, I had to go to the project principal investigator and explain all the idiotic ways in which I had managed to waste our money in order to ask for some additional funds. Even though Mturk tasks cost about a penny each, when you purchase several dozen or hundreds of them at a time, and the workers respond to your tasks within seconds, the pennies add up fast.

All of these failures forced me to confront the fact that even a crowd will not do your work for you automatically. Crowdsourcing is only as fast and effective as the person designing and managing the task. The idea that you can conduct large-scale projects more efficiently simply by hiring more people falls apart if you do not know how to take advantage of their attention and intelligence. Indeed, my experience suggests a social science corollary to Brook's Law[10]: adding manpower to a large research project likely makes it slower.

9386_007.indd 161 7/6/2015 3:18:02 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

Communication Breakdowns

Eventually, after completing a few successful pilots, I ran a pair of controlled experiments on Mechanical Turk. For both of these studies, I worked in collaboration with multiple researchers across different disciplines and organizations, using Mechanical Turk as a subject pool to recruit participants (Shaw, Horton, and Chen 2011; Antin and Shaw 2011). At the same time, my coauthors and I conducted the experiments using different platforms. The first involved a Web-based application that software developers and researchers in a university setting had designed to handle traditional survey and content analysis tasks. The second study ran on the CrowdFlower platform, originally built for the purpose of commercial, enterprise-scale crowdsourcing jobs. In both cases, there were several unanticipated hang-ups in the process of actually getting to the point where we could run the study. Some of these problems stemmed from using two relatively new pieces of software that still had numerous bugs.[11] Others emerged later, the products of communication breakdowns between myself and the software engineers.

In both experiments, the biggest communication breakdowns concerned the process of assigning participants into groups. It is only a slight exaggeration to say that *random* and *unique* treatment assignment represent the most fundamental prerequisites for a valid, controlled experiment in the modern sense of the term (see Fisher 1935). Basically, this means that each subject in the study gets randomly assigned and exposed to one, and only one, of the experimental conditions. Meeting these criteria help ensure that the study does not violate the stable unit treatment value assumption (a.k.a. "SUTVA").[12] The bottom line here is that if you mess up either the random or the unique assignment of treatment and control conditions, your experiment cannot generate valid inferences about either any of the differences between the effects of the experimental conditions or any effects the treatment might or might not have on the population at large. In other words, you cannot claim that your experimental treatment caused the results that you found. You can only claim to understand what the results of a controlled experiment mean if you handle the treatment assignment in a systematic way. No exceptions.

9386_007.indd 162 7/6/2015 3:18:02 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

Not surprisingly, my collaborators and I wanted to ensure that our engineer collaborators really understood what we meant when we asked them whether the software really met the requirements of random and unique treatment assignment. For the most part, we, the social scientists, did not really try to understand exactly how the Web applications did what they did; however, given the stakes of getting treatment assignment right, we felt we needed to be absolutely certain we understood the steps involved. We would not drop the issue until we were 100% certain that everything worked exactly as we expected 100% of the time.

Needless to say, our software engineer colleagues on both projects found this sort of nit-picky insistence tiresome. On the first experiment with the software developed in an academic research center, hashing out the details of the randomization function generated epic email threads, spanning several weeks and incorporating dozens of examples as all of us tried repeatedly to illustrate our ideas in jargon-free terms. It also resulted in a few more failed pilot studies as we tested and retested the software to make sure it behaved as planned. One afternoon, when we were painfully close to running the study, the lead developer and I spent three hours on the phone, walking through every step of the experiment (as I wanted it to run) and every block of the treatment assignment code (as he had written it). This sort of uncompromising attention to detail allowed the study to succeed eventually, but I still came away wondering why it had taken so incredibly long to do so.

On the surface, my experience working with the private startup *looked* totally different. In that case, I had reached an informal agreement with one of the co-founders to help the company implement some survey work on behalf a prospective client. At the time, the company was extremely small (about five employees), and one of the perks of working with them was that they were excited to help me run experiments using their platform.[13] The three members of the engineering team all had experience building commercial Web applications, and the platform already had an extremely robust set of features as well as an elegant user interface. Everything suggested that the process would be much smoother and faster than my experience with the small academic research center's software developers.

9386_007.indd 163 7/6/2015 3:18:02 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

Despite the platform's powerful ability to scale and all the amazing features that were already built in, it still took almost three months of back and forth with one particularly friendly engineer for me to actually run a controlled experiment. Once again, I spent a good part of this time struggling to explain exactly what I meant by the concept of treatment assignment. Instead of randomization, though, the hangup was around ensuring that each worker saw *one and only one* of the treatments. As with the my first Mechanical Turk pilots, fuzzy distinctions between units of analysis, tasks, and groups of tasks made it hard for the engineer and I to figure out exactly what the other person was talking about. As with the first study, we worked through the same conversations three or four separate times.

Eventually, my engineer friend told me about an entire open-source scripting language that had been built into the platform at an earlier point in the development, but that was not included in any of the formal documentation. He thought that I might be able to use the language to make the software do what I wanted. After reading the scripting language documentation and fiddling around with a few more pilot studies, it did in fact turn out that this language could manage the sort of randomization I needed without any customizations.

Unique assignment was harder this time because it undermined some of the assumptions the engineers held about how crowdsourcing tasks tended to work and which they had therefore written into the code. Luckily, I was able to convince them, in the context of a separate discussion with clients interested in using the software to run surveys, that the feature would actually be useful for a wide range of applications, and so the lead engineer felt he could justify putting a couple of hours into implementing the new feature late one night.

In hindsight, nothing about either of these experiences—the one with the developers at the academic research center or the one with the engineers at the startup—strikes me as especially surprising. Perhaps the only part of either story that really defies explanation is why I ever thought the process would be faster, simpler, or less challenging than it turned out to be. As a social scientist with relatively minimal programming experience, I had not anticipated the complexity of software development as a process. I had also failed to explain our needs and concerns in a way that made intuitive sense to software developers, who likewise struggled to

9386_007.indd 164 7/6/2015 3:18:02 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

explain the inner workings of their code in a way that a nondeveloper could easily follow.

Hired Hands Revisited

In the end, both experiments worked out well, and the experience helped me apply crowdsourcing tools for the purpose I originally intended: distributed content analysis. I have now completed multiple pilots and small-scale side projects applying these techniques and am designing several larger studies. When colleagues learn that some of my research involves data collected through Mechanical Turk, they tend to express disbelief that anybody would actually do crowdsourcing work for so little money. They also often voice doubts that the workers on Mturk and other online labor markets answer questions honestly or perform the tasks requested of them in earnest. They never suspect that my early failures with crowdsourced data collection would derive from my own mistakes or my inability to communicate effectively with software engineers. Such responses overlook the practical challenges of learning how to work with unfamiliar digital tools or the process of software development. They also betray an underlying mistrust of "hired hand research" more generally. While this mistrust is warranted in the context of crowdsourced data, I find it overstated and somewhat ironic. Researchers engaged in crowdsourcing data collection and analysis have developed methods that account for the inevitability of low-quality data with robust techniques to maximize the precision of their results. At the same time, researchers performing more traditional forms of content analysis, interviewing, and survey research have few corresponding solutions to correct for the identical problems.

A little background on crowdsourcing workers provides a useful entry point to a broader discussion of reliability and data quality in the context of crowdsourcing. A diverse population of individuals participate in crowdsourcing as workers in online labor markets or as volunteers in noncommercial projects. For example, while Amazon refuses to make representative data about Mechanical Turk workers publicly available, several informal surveys and qualitative studies provide evidence that the population includes a moderate gender balance as well as individuals from a range of socioeconomic, linguistic, and cultural backgrounds (Ipeirotis 2010; Khanna et al. 2010; Ross et al. 2010). In the experiment

9386_007.indd 165 7/6/2015 3:18:03 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

I conducted with John Horton and Daniel Chen, we also found evidence of substantial variation in Internet skills among the participants in our study. Some of this diversity makes Mturk a *more* attractive site for comparative research than traditional laboratories or some other, less diverse online environments. For example, Judd Antin and I found evidence of different motivational patterns across Mturk workers in India when compared with those in the United States. While it does not make sense to claim that such findings are representative of broader national trends in either country, the experiment incorporates the variations in the site's user population into the research design, taking advantage of the fact that Mturk makes it relatively affordable to run a cross-national research project.

The diversity of Mturk workers creates some interesting challenges in the day-to-day process of running a study. Earlier, I alluded to the fact that workers can email you about any task that you post to the site. In that case, a worker alerted me to a problem with one of my tasks, saving me some additional frustration and research funds. In the course of piloting and running the rest of that same experiment, I received close to 100 other emails from workers. Most of them were not nearly so helpful, polite, or grammatically coherent. Many reported problems viewing my task (the majority of these were caused by old versions of Internet Explorer failing to render the site correctly) or simply requested that I submit payments for the task more quickly (I tended to process the payments in monthly batches, which was nowhere near fast enough for some participants). I set up auto-responses and stock replies to common inquires in order to alleviate the inbox traffic, but I had effectively become the payroll and technical support department of a very small, transnational research organization overnight. Somehow, I had wound up at both ends of an outsourcing supply chain at the same time.

In terms of assessing the validity and reliability of data, the diversity of skills, languages, and educational backgrounds among Mturk workers has given rise to some creative solutions that improve upon traditional content analysis techniques. Historically, content analysis research has employed a multistage process to ensure that the results are both valid and reliable (see Krippendorff 2004; Neuendorf 2002). The basic idea behind reliability in this context is that you want to be sure that your results are not simply the product of an idiosyncratic interpretation of the

9386_007.indd 166 7/6/2015 3:18:03 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

world, but rather correspond to a widely intelligible and acceptable interpretation. Methodologically, the state of the art way to test this is pretty simple: you have multiple individuals apply the content analysis instrument (or codes) to a bunch of the same content (text, images, videos, or whatever) and then run some statistical tests on the results to calculate the rate of agreement between the coders.

Nothing about crowdsourcing changes the underlying objective here: you still want your results to capture real concepts, and you still need to get some sort of confirmation that these concepts hold up at an intersubjective level. The core difference just lies in the process by which you go about measuring reliability. The problem with the standard approaches to reliability in the context of crowdsourcing is severalfold. Unlike a more typical research project employing undergraduate research assistants, crowdsourcing, by definition, opens up the project to a wider population of people about whom you, the researcher, know pretty much nothing and over whom you have very little influence. In addition, the scale of participation possible with crowdsourcing makes it feasible for you to incorporate results from even more coders about even more questions than a traditional content analysis study. In this sense, Roth's aforementioned concerns about "dubious guesses" take on a new urgency. The messy data generated by a crowd is, by definition, more dubious than most.

Luckily, a lot of smart statisticians and computer scientists have come up with creative responses to processing unreliable data from multiple coders.[14] The key involves approaching the raw, messy data as an optimization problem: for every data point, we want to maximize the likelihood that we select the correct answer on the basis of the coders' responses. We (or, more accurately, our statistical software), can then treat each coder as possessing an underlying probability of providing a correct response to any question, and can estimate this probability for each coder by asking them to answer a few questions to which we know the answer ahead of time. After establishing the workers' rate of accuracy, we collect multiple responses to every question from multiple coders and weight these observations by the coder's accuracy rate to estimate the answer that is most likely to be correct.

The elegance of this approach stems from the use of "bad" information to produce valid, reliable inferences about the best answers to any

9386_007.indd 167 7/6/2015 3:18:03 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

particular question. In Roth's terms, you take a lot of dubious guesses, try to figure out how dubious they are, and then use that knowledge to estimate the least dubious guess. Instead of fighting or denying the hired hands problem, this technique takes advantage of it in a clever way.

No Silver Bullets

If you step back from the details, these sorts of algorithmic approaches to maximizing data quality contain more general implications for research design and methods. As I said in the introduction, I think all research work entails a process of combining tools and perspectives in an effort to extract accurate and reliable knowledge about a particular topic. The extent to which this process involves hired hands, online participants, collaborators unfamiliar with social scientific methods, or computational techniques varies from project to project. In the present case, all of these factors introduced sources of error, bias, and uncertainty into my work. As with the failed pilots and bad-quality data that came from some of the respondents in my Mturk studies, I think these sorts of challenges constitute an important part of the learning process that eventually allowed me to produce better studies. In retrospect, it does not make sense to me to even pretend that I could have completely eliminated all the mistakes along the way. Instead, the best I could do was to incorporate these mistakes into my knowledge of the process and try to optimize the process within the limits of that knowledge.

If you think of research as a constructive learning process involving multiple people, procedures, and tools, each of which introduces some sort of bias and error along the way, I think the question of hired hands opens up a broader discussion. The big point here is that any kind of wrong information is still informative—whether about the quality of the data, the reliability of the data collection processes, or the nature of the people participating in the research project. As a result, every dubious guess provides an opportunity to learn and to refine and improve the research project and the finished product it generates.

In this sense, I think there comes a time in any research project when each of us, intentionally or not, undermines the validity, reliability, or accuracy of the work. As researchers, we may do this by imposing an overly idealized vision of the research process on the reality of data

9386_007.indd 168 7/6/2015 3:18:03 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

collection (which tends to be messy, unpredictable, and biased in some way or another). We may also do this by naively expecting our research assistants and collaborators to share our assumptions, objectives, and desires.

For classic "hired hands" like Roth's research assistants or the typical worker on Mechanical Turk, the reasons behind this have to do with the structural constraints imposed by the division of labor and the dynamics of paid work in labor markets. In these markets, the exigencies of wage earning too often result in people doing a task they find uninspiring, boring, or stupid. In the context of Mechanical Turk and other similar online platforms, these dynamics tend to be exacerbated by extraordinarily atomized and antisocial work environments that do little to counterbalance the intrinsic inequalities between workers and employers.

For collaborators like my software engineer friends, the story gets more complicated: the ways in which they undermined the project's progress and objectives did not seem to stem from the desire to shirk responsibility for financial or other reasons. Instead, there was an underlying misunderstanding that simply took a lot of work and communication to correct. Once everybody was on the same page, they quickly and happily helped me execute randomization in both of my studies.

Needless to say, I am still quite far removed from the vision of crowdsourced data collection that my colleagues and I imagined around that conference table. In fact, almost three years later, I am just beginning to realize some of the plans we laid out. It turns out that even when you can do research on and about a networked environment where the cost of attracting thousands of people to participate in your study approaches zero, nobody and no technology will ever truly do all the work for you.

Notes

1. As with many neologisms related to the Internet, competing definitions for crowdsourcing exist. Mine is derived from a combination of the original description by Howe (2008), who coined the term, and my own observations.

2. As it happens, this is roughly the number of photos uploaded to Facebook *every single day*. Source: http://en-gb.facebook.com/notes/facebook-engineering/developing-facebooks-new-photo-viewer/499447633919. To give you some tangible idea of how many photos this is, consider that 100 million sheets of regular thickness (0.0004 in. or 0.01 cm) paper would create a stack roughly 30,000 feet (or 10 km) high.

9386_007.indd 169 7/6/2015 3:18:03 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

3. This kind of arrangement raises all kinds of questions, like how you know that the photo reviewers are not also the kind of people who like to post porn everywhere they go on the Internet; how you figure out what to do when two reviewers of the same photo disagree about what constitutes porn in the first place; and whether it's ethically a good thing for anybody to be doing mind-numbing work like this in exchange for such small amounts of money; etc. These are hard questions that are fundamental to crowdsourcing as both a social technology and an economic phenomenon. I'll try to address some of them below.

4. Thanks to the editors of this volume for pointing me to Roth's essay.

5. My account here leans heavily on David Alan Grier's *When Computers Were Human* (2005).

6. It is worth noting that from the late nineteenth century until the advent of the mainframe computing era, the majority of *human* computers were women (Grier 2005). This kind of academic grunt work provided one of the few avenues of scientific achievement accessible to (and culturally acceptable for) women at the time. In all likelihood, the relative invisibility of human computers in the history of science and mathematics derives from this *feminization* of computational labor. In addition, the similarly forgotten fact that the earliest computer programmers were women resulted directly from the legacy of women in human computation. Just as the equipment and research budgets for computing got bigger, the number of women involved in the field declined.

7. In that case, three NASA researchers, Bob Kanefsky, Nadine Barlow, and Virginia Gulick, wanted to identify craters in photographs of the surface of Mars taken by the Viking orbiter and decided to see whether volunteers recruited over the Internet could perform the task as well as a physicist with a PhD (Barlow herself). The results of their study (which they conceived as a side project) suggested that with a little bit of filtering, the accuracy of the results provided by the volunteers was about the same as that of the physicist (Kanefsky, Barlow, and Gulick 2001).

8. I should point out that some of the creators of these projects do not call what they do crowdsourcing and argue that there are fundamental differences between crowdsourcing and citizen science. I think this is fine, but since I prefer a broader definition of crowdsourcing (see above), I tend not to observe such fine-grained distinctions.

9. Full disclosure: I have done paid consulting work with CrowdFlower, which is based in San Francisco, CA.

10. Brooks (1995) law states: "Adding manpower to a late software project only makes it later."

11. A very distinguished computer science professor only recently explained to me that even the greatest software only has *fewer* bugs; it is never bug-free. I didn't fully understand this until I started collaborating with developers in the process of designing, testing, and improving real code.

12. Since I find the term completely obfuscatory and unhelpful, I avoid using it, but I feel sort of obligated to mention it in case you want to look into this kind of

9386_007.indd 170 7/6/2015 3:18:03 PM

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

thing further or intimidate your friends with academic jargon. The assumptions underlying SUTVA get pretty abstruse and discussing them would take me deep into the methodological weeds, so I'm going to just refer you to some really clear articles that explain how it works in more detail: see Holland, 1986; Little & Rubin, 2000.

13. In this regard, my experience collaborating on research with a private-sector organization had some interesting parallels with the discussion by Williams and Xiong (2009). Like them, I found it necessary to consciously frame my work to be more interesting to an audience oriented toward the very practical problems of running a small company. At the same time, the reduced scale of the company also made it much easier to approach anybody involved and start a casual conversation about the company's work and my interests.

14. Everything I am about to say comes directly from previous research in maximum likelihood estimation and human computation (Dawid and Skene 1979; Sheng et al. 2008; Snow et al. 2008).

References

Antin, J., and A. Shaw. 2011. Social desirability bias and self-reports of motivation: A cross-cultural study of Amazon Mechanical Turk in the US and India. Presented at the 2011 ACM Conference on Computer-Supported Cooperative Work. Hangzhou, China.

Benkler, Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.

Brooks, F. P. 1995. *The Mythical Man-Month: Essays on Software Engineering*. Reading, MA: Addison-Wesley.

Dawid, A. P., and A. M. Skene. 1979. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Journal of the Royal Statistical Society: Series C, Applied Statistics 28* (1): 20–28.

Fisher, R. A. 1935. *The Design of Experiments*. Edinburgh, Scotland: Oliver and Boyd.

Grier, D. A. 2005. *When Computers Were Human*. Princeton: Princeton University Press.

Holland, P. W. 1986. Statistics and causal inference. *Journal of the American Statistical Association 81* (396): 945–960.

Horton, J. J., D. Rand, and R. J. Zeckhauser. 2011. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics 14* (3): 399–425.

Howe, J. 2008. *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. New York: Crown Business.

Ipeirotis, P. 2010. Demographics of Mechanical Turk. New York University working paper.

PROPERTY OF THE MIT PRESS
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY

Kanefsky, B., N. G. Barlow, and V. C. Gulick. 2001. Can distributed volunteers accomplish massive data analysis tasks? Proceedings of the Lunar and Planetary Institute Science Conference 32: 1272.

Khanna, S., A. Ratan, J. Davis, and W. Thies. 2010. Evaluating and improving the usability of Mechanical Turk for low-income workers in India. Proceedings of the First ACM Symposium on Computing for Development, 12: 1–10. London, UK: Association for Computing Machinery.

Krippendorff, K. 2004. *Content Analysis: An Introduction to Its Methodology*. 2nd ed. Thousand Oaks, CA: Sage.

Little, R. J., and D. B. Rubin. 2000. Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health 21*: 121–145.

Neuendorf, K. A. 2002. *The Content Analysis Guidebook*. Atlanta: Sage.

Quinn, A. J., & Bederson, B. B. 2011. Human Computation: A Survey and Taxonomy of a Growing Field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1403–1412). New York, NY, USA: ACM. http://doi.org/10.1145/1978942.1979148.

Ross, J., L. Irani, M. S. Silberman, A. Zaldivar, and B. Tomlinson. 2010. Who are the crowdworkers? Shifting demographics in Mechanical Turk. Proceedings of the 28th international conference extended abstracts on human factors in computing systems, 2863–2872. Atlanta, GA.

Roth, J. A. 1966. Hired hand research. *American Sociologist 1* (4): 190–196.

Shaw, A., J. J. Horton, and D. L. Chen. 2011. Designing incentives for inexpert human raters. Proceedings of the 2011 ACM Conference on Computer-Supported Cooperative Work. Hangzhou, China.

Sheng, V. S., F. Provost, and P. G. Ipeirotis. 2008. Get another label? Improving data quality and data mining using multiple, noisy labelers. Proceedings of the Conference on Knowledge Discovery and Data Mining.

Snow, R., B. O'Connor, D. Jurafsky, and A. Y. Ng. 2008. Cheap and fast—but is it good? Evaluating non-expert annotations for natural language tasks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Williams, D., and L. Xiong. 2009. Herding cats online: Real studies of virtual communities. In *Research Confidential: Solutions to Problems Most Social Scientists Pretend They Never Have*, ed. E. Hargittai, 122–140. Ann Arbor, MI: University of Michigan Press.

Winner, L. 1986. Do artifacts have politics? in *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, 19–39. Chicago: University of Chicago Press.

9386_007.indd 172 7/6/2015 3:18:03 PM

Evaluating Open Collaboration Opportunities in the Fire Service with FireCrowd

Eleanor R. Burgess
University College London
eleanor.burgess.15@ucl.ac.uk

Aaron Shaw
Northwestern University
aaronshaw@northwestern.edu

ABSTRACT

In emergency response organizations like the fire service, personnel require easy access to reliable, up-to-date safety protocols. Systems for creating and managing Standard Operating Procedures (SOPs) within these command and control organizations are often rigid, inaccessible, and siloed. Open collaboration systems like wikis and social computing tools have the potential to address these limitations, but have not been analyzed for intra-organizational use in emergency services. In response to a request from the Fire Protection Research Foundation (FPRF) we evaluated a high-fidelity open collaboration system prototype, FireCrowd, that was designed to manage SOPs within the U.S. fire service. We use the prototype as a technology probe and apply human-centered design methods in a suburban fire department in the Chicago area. We find that organizational factors would inhibit the adoption of some open collaboration practices and identify points in current practices that offer opportunities for open collaboration in the future.

Keywords

Organizations; emergency response; open collaboration; knowledge management

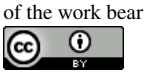
Categories and Subject Descriptors

H.5.m. [**Information Interfaces and Presentation (e.g. HCI)**]: Miscellaneous

1. INTRODUCTION

We investigate whether open collaboration systems can enhance knowledge management around safety protocols within the U.S. fire service. Fire fighters and emergency response personnel need fast, reliable access to up-to-date safety protocols in the field. However, in the fire service, Standard Operating Procedures (SOPs) are written and approved exclusively within individual departments even though many departments confront similar situations. Protocols are infrequently updated and are usually stored in paper files or desktop computers that never leave the firehouse (see Figure 1).

This work is licensed under a Creative-Commons Attribution International 4.0 License. The authors request that copies, derivatives, and other reuses of the work bear the full citation on the first page.



OpenSym '16, August 17–19, 2016, Berlin, Germany
© 2016 Copyright held by the authors.
ACM ISBN 978-1-4503-4451-7/16/08.
DOI: http://dx.doi.org/10.1145/2957792.2957794



Figure 1: The SOP binder in the fire department where we conducted this study.

These challenges pose an urgent threat to the health and safety of fire service personnel.

Open collaboration systems such as wikis or social Q&A platforms support collaborative, distributed knowledge management [6]. Prior work has shown how open collaboration systems support cooperative work in firms [2, 12] and other settings, but has not evaluated these systems within emergency response contexts. Earlier studies of collaboration technologies in emergency response emphasize how cultural tensions between different organizations (e.g., fire service, police, and emergency medical services) prevent effective cooperation [9]. In contrast, we explore the opportunities for open collaboration in a critical knowledge management system *within* a single branch of emergency response: the fire service.

At the invitation of the FPRF, we evaluated FireCrowd, a prototype open collaboration system for SOP creation and revision in the U.S. fire service.[1] We used the FireCrowd prototype as a technology probe, applying qualitative, human-centered design methods to assess opportunities for open collaboration in a single fire department. We extend prior HCI research on technology adoption in organizations and open collaboration systems by focusing on a novel domain: knowledge management within a command and control emergency response organization. Our findings confirm the potential of open collaboration systems to overcome knowledge management inefficiencies, but underscore how organizational structures and routines would also inhibit collaboration and openness.

[1]See http://firecrowd.com

2. PRIOR WORK & CONTEXT

Open collaboration and social computing systems offer unique advantages for facilitating knowledge sharing, transactive memory, coordination, and information diffusion across a number of domains [2, 6, 12]. However, introducing open collaboration into the safety protocol management of a command and control organization such as the fire service presents specific design challenges. On the one hand, systematic SOP review is necessary to ensure high quality, reliable protocols. On the other, protocols should be updated quickly and take advantage of the knowledge spread across individuals in many departments. These competing demands create tensions between hierarchy and openness that existing research has not explored.



Figure 2: Diagram of a typical fire department SOP creation process. The FPRF designed FireCrowd to intervene in the steps represented in orange boxes.

HCI scholars have analyzed systems to support the operations of public safety organizations with command and control structures, including fire and police departments [4, 10, 9]. Some have examined the adoption of mobile computing systems in the fire service (e.g., [10]) and intelligence communities [1, 3]. Recent studies investigate organizational uses of social media systems in the context of crisis response [7, 13]. This prior work demonstrates how organizational factors such as boundary management, professional closure, norms, and organizational structure can inhibit information sharing and collaboration. A related study by Ley and colleagues [9] shows that technological systems that support context awareness can facilitate coordination and communication across emergency response organizations. We are not aware of studies examining opportunities for open collaboration systems *within* command and control organizations where poor knowledge management threatens the lives and safety of system users.

2.1 Standard Operating Procedures

We focus on a critical intra-organizational knowledge management system in the United States fire service: Standard Operating Procedures. SOPs are prescriptive protocols that fire fighters follow when responding to calls. The protocols vary widely and cover topics from high-rise building fires to proper hose operations. Within the United States, fire fighters are trained using SOPs and held accountable to them on the job. In order to ensure SOP quality, departments typically follow a multi-step process to draft, review, authorize, implement, and revise their procedures (see Figure 2). Many fire fighters provide writing and editing input and the chief provides final signoff before the SOP is put into action.

Fire service SOP creation and revision routines lack optimal mechanisms to incorporate new knowledge quickly, leading to avoidable safety risks [5]. A striking example of this occurred when electric and hybrid vehicles first became popular. During incidents involving the near-silent motors of these vehicles, fire fighters attempted to extinguish fires without realizing that the cars were still running and unoccupied, burning cars would start driving away [5]. It took a long time for this knowledge to diffuse into individual fire department's SOPs.

These limitations around SOP creation and management led the FPRF, a research body serving the National Fire Protection Association, to approach a member of our research team about evaluating an open online collaboration system to manage SOPs at a national level within the fire service. The foundation funded the creation of the prototype, which we evaluated in the current study. Below we briefly describe the prototype, research methods, and setting for our evaluation.

2.2 The FireCrowd Prototype

FireCrowd is a high-fidelity prototype open collaboration system. The prototype was created under the direction of an FPRF advisory board of fire service experts. It uses a customized Drupal template that allows collaborative editing, manages user roles and authentication, and facilitates the aggregation of collaboratively-authored documents (see Figure 3).

FireCrowd enables U.S. fire service professionals to collaboratively author, share, and access SOPs at a national level. The prototype assumes users are integrated into an existing department SOP workflow and introduces open collaboration at strategic points (see Figure 2). Following FPRF specifications, the system presents users with a repository of editable SOPs in-progress. Changes to the SOPs in-progress are reviewed on a weekly basis by a fire fighter safety expert, who then implements approved changes in a parallel repository of vetted SOPs. Vetted SOPs are made available for download and further department-specific revision.

3. METHODS

We apply qualitative, human-centered design methods to analyze current fire service practices around SOPs and the viability of the FireCrowd prototype. Using the FireCrowd prototype as a technology probe [8], we explore opportunities for open collaboration and knowledge management in the fire service through usability tests, contextual inquiries, and a large focus group. Presenting fire fighters with the FPRF's prototype system elicited direct feedback and responses in a way that interviewing or observations alone (without the prototype) would not have allowed.

Our empirical work took place through interviews with fire service safety experts and in a fire department we partnered with for the purpose of the study. Our partner fire department serves a suburban, primarily residential community of about 75,000 people over

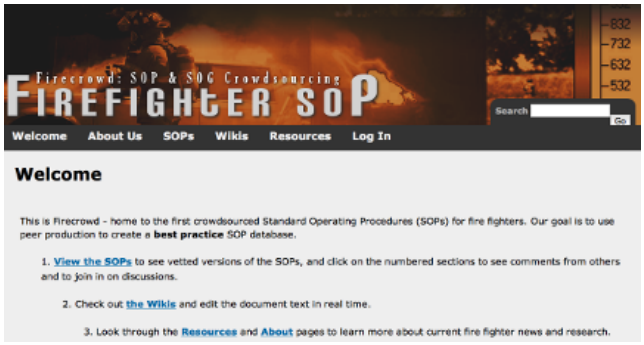


Figure 3: FireCrowd Landing Page

7.8 square miles in the Chicago area. The department maintains a full-time staff of 110 members at five stations.

We conducted usability tests of the FireCrowd prototype with three fire fighter participants each working at a different fire station within our partner department. These tests were completed using fire station desktop computers. Following a study protocol which approximated a typical user set of site actions, we prompted participants to complete example tasks using FireCrowd, (e.g. "log in, find [X] SOP, edit [X] SOP, etc."). We then gathered design feedback and contextual information about technology use in the fire house environment. We audio recorded these usability tests and took written field notes.

In addition, we conducted contextual inquiries with two fire safety chiefs; one with the Division Chief of Operations and Training at our partner department, and one with the safety chief of an East Coast department well-regarded for their quality SOPs. We took written notes regarding their department processes of new SOP development, implementation, and revision. We also held follow-up phone conversations to clarify details in response to topics that emerged as we analyzed their original responses.

Finally, we facilitated a focus group of the 33 SOP Committee members in our partner department. We presented FireCrowd to the group and then solicited feedback about SOP procedures in writing and discussion following a study protocol (available upon request). We collected all written responses, audio recorded the meeting, and took detailed written notes.

Data collection took place over a 12 week period in 2014, yielding field notes, interview recordings and transcripts, photographs, and fire fighters' written responses to focus group questions. We reviewed, summarized, and analyzed all of these materials. We developed inductive codes based on prior literature and emergent themes. Then, over three rounds of analysis, we elaborated the codes in written memos, refined them, and, re-applied the refined codes to the study documents and notes.

4. FINDINGS

Our observations confirm that SOPs are important political and social artifacts that organize work in the fire service. Probing the existing SOP system with FireCrowd, we find that fire fighters perceive open collaboration as a threat to the norms of SOP creation and implementation. However, we also find opportunities for open collaboration to solve urgent knowledge management challenges.

4.1 Open Collaboration as a Threat

The routines and command hierarchy of the fire service shape fire fighter attitudes toward the control of SOPs. Because the creation and validation of SOPs traditionally reside at the level of each fire department, fire fighters express a strong sense of ownership over the content of "their" SOPs. However, in the FireCrowd prototype, individuals from many departments are encouraged to edit SOP content. This provokes apprehension among would-be users. For example, several members the fire fighter focus group opposed the idea of outsiders changing department work. One standard-rank SOP Committee fire fighter articulated his concern:

> That's almost a deal-breaker for us. Having a third party be able to go through our process and our SOPs...the reason we have this whole [SOP] Committee is that our process is for our department based on our needs and makes us unique.

Department-level identity, boundaries, and SOP content ownership would inhibit collaboration.

Departments also use SOPs as tools to describe expectations and to evaluate fire fighter job performance. The degree of individual fire fighter accountability to the SOP often depends on their Batallion Chief. One department safety chief argued for strict adherence to SOPs: "Fire fighters should 'stay in the box' and use [the SOPs] they learned in training. If a fire fighter does deviate with good reason, fine, but you have to have a good reason." In such a case, the SOPs express departmental norms directly and are not suitable for editing across department boundaries. More open collaboration threatens to undermine department-level hierarchy and rules.

4.2 Open Collaboration as Opportunity

Conversations with fire fighters and chiefs also revealed several failure points in existing departmental knowledge management routines that open collaboration systems can ameliorate. These failure points include the length and constraints of the SOP creation process as well as general problems with SOP accessibility.

Fire fighters and chiefs perceive SOP creation as time consuming, slow, and inefficient. The current process for our partner department looks like this: one or two fire fighters write a new SOP on their own (unpaid) time. Then the document is edited by the SOP Committee, often going through several rounds of review before final signoff from the chief. Fire fighters must be trained on the new SOP, which takes additional time because, as the department safety chief said, "firemen don't accept change very well." As a result of these delays, SOPs are often poorly maintained and, in the words of one focus group participant, "reactive instead of proactive." Department members recognize the need to revise old SOPs, but fail to do so in a timely manner because of the difficult process. The safety chief at our partner department reiterated that relying on outdated SOPs introduces unnecessary risks.

Our study participants felt that open, collaborative editing would facilitate faster rounds of SOP drafting and revision, better version control, and easier review. The focus group became excited about the ability to support rapid, collaborative intra-departmental edits to SOPs without waiting for the next committee meeting. Another focus group participant spoke about specific collaborative editing features that he would find useful: "When I make changes to an old [SOP], they are hard for other people to identify...Using color...anything added or removed is easy to see." Basic collaborative text editing functions such as tracked changes and version control would address this need.

Respondents also underscored that fire fighters struggle to access and stay up-to-date on SOPs. In our partner department, fire fighters access SOPs in three ways: a large SOP binder in each station (see Figure 1); an intranet SOP database on fire house computers; or, in some cases, a small computer mounted in the fire truck (rig).

All of these methods have limitations identified by the fire fighters, including poor searchability, mobility, and version control.

Study participants saw online and mobile access to SOPs as potential solutions to these problems. One standard-level fireman spoke about burning mailbox calls that are rare, but require a special procedure. He suggested that a smartphone interface could be used to check the SOP in the field. Another recommended that cross-references and suggested searches would be valuable.

5. DISCUSSION

Our findings illustrate how tensions between openness and hierarchy play out in the context of fire service knowledge management. Fire service personnel want higher-tempo forms of collaboration and openness to overcome limitations of existing knowledge management infrastructure. They perceive several ways open collaboration systems might help achieve these goals. However, they also make it clear that some technical affordances of open collaboration contradict key organizational practices. They reject forms of openness that undermine their department command structure.

These findings suggest that open collaboration platforms for SOP management need to be designed and adapted to suit the organizational setting and structure of the fire service more than the FPRF anticipated. For example, any open collaboration system adopted in such a setting needs to afford accountability to versions of the text that correspond to the specific needs, norms, and routines of individual departments. In many cases, fire fighter concerns could be addressed by providing open and flexible access to the SOPs rather than fundamentally changing how SOPs are created. Additional prototyping and evaluation in a larger, more diverse sample of departments can test whether these observations generalize.

The results extend prior literature by probing open collaboration opportunities within the fire service, a command and control public safety organization. Whereas prior work emphasized the advantages of open collaboration for knowledge management, the FireCrowd prototype reveals how open collaboration systems contradict fire service routines and structures. These contradictions affect both fire fighter understandings of such systems as well as their attitudes toward system adoption. Our work confirms that organizational environments shape technology adoption.

6. CONCLUSION

The current study investigates open collaboration and knowledge management within command and control organizations where information reliability and safety are critical concerns. Many emergency response organizations utilize SOPs and the findings we report may be applicable to other contexts including law enforcement. Future research should expand this important area of open collaboration design and evaluation.

7. ACKNOWLEDGEMENTS

Thanks to our partner fire department, Casey Grant and the FPRF, Madhu Reddy, Darren Gergle, Anne Marie Piper, Haoqi Zhang, Liz Gerber, and others for their support and help with this project.

8. REFERENCES

[1] D. C. Andrus. Toward a Complex Adaptive Intelligence Community: The Wiki and the Blog. *Studies in Intelligence*, 49(3):2005–2006, Apr. 2007.

[2] O. Arazy, I. Gellatly, S. Jang, and R. Patterson. Wiki deployment in corporate settings. *IEEE Technology and Society Magazine*, 28(2):57–64, 2009.

[3] R. A. Best Jr. *Intelligence Information: Need-To-Know Vs. Need-to-Share*. Congressional Research Service. Diane Publishing, Darby, PA, July 2011.

[4] L. Carver and M. Turoff. Human-computer Interaction: The Human and Computer As a Team in Emergency Management Information Systems. *Commun. ACM*, 50(3):33–38, Mar. 2007.

[5] J. D. Emery. Hybrid Vehicles: Separating Fact from Fiction. *Fire Engineering*, 162(7):73–82, July 2009.

[6] A. Forte and C. Lampe. Defining, Understanding, and Supporting Open Collaboration Lessons From the Literature. *American Behavioral Scientist*, 57(5):535–547, May 2013.

[7] A. L. Hughes, L. A. A. St. Denis, L. Palen, and K. M. Anderson. Online Public Communications by Police and Fire Services During the 2012 Hurricane Sandy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14. ACM, 2014.

[8] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, and B. Eiderbäck. Technology Probes: Inspiring Design for and with Families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03. ACM, 2003.

[9] B. Ley, V. Pipek, C. Reuter, and T. Wiedenhoefer. Supporting Improvisation Work in Inter-organizational Crisis Management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12. ACM, 2012.

[10] Á. Monares, S. F. Ochoa, J. A. Pino, V. Herskovic, J. Rodriguez-Covili, and A. Neyem. Mobile computing in urban emergency situations: Improving the support to firefighters in the field. *Expert Systems with Applications*, 38(2):1255–1267, Feb. 2011.

[11] W. J. Orlikowski. Learning from Notes: Organizational Issues in Groupware Implementation. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work*, CSCW '92. ACM, 1992.

[12] D. Riehle, J. Ellenberger, T. Menahem, B. Mikhailovski, Y. Natchetoi, B. Naveh, and T. Odenwald. Open Collaboration within Corporations Using Software Forges. *IEEE Software*, 26(2):52–58, Mar. 2009.

[13] A. H. Tapia and K. Moore. Good Enough is Good Enough: Overcoming Disaster Response Organizations' Slow Social Media Data Adoption. *Computer Supported Cooperative Work (CSCW)*, 23(4-6):483–512, July 2014.

CHS: Small: Collaborative Research: Community Success: Advancing Knowledge of Collaborative Organization Through Large-Scale Empirical Comparison of Peer Production Systems

Peer production – the form of community-based online collaboration used to create public information goods like Wikipedia and Linux – has transformed knowledge production, management, and innovation (Benkler, 2006; Benkler et al., 2015). Inspired by large, successful peer production projects, thousands of wikis and free/libre open source software (FLOSS) projects are started each year. The vast majority of these efforts – and even those that build sustainable and useful information goods – never become more than the work of a single individual (Healy and Schussman, 2003; Schweik and English, 2012).

Why do some peer production systems mobilize large communities of contributors and create valuable information goods while most do not? Although there has been an enormous amount of research on peer production (see reviews by Benkler et al., 2015; Crowston et al., 2010; Jullien, 2012; von Krogh et al., 2012), most has described the inner workings of projects that are among the largest and most collaborative communities and can only provide limited insight into what leads projects to thrive in the first place.

We propose a systematic comparative approach to the study of peer production communities that will test core theories of organizational behavior and advance both theoretical and practical understanding of collaboration in peer production systems. We will pursue this analysis through large-scale empirical comparisons across the wikis hosted by the Wikimedia Foundation and Wikia Inc., the largest populations of peer production wikis. In the process, we will combine insights from socio-technical systems and organization science to understand the determinants of community success.

Our empirical research will test three of the most pervasive claims about the organization of peer production: (1) that peer production is facilitated by low transaction costs; (2) that peer production is limited by competition for volunteer resources; and (3) that social interaction drives sustained participation and growth of peer production projects. Contrary to received wisdom, we anticipate that peer production organizations will be robust to small increases in transaction costs, will benefit from spillovers in volunteer effort across similar projects, and that increased social interaction will drive greater quantity and quality of participation. In testing these claims, we will design, build, and publish software to engage in organizational-level analysis of wikis and a unique research dataset of nearly 80,000 wiki projects.

The **intellectual merit** of this project is that it advances knowledge of collaborative organization through the empirical comparison of many public peer production systems using a common software platform. We contribute to knowledge at the intersection of social computing and human collaboration by using organizational theory to draw inferences about the factors that predict the growth and survival of peer production systems. We also contribute to the study of organizational behavior by capitalizing on the extraordinarily granular sources of data generated by peer production and comparing across many thousands of communities.

The **broader impact** of this project is two-fold. First, we will develop actionable insights that communities, system designers, firms, and movements engaged in online collaboration can use to achieve their collaborative goals. Additionally, we will generate a set of freely licensed and publicly available computational research systems and datasets. In both ways, we will contribute to the ability to design for better peer production projects.

1. Project Goals and Prior Work

Peer production has given rise to new types of organizations engaged in the creation of culturally and economically valuable information goods. For example, free/libre open source software (FLOSS) now accounts for some of the foundational infrastructure of the Internet and has become a mainstay of embedded systems and enterprise computing environments. Wikipedia, perceived as Utopian and unrealistic when it began in 2001, has transformed the encyclopedia industry, eliminating most for-profit encyclopedia producers and building a body of articles found to be of equal or higher quality than the leading for-profit competitor (e.g., Giles, 2005; Benkler et al., 2015). The success and impact of these projects have inspired many thousands of attempts to create peer production communities and harness the potential of distributed online collaboration (Crowston et al., 2010). However, the vast majority of these efforts fail to mobilize contributors or produce collaboration (e.g., Hill and Monroy-Hernandez, 2013b; Schweik and English, 2012). Our work tests three of the most influential explanations of why some attempts at peer production succeed while most do not.

In previous research, scholars have sought to understand the foundations of the success and failure of peer production projects through a variety of approaches and methods. Some have approached the study of peer production primarily through the lens of socio-technical systems research, while others have oriented their work toward testing and elaborating theories from organization science. At the same time, most studies have focused on analyzing a single community. In Figure 1, we provide a schematic visualization of these previous approaches along two dimensions in order to illustrate the contributions of our work.

Our work is strongly situated in the relatively underexplored left side of the graph. Like many of the most successful examples of peer production research, but unlike almost all comparative multi-project research, our proposal draws heavily both on socio-technical systems and organization science theories. To understand community success, we will compare many thousands of peer production communities and test the role of three of the most widespread explanations for the success of peer production systems drawn from both socio-technical systems research and classical organizational theory: transaction costs, resource competition, and social interaction.

1.1. Building a Comparative Organization Science of Peer Production Systems

In order to understand why peer production organizations succeed or fail, researchers must compare successful and failed projects. A major part of organization science in the twentieth century turned away from models of organizations as self-contained entities to models that placed organizations in their social environments (Scott and Davis, 2006). One important empirical consequence was a move from intra-organizational studies of internal processes to cross-organizational comparisons of similar organizations. This has led to a rich tradition of testing theories at the level of populations of comparable firms (Hannan and Freeman, 1977; Hannan and Carroll, 1992; Steinfield et al., 2005) and social movements (McCarthy and Zald, 1977; Soule and King, 2008). With the large-scale, public, and exhaustive data available from many peer production projects, socio-technical systems researchers can increasingly embrace similar, comparative approaches. Because peer production communities frequently use common software platforms to mediate their work and communication, they will often provide data that is deeper, more granular, and more directly comparable across projects than is possible in traditional organizational analyses.

Both socio-technical systems research and organization science have produced numerous insights into organizational aspects of peer production communities and wikis (especially Wikipedia). We have reviewed this extensive literature in prior work (Benkler et al., 2015) and elaborated how existing accounts document Wikipedia's early growth, more recent decline, editor contribution patterns

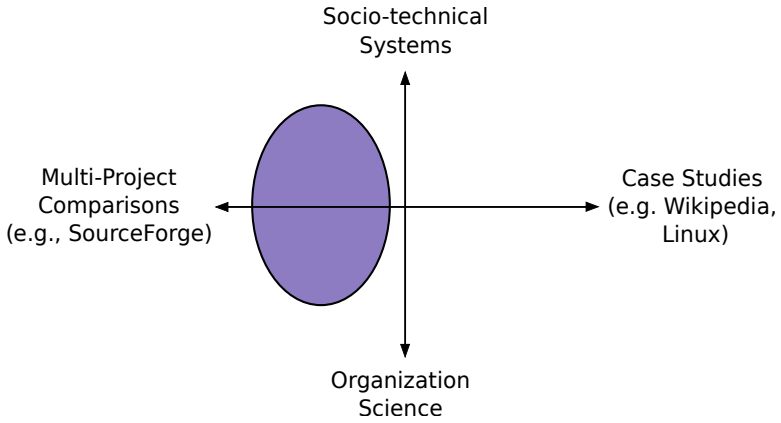


Figure 1: Schematic visualization of prior organizational research on peer production. We combine socio-technical systems approaches with organization science in large-scale comparative analysis. Most existing work has occurred in the upper and lower right quadrants. The purple oval locates our proposed research.

socialization practices, leadership dynamics, bureaucratic structure, governance procedures, community culture, gender relations, coordination dynamics, social roles, and more.

Many of these studies have focused on elaborating the organizational practices and structure within a single community (the right half of Figure 1), whether from a perspective more rooted in the traditions of organization science (e.g., Ransbotham and Kane, 2011; Shaw, 2012), the approaches of socio-technical systems work (e.g., Kittur et al., 2009; Kriplean et al., 2008; Priedhorsky et al., 2007; Welser et al., 2011), or hybrid approaches (e.g., Arazy et al., 2015; Butler et al., 2008; Mugar et al., 2014; Zhu et al., 2012). The vast majority of the previous work on organizational aspects of peer production and online groups has not adopted a cross-organizational perspective.

Our project extends a comparative, inter-organizational approach to the study of peer production, social computing, and online communities. Several reviews of this literature – including ours and one by Crowston et al. (2010) – have highlighted the need for more comparative analysis that includes less successful projects. Only a handful of studies have considered wikis other than the English language version of Wikipedia. In one example, Kittur and Kraut (2010) look at a large group of wikis primarily as a way of understanding the generalizability of findings from Wikipedia. Another small group of studies by Bao et al. (2012) and colleagues has compared the content of different language versions of Wikipedia and focused on cultural and linguistic differentiation. Other work has tested theories of common pool resource management (Schweik and English, 2012) and described communication practices across a large sample of software development teams in SourceForge (Østerlund and Crowston, 2011). In prior work, we have tested a central finding from organization science across a comparative sample of over 600 wikis (Shaw and Hill, 2014).

In the project we propose here, we will design and execute a series of comparative analyses on a large population of wikis to understand how core organizational mechanisms impact peer production communities' growth and survival over time. In ways we have described, research on a large population of projects promises greater external validity than much previous work. We will also be able to test findings from previous peer production research in organization science and socio-technical systems research. This research can inform the design of future peer production systems.

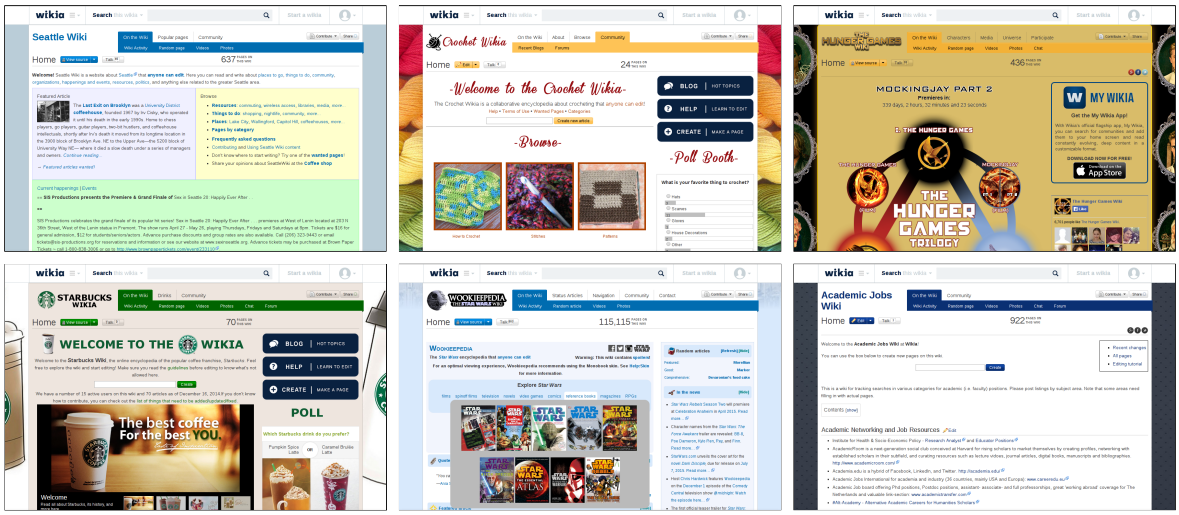


Figure 2: Examples of six wikis hosted by Wikia Inc. from our dataset including (clockwise from top left) Seattle Wiki, Crochet Wiki, The Hunger Games Wiki, the Academic Jobs Wikis, Wookiepedia (a Star Wars encyclopedia) and Starbucks wiki.

1.2. Comparing Peer Production Wikis from Wikia and the Wikimedia Foundation

We propose to build a large-scale comparative dataset of peer production wikis from the two single-largest public wiki-hosting platforms: the Wikimedia Foundation (WMF) and Wikia, Inc. WMF is the the non-profit organization that supports English Wikipedia and more than 800 different wiki communities that include efforts to build encyclopedias, dictionaries, travel guides, text books, biological taxonomic databases, and other reference works in more than 280 languages. Wikia, founded in 2004, sought to apply the Wikipedia model beyond the education-based scope of WMF. Wikia was founded by Jimmy Wales, Wikipedia's founder, and Angela Beesley, one of the most active and respected contributors to Wikipedia in its early years. Wikia's policies, structures, and technologies have been heavily influenced by Wikipedia. Although many private firms host wikis (e.g., PBWiki, WikiSpaces, and SocialText), Wikia is unique in that it only hosts publicly accessible, volunteer-produced, peer production projects and never restricts viewership or participation except to combat spam or vandalism.

Several aspects of Wikia and WMF wikis make them ideal settings to compare organizational behavior and outcomes in peer production communities. First, they are both peer production. Like Wikipedia, anybody can create an account on any WMF or Wikia wiki. The vast majority of these wikis allow contributions even without accounts. Like FLOSS, and other peer production projects, all WMF and Wikia content is distributed freely to the public.[1] More importantly, all WMF and Wikia wikis run on the same software platform – MediaWiki – making them ideally suited to a large-scale comparative analysis. As Figure 2 illustrates, Wikia wikis in particular cover diverse topics and attract widely varying degrees of participation within a common infrastructure. The majority, like Starbucks wiki, only have a handful of contributors and a few dozen pages at most. Others, such as Seattle Wiki, Crochet Wiki, or the Academic Jobs Wiki, become vibrant, medium-sized communities with hundreds or thousands of editors. A very few, like Wookiepedia (a wiki about Star Wars), attract tens of thousands of contributors and contain hundreds of thousands of pages – placing them

[1] As is common in peer production, ownership of the copyright on wiki content remains with the contributors but all material is licensed freely to the public as a condition of contribution. All WMF and Wikia content is released under the Creative Commons Attribution-ShareAlike license and is made publicly available for download.

among the largest and most active peer production communities anywhere.

1.3. Testing Key Theories of Peer Production Community Success

In order to formally test theories at a comparative, inter-organizational level, scholars of peer production need to continue to embrace the opportunity presented by the existence of many thousands of otherwise similar projects that vary in terms of the size, collaborativeness, and behavior. We propose to advance the science of organizations and socio-technical systems through just such an approach. Specifically, we will test three core theories of peer production community success:

> **Project 1 — Transaction costs:** Do small changes in the effort required for contribution to a peer production community positively affect the growth of the resource being produced? Divergent claims suggest that "transaction costs" might either facilitate peer production by lowering barriers to entry (Benkler, 2006) or open the door to enhanced abuses and undermine community growth (Resnick et al., 2000; Friedman and Resnick, 2001).
>
> **Project 2 — Resource mobilization:** Is volunteer effort in the context of peer production communities a "fixed and finite" resource for which voluntary organizations compete (McCarthy and Zald, 1977; Soule and King, 2008; Wang et al., 2013)? Or do we see patterns consistent with theories predicting a positive feedback cycle as contributors build on each others' work and interest and generate positive spillovers (Meyer and Whittier, 1994)?
>
> **Project 3 — Social interaction:** Do increases in the frequency and visibility of social interactions among contributors to peer production projects drive increases in contributions to the information resource being produced? A group of studies have proposed that such interactions will generate high quality participation in the future (Arguello et al., 2006; Butler, 2001; Cheshire and Antin, 2008; Erickson and Kellogg, 2000) and these claims mirror findings in research on the importance of interactions and social participation in collective action (Bimber et al., 2012; Bennett and Segerberg, 2012); corporate intranets (Fulk et al., 2004); and labor unions (Lipset et al., 1956).

We describe the research design and methods for these studies in detail in the corresponding subsections of the project plan in Section 2. We also elaborate how, in the process of doing this work, we will create software for parsing and extracting data from tens of thousands of wiki database files and tools to extract metadata from large numbers of public wiki application programming interfaces (APIs). We will publish these tools under FLOSS licenses and will use them to create and publish comprehensive comparative research datasets.

2. Project Plan

As we have suggested, few organization-level theories about peer production have been formalized and tested empirically across a large population of projects. Among the claims that have emerged in previous peer production and socio-technical systems research, three stand out for their clarity, tractability, and resonance with earlier perspectives in organizational analysis: Projects that (1) offer low barriers to entry, (2) successfully compete with other projects to recruit and mobilize volunteers, and (3) create an environment where social interaction is more frequent and visible, are more likely to grow and survive than projects that do not. We propose to test all three theories in comparative quantitative analyses.

Across all three of the proposed studies, we will use multilevel regression to evaluate empirical models. Where appropriate, we will combine these techniques with "quasi-experimental" methods to identify and estimate the causal impact of the different organizational mechanisms under consideration (Shadish et al., 2002). In each case, we will publish documentation and statistical code used

to conduct analyses in conformance with Stodden and Miguez's (2013) best practices for ensuring reproducible computational research results.

2.1. Project 1: Testing The Effect of Transaction Costs on Participation in Peer Production

Several scholars suggest that extremely low costs of contribution are one of the central characteristics of peer production systems, and that, all else being equal, a project with lower barriers to contribution will attract a higher volume of contributions (e.g., Antin et al., 2012; Benkler, 2002; Preece and Shneiderman, 2009). Others have suggested that eliminating barriers to contribution entirely invites low quality participation and can create a negative feedback cycle driving subsequent contributors away (e.g., Halfaker et al., 2013; Hill and Monroy-Hernandez, 2013b). In support of this approach, a large body of research on reputation systems in online groups suggests that requiring users to adopt consistent identifiers can help recruit newcomers and elicit commitment (Kraut and Resnick, 2012) while also potentially deterring negative and disruptive behavior (Resnick et al., 2000). There have been almost no systematic attempts to compare the benefits and costs of either approach.

We propose to adjudicate between these perspectives on the effect of small barriers to entry in peer production communities by analyzing a series of events where 182 wikis suddenly had anonymous editing enabled or disabled over relatively short periods of time. The unannounced changes are exogenous, discontinuous shocks to the projects, and therefore natural experiments (Shadish et al., 2002). The fact that the change was implemented in software means that end-users and potential participants in the wikis had no way of avoiding exposure to the "treatment" (in this case, the absence of non-registered editing opportunities). We will estimate the effect of the discontinuous increase in transaction costs on the overall quantity and quality of incoming contributions to peer production projects by comparing edits immediately before and after the unanticipated feature change in a "regression discontinuity design" (Lee and Lemieux, 2010).

To assess contribution quality, we draw on an extensive body of research applying computational approaches to the analysis of natural language and network-based measures in the specific context of collaborative wikis (Adler and de Alfaro, 2007; Adler, 2012). Our unit of analysis is the individual wiki and we will examine the effect of the changes on the number of high and low quality contributions by comparing the five weeks immediately preceding and following each change using multilevel longitudinal models that allow for flexible specifications of time and incorporate wiki level effects (Singer and Willett, 2003).

Hypotheses While the theoretical and empirical findings described above imply divergent outcomes, we frame our hypotheses to correspond to the perspective that increased transaction costs will drive a decline in contributions across the board. Specifically, we hypothesize that:

> *$H1_A$: When wikis remove the ability to contribute without accounts, the number of low quality edits will decrease.*
>
> *$H2_A$: When wikis remove the ability to contribute without accounts, the number of high quality edits will decrease.*

In this sense, $H1_A$ represents the intentional effect of reducing less valuable contributions while $H2_B$ reflects the "collateral damage" of also deterring good contributions.

Measures We will classify edits by quality using Adler and de Alfaro's (2007) widely-adopted measure *WikiTrust*. In the WikiTrust model, an edit that is entirely preserved while an article is revised is considered to be of high quality while an article is that is entirely or largely removed would be

classified as being of low quality. Because our hypotheses are stated in terms of the number of high quality or low quality edits, our dependent variables will be binned at the week level. To capture the number of low quality edits for every wiki in each week, we will count the number of edits made to the wiki where less than 20% of the text added in the contribution remains after six subsequent contributions.[2] Our measure of the number of high quality edits is the number of edits made to the wiki where more than 80% of the text added in the contribution is preserved over six subsequent contributions.

Because wikis might see higher or lower quality edits at different points in their lives, we will construct and include a control for each wiki's age in years at each time period. Because the calendar date can affect the way that individuals contribute to wikis (e.g., the creation of spamming robots might lead to more low quality edits), we will create a control for this as well. Finally, we will include additional controls for wiki subject area and topic that we will create by working with a team of graduate student researchers to classify the wikis in our dataset. As part of the creation and validation of our measures, we will also conduct open-ended interviews with a small, random sample of wiki administrators to evaluate the process and results through which the feature change occurred.

Anticipated Findings Based on informal qualitative observation of several of the wikis we plan to measure, we anticipate that higher transaction costs will result in less contributions from a smaller number of contributors across the wikis in our population in support of $H1_A$. However, we do not expect to find less high quality contributions and anticipate finding little or no support for $H1_B$. Regardless of the findings, we will design subsequent analyses and follow-up studies based on the initial results to identify mechanisms that support the behaviors we observe. If we find evidence of heterogeneous project behaviors (e.g., some projects with reduced barriers to participation see an increase in high-quality contributions, whereas others only encounter an influx of spam), we plan to pursue multiple case study analyses (Eisenhardt, 1989) in order to better understand the processes driving these results. Any result will provide an important first test of Benkler's (2006) basic theory of peer production.

2.2. Project 2: Evaluating Theories from Resources Mobilization in Peer Production

According to ecological models of organizational resource constraint, voluntary organizations operating in the same field seek members to engage in similar work and therefore compete for their time and effort (McCarthy and Zald, 1977, 2001; Minkoff, 1997; Soule and King, 2008). Using data from online discussion groups, scholars of online communities have shown that discussion groups covering similar topics compete for participants (Wang et al., 2013; Zhu et al., 2014b) and succeed when they have found effective niches (Zhu et al., 2014a). But volunteer labor is different from many other resources in that organizations may generate interest in volunteering (McCarthy and Zald, 1977, 2001). Indeed, volunteer mobilization efforts by one organization may even generate positive "spillovers" in other organizations (Marwell and Oliver, 1993; Meyer and Whittier, 1994). These spillovers may be even more likely in the context of peer production because potential contributors can immediately utilize and re-purpose earlier contributions due to the systematic free licensing of work products in peer production allowing content reuse.

We will compare volunteer activity across different wikis on the same topic, and articles on same topics in Wikipedia, to test whether volunteer labor behaves like the "fixed and finite" resource assumed in ecological theories. Drawing on one of the example wikis in Figure 2, we might ask whether contributions to *The Hunger Games* wiki go up, down, or remain the same when contributions to the *The Hunger Games* articles on Wikipedia pages increase.

[2] See Adler and de Alfaro (2007) and Adler et al. (2011) for details on specifics of the WikiTrust algorithm and its validation.

Hypotheses Consistent with the perspective adopted in the majority of the literature on resource mobilization, we will frame our hypotheses around the claim that volunteer effort functions like a "fixed and finite" resource:

> *$H1_B$: Contributions to active Wikia wikis on a given topic will decline as contributions to articles in the corresponding topic areas on Wikipedia go up.*

Measures To compare across active Wikia wikis and the corresponding topic areas of Wikipedia, we plan to categorize active Wikia wikis according to their subject matter (measures we describe briefly in Project A). Using this information, we will identify corresponding categories for Wikipedia articles for these topics. Our key dependent and independent variables will then consist of counts of the number of edits within a given topic on Wikipedia and Wikia for every week in our dataset.

Once again, we will include controls for the wiki's age in years at each time period. Also, because the underlying interest in a given topic might fluctuate from one week to the next, we include a control for the calendar date as well as a baseline measure of interest in a topic measured by relative search engine traffic for each topic. As part of our measure development and validation, we will conduct open-ended interviews with a small sample of contributors who work across Wikia and WMF projects in order to understand how they engage in the different communities.

Anticipated Findings In exploratory work using twenty-six wikis, we have found evidence suggesting that even during a period during which volunteer contributions to Wikipedia are shrinking overall, contributions to Wikia wikis are positively related to contributions to the overlapping topics in Wikipedia. We also find a positive relationship within the activity of editors we identify as participants in both projects, providing some evidence of spillovers.

Assuming the pattern of spillovers we have observed in our pilot analysis holds up in a larger number of wikis and over longer periods of time, we plan to investigate the causes of spillovers in greater depth. For example, spillovers may be higher among particular kinds of topics or genres of collaborative production. We will pursue follow-up analyses of wiki projects that either represent or resist the central tendencies observed across the rest of the dataset to refine and build our analyses.

2.3. Project 3: Analyzing the Impact of Increased Social Interaction on Public Goods Production

Overlapping bodies of organization science and socio-technical systems research have emphasized the role of group dynamics and social interactions in shaping organizational behavior. We will seek to understand the effect of sudden changes in the frequency, visibility, and quality of social interactions on participation patterns in peer production organizations. Existing work suggests that such sudden increases in the frequency and visibility of social interactions will drive increased participation in collective action (Bimber et al., 2012) and online communities (Arguello et al., 2006; Erickson and Kellogg, 2000), as well as enhanced performance in other kinds of organizations (Kellogg, 2009), including those engaged in the production of common pools of shared information (Butler, 2001; Cheshire and Antin, 2008; Fulk et al., 1996, 2004). To our knowledge, the causal form of this claim has not been previously tested across a large sample of organizations of any kind.

On many Wikia wikis, a threaded, forum-style discussion feature called "message walls" was suddenly switched on by project administrators. WMF is currently developing a very similar tool called "Flow"[3] which it has already rolled out on a very small number of pages and projects. In preliminary qualitative observations, we have found that the introduction of these tools facilitated more

[3] https://www.mediawiki.org/wiki/Flow

sustained conversations between participants and made these conversations more visible. We will treat the adoption of these interpersonal communication tools as another "natural experiment" and test whether the shift in social interactions driven by the introduction of these tools catalyzed different contribution patterns among new editors. Specifically, we will examine the impact of the tools on the frequency and tone of interactions between participants, and on the frequency and quality of contributions from users who join immediately before or after the introduction of the feature.

Hypotheses We plan to test four hypotheses that operationalize the findings of prior literature suggesting that enhanced ease and visibility of social interactions will result in more frequent, positive, and high quality contributions:

> *$H1_C$: The introduction of improved tools for interpersonal communication will cause the number of newcomer social interactions on wikis to increase.*
>
> *$H2_C$: The introduction of improved tools for interpersonal communication will cause the tone of newcomer social interactions on wikis to become more positive.*
>
> *$H3_C$: The introduction of improved tools for interpersonal communication will cause the number of newcomer contributions to wiki articles to increase.*
>
> *$H4_C$: The introduction of improved tools for interpersonal communication will cause the quality of newcomer contributions to wiki articles to increase.*

Measures The dependent variables in this project will all be attributes of newcomer edits to wikis that turned on the improved interpersonal communication feature. To measure the number of newcomer social interactions, we will construct a count of edits to "message wall" pages or "talk" pages (the new and the old systems for interpersonal communication) by users who created new accounts shortly before and shortly after the transition to the new system. To measure the tone of newcomer social interactions, we will apply automated sentiment analysis using the Linguistic Inquiry and Word Count (LIWC) system (Pennebaker et al., 2001) to the text of all messages sent by newcomers. To examine contributions to articles, we will construct a count of the number of edits the same newcomers have made to article pages. Finally, we will measure the quality of these article edits using the WikiTrust measure we described in Project A above. As in the first two studies described above, we also include a set of controls to account for the likelihood that wikis, as well as individual editors, may vary substantially in ways unrelated to either the introduction of message walls or our outcomes of interest, including controls for the wiki's age at each time period, calendar date, and wiki subject area. As part of our measure development and validation, we will also conduct interviews with a sample of project contributors and administrators to understand their experience of working in wikis with and without augmented interpersonal communication tools like message walls.

Anticipated Findings Based on the findings of previous research and an informal qualitative analyses of two communities' transition to message walls we have already conducted in preparation for this proposal, we anticipate that message walls likely drive a spike in the frequency of social interactions on the wikis, and result in a larger proportion of contributors engaging in on-site conversation. We also predict that the message walls increase the positive tone of these conversations. Finally, we expect that the introduction of message walls caused a positive increase in both the number of contributions by new members and the quality of those contributions. We plan to investigate whether these expectations stand up to systematic testing, and, depending on the outcomes of our initial findings, we will pursue follow-up research to uncover additional details about the mechanisms involved.

2.4. Creating Research Tools and Datasets

To complete the three studies described above, we will need to construct a set of comparative longitudinal datasets that include summaries of activity across many Wikia and WMF wikis. We will write software to build these datasets and will then document and distribute these tools for comparative wiki research under a FLOSS license in order to facilitate adoption and improvement. Currently, multiple tools exist to process Wikipedia data, but none effectively support processing multiple MediaWiki databases from multiple sources in order to create comparative, multi-community datasets. Because both WMF and Wikia use the MediaWiki software, their database dumps use an identical format and creating such datasets is realistic and achievable.

Although specific variables will differ across the three projects we have proposed, each will require a dataset that includes a variety of data and metadata that capture the rate and type of activity within wikis over periods of time (e.g., where the unit of analysis is the "wiki day" or "wiki week"). We have already described several key variables we will use and we have used other variables in our previous research. Although we will not detail every possible variable here, examples of common variables we will create include the number of contributions, the number of unique contributors, the number of contribution to administrative pages, and the number of "talk messages" on a given wiki during a given period.

In preliminary work, we have downloaded database dumps from more than 77,000 publicly available wikis hosted by Wikia Inc. The data were made available for every Wikia wiki routinely through April 2010 when we collected the raw data. We have also downloaded database dumps for WMF hosted wikis. Together, these datasets contain many terabytes of "raw" XML data including, for example, the full text of every revision of every page in every wiki and detailed metadata on contributors. Additionally, we have downloaded a complete archive of viewership logs from WMF. All of these dataset are currently stored in *Hyak*, a shared high performance computing environment at the University of Washington. In some cases, information is not available in these dumps (e.g., data on administrative rights, blocking, banning, and deleting). To collect these data, we will construct tools to query public wiki application programming interfaces (APIs) and to merge the results of these queries with the data from the database dumps.

To create comparative datasets, we will (1) parse the raw XML dataset to create tables of metadata for each individual contribution to each wiki; (2) summarize these edit histories to create datasets with summaries of each wiki; (3) bin the per-wiki edit histories by period (i.e., by days, weeks, months, and years); and (3) merge these binned and summary datasets into large comparative datasets of various size and longitudinal granularity.

As we have done in previous work, we will build software to parse and analyze these datasets in Python and R and will reuse and extend existing tools in the process. For example, we will make extensive use of the freely licensed and open source MediaWiki-Utilities library created and published by the Wikimedia Foundation.[4] Were possible, we will contribute code back to the library where we have already made contributions as part of pilot work for this proposal.

We will work together with the Wikimedia Foundation in developing this aspect of the project. First, they will host our research team for a month each year over the course of the proposed research. Second, they will provide us with access to computational resources designed for analyzing wiki data, including their Hadoop computing cluster to assist in preparation of research datasets. Third, they will work with us to arrange for access to private data stored by the Foundation but not systematically made available to the public. This access will allow us to answer substantive research questions about community success and to collaborate with them to prepare aggregate and/or anonymized

[4] https://github.com/halfak/mediawiki-utilities

versions of these data for public release. Where appropriate, WMF will also provide a venue for us to publish comparative datasets and tools to facilitate the widest dissemination and impact of our research products.

In addition to computational resources made available through WMF, we will purchase capacity in UW's Hyak high performance computing cluster to ensure that we have the processing power and memory necessary to complete this work. Finally, we will purchase a server and storage array to be hosted at Northwestern University to complete additional analysis not possible in Hyak because it is firewalled from the Internet and to host the processed public datasets for other researchers to download.

Although building software to create and analyze these data are an important challenge, we have already successfully built and used two prototype versions of the tools we describe in this proposal. An prototype used to create a dataset for an earlier paper (Shaw and Hill, 2014) was written in C++ and is published online.[5] A second prototype using the Python architecture we will use to complete the work described in this proposal was created as part of our study on redirects (Hill and Shaw, 2014) and is more flexible and adaptable. In unpublished work, we have bench-marked and tested the new system and are confident that it can adapted for building large comparative datasets with the computational resources we are requesting.

Finally, we will make our research datasets available to the public. We will release all research data under the GNU GFDL and Creative Commons Attribution Share-Alike license (the free/open source licenses used for content by Wikia and WMF). Our goal is to produce material analogous to the dataset created by the FLOSSmole project (Howison et al., 2006), which have supported a large proportion of comparative research into peer production to date.

In pilot work on redirects, we have released tools and technical documentation for processing wiki datasets.[6] Because this is a much larger project, we will issue a preliminary release of tools, documentation, and data after two years of work and will announce this release on our blogs and several popular mailing lists for Internet and peer production research (e.g., AIR-l, wiki-research-l). We will solicit feedback, suggestions, and patches to prepare an updated version of the software, documentation, and dataset for release in the final year of the project. We will also place this version of the dataset in a public archive for scientific data (e.g., the Harvard Dataverse) and will seek to publish documentation for the data in a peer reviewed venue (e.g., *Nature Scientific Data* or *PLOS ONE*).

3. Dissemination Plan, Education, and Outreach

We will disseminate the outcomes of this project through scholarly communication channels including scientific workshops, conferences, and journals. We have a history of, and a commitment to, publication in venues including Computer-Supported Cooperative Work (CSCW), Human-Computer Interaction (CHI), the International Conference on Weblogs and Social Media (ICWSM), the International Symposium on Open Collaboration (OpenSym, formerly WikiSym), as well as peer-reviewed journals on communication and sociology (e.g., the *Journal of Communication*; *Information, Communication and Society*; *Politics and Society*; and *American Behavioral Scientist)*. Whenever possible, we will also release public and freely licensed versions of our research products. In the past, we have published in open access scientific journals (*PLOS ONE*), released work under open access licenses (e.g. Hill and Shaw, 2014), and published pre-print versions on our professional websites. As described in preceding sections of this proposal, we will also disseminate the tools and datasets used in our

[5] Available at: https://github.com/makoshark/wikiq

[6] http://communitydata.cc/wiki-redirects

research under free and open licenses. These activities will both ensure the reproducibility of our findings, and facilitate the widest dissemination and scientific impact of our work.

The PIs will both include methods and results from this research directly in the courses they teach at Northwestern University and the University of Washington. PI Shaw teaches an undergraduate course on online communities and graduate seminars on peer production and research design which will both incorporate the results of the projects proposed here. PI Hill also teaches a class on online communities and graduate courses on designing Internet research and data science with online community data. All of his teaching has a strong conceptual and practical engagement with peer production and he will incorporate findings, tools, and data into all three courses.

As both PIs have done in their previous research, the research team will work closely with organizations supporting peer production in ways that go beyond simply using these communities as a source of data. Both PIs are active contributors to Wikipedia and PI Hill has been a member of the Wikimedia Foundation advisory board since 2007. In collaboration with PI Shaw, Hill also delivers an annual talk on academic research at Wikipedia's yearly International conference, Wikimania. Both PIs have regularly given talks on their research at Wikimania, the Wikimedia Foundation, Wikia Inc., and WikiHow (another commercial wiki hosting firm). As part of their support for this project, WMF has committed to hosting talks on findings of this research for the WMF staff. Hill is also a founding board member of the Wikimedia Cascadia User Group, and a regular attendee and organizer of events related to Wikipedia in the Seattle area. PI Shaw serves on the Citizen Science Advisory Board of the Adler Planetarium in Chicago, where he assists in the design and assessment of the Zooniverse, another widely successful peer production platform. We will use these relationships with the peer production contributor and business communities to disseminate our research and to ensure that it is relevant, useful, and usable.

In this project, we will work particularly closely with the research, analytics, and development teams at the WMF. In support of this proposal, the WMF has offered to host both PIs as well as graduate student research assistants for multiple weeks in their San Francisco offices. Our relationship with WMF will ensure that we access to knowledge of the MediaWiki software (of which WMF is the primary author and will ensure that our work is integrated into the technical analytics and research infrastructure used by peer production and wiki researchers worldwide.

At both institutions, graduate students will be involved in the project as research assistants, collaborators, and advisees, and will interact with both PIs through research activities and monthly team meetings over video conference. These experiences will provide them with a supportive learning environment incorporating personnel from both Northwestern and the University of Washington.

4. Project Timeline

The timeline in Figure 4 describes our empirical research in terms of the three projects proposed in Section 2 as well as a supporting work in terms of software systems and dataset preparation.

5. Intellectual Merit

The **intellectual merit** of the proposed work stems from the way our research will will bridge the rich traditions of computing research and organization science and use comparative data from populations of peer production communities to test three of the most important theories explaining community success. In particular, we will test competing theories about the effect of small transaction costs in the form of required accounts, evaluate resource mobilization theory and ecological models of competition in the context of online communities, and consider the effect of socializing on

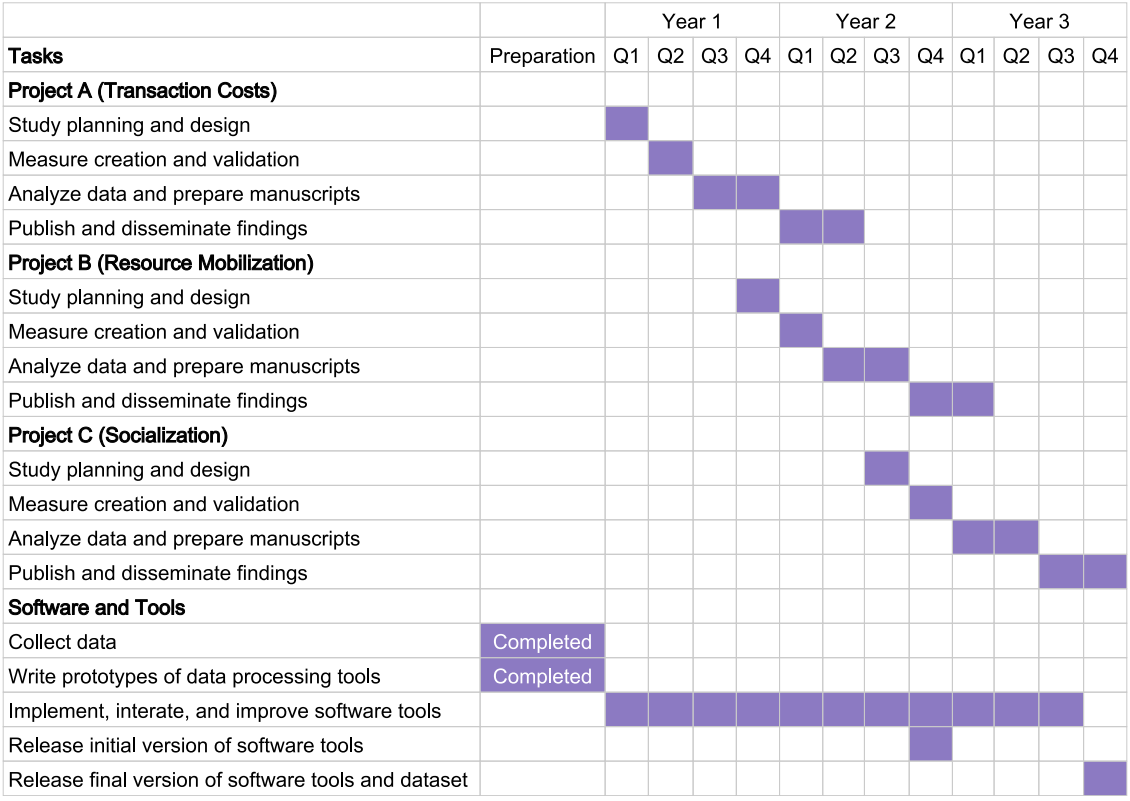


Figure 3: Project timeline broken up in the quarter system used by both Northwestern and UW. Q1 is the start of the academic year on the quarter system (i.e., Oct-Dec).

communal public good production. In doing so, we will offer tests of three of the most important theories from social computing, organization science, social movement research, and collective behavior in the context of a large population of online communities. Contrary to prior expectations, we anticipate finding that peer production will be robust to small increases in transaction costs and will overcome resource constraints experienced by traditional volunteer organizations. We also anticipate that social interaction and communication will play a key role in support the quality and quantity of contributions to peer production projects. Testing these theories will build toward a broader theory of community success. This work will facilitate subsequent comparisons with other kinds of organizations, including offline organizations that pre-date the rise of peer production.

6. Broader Impacts of the Proposed Work

The **broader impacts** of this work focus on the initiators and managers of virtual communities and the designers of community software and systems. A growing number of organizations rely on distributed and often ad-hoc teams whose works is organized into virtual communities. Moreover, a growing number of organizations rely on new online communities for content, quality assurance, support, and innovation. Testing three of the key theories of community success will have direct impact on firms building these communities. In particular, it will provide prescriptions on whether and when to require accounts, how to approach new communities on popular topics, and how to invest in technological systems for improving social communication around work products. Our goal is for these results to shape the decisions of firms, non-profits, and governments working with

social media and user generated content and to improve the experience of every day users of these systems. In this process, our project may help make virtual work more effective resulting in general economic benefits and improved satisfaction among knowledge workers.

7. PI Preparation and Preliminary Results

7.1. Personnel

Aaron Shaw, Ph.D. (PI), is an Assistant Professor in the Department of Communication Studies at Northwestern University and a Faculty Associate of the Berkman Center for Internet and Society at Harvard University. At Northwestern, he is also affiliated with the Sociology Department, the Institute for Policy Research, and the Buffett Center. His research has been supported by FUSE Labs at Microsoft Research, the Ford Foundation, the Ewing and Marion Kaufman Foundation, the Berkman Center for Internet and Society, the University of California Office of the President, and the United States Department of Education. Shaw holds graduate degrees in Humanities and Sociology from both Stanford University and UC Berkeley.

Shaw studies collective action, collaboration, and mobilization in peer production and crowds, and has published widely on these topics (e.g., Antin and Shaw, 2012; Hargittai and Shaw, 2013; Kittur et al., 2013; Shaw et al., 2011; Shaw and Benkler, 2012; Shaw, 2012). Shaw has received awards from the American Political Science Association, American Sociological Association, the International Communication Association, and the Association for Computer Machinery Conference on Computer Supported Cooperative Work (CSCW) for several papers and his dissertation research. He brings expertise in organizational research, peer production, quantitative and quasi-experimental methods, empirical analysis of online communities, and research design.

Benjamin Mako Hill, Ph.D. (PI), is an Assistant Professor of Communication at the University of Washington. He is also a faculty affiliate at the Berkman Center for Internet and Society and an affiliate at the Institute for Quantitative Social Science – both at Harvard University. Hill holds a Masters degree from the MIT Media Lab and a Ph.D. from MIT in Management and Media Arts and Science from an interdepartmental program overseen by HCI faculty at the MIT Media Lab and social science faculty at the MIT Sloan School of Management. He has published numerous articles in peer reviewed journals and conference proceedings (e.g., Buechley and Hill, 2010; Hill and Monroy-Hernandez, 2013a,b; Monroy-Hernández et al., 2011), and has received awards from the International Communication Association, the Association for Computing Machinery Conferences on Computer Supported Cooperative Work (CSCW) and Human Factors in Computing Systems (CHI), MTV, and Cisco.

Prior to his graduate studies, PI Hill worked full time as a software engineer and has received his masters degree from MIT for software development and HCI related research. Hill has a strong background in technology management, data-driven statistical analyses of online communities, computational research, management science, and peer production. In particular, Hill has been a leader, developer, and contributor to the free and open source software community for more than a decade as part of the Debian and Ubuntu projects, two of the most popular Linux distributions with millions of users worldwide and is the author of several best-selling technical books (e.g., Hill et al., 2008). He is a member of the Free Software Foundation board of directors, and is has been on the Advisory Board of the Wikimedia Foundation since 2007.

If funded, PI Hill will be assisted by a graduate research assistant at the University of Washington. PI Shaw is already assisted by graduate student (funded by Northwestern) who holds expertise in peer production, quantitative research methods, and computational analysis.

7.2. *Related Experience and Prior Work*

Both PIs possess advanced training in organizational research, statistical analysis, and quantitative methodology and have published peer reviewed research employing large-scale, empirical data analysis techniques. Over the past four years, we have been working collaboratively on research projects and pilot studies to build the experience and skills to successfully complete these projects. As collaborators, we have undertaken multiple research projects on peer production, resulting in three co-authored peer reviewed publications (Hill and Shaw, 2013; Shaw and Hill, 2014; Hill and Shaw, 2014) and multiple co-authored working papers and conference presentations (e.g., Hill et al., 2012; Shaw et al., 2014). In collaboration with Yochai Benkler at Harvard Law School, we have co-authored a book chapter for a forthcoming edited volume from MIT Press that lays out out the need for the kind of comparative organizational research described in this proposal (Benkler et al., 2015).

As discussed in Section 2.4, we have already collected the raw data necessary to complete the studies described in this proposal. Although analyzing these data in the way have described is challenging, we have the preparation to complete this research with the resources requested. Using the prototype tools we describe in Section 2.4, we have already created a dataset of the largest 1% of Wikia wikis which we have used to test several core theories of organizational democracy. The results of this research have been published in a 2014 issue of the *Journal of Communication* on "Big Data" research (Shaw and Hill, 2014). In another pilot project, we wrote software to analyze the full text history of English Wikipedia (the largest and most complicated project we will analyze) to create a comphensive longitudinal dataset of "redirects" and to use the dataset to show that conclusions from several of the most highly cited papers on peer production are subtantively changed when our dataset is taken into account (Hill and Shaw, 2014). We have published both our software and dataset alongside the substantive findings in the published manuscript.[7]

8. Results from Prior NSF Support

PI Hill has one active NSF award (Collaborative Research: New Pathways into Data Science: Extending the Scratch Programming Language to Enable Youth to Analyze and Visualize Their Own Learning; DRL-1417663; $124,374.00; September 1, 2014 – August 31, 2016) studying the impact of data and analytic tools on learning in the Scratch online community. The award is too recent to report results. PI Hill has no other prior NSF awards. PI Shaw has received no previous support from NSF.

[7] http://communitydata.cc/wiki-redirects/

References

Adler, B. T. (2012). *Wikitrust: Content-driven Reputation for the Wikipedia*. Ph.d. dissertation, University of California at Santa Cruz, Santa Cruz, CA, USA. AAI3521730.

Adler, B. T., Alfaro, L. d., Mola-Velasco, S. M., Rosso, P., and West, A. G. (2011). Wikipedia vandalism detection: Combining natural language, metadata, and reputation features. In Gelbukh, A., editor, *Computational Linguistics and Intelligent Text Processing*, number 6609 in Lecture Notes in Computer Science, pages 277–288. Springer Berlin Heidelberg.

Adler, B. T. and de Alfaro, L. (2007). A content-driven reputation system for the wikipedia. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 261–270, New York, NY, USA. ACM.

Antin, J., Cheshire, C., and Nov, O. (2012). Technology-mediated contributions: editing behaviors among new wikipedians. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 373–382, New York, NY, USA. ACM.

Antin, J. and Shaw, A. (2012). Social desirability bias and self-reports of motivation: A study of amazon mechanical turk in the US and india. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2925–2934, New York, NY, USA. ACM.

Arazy, O., Ortega, F., Nov, O., Yeo, L., and Balila, A. (2015). Determinants of wikipedia quality: the roles of global and local contribution inequality. In *Proceedings of the 2015 ACM conference on Computer supported cooperative work*, CSCW '15, pages 233–236, New York, NY, USA. ACM.

Arguello, J., Butler, B. S., Joyce, E., Kraut, R., Ling, K. S., Rosé, C., and Wang, X. (2006). Talk to me: Foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 959–968, New York, NY, USA. ACM.

Bao, P., Hecht, B., Carton, S., Quaderi, M., Horn, M., and Gergle, D. (2012). Omnipedia: bridging the wikipedia language gap. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 1075–1084, New York, NY, USA. ACM.

Benkler, Y. (2002). Coase's penguin, or, linux and the nature of the firm. *Yale Law Journal*, 112(3):369–446.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.

Benkler, Y., Shaw, A., and Hill, B. M. (2015). Peer production: A form of collective intelligence. In Malone, T. and Bernstein, M., editors, *Handbook of Collective Intelligence*. MIT Press, Cambridge, MA.

Bennett, W. L. and Segerberg, A. (2012). The logic of connective action. *Information, Communication & Society*, 15(5):739–768.

Bimber, B. A., Flanagin, A. J., and Stohl, C. (2012). *Collective Action in Organizations: Interaction and engagement in an Era of Technological Change*. Communication, society and politics. Cambridge University Press, New York.

Buechley, L. and Hill, B. M. (2010). LilyPad in the wild: how hardware's long tail is supporting new engineering and design communities. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, DIS '10, pages 199–207, New York, NY, USA. ACM.

Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, 12(4):346–362.

Butler, B. S., Joyce, E., and Pike, J. (2008). Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1101–1110, New York, NY, USA. ACM.

Cheshire, C. and Antin, J. (2008). The social psychological effects of feedback on the production of internet information pools. *Journal of Computer–Mediated Communication*, 13(3):705–727.

Crowston, K., Wei, K., Howison, J., and Wiggins, A. (2010). Free/libre open source software: What we know and what we do not know. *ACM Computing Surveys*, 44(2):2012.

Eisenhardt, K. M. (1989). Building theories from case study research. *The Academy of Management Review*, 14(4):532–550.

Erickson, T. and Kellogg, W. A. (2000). Social translucence: An approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1):59–83.

Friedman, E. J. and Resnick, P. (2001). The social cost of cheap pseudonyms. *Journal of Economics & Management Strategy*, 10(2):173–199.

Fulk, J., Flanagin, A. J., Kalman, M. E., Monge, P. R., and Ryan, T. (1996). Connective and communal public goods in interactive communication systems. *Communication Theory*, 6(1):60–87.

Fulk, J., Heino, R., Flanagin, A. J., Monge, P. R., and Bar, F. (2004). A test of the individual action model for organizational information commons. *Organization Science*, 15(5):569–585.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901.

Halfaker, A., Geiger, R. S., Morgan, J. T., and Riedl, J. (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):664–688.

Hannan, M. T. and Carroll, G. (1992). *Dynamics of organizational populations density, legitimation, and competition*. Oxford University Press, New York.

Hannan, M. T. and Freeman, J. (1977). The population ecology of organizations. *American Journal of Sociology*, 82(5):929–964.

Hargittai, E. and Shaw, A. (2013). Digitally savvy citizenship: The role of internet skills and engagement in young adults' political participation around the 2008 presidential election. *Journal of Broadcasting & Electronic Media*, 57(2):115–134.

Healy, K. and Schussman, A. (2003). The ecology of open-source software development.

Hill, B. M., Burger, C., Jesse, J., and Bacon, J. (2008). *Official Ubuntu Book*. Prentice Hall, 3 edition.

Hill, B. M. and Monroy-Hernandez, A. (2013a). The cost of collaboration for code and art: evidence from a remixing community. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 1035–1046, New York, NY, USA. ACM.

Hill, B. M. and Monroy-Hernandez, A. (2013b). The remixing dilemma the trade-off between generativity and originality. *American Behavioral Scientist*, 57(5):643–663.

Hill, B. M. and Shaw, A. (2013). The wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PLoS ONE*, 8(6):e65782.

Hill, B. M. and Shaw, A. (2014). Consider the redirect: A missing dimension of wikipedia research. In *Proceedings of The International Symposium on Open Collaboration*, OpenSym '14, pages 28:1–28:4, New York, NY, USA. ACM.

Hill, B. M., Shaw, A., and Benkler, Y. (2012). Status, social signalling and collective action in a peer production community. Working Paper.

Howison, J., Conklin, M., and Crowston, K. (2006). FLOSSmole. *International Journal of Information Technology and Web Engineering*, 1(3):17–26.

Jullien, N. (2012). What we know about wikipedia: A review of the literature analyzing the project(s). SSRN Scholarly Paper ID 2053597, Social Science Research Network, Rochester, NY.

Kellogg, K. (2009). Operating room: Relational spaces and micro-institutional change inside two surgical teaching hospitals. *American Journal of Sociology*, (Forthcoming).

Kittur, A. and Kraut, R. E. (2010). Beyond wikipedia: coordination and conflict in online production groups. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW)*, pages 215–224, Savannah, Georgia, USA. ACM.

Kittur, A., Lee, B., and Kraut, R. E. (2009). Coordination in collective intelligence: the role of team structure and task interdependence. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 1495–1504.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., and Horton, J. (2013). The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 1301–1318, New York, NY, USA. ACM.

Kraut, R. E. and Resnick, P. (2012). *Building Successful Online Communities: Evidence-Based Social Design*. The MIT Press.

Kriplean, T., Beschastnikh, I., and McDonald, D. W. (2008). Articulations of wikiwork: uncovering valued work in wikipedia through barnstars. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW2008)*, pages 47–56, San Diego, CA, USA. ACM.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Lipset, S. M., Trow, M. A., and Coleman, J. S. (1956). *Union democracy: The Internal Politics of the International Typographical Union*. Free Press, Glencoe, Ill.

Marwell, G. and Oliver, P. (1993). *The critical mass in collective action : a micro-social theory*. Cambridge University Press, Cambridge.

McCarthy, J. D. and Zald, M. N. (1977). Resource mobilization and social movements: A partial theory. *The American Journal of Sociology*, 82(6):1212–1241.

McCarthy, J. D. and Zald, M. N. (2001). The enduring vitality of the resource mobilization theory of social movements. In Turner, J. H., editor, *Handbook of Sociological Theory*, Handbooks of Sociology and Social Research, pages 533–565. Springer US. 10.1007/0-387-36274-6_25.

Meyer, D. S. and Whittier, N. (1994). Social movement spillover. *Social Problems*, 41(2):277–298.

Minkoff, D. C. (1997). The sequencing of social movements. *American Sociological Review*, 62(5):779–799.

Monroy-Hernández, A., Hill, B. M., Gonzalez-Rivero, J., and boyd, d. (2011). Computers can't give credit: How automatic attribution falls short in an online remixing community. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 3421–3430, New York, NY, USA. ACM.

Mugar, G., Østerlund, C., Hassman, K. D., Crowston, K., and Jackson, C. B. (2014). Planet hunters and seafloor explorers: Legitimate peripheral participation through practice proxies in online citizen science. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 109–119, New York, NY, USA. ACM.

Østerlund, C. and Crowston, K. (2011). What characterize documents that bridge boundaries compared to documents that do not? an exploratory study of documentation in FLOSS teams. In *2011 44th Hawaii International Conference on System Sciences (HICSS)*, pages 1–10.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001.

Preece, J. and Shneiderman, B. (2009). The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1):13–32.

Priedhorsky, R., Chen, J., Lam, S. T. K., Panciera, K., Terveen, L., and Riedl, J. (2007). Creating, destroying, and restoring value in wikipedia. In *Proceedings of the 2007 international ACM conference on Supporting group work*, GROUP '07, pages 259–268, New York, NY, USA. ACM.

Ransbotham, S. and Kane, G. C. (2011). Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in wikipedia. *MIS Quarterly-Management Information Systems*, 35(3):613.

Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation systems. *Communications of the ACM*, 43(12):45–48.

Schweik, C. M. and English, R. C. (2012). *Internet success: a study of open-source software commons*. MIT Press, Cambridge, Mass.

Scott, W. R. and Davis, G. F. (2006). *Organizations and organizing: Rational, natural and open systems perspectives*. Prentice Hall.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin, New York, NY.

Shaw, A. (2012). Centralized and decentralized gatekeeping in an open online collective. *Politics & Society*, 40(3):349–388.

Shaw, A. and Benkler, Y. (2012). A tale of two blogospheres: Discursive practices on the left and right. *American Behavioral Scientist*, 56(4):459–487.

Shaw, A. and Hill, B. M. (2014). Laboratories of oligarchy? how the iron law applies to peer production. *Journal of Communication*, 64(2):215–238.

Shaw, A., Zhang, H., Monroy-Hernández, A., Munson, S., Hill, B. M., Gerber, E., Kinnaird, P., and Minder, P. (2014). Computer supported collective action. *interactions*, 21(2):74–77.

Shaw, A. D., Horton, J. J., and Chen, D. L. (2011). Designing incentives for inexpert human raters. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, CSCW '11, pages 275–284, New York, NY, USA. ACM.

Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press, USA, 1 edition.

Soule, S. A. and King, B. G. (2008). Competition and resource partitioning in three social movement industries. *The American Journal of Sociology*, 113(6):1568–1610.

Steinfield, C. W., Markus, M. L., and Wigand, R. T. (2005). Exploring interorganizational systems at the industry level of analysis: evidence from the US home mortgage industry. *Journal of Information Technology*, 20(4):224–233.

Stodden, V. and Miguez, S. (2013). Best practices for computational science: Software infrastructure and environments for reproducible and extensible research. SSRN Scholarly Paper ID 2322276, Social Science Research Network, Rochester, NY.

von Krogh, G., Haefliger, S., Spaeth, S., and Wallin, M. W. (2012). Carrots and rainbows: Motivation and social practice in open source software development. *MIS Quarterly*, 36(2):649–676.

Wang, X., Butler, B. S., and Ren, Y. (2013). The impact of membership overlap on growth: An ecological competition view of online groups. *Organization Science*, 24(2):414–431.

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., and Smith, M. (2011). Finding social roles in wikipedia. In *Proceedings of the 2011 iConference*, iConference '11, pages 122–129, New York, NY, USA. ACM.

Zhu, H., Chen, J., Matthews, T., Pal, A., Badenes, H., and Kraut, R. E. (2014a). Selecting an effective niche: An ecological view of the success of online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 301–310, New York, NY, USA. ACM.

Zhu, H., Kraut, R., and Kittur, A. (2012). Organizing without formal organization: group identification, goal setting and social modeling in directing online production. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 935–944, New York, NY, USA. ACM.

Zhu, H., Kraut, R. E., and Kittur, A. (2014b). The impact of membership overlap on the survival of online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 281–290, New York, NY, USA. ACM.

Proceedings, The Fourth AAAI Conference on
Human Computation and Crowdsourcing

Studying the Effects of Task Notification Policies on Participation and Outcomes in On-the-Go Crowdsourcing

**Yongsung Kim, Emily Harburg, Shana Azria, Aaron Shaw,
Elizabeth Gerber, Darren Gergle, Haoqi Zhang**
Northwestern University, Evanston, IL USA
{yk, eharburg, shanaazria2016}@u.northwestern.edu, {aaronshaw, egerber, dgergle, hq}@northwestern.edu

Abstract

Recent years have seen the growth of physical crowdsourcing systems (e.g., Uber; TaskRabbit) that motivate large numbers of people to provide new and improved physical tasking and delivery services on-demand. In these systems, opportunistically relying on people to make convenient contributions may lead to incomplete solutions, while directing people to do inconvenient tasks requires high incentives. To increase people's willingness to participate and reduce the need to incentivize participation, we study *on-the-go crowdsourcing* as an alternative approach that suggests tasks along people's existing routes that are conveniently on their way. We explore as a first step in this paper the design of *task notification policies* that decide when, where, and to whom to suggest tasks. Situating our work in the context of practical problems such as package delivery and lost-and-found searches, we conducted controlled experiments that show how small changes in task notification policy can influence individual participation and actions in significant ways that in turn affect system outcomes. We discuss the implications of our findings on the design of future on-the-go crowdsourcing technologies.

Introduction

Recent years have seen the growth of physical crowdsourcing systems that motivate large numbers of people to provide new and improved physical tasking and delivery services on-demand. With a few button clicks, people can connect to workers who provide rides (Uber, Lyft), deliver groceries or meals (Instacart, Postmates, DoorDash), complete errands (TaskRabbit), and walk dogs (Wag). Enabled by network-connected mobile devices, mobile apps help connect people to tasks and lower barriers to participation, and market algorithms effectively coordinate the provision of physical resources on-demand. Together, these system components facilitate transactions and interactions to scale services to a broader market and worker population in ways that are transforming entire service sectors.

While physical crowdsourcing systems may change who can participate and what tasks can be done, applications are still limited by the fact that contributions are gathered via one of two approaches: *opportunistic* or *directed*. Opportunistic approaches rely on contributions from workers in situations where the workers themselves decide when and where to contribute. While convenient for workers, opportunistic tasking behaviors can lead to incomplete tasks or slow task completion (Han et al. 2015; Thebault-Spieker, Terveen, and Hecht 2015). As a result, commercial physical crowdsourcing systems mainly use a directed approach to assign tasks to workers that best address system needs. However, a directed approach often requires contributions that are largely outside workers' existing routines and as a result strong incentives are required (Thebault-Spieker, Terveen, and Hecht 2015). For example, a directed approach may lead to the assignment of tasks that require significant travel and incentives in order to increase workers' willingness to complete the task (Teodoro et al. 2014).

We study *on-the-go crowdsourcing* as a hybrid approach that seeks to bring about the best elements of both opportunistic and directed approaches. On-the-go crowdsourcing systems use people's existing routes to complete physical tasks that are conveniently on their route, but do so in ways that also address system needs.

We focus in this paper on understanding the challenges in designing effective *task notification policies* that determine when, where, and to whom to suggest tasks. Unlike directed approaches that assume workers are insensitive to when and where task assignments are sent so long as the tasks are generally nearby, we hypothesize that people contributing on-the-go may be quite sensitive to exactly when, where, and how task suggestions are presented. Task notification policies can thus affect individual participation and actions that in turn affect global tasking outcomes and can influence the success of an on-the-go crowdsourcing system.

To better understand such challenges, we design task notification policies to study 1) how slight changes in notification radius can affect individual participation and user recruitment, and 2) how slight changes in the timing of notifications can affect individual task actions and system outcomes. We conducted controlled experiments and the results show that small changes in task notification policies can affect individual participation and actions in significant ways that ultimately affect system outcomes. Specifically, we found that small changes in the notification radius had a drastic effect on user recruitment and user participation in package delivery settings. In addition, we found that small changes in the timing of notifications had a significant im-

Copyright © 2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

pact on individual actions, which in turn affected global search coverage in lost-and-found settings.

This paper makes the following contributions:

- We design and experimentally assess different task notification policies in order to understand how small policy changes can significantly affect participation and actions in ways that influence system outcomes.
- We identify and report on situational factors that affect on-the-go participation that are not prominent in existing physical crowdsourcing systems, such as *providing little or no travel detour*, *limiting the need to walk back*, and *assessing the likelihood of a user's hands availability*.
- We discuss the implications of our findings for the design of on-the-go crowdsourcing systems, and highlight the need for new design and algorithmic work to advance this promising new model for physical crowdsourcing.

Related work

Existing physical crowdsourcing systems rely on one of two approaches for soliciting contributions: directed and opportunistic. The directed approach routes tasks to workers who are either nearby or who fit the task request criteria. Many existing on-demand services use the directed approach to ensure task completion and task quality. However, the directed approach requires high incentive for inconvenient tasks, such as tasks that are at inconvenient locations (Teodoro et al. 2014) or that require large travel distances (Thebault-Spieker, Terveen, and Hecht 2015). Long travel distances not only deter workers from participating (Teodoro et al. 2014; Musthag and Ganesan 2013), but also prevent requesters who cannot afford the associated higher cost for the service. The opportunistic approach relies on workers to contribute as they wish, e.g., by browsing and selecting tasks from a list. Some community-driven systems, such as *time-banking* systems which support members of a community exchanging tasking services, use the opportunistic approach because it provides workers freedom and convenience to choose when to complete tasks and what tasks to complete. Since this relies solely on worker's task selection behavior, it often leads to task incompletion (Han et al. 2015) and can leave requesters whose tasks have not been completed unsatisfied with the service. On-the-go crowdsourcing seeks to bring about the best elements of both opportunistic and directed approaches to permit access to a large number of people who can contribute through their existing routines in ways that better achieve system goals.

Prior theoretical work highlights the possibility of coordinating physical crowd work to enable new applications by effectively leveraging people's existing routines (Sadilek, Krumm, and Horvitz 2013; Chen et al. 2014). Sadilek et al. introduced Crowdphysics and conducted empirical study with geo-tagged tweets to demonstrate that it may be possible to coordinate long-distance package delivery by leveraging people's existing routines with minimal diversion and short wait time (Sadilek, Krumm, and Horvitz 2013). Chen et al. introduced an approach for optimizing the sequence of physical tasks to assign to workers so as to minimize travel detours (Chen et al. 2014). While similarly focused on the idea of promoting convenient contributions, the methods in these prior works broadly assume that tasks will be accepted whenever assigned, which makes them a poor match for on-the-go crowdsourcing in which helpers may or may not perform tasks presented to them. We find through our studies on task notification policies that how we suggest tasks can significantly affect system outcomes, and present later in the paper the need for further technical work to support on-the-go crowdsourcing systems in practice.

Existing physical crowdsourcing applications that leverage people's existing routines focus primarily on data collection and rely on the opportunistic approach. One approach is to attach sensors onto people and objects that travel regular routes to passively collect sensing data, e.g., on bikes (Eisenman et al. 2007), street sweepers (Aoki et al. 2008), and public transportation (Aberer et al. 2010). This approach does not require active user involvement but can only collect data using machine sensors. Another approach is to opportunistically collect data from active user. For example, Tiramisu crowdsources bus information from bus riders to provide real-time arrival time (Zimmerman et al. 2011), and Twitch uses smartphone unlocks to collect coarse-grained census data (Vaish et al. 2014). This approach leverages human sensors, but are limited to the situations and locations in which people opportunistically contribute data. In contrast, our on-the-go approach actively suggests tasks to potential contributors using task notification policies that must balance the goals of promoting convenient and valuable contributions to advance individual and system goals.

When delivering tasks to potential helpers, task notification policies need not only consider interruptible moments in which people are more likely to accept notifications (Ho and Intille 2005; Fischer, Greenhalgh, and Benford 2011) but also situational factors that can significantly affect one's interest in contributing to a task. In this direction, our studies contribute a set of small but significant factors, or *channel factors* (Ross and Nisbett 2011), that affect participation in on-the-go crowdsourcing systems. Understanding these factors can help us to more accurately identify situations in which someone on-the-go may be able to help and be used to extend models that consider only distance-based measures for the cost of diversion (Horvitz and Krumm 2012).

Task Notification Policies for On-the-go Crowdsourcing

We envision on-the-go crowdsourcing systems that suggest tasks to potential helpers who happen to pass by task locations. These systems implement *task notification policies* that decide when, where, and to whom to suggest tasks in a way that balances individual convenience with the system's needs for timely, accurate, and complete solutions. To make decisions, a task notification policy may consider input factors such as a potential helper's location, their distance to tasks, their likely future routes, the urgency of tasks, the likelihood that others will become available in the near future, and so on. Based on this input, a policy specifies the conditions under which to notify a potential helper about a task opportunity. For example, a policy may specify a task radius

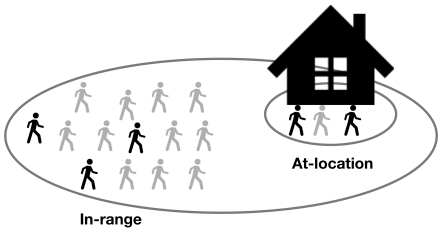


Figure 1: We study two task notification techniques, *in-range* and *at-location*, that differ only in the radius within which to notify potential helpers of task needs. In this illustration, black and grey colors respectively denote people's decision to help or not help when notified.

within which to notify people about tasks, which in turn affects the size and composition of the pool of helpers who receive notifications. A policy may also specify the precise timing of notifications, affecting when and where the helper sees the notification and in turn may influence their decisions on whether to contribute and which tasks they perform.

We describe below two core challenges in designing task notification policies with illustrative scenarios drawn from practical problems such as package delivery and lost-and-found searches. We introduce and compare a number of notification techniques to help to illustrate these challenges.

Balancing individual disruption and the quality-of-service

Scaling a physical tasking service with an on-the-go crowd requires balancing desired qualities of service with the disruption to individual routines in order to promote efficiency and ensure long-term community viability. Given a set of task requests and incomplete knowledge of who might be willing and able to help, a system must decide whom to suggest tasks to from among the pool of potential helpers who pass by a task location. Setting an aggressive notification policy that notifies many potential helpers can quickly identify those who agree to help but may disrupt numerous others who cannot help. While this may help satisfy demand right now, it risks notification blindness over time that will eventually reduce the pool of potential helpers. Setting a conservative notification policy that suggests tasks only to those who are deemed most likely to help may risk recruiting too few helpers, which can lead to task delays or incomplete tasks. Over time, this approach risks overburdening the best helpers and can lead to temporal burnout that causes the best helpers today to eventually exit the system for good.

Illustrative scenario: Community package delivery To illustrate the challenges of managing the supply of helpers and deciding whom to notify of task opportunities, consider a package delivery scenario where the goal is to leverage people's existing travel for picking up packages from a mail center and delivering them to others who could not pick up packages on their own. The system allows people to submit requests for pickup and then notifies potential helpers passing by to perform the pickup. As potential helpers approach the mail center, the system must decide which sets of helpers to notify, for example, based on how likely they are able to help and on whether deliveries are time critical.

In such a scenario, identifying an effective task notification policy is non-trivial. Using an aggressive task notification policy allows us to reach people who are able and willing to help, such as someone who was already going to the package center or someone who wasn't going to the center but is willing to help after a notification. However, at the same time, we are also sending notifications to many people who are not able and willing to help right now (e.g. they don't currently have physical ability to carry the packages or they are too far from the package center). If significant disruption continues, it may lead this group of people to eventually ignore notifications, and result in their leaving the pool of likely helpers in the future.

On the other hand, using a conservative task notification policy reaches people who are already going to or at a package center. This may lead to higher percentage of pick-ups since it's likely to reach people who are able and willing to help, and would alleviate issues of notifying people who can't or don't want to help right now. However, it also misses opportunities to notify likely helpers who would have helped if we notified them, and as a result we may overly rely on people who often go to package center and cause them to burn out eventually.

In-range and at-location notification techniques To better understand the challenges in setting task notification policies to balance the goals of reaching willing helpers and disrupting others, we introduce and compare *in-range* and *at-location* notification techniques to study how notification radius might affect user recruitment and individual participation (see Figure 1). In-range notifications suggest tasks to potential helpers who pass by regions that are within a certain range of the task location. At-location notifications monitor a user's distance to the task location and only notifies them of a task when they are at the location. As the notification radius shrinks, we expect to reach fewer potential helpers but see that those receiving notifications complete a higher percentage of tasks suggested to them. In our first experiment, we use these two notification techniques in the context of package delivery to study users' willingness to pick up packages and overall user recruitment with regard to small distance changes.

Promoting contributions where they are most needed

Another core challenge in setting task notification policies is deciding when to suggest tasks. We assume that on-the-go crowdsourcing should minimally disrupt people's routines, and not require the attention of potential helpers until a task request is made. The timing of task requests may affect where helpers complete tasks, which tasks are completed, and which tasks are skipped. Contributions may be valued more or less depending on task's importance and time-criticality, as well as how many people can complete the task and are likely to come across it. A task notification policy risks either (1) sending a notification too early, which

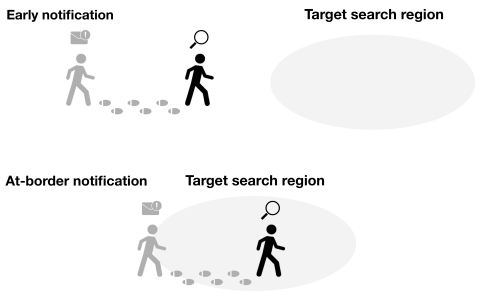


Figure 2: We study two task notification techniques, *early notification* and *at-border notification*, that differ only in the timing of when notifications are sent to helpers to alert them of a task need.

can lead to helpers completing a low-valued task when their routes would have led them to a nearby higher-valued task; or (2) passing on an opportunity to send a task or sending a task too late, which may lead to lost opportunities.

Illustrative scenario: Lost and found searches To illustrate the challenges of determining when to suggest tasks to potential helpers, we consider a lost & found scenario where the goal is to effectively coordinate the search for lost items with people who can help look for lost items along their existing routes. A requester who has lost an item marks the region(s) where it may have been lost, and the system notifies potential helpers who pass by the marked region. As a helper walks through a region, the system must decide when to notify the helper with the goal of steering their attention toward areas that most require their search efforts.

A general challenge in timing notifications is to avoid wasting search efforts while at the same time avoiding missed search opportunities. For example, if we notify helpers too early before they reach the target search region, the helpers may search for lost items in an unintended region and thus their searches will be wasted. If we notify people right at the beginning of the search region, by the time they respond to the notification and start looking for the item, they might have already passed some parts of the target search region, thereby leaving some areas unsearched. In either case, individual search efforts affect overall search coverage, and the timing of notifications thus affect search coverage globally and the overall efficiency of the system.

Pre-tracking and early notification techniques Many existing systems rely on *geofencing* for location-based notifications that trigger notifications anywhere around 100 meters from a centroid of a geo-fence (Rodriguez Garzon and Deva 2014). We posit that geofencing is inadequate for setting task notification policies because it is limited in its precision and ability to steer users to start searching at precise locations. To overcome this, we introduce *pretracking* as a general method for precisely notifying helpers in a more timely manner. Pre-tracking triggers fine-grained GPS monitoring once a potential helper approaches a region of interest. Upon entering the region, it continuously makes decisions based on precise tracking of a helper's location and uses this information to time when to deliver tasks.

Using a pre-tracking technique, we design two notification techniques, *early notification* and *at-border notification*, to study how the timing of notifications affects individual actions and global system outcomes. The early notification technique sends notifications ahead of time, by taking into account factors such as the helpers' interaction time with notifications and their walking rate, so that they might start a task at the beginning of a target region by the time they decide to help. The at-border notification technique sends notifications right at the border of the target region where the system wants the helper to start performing the task. In our second experiment, in the context of lost & found, we study how slight changes in the timing of notifications can affect where helpers' searches are made in ways that affect the value of their individual searches and overall search outcomes.

Controlled Experiments

As a first step to explore the challenges mentioned in the previous section, we designed two controlled experiments to study the challenges in 1) balancing individual participation and user recruitment, and 2) promoting convenient contributions that are most valued. We will also discuss on-the-go situational factors we discovered in the experiments.

Experiment 1: The effect of notification radius on user recruitment and participation

Methods We conducted a two-week long, within-subjects experiment for a package delivery scenario in order to understand 1) how sensitive user recruitment and participation are to notification radius changes, and 2) situational factors influence helping behaviors. We measured the number of notifications sent for user recruitment and the task pickup rate for user participation. We used a within-subject design to limit the effect that large differences in participant's routines could have on our findings; some participants may frequent task location areas more than others, which could unduly influence the number of notifications being sent.

For the experiment, we developed a prototype of Libero, a mobile application that collects package delivery requests and routes them to potential helpers. In our setting, since requesters receive package notifications from the package center through their email, they can either forward the email to the system or take a screenshot of the email if they want to request that their package be picked up. First, a helper receives package pick-up request notifications (Figure 3a) as they pass by the package center. Next, if the helper decides to pick up the package, she is redirected to "Others' Requests" where she sees a list of the packages with package details (Figure 3b). To pick up a package, the helper selects the package she wants to pickup and, after confirming the pickup, she is provided with the image of the requester's package notification email. As this occurs, the requester whose package was just picked up, receives an email and an in-app notification updating her of the status of her request, revealing the helper's name as well as the date and time of pickup.

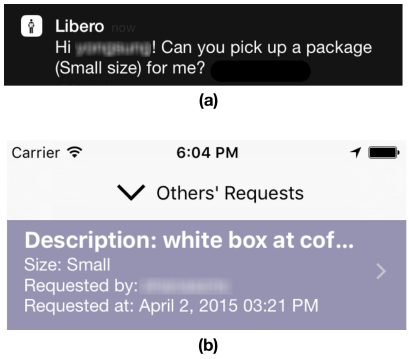


Figure 3: (a) Libero sends notifications to people who pass by a package center; (b) potential helpers can see the details of the package.

We used two notification conditions: *at-location notifications* that notify users when they enter a task location, and *in-range notifications* that notify users when they pass by a region that is within 100 meters of the task location. To implement at-location notification, we placed a bluetooth low energy beacon with the broadcasting signal power of 4dB (which translates into 1-3 meters in distance) and 1200ms as the interval. To implement in-range notification, we used a geo-fence with 100 meters as the radius.

We recruited 16 people who had iPhone 5 or above via flyers and local university mailing lists. The average participant age was 25 (SD: 3) with 9 male and 7 female participants. They were randomly assigned to one notification condition and asked to switch to the other condition after one week. The order of conditions for each person was counterbalanced. One participant stopped participating during the second week of the study due to personal reason, so we excluded that participant's data from the quantitative analysis.

In both conditions, when the participants accepted a request notification they were asked to deliver packages from a specified pick-up location and take them to a specified drop-off location. We chose a coffee shop that is close to the nearest train station from a school building where many classes take place as a pick-up location, and set up a drop-off location in a building that is next to the school building. We chose these two locations as pick-up and drop-off locations because the participants frequently visit places that are nearby and as a result they would not have to deviate much from their travel routine.

An author served as a requester during the experiment. Since we did not want the package size to affect willingness to help, we only requested packages that were small enough to be carried by one hand.

The participants were also asked to rate statements related to the perceived cost of disruption at two times during the study: once after the first week and another time after the second week. We also interviewed participants who delivered packages at least once in order to better understand what were the factors affecting their willingness to pick up. The study lasted for two weeks and the participants received a $25 gift card as compensation.

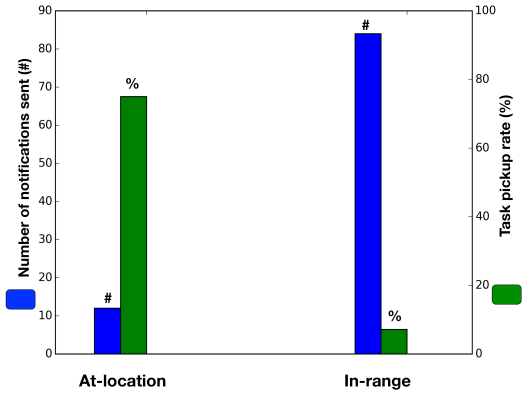


Figure 4: Comparison between at-location and in-range condition in terms of number of notifications sent and task pickup rate.

Results Our results showed that even the small distance changes that occurred across our two notification conditions had a drastic and inverse effect on both user recruitment and user participation (Figure 4). There were 7 times more opportunities presented in the in-range condition than in the at-location condition, with 84 notifications sent to the participants in the in-range condition, while only 12 notifications were sent in the at-location condition. On average, the participants in the in-range condition received 5.6 notifications (SD: 4.37), while the participants in the at-location received 0.8 notifications (SD: 0.8). A Wilcoxon Signed-Ranks test shows that there is a significant difference between the two conditions ($Z=3.49$, $p<0.001$).

We saw a significant effect in the reverse direction for task pickup rate, which was almost 10 times lower in the in-range notification condition than the at-location condition with an average of 7.13% (SD: 14.79) task pickup rate in the in-range condition and the average of 75% (SD: 28.87) in the at-location condition. Wilcoxon Signed-Ranks test results show that there is a significant difference between two condition ($Z=3.01$, $p=0.003$).

Although there is a significant difference in task pickup rate and number of notifications sent, we found that participants' perceived cost of disruption was low in both conditions. The participants' response on a 5-point likert scale asking about the perceived cost of disruption had a mean response of 1.8 (SD: 0.77) for the at-location condition and 2.33 (SD: 1.11) for the in-range condition (1 indicates not disruptive at all while 5 indicates very disruptive), and there was no significant difference between the conditions ($Z=1.29$, $p=0.2$).

While it is limited to our experiment, our results show the potential for leveraging on-the-go crowds to deliver packages in which people are willing and able to pick up tasks when they are suggested at the right time. During the study, there were 13 packages delivered by 5 different participants. Three participants delivered packages in both conditions while 2 participants only delivered in one condition (one in

at-location, one in in-range). On average, the participants received 0.79 notifications (SD:0.56) per day.

Discussion Our exploration with in-range and at-location notifications showed that small changes in distance for task notification policy resulted in significant opposing effects on individual participation and user recruitment. It resulted in high task pickup rate and low numbers in user recruitment for the at-location condition, while resulting in low task pickup rate and large numbers in user recruitment for the in-range condition.

The results highlight the trade-offs of setting task notification policies to achieve desired outcomes; they do not to show that either notifying at-location or in-range is better than the other. Instead, our results show that small changes in task notification policies can affect individual participation in significant ways. It should also be noted that the significance of such a shift may vary depending on who the users are. For example, some people may be less sensitive to task distance, while others would be more sensitive to it.

As on-the-go crowdsourcing systems designers, we have to consider the trade-offs in possibly overburdening a small subset of people who are more likely to help with disrupting larger crowds who might not be able to help. In either case, there is a potential risk of losing helpers in the long run due to burnout or disruptions. Since the cost of disruption was considered low at the two ends of the distance spectrum where we were able to find potential helpers, there might exist a "goldilocks" zone where we can get enough potential helpers and low enough disruptions to mitigate overburdening or disrupting participants at both the individual and aggregate levels.

Limitations Due to the small sample size, we used a within-subject experimental design that may lead to fatigue and learning bias. Regarding the potential fatigue issue in a within-in subject design, we saw that two people who helped in the first week did not help in the second week. There was one dropout in each of the counterbalanced conditions (i.e., one in the in-range, the other in the at-location). For the other three participants, two people picked up more packages and one picked up same amount of packages compared to first week. Based on these empirical results, it appears unlikely that fatigue was an issue that influenced our results.

We also used the task pick-up rate as a measure for individual participation and the number of notifications sent for user recruitment, which served as an intermediary measure for system outcomes. In the future, we may conduct a large-scale, between subject, longitudinal study to explore actual number of pickups and timeliness of delivery to better measure individual participation and overall system outcomes, and we can further assess user attrition rate to gain insight into possible burnout that is caused by too much disruption.

Unlike other physical tasking services which use monetary incentives, we did not use payment per task as an incentive mechanism since it would be a confounding factor for understanding how people's participation varies as its convenience (i.e. distance to task locations) varies.

Another potential limitation of this experiment is the simulated package center, a coffee shop in a region where our participants frequent, which might have influenced people's willingness to help. Some people reported feeling awkward about picking up packages and not purchasing a coffee. In the future, we plan to conduct a field deployment with an actual college package center in order to eliminate such a deterrent to participation.

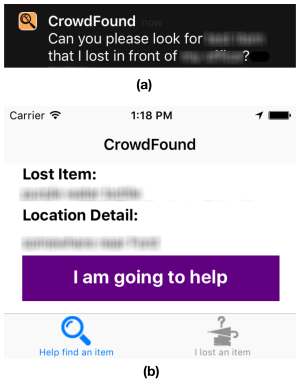


Figure 5: (a) CrowdFound sends notification to people who pass by possible lost item location; (b) the details of the lost item are also provided to the helpers.

Experiment 2: The effect of notification timing on most valued contributions

Methods We conducted a two-week long, within-subjects experiment in order to understand 1) how the particular timing of a notification can affect an individual's actions and the global system outcomes, and 2) to help reveal what the situational factors that affect an individual's search behavior.

To facilitate the experiment, we developed a prototype of CrowdFound, a mobile application where users can post lost item search requests that then notifies people who pass by possible lost item locations. A user who lost an item can post a request by providing a lost item type, a detailed description of the item, an approximate location where she believes she lost the item on a map, additional location details, and a picture of the item if she has one. When a potential helper is in the vicinity of the tagged location of a lost item, she receives a notification (Figure 5a) asking if she can help look for the missing item. Once she clicks the notification, she is shown the relevant description of the item (Figure 5b). If the user decides to help, she clicks "I am going to help" and a 30 second countdown is shown on the screen. After 30 seconds, a thank-you message is shown, along with the option to either select: "I found the item" or "I couldn't find it". If the user clicks "I found the item", an e-mail is sent to the requester informing her that her lost item has been found and connecting her to the helper so that she may retrieve her item. For this experiment, we set the default search time as 30 seconds and the diameter of the search region as 40 meters.

We used two notification conditions: in the *early notification* condition we notified the users 20 meters in advance of the target search region, and for the *at-border* condition we notified users immediately upon entering to the region. We chose 20 meters based on a calculation that added the mean

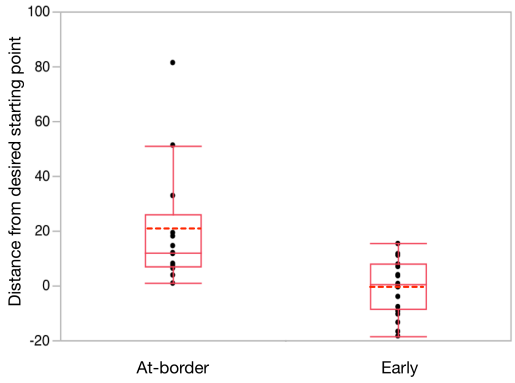


Figure 6: Distance from where they were supposed to start searching and where the participants actually searched in at-border and early notification. Means are shown as dotted lines and medians as solid lines. The box whiskers indicate range including outliers.

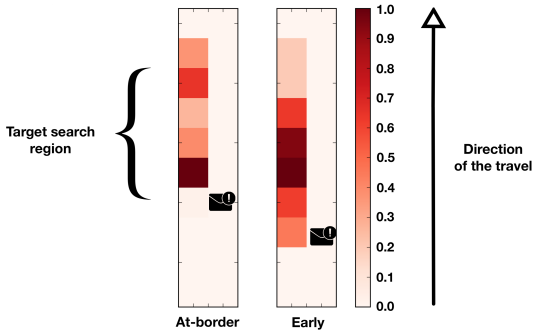


Figure 7: A heatmap shows the relative search distribution in each condition when participants were heading north in 30 seconds time window. In at-border condition people missed some of the areas at the beginning of the region, while in early notification some of the searches were wasted outside of the intended search region.

interaction time (7 seconds) for reading and opening notifications (numbers derived from a pilot study) to the mean time it takes for people to read the descriptions in the lost item description view (6 seconds, as derived from a usability testing). We then multiplied this sum by the average walking rate (1.4m/s) to set the early notification distance.

We recruited 15 people (8 male and 7 female) who had iPhone 5 mobile devices or above from undergraduate class and convenient samples. The participants' age ranged from 19 to 34 (mean: 23.33, SD: 3.64). They were randomly assigned to one condition and then switched to the second condition after one week. The order of conditions for each person was counterbalanced.

One of the authors served as a requester during the experiment, and the participants were asked to find lost items in both conditions. Each participant received 8 requests from among 5 different types of items (item types were based on actual posts on Northwestern University's lost&found Facebook group). The items ranged in size, but we tried to avoid bulky items which might be too obvious to spot. Descriptions of the item locations used similar prepositional phrases (e.g. somewhere, near, in front of) to those found in common descriptions posted on the lost&found group.

After the experiment, the participants were asked to rate statements on a 5-point likert scale about the perceived cost of disruption. Since we believe that there will be no difference in terms of the number of notifications sent between the two conditions and its effect on the perceived cost of disruption, we only sent out one post-study survey at the conclusion of the experiment. We also performed post-interviews with participants who searched for lost items and asked them about the times they helped or didn't help, how they responded to the notifications, what were their search strategies, and what features might have helped them to search.

In our analysis, we use distance from the actual search area to the intended search point as a measure of individual behavior. We use coverage of the search area as a way to measure system outcomes. Wasted efforts mean that the searches happened outside of the target search region, and missed opportunities means that helpers were in the target search region but did not cover some areas.

Results Our results show that small changes in notification timing have a significant effect on where individual searches occurred and they also affect the aggregate area covered by all participants.

Figure 6 shows the distance from where the participants started searching to the border of the target search region where they were supposed to search. The distance was closer to the desired search border in the early notification condition than in the at-border notification condition.[1] Wilcoxon test shows that there is a significant difference ($Z=3.4$, $p<.001$) between the at-border (mean: 20.55 meters, SD: 22.70) and the early notification condition (mean: -0.42 meters, SD: 10.18). This indicates that, on average, by the time the helpers searched in the at-border condition they had already passed almost half of the desired search region, while those in the early notification condition started 0.42 meters before the border of the desired search region.

Locations where the individual helper started searching also influenced the areas they covered within the region. In the at-border notification condition the participants missed some of the desired search areas located near the beginning of the target area, while in the early notification condition some people started searching too early and outside of the desired target region. Figure 7 shows the relative search distribution surrounding the target search region in each condition when participants were heading on a northerly trajectory. We could not clearly see the difference in the relative coverage distribution between the conditions when the participants were heading south, since there were only five searches in the at-border condition in total and we had to discard two searches due to GPS inaccuracy.

[1]We excluded 4 searches in which app open time exceeded 50 seconds from when they received notifications.

An individual's on-the-go interaction with the mobile devices also influenced their participation. Most of the people would open the app via notifications and read the details of lost items as they were walking. It took 21.62 seconds (SD: 15.12) on average from the time the notification was received to the time the participants started helping, and they walked an average of 22.41 (SD:18.75) meters. However, a few participants would stop and read the notifications because they were afraid that they might miss the region they were supposed to search and didn't want to circle back.

Although limited to our experiment, our results show that it was possible to find lost items with relatively little effort and little disruption on the part of the on-the-go helpers. At least six items (out of eight) were found by six different participants. It took 15.38 hours (SD: 17.22) on average from the time the item was posted until it was found. It also took 10.65 hours (SD: 11.31) on average to find an item since the time the first notification was sent to a helper. The helpers searched 38 times out of the 68 notifications sent to them, resulting in a 55.88% task acceptance rate. Also, the mean rating for the perceived cost of disruption on 5-point likert scale was 2.4 (SD: 0.91), indicating that the notifications were not disruptive (1 indicates not disruptive at all while 5 indicates very disruptive).

Discussion Our experiment with early and at-border notifications shows that small changes in the timing of a notification can have significant effects on values of individual actions and overall system outcomes. While early notifications permitted people to start their search for lost items without missing areas at the beginning, some of their searches happened prior to the intended search region. On the other hand, in the at-border condition, some people missed areas at the beginning of the region and as a result those areas were never covered.

Limitations Although in our experiment we only focused on individual searches within a single small region, in a real world setting the size of a lost item region may be large and it will require systems to break down a large search region into smaller regions to solicit help. In such cases, at a system-level, we need to reason about helper's future routes, value of asking help in a certain region, and coordinate searches across the regions in order to minimize the wasted efforts or missed opportunities to accomplish desired system outcomes. That being said, in the future, we could conduct a large-scale study with multiple lost item locations and a larger number of participants.

On-the-go situational factors

We now discuss some of the situational factors that influenced a helper's willingness to contribute. We identified on-the-go situational factors such as *little or no travel detour, having to walk back, and the availability of the worker's hands* that are not prominent in existing physical crowdsourcing systems.

Existing regular route, having to walk back, missing the right moment In lost and found searches, most of the people tended to search for items along their current routes and they mainly covered areas that correspond to those routes, and they were unwilling to walk back in the opposite direction of their destination. As one participant said: *"It just seemed more natural to look, or not natural, convenient for me to look in the direction I was already walking especially. Because it's usually when I am walking I have a destination so it doesn't make much sense to me to turn around and look in the other way I just came. It makes more sense to make that sweep when I am already walking that way."*

We found similar results in our package delivery settings. Participants mentioned that sometimes they missed the notifications or checked the notifications too late and didn't want to walk back and pick up packages. One participant said: *"Sometimes I couldn't feel the notification, and sometimes I checked the notification after passed the [task location], then I was like 'oh well...' and I kept walking."*

Factors like having to walk back served as a channel factor for participants, in which situations they were dissuaded to participate.

Hands Availability One of the main reasons that the helpers didn't pick up packages was that they couldn't carry the package. One participant mentioned that he had to bring both his lunchbox and a coffee so he couldn't deliver the package. Another participant said: *"I did not have hands to carry [so I didn't pick up]. Also one time, I had to put a package in my backpack [to help]..."* This finding showed that for physical tasks such as delivery it is important for systems to understand whether or not the potential helpers have enough hands to do the task.

Personal situation, time availability and temporal preference Participants' schedules and mood also influenced their willingness to pick up items. Sometimes they were in a hurry or they just didn't feel like doing it. One participant stated: *"It also depends on my schedule and mood. If I have a meeting with my advisor and get nervous, I wouldn't pick up even if I am there. If there is something important or you are nervous, you don't care about other things."*

In lost and found searches, the common reason for not being able to help was the lack of time. Participants were not willing to help or even pull out their phone to check the notification when they were rushing to classes or meetings. One participant said: *"Most of the time if I am rushing for class [and] I know I am already late, so I wouldn't check my phone."*

Although it differs individually, people have different time periods when they are not willing to help for various reasons. For instance, one participant stated that he never searched after class during lunch time, because he was hungry and didn't feel like helping. Another participant was not willing to search in the morning because there were many people heading to classes and work, and he didn't want to stop and block the crowds in the busy morning: *"I am always rushing to work in the morning...There are a lot of people walking, [so] it will be weird if I just stop and look around."*

Being with other people We found that whether or not participants were alone or with other people influenced their willingness to help. For some people, being with others mo-

tivated them to participate together: *"This was the time when I was looking for the gloves...I just looked around like the same route, but my friend who was more curious and went to the grass area...He looked around a little bit here and there, and he gave up...It wasn't about helping, it was more like a game that I should find something. It was very amusing to me, and it was a challenge, and I really wanted find this."* In contrast, some participants felt it was odd or socially unacceptable to search for items when they were with other people. One participant said: *"If I am walking with other people then I wouldn't, this is kind of weird."*

Weather conditions Current weather condition influenced not only participants' willingness to search but also their search time and search regions. Many participants didn't even want to check their phone because it was too cold on some of the days during the experiment period. Even if they were willing to help despite the weather condition, people's search behavior changed: *"Oh, one time it rained and this part [lawn area] was really muddy, and I didn't wanna go look in there. So I just kept looking on the sidewalk while I walked, but I didn't like go search for a minute."*

Discussion We found that situational factors such as time availability, the timing of the task, weather conditions, and a convenient physical location were all important situational factors, and this result is in line with previous findings in mobile crowdsourcing services (Teodoro et al. 2014). We also found that there are other on-the-go specific factors such as having to walk back or re-trace one's prior route, taking a detour that is in the opposite direction of the destination, and having one's hands available, are important situational factors for on-the-go helpers. Extending the work from Horvitz and Krumm (Horvitz and Krumm 2012), in which distance was taken into account for calculating the cost of diversion in opportunistic routing, in on-the-go settings we might account for individual differences and the situational factors identified in this study when calculating the cost of diversion and incorporate it into task suggestion decisions.

Design Implications

Our research seeks to broadly realize the vision of leveraging people's planned travel to create previously unavailable opportunities for physical crowdsourcing over large geographic areas (Sadilek, Krumm, and Horvitz 2013). In this paper we studied the challenges in designing task notification policies for on-the-go crowdsourcing systems. From two controlled experiments in package delivery and lost-and-found settings, we found that small changes in notification radius and timing can have a significant effect on individual participation and actions that in turn affect global outcomes. In this section, we discuss the implications of our findings on the design of future on-the-go crowdsourcing technologies and applications.

Technical Advancements

Focusing on a general population of people on-the-go necessarily implies that helpers may or may not perform tasks presented to them at any given moment. Ensuring that system goals are achieved thus requires managing the constantly changing levels of availability, attention, and interests of people on the move. This motivates the need to plan and execute actions *opportunistically* based on available mobility resources (Horvitz, Koch, and Subramani 2007; Kamar, Horvitz, and Meek 2008), and to make adjustments over the course of problem solving as partial solutions are made available (Malone and Crowston 1994). Future work can fill a void in the existing literature on task planning and routing (Hinds 2002; Shahaf and Horvitz 2010; Wooldridge 2001; Grosz and Kraus 1996) by providing flexible frameworks that will enable on-the-go crowds to achieve a wide range of tasks and objectives.

Since the precise means for supporting system goals with an on-the-go crowd is often unknown a priori, coordinating complex physical behaviors will rely upon dynamic monitoring and flexible utilization of resources. First, we need models for monitoring and predicting the supply of potential helpers built not only on our understanding of people's mobility routines but also their willingness to take on a task in a given situation, taking into consideration the costs of diversion and disruption as well as any channel factors that affect participation. Second, we need decision-theoretic frameworks that use such models to optimize and adjust task notification policies to best support desired individual, community, and system outcomes.

To balance individual disruption and the quality of service, we need a *supply management framework* for on-the-go crowdsourcing that can simultaneously reason over people, tasks, and time to opportunistically govern who to send tasks to based on needs and the situation on the ground. Based on the urgency of tasks and the availability of helpers, such a framework can help us reason about which tasks to suggest to whom so as to maintain a healthy pool of helpers by avoiding unnecessarily disrupting or overburdening helpers, while also being sensitive to system goals to ensure the timely completion of tasks.

To promote contributions where they are most needed, we need a *hit-or-wait framework* that can reason over tasking opportunities a person can encounter on their route and devise policies that decide whether to *hit* (i.e. send a task request) or to *wait* (i.e. to hold off on a request) for a better opportunity. Based on the importance of tasks and people's likelihood of reaching them and helping, such a framework can help us make informed decisions about when to suggest tasks to potential helpers en-route so as to best leverage individual efforts in ways that advance system goals.

Design Opportunities

On-the-go crowdsourcing can open up new ways for people to help others in their communities and neighborhoods. By surfacing task needs in situations where potential helpers are likely able and ready to help (e.g., in package pickup), on-the-go crowdsourcing can facilitate more connections between helpers and requesters to bring significant benefits through minimal efforts. By coordinating opportunistic contributions from willing helpers each contributing a small part to completing a larger task (e.g., in lost-and-found), on-the-

go crowdsourcing can lower the barrier to participation and connect requesters to a larger population of helpers. Tapping into a large crowd of potential helpers and facilitating convenient help should make existing community support systems more effective and efficient. By scaling the efforts of an engaged community, on-the-go crowdsourcing can also spur new community-based services designed to help people in need (e.g., the elderly, people with disabilities, new moms, etc) that may otherwise be difficult to support.

On-the-go crowdsourcing can also open up new ways of providing commercial physical tasking services. First, on-the-go crowdsourcing can enable more flexible ways of working that fit into workers' existing routines and mobility. For example, this can allow a driver to pickup and drop off a passenger on their way to picking up a child from school. Second, on-the-go crowdsourcing systems can reduce travel costs to make affordable services that would otherwise have been prohibitively expensive to scale. This can enable new services that, by effectively coordinating convenient contributions from on-the-go workers, can provide a quality of service that matches or exceeds existing models but at just a fraction of the cost.

In order to realize these and other design opportunities, future work should seek to better understand the motivations required for people to contribute on-the-go and design solutions that address practical issues such as establishing trust. For example, there is a need to better understand how physical tasks can conveniently fit into people's routines in ways that enrich their lives. There is also a need for designs that assure requesters that tasks will be completed well and on-time, given a service that is supported with an on-the-go crowd instead of workers who respond on-demand. We look forward to designing new on-the-go crowdsourcing systems that build on people's motivations and that address such practical challenges in future work.

Acknowledgement

We thank members of the Design, Technology, and Research program and the Delta Lab for their valuable feedback and helpful discussions. This work was funded by National Science Foundation grant #1618096, a Microsoft Research FUSE Labs Award, and a Segal Design Cluster fellowship.

References

Aberer, K.; Sathe, S.; Chakraborty, D.; Martinoli, A.; Barrenetxea, G.; Faltings, B.; and Thiele, L. 2010. OpenSense: Open Community Driven Sensing of Environment. In *SIGSPATIAL International Workshop on GeoStreaming*.

Aoki, P. M.; Honicky, R. J.; Mainwaring, A.; Myers, C.; Paulos, E.; Subramanian, S.; and Woodruff, A. 2008. Common Sense: Mobile Environmental Sensing Platforms to Support Community. In *Ubicomp*.

Chen, C.; Cheng, S.-F.; Gunawan, A.; Misra, A.; Dasgupta, K.; and Chander, D. 2014. TRACCS: A Framework for Trajectory-Aware Coordinated Urban Crowd-Sourcing. In *HCOMP*.

Eisenman, S. B.; Miluzzo, E.; Lane, N. D.; Peterson, R. A.; Ahn, G.-S.; and Campbell, A. T. 2007. The BikeNet Mobile Sensing System for Cyclist Experience Mapping. In *SenSys*.

Fischer, J. E.; Greenhalgh, C.; and Benford, S. 2011. Investigating Episodes of Mobile Phone Activity As Indicators of Opportune Moments to Deliver Notifications. In *MobileHCI*.

Grosz, B. J., and Kraus, S. 1996. Collaborative plans for complex group action. *Artificial Intelligence* 86(2):269–357.

Han, K.; Shih, P. C.; Bellotti, V.; and Carroll, J. M. 2015. It's Time There Was an App for That Too:: A Usability Study of Mobile Timebanking. *IJMHCI* 7(2):1–22.

Hinds, P. 2002. *Distributed work*. MIT Press.

Ho, J., and Intille, S. S. 2005. Using Context-aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. In *CHI*.

Horvitz, E., and Krumm, J. 2012. Some help on the way: Opportunistic routing under uncertainty. In *Ubicomp*.

Horvitz, E.; Koch, P.; and Subramani, M. 2007. Mobile opportunistic planning: methods and models. In *User Modeling*.

Kamar, E.; Horvitz, E.; and Meek, C. 2008. Mobile opportunistic commerce: mechanisms, architecture, and application. In *AAMAS*.

Malone, T. W., and Crowston, K. 1994. The interdisciplinary study of coordination. *ACM Computing Surveys (CSUR)* 26(1):87–119.

Musthag, M., and Ganesan, D. 2013. Labor Dynamics in a Mobile Micro-task Market. In *CHI*.

Rodriguez Garzon, S., and Deva, B. 2014. Geofencing 2.0: Taking Location-based Notifications to the Next Level. In *Ubicomp*.

Ross, L., and Nisbett, R. E. 2011. *The person and the situation: Perspectives of social psychology*. Pinter & Martin Publishers.

Sadilek, A.; Krumm, J.; and Horvitz, E. 2013. Crowdphysics: Planned and Opportunistic Crowdsourcing for Physical Tasks. In *ICWSM*.

Shahaf, D., and Horvitz, E. 2010. Generalized task markets for human and machine computation. In *AAAI*.

Teodoro, R.; Ozturk, P.; Naaman, M.; Mason, W.; and Lindqvist, J. 2014. The Motivations and Experiences of the On-demand Mobile Workforce. In *CSCW*.

Thebault-Spieker, J.; Terveen, L. G.; and Hecht, B. 2015. Avoiding the South Side and the Suburbs: The Geography of Mobile Crowdsourcing Markets. In *CSCW*.

Vaish, R.; Wyngarden, K.; Chen, J.; Cheung, B.; and Bernstein, M. S. 2014. Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time. In *CHI*.

Wooldridge, M. J. 2001. *Multi-agent systems: an introduction*. Wiley Chichester.

Zimmerman, J.; Tomasic, A.; Garrod, C.; Yoo, D.; Hiruncharoenvate, C.; Aziz, R.; Thiruvengadam, N. R.; Huang, Y.; and Steinfeld, A. 2011. Field trial of tiramisu: crowdsourcing bus arrival times to spur co-design. In *CHI*.

Starting Online Communities: Motivations and Goals of Wiki Founders

Jeremy Foote
Northwestern University
Evanston, IL
jdfoote@u.northwestern.edu

Darren Gergle
Northwestern University
Evanston, IL
dgergle@northwestern.edu

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

ABSTRACT
Why do people start new online communities? Previous research has studied what helps communities to grow and what motivates contributors, but the reasons that people create new communities in the first place remain unclear. We present the results of a survey of over 300 founders of new communities on the online wiki hosting site Wikia.com. We analyze the motivations and goals of wiki creators, finding that founders have diverse reasons for starting wikis and diverse ways of defining their success. Many founders see their communities as occupying narrow topics, and neither seek nor expect a large group of contributors. We also find that founders with differing goals approach community building differently. We argue that community platform designers can create interfaces that support the diverse goals of founders more effectively.

ACM Classification Keywords
H.5.3 Information Interfaces and Presentation (e.g. HCI): Group and Organization Interfaces—Computer-supported cooperative work

Author Keywords
Online communities; motivation; wikis; survey; peer production

INTRODUCTION
The people who found new organizations are in a unique position to exert influence on the way that the organization develops. For example, research on firms has shown that the attributes of a company's founder, such as their education, industry experience, and position in a social network, can influence the firm's growth and survival [3, 12]. Founders can also influence the fate of online communities. Communities which have active, well-connected, experienced founders are more likely to experience continued activity [8].

Prior research on firm and community founders makes an implicit assumption: that all founders are hoping for their organizations to grow and survive. This may be a reasonable assumption for entrepreneurs, but online community founders are likely to have more varied motivations and project goals. Learning more about founders' motivations and perspectives on community success is important for understanding how this new form of organizaing works.

To learn more about why founders start communities, we surveyed users of the online community hosting site Wikia.com immediately after they founded a new wiki. We focus on founders' *motivations* for founding new wikis and their *project goals*. Motivations encompass more immediate incentives in comparison to project goals, which are more forward-looking and relate to how founders will evaluate the success of their community.

While some founders seek to build large, productive online communities, only 18% of our respondents rank content growth as the top goal. The most common primary goal is the creation of high-quality information. Overall, motivations and goals are very diverse, but many point to the creation of communities which are modest in scope and in aim. Wikis are often created on a whim, with nearly half of respondents reporting that they had considered starting their wiki for only a few hours or less before founding it. The primary goal that a founder has for a project is related to how and whether they engage in community building activities, suggesting a path whereby community outcomes may relate to founder goals.

Our results suggest that online community researchers and community platform designers can do more to understand and support attempts to build smaller and shorter-term communities, and that focusing on growth or longevity creates a limited view of success in online communities.

BACKGROUND
Prior research on motivations in online communities has focused primarily on *contributors*, finding that online community participants have heterogeneous motivations. For example, Wikipedia editors have motivations like enjoyment, learning, and ideology [11, 13]. Similarly diverse motivations have been found for contributors to open source software [6] and other online communities (e.g., [10]). We expect that community founders also have diverse motivations, but lack empirical evidence of what those motivations are.

Existing research also suggests diversity in how community members assess whether an existing community is successful. For example, open source programmers use many different measures to determine success, such as user satisfaction, developer satisfaction, code quality, and popularity [4, 14]. Just as

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
CHI 2017, May 6-11, 2017, Denver, CO, USA.
ACM ISBN 978-1-4503-4655-9/17/05.
http://dx.doi.org/10.1145/3025453.3025639

contributors see current success differently, we anticipate that founders will have different understandings of future success.

We might also expect to see diversity in founder goals due to different levels of experience on a specific platform. As users become more familiar with a tool or ecosystem of projects (e.g., Wikipedia, Wikia, GitHub), they may recognize opportunities for community and interaction that are not visible at first [2]. As users gain experience and their perspective of the existing projects in a particular domain changes, their goals for new communities may also change.

Differences in the motivations and goals of founders may also lead to different community outcomes. For example, the way participants in a community understand a project helps explain their subsequent levels of participation [1, 15], and greater founder activity is associated with community survival [8].

Founders' approach to community building presents one mechanism that could produce divergent outcomes. The people in charge of a community can take practical steps to help the community grow and be productive, such as attracting newcomers, encouraging contributions, and regulating bad behavior [9]. The way that founders apply these tactics (or whether they apply them at all) may depend on the goals they have for the community. For example, founders who create a community as a place to communicate with a small group of friends will be unlikely to take steps to attract newcomers, but may have plans for encouraging contributions.

METHOD

To better understand community founder motivations, we conducted a survey of founders on the Wikia platform. We chose Wikia as it is a widely used platform for the creation of topic-based wikis. Users on Wikia can start new wikis about any topic they choose, and users are free to edit any wikis on the site. Wikis represent a distinct type of online community, based on the creation and maintenance of a shared knowledge artifact. There are, of course, many other types of online communities, whose founders may have a different set of motivations and goals.

We used an online survey to gather data from founders because it facilitated the collection of a broad range and large number of responses. We worked with representatives at Wikia to distribute the survey to founders of new communities. After users initiate and complete the new wiki creation process, an automated message is posted on the user's "Message Wall" (a threaded discussion interface for user-to-user interactions). By default, messages posted on a user's Message Wall are also emailed to them. During our study, an invitation to participate in our online survey was included at the bottom of this automated message. The data included in this paper is from surveys taken between 5 April and 31 August 2016. By surveying founders immediately after they created a new wiki, we capture the motivations and goals they had at the time of founding, before any community outcomes could distort their memory of events.

The survey consisted of 70 items, including items about demographic background, Internet use skills[1], and past experience on Wikia and other online communities. Our items about motivations included an open-ended question ("What made you want to start this wiki?") and a set of thirteen Likert-scale items based on the contributor motivation literature[2]. To discover goals, we asked participants to think about how they would determine whether their wiki is successful, and to rank the importance of seven success metrics. These included both instrumental and terminal goals, which were drawn from prior literature (e.g., [4, 14]) and expanded to incorporate additional community-oriented and interaction-oriented goals[3]. For both motivations and goals, we allowed respondents to enter free text "Other" responses. The first author reviewed the text responses and selected characteristic examples that we report below. Finally, to determine a founder's plans for community building, we asked them whether they would implement community building strategies suggested by [9].

RESULTS

During the period of data collection, 46,828 total new wikis were created by 35,749 users. A total of 720 respondents started the survey. We removed 91 respondents who failed our attentiveness check, were obviously fraudulent[4], or who reported being minors. This left us with 629 respondents. Respondents were allowed to skip items they did not want to answer. For example, 312 founders filled out the motivation section and 254 ranked their goals. In order to maximize the number of participants included in each analysis, we used responses that were complete for that portion, even if the respondent skipped other portions. We report the *N* for each analysis in the table captions.

Founding Motivations

To simplify the analysis of motivations, we reduced the dimensionality of our thirteen motivation items. Factor analysis suggested that there were five dimensions[5]: spreading information and building community, problems with existing wikis, fun and learning, creating and publicizing personal content, and the "Other" item. We summarize the distribution of responses in Table 2. The factor representing the desire to build a community and to spread information about a topic has the highest overall popularity, although each of the motivations was identified as a primary motivation by a number of respondents.

[1] We used a modified version of the 10 item instrument developed by [5].

[2] Items were ranked from 1 ("Not a motivation") to 5 ("A primary motivation"). If a respondent rated more than half of the motivation items then we imputed "Not a motivation" for any missing items.

[3] We started with seven project goals (see Table 1), and collapsed two items representing facets of community and two measures representing facets of growth into one measure each.

[4] For example, those who reported having been born before 1915.

[5] A KMO test [7] showed that exploratory factor analysis was appropriate (MSA=.69), and a scree plot suggested four factors was optimal. All of the items loaded onto at least one factor with a cut-off of .3, with the exception of the "Other" item, which we treat as its own factor. For cross-loaded items, we included the item in the primary factor, all of which loaded at .4 or higher. Combined, the factors explained 34% of the variance.

Topic	Survey Items
Motivation	"What made you want to start this wiki?" "On a scale of 1-5, how much did each of these motivations contribute to your decision to start the wiki? 'Playing around / Learning how the software works,' 'As a joke,' 'A place to organize personal material,' 'To communicate with friends,' 'To provide publicity for myself or my company,' 'A new product was released that I was interested in,' 'There wasn't an existing wiki about the topic,' 'Exisiting wikis about the topic were of poor quality,' 'I wanted to spread information about the topic,' 'I disagreed with how existing wikis were being run,' 'I wanted to build a community,' 'I wanted to be more involved in governing the conversation around this topic,' 'Other (please explain)' "
Project Goals	"Think about how you will assess whether or not your wiki is successful, and rank which of the following are the most important measures of success for you: 'A large number of contributors', 'A large amount of information about the topic,' 'High quality information about the topic,' 'A community that remains active for a long time,' 'A highly active community,' 'Meeting new people,' 'Other (please explain)' "

Table 1. Key survey topics and abbreviated questions

	Mean	SD	Primary Motivation
Info and Community	3.12	1.04	70
Bad Existing Wikis	1.89	1.28	28
Fun and Learning	2.32	0.96	15
Personal Material	2.71	1.26	43
Other	1.84	1.59	60

Table 2. Distribution of motivations. The final column shows how many respondents chose that as a primary motivation (mean rating > 4), N=312.

While some additional motivations appeared in the free-text responses, none of these was reported by more than a few users. This suggests that our original items represent typical motivations of respondents. That said, some of the new motivations identified are interesting. A few respondents wrote about teaching others about a topic. For example, one wrote, "I am a huge fan of the books and I wanted to create a wiki where other fans can read about and learn about the series." Another expressed an interest in gathering feedback: "I run a gaming community that is currently in development, I want users to be [...] able to contribute and add information about [...] our game." Another reported using the wiki to organize and store content: "[I] wanted a place to organize content as [my] web serial was written." Many examples are inconsistent with the goal of building a large and active community with many contributors.

Motivations need not be particularly strong because the barrier to founding a new community is so low within a platform like Wikia. Among our respondents, only 32% had planned to start a wiki for a few weeks or longer, while 46% of respondents reported that they had only thought about it for "a few hours" or "a few minutes" beforehand. Founding new communities appears to often be impulsive and not the result of delibration or careful planning.

Project Goals

Next, we look at the goals that founders have for their projects. Each of the project goals was ranked as the top goal by a substantial proportion of respondents. The creation of high-quality information was the most prevalent top goal, with 47% of respondents selecting that option. The next two most popular goals, "community" and "growth", were nominated as the top goal by 20% and 18% of respondents, respectively.

While many studies focus on the longevity and activity of a community as markers of success, this finding suggests that many wiki founders care most about creating a high-quality repository of information rather than community growth or other outcomes.

Indeed, most founders do not expect that their projects will become large and popular. When we asked founders about the potential audience for their projects, 50% chose the response that "only a handful of people" would be interested in the topic of their wiki. When we asked how many contributors they expected in the first 30 days, the median response was 4.

One additional goal emerged from the "Other" free text responses. Multiple respondents expressed the desire for external usefulness and popularity, independent of community-building. For example, one respondent's goal was, "Seeing the material actually used by [...] groups and individuals." Another's was, "People that actually benefit from my information." Various other goals were expressed by one or two respondents, such as strengthening relationships.

Comparing Founders with Different Goals

We next explore whether there are differences between founders with different project goals. We group respondents based on their top-ranked goal and compare those who selected each of the two most common top goals ("Information Quality" and "Community").

Figure 1 compares how motivations differ between those who chose each of the two top goals. t-tests at $\alpha < .05$ indicate that those whose top goal was community-based ("a community that remains active for a long time" or "a highly active community") were more likely to be motivated by fun and learning motivations and "other" motivations than those whose top goal was "high quality information about the topic."[6]

These differences in goals may result from pre-existing differences between founders, such as demographic attributes or experiences. For example, perhaps long-term users value community more since they are more likely to perceive the role of community on a site [2]. However, we performed χ^2 tests and found that community-oriented founders do not differ from quality-oriented founders by gender ($p = 0.36$), employment status ($p = 0.56$), how often they visit Wikia ($p = 0.21$), or how long they have been active on the site ($p = 0.5$). A t-test of founders' self-reported technical knowledge also showed no significant difference between the groups ($p = 0.1$). Based

[6]The differences in the "Other" motivation in Figure 1 and the "Recurit Members" strategy in Figure 2 were only marginally significant after applying a Holm-Bonferroni correction and should be interpreted with caution.

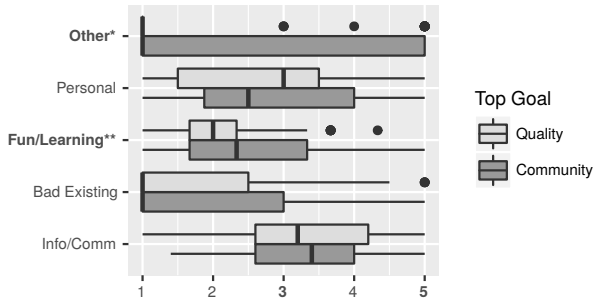


Figure 1. Differences in the motivations for respondents on a scale of 1 ("Not a motivation") to 5 ("A primary motivation"), based on their top goal for the wiki, N=248. * p < .05, ** p < .01

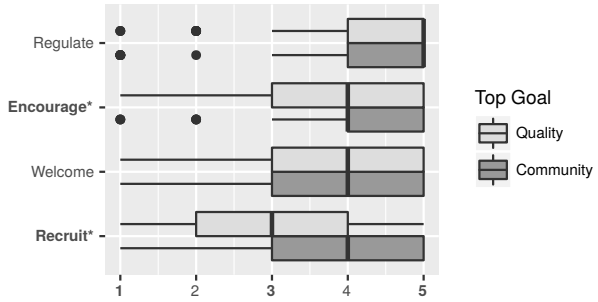


Figure 2. Differences in whether a respondent plans to implement each community building strategy, on a scale of 1 ("Definitely not") to 5 ("Definitely yes"), based on their primary goal for the wiki, N=251. * p < .05

on these basic measures, we find no relationship between a founder's attributes or experience and her project goals.

We also investigate whether having different goals changes the way that founders approach managing their community, finding mixed results. Prior research has shown that groups whose founders add content to the site and who are engaged and active are more likely to survive [8]. We asked founders about the time they were planning to devote to their wiki, but a χ^2 test shows no difference between community-oriented and quality-oriented founders in the ratio of those planning to dedicate at least 3 hours per week (p = 1).

However, we do find differences in the specific plans founders have for community building, encouragement, and regulation [9]. As with motivations, we use t-tests to compare the responses based on whether a founder's top goal is community or information quality. The results appear in Figure 2.

Those who rate community their top priority are more likely to have plans to recruit members and to encourage contributions. Both groups are equally likely to plan to welcome new members and regulate behavior.

DISCUSSION

The results of our survey provide evidence that wiki founders have varied motivations for starting their communities and varied goals for what those communities will eventually become. Even along dimensions that many founders care about (e.g., information quality and community activity), founders differ in how they prioritize these goals. While some respondents sought to build large, popular communities, we learn that most founded their wikis with very little foresight or planning and with modest goals. Instead of aspiring to lead and influence the masses, many founders seek to collect or distribute information about niche topics among a small group.

We suggest that the prevailing focus on community growth and longevity obscures the importance of small, niche communities. Small communities are not necessarily failures or cautionary tales. They may be meeting their founders' goals and deserve to be studied on their own terms. On Wikia, where communities are built around a shared, public artifact, many founders see creating that artifact as more important than building a community. Other community platforms like reddit or Facebook Groups have their own affordances and their founders likely have a different distribution of motivations and goals. However, growth-based measures of success are unlikely to fully represent founders' objectives in any of these venues.

By learning more about the founders of communities, platform designers gain insight into how people aspire to use a system and can better support those goals. For example, designers could add dashboards to measure and display the metrics of success that founders care most about. They could also build tools to support those goals. In the case of Wikia, we find a number of founders use wikis to create and distribute personal material. Wikia might enhance the user page interface or develop templates to facilitate this behavior.

LIMITATIONS & FUTURE DIRECTIONS

Future studies can assess the generalizability of these findings on Wikia and other platforms. Running a similar study across a set of different online community platforms would allow us to learn whether founders on other platforms have similar motivations and goals. Also, as with any survey, non-response may bias our results. However, we anticipate that a more representative sample would strengthen our key findings about the modest goals of Wikia founders, since the least dedicated founders would presumably be less likely to take the survey.

This study indicates that researchers and designers can do more to attend to founders. We examined how founders' motivations and goals relate to their plans for community building, but future research should also analyze how motivations and goals relate to the actions founders take. In this way, different motivations and goals may explain variations in community outcomes. On the other hand, future research may find that founders have limited influence and communities can transcend their founders' intentions as well as their actions.

ACKNOWLEDGEMENTS & ACCESS TO DATA

The authors would like to thank Trevor Bolliger at Wikia for his help deploying the survey. We are also grateful for the comments and suggestions of our reviewers. Financial support for this work came from the National Science Foundation (grant IIS-1617468) and Northwestern University. A replication dataset has been archived in the Harvard Dataverse and is available at http://dx.doi.org/10.7910/DVN/YG9IID.

REFERENCES

1. Judd Antin, Coye Cheshire, and Oded Nov. 2012. Technology-mediated contributions: editing behaviors among new wikipedians. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 373–382. DOI:http://dx.doi.org/10.1145/2145204.2145264
2. Susan L. Bryant, Andrea Forte, and Amy Bruckman. 2005. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work (GROUP '05)*. ACM, New York, NY, USA, 1–10. DOI: http://dx.doi.org/10.1145/1099203.1099205
3. Arnold C. Cooper, F. Javier Gimeno-Gascon, and Carolyn Y. Woo. 1994. Initial human and financial capital as predictors of new venture performance. *Journal of Business Venturing* 9, 5 (Sept. 1994), 371–395. DOI: http://dx.doi.org/10.1016/0883-9026(94)90013-2
4. Kevin Crowston, Hala Annabi, and James Howison. 2003. Defining Open Source Software Project Success. In *ICIS 2003 Proceedings*. Seattle, Washington, USA, 327–340. http://aisel.aisnet.org/icis2003/28
5. Eszter Hargittai and Yuli Patrick Hsieh. 2012. Succinct Survey Measures of Web-Use Skills. *Social Science Computer Review* 30, 1 (Feb. 2012), 95–107. DOI: http://dx.doi.org/10.1177/0894439310397146
6. Guido Hertel, Sven Niedner, and Stefanie Herrmann. 2003. Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. *Research policy* 32, 7 (2003), 1159–1177. http://www.sciencedirect.com/science/article/pii/S0048733303000477
7. Henry F. Kaiser and John Rice. 1974. Little Jiffy, Mark IV. *Educational and Psychological Measurement* 34, 1 (1974), 111–117. DOI: http://dx.doi.org/10.1177/001316447403400115
8. Robert E. Kraut and Andrew T. Fiore. 2014. The Role of Founders in Building Online Groups. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, Baltimore, Maryland, USA, 722–732. DOI: http://dx.doi.org/10.1145/2531602.2531648
9. Robert E Kraut, Paul Resnick, and Sara Kiesler. 2011. *Building successful online communities evidence-based social design*. MIT Press, Cambridge, Mass. http://site.ebrary.com/id/10539250
10. Cliff Lampe, Rick Wash, Alcides Velasquez, and Elif Ozkaya. 2010. Motivations to participate in online communities. In *Proceedings of the 28th international conference on Human factors in computing systems*. ACM, Atlanta, Georgia, USA, 1927–1936. DOI: http://dx.doi.org/10.1145/1753326.1753616
11. Oded Nov. 2007. What motivates Wikipedians? *Commun. ACM* 50, 11 (2007), 60–64. DOI: http://dx.doi.org/10.1145/1297797.1297798
12. Tone A. Ostgaard and Sue Birley. 1996. New venture growth and personal networks. *Journal of Business Research* 36, 1 (May 1996), 37–50. DOI: http://dx.doi.org/10.1016/0148-2963(95)00161-1
13. Joachim Schroer and Guido Hertel. 2009. Voluntary Engagement in an Open Web-Based Encyclopedia: Wikipedians and Why They Do It. *Media Psychology* 12 (Jan. 2009), 96–120(25). DOI: http://dx.doi.org/10.1080/15213260802669466
14. Charles M. Schweik and Robert C. English. 2012. *Internet Success: A Study of Open-source Software Commons*. MIT Press, Cambridge, MA.
15. Georg von Krogh, Stefan Haefliger, Sebastian Spaeth, and Martin W. Wallin. 2012. Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development. *MIS Quarterly* 36, 2 (June 2012), 649–676.

THIS IS THE FINAL AUTHORS' VERSION OF A PUBLISHED BOOK CHAPTER. WHEN CITING THIS WORK, PLEASE CITE THE PUBLISHED VERSION: Foote, Jeremy D., Aaron Shaw, and Benjamin Mako Hill. 2017. "A Computational Analysis of Social Media Scholarship." In *The SAGE Handbook of Social Media*, edited by Jean Burgess, Alice Marwick, and Thomas Poell, 111–34. London, UK: SAGE.

A Computational Analysis of Social Media Scholarship

Jeremy Foote (jdfoote@u.northwestern.edu)

Aaron Shaw (aaronshaw@northwestern.edu)

Benjamin Mako Hill (makohill@uw.edu)

Abstract Data from social media platforms and online communities have fueled the growth of computational social science. In this chapter, we use computational analysis to characterize the state of research on social media and demonstrate the utility of such methods. First, we discuss how to obtain datasets from the APIs published by many social media platforms. Then, we perform some of the most widely used computational analyses on a dataset of social media scholarship we extract from the Scopus bibliographic database's API. We apply three methods: network analysis, topic modeling using latent Dirichlet allocation, and statistical prediction using machine learning. For each technique, we explain the method and demonstrate how it can be used to draw insights from our dataset. Our analyses reveal overlapping scholarly communities studying social media. We find that early social media research applied social network analysis and quantitative methods, but the most cited and influential work has come from marketing and medical research. We also find that publication venue and, to a lesser degree, textual features of papers explain the largest variation in incoming citations. We conclude with some consideration of the limitations of computational research and future directions.

INTRODUCTION

The combination of large-scale trace data generated through social media with a series of advances in computing and statistics have enabled the growth of

'computational social science' (Lazer et al., 2009). This turn presents an unprecedented opportunity for researchers who can now test social theories using massive datasets of fine-grained, unobtrusively collected behavioral data. In this chapter, we aim to introduce non-technical readers to the promise of these computational social science techniques by applying three of the most common approaches to a bibliographic dataset of social media scholarship. We use our analyses as a context for discussing the benefits of each approach as well as some of the common pitfalls and dangers of computational approaches.

The chapter walks through the entire process of computational analysis, beginning with data collection. We explain how we gather a large-scale dataset about social media research from the *Scopus* website's application programming interface. The dataset we collect contains metadata about every article in the Scopus database that includes the term 'social media' in its title, abstract, or keywords. Using this dataset, we perform multiple computational analyses. First, we use network analysis (Wasserman & Faust, 1994) on article citation metadata to understand the structure of references between the articles. Second, we use topic models (Blei, 2012), an unsupervised natural language processing technique, to describe the distribution of topics within the sample of articles included in our analysis. Third, we perform statistical prediction (James, Witten, Hastie, & Tibshirani, 2013) in order to understand what characteristics of articles best predict subsequent citations. For each analysis, we describe the method we use in detail and discuss some of its benefits and limitations.

Our results reveal several patterns in social media scholarship. Bibliometric network data reveals disparities in the degree that disciplines cite each other and illustrate that marketing and medical research each enjoy surprisingly large influence. Through descriptive analysis and topic modeling, we find evidence of the early influence of social network research. When we use papers' characteristics to predict which work gets cited, we find that publication venues and linguistic features provide the most explanatory power.

In carrying out our work in this chapter, we seek to exemplify several current best practices in computational research. We use data collected in a manner consistent with the expectations of privacy and access held by the subjects of our analysis as well as the publishers of the data source. We also make our analysis fully reproducible from start to finish. In an online supplement, we provide the full source code for all aspects of this project – from the beginning of data collection to the creation of the figures and the chapter text itself – as

a resource for future researchers.

Collecting and describing data from the web

A major part of computational research consists of obtaining data, preparing it for analysis, and generating initial descriptions that can help guide subsequent inquiry. Social media datasets vary in how they make it into researchers' hands. There are several sources of social media data which are provided in a form that is pre-processed and ready for analysis. For example, The Stanford Large Network Dataset Collection (Leskovec & Krevl, 2014) contains pre-formatted and processed data from a variety of social media platforms. Typically, prepared datasets come formatted as 'flat files' such as comma-separated value (CSV) tables, which many types of statistical software and programming tools can import directly.

More typically, researchers retrieve data directly from social media platforms or other web-based sources. These 'primary' sources provide more extensive, dynamic, and up-to-date datasets, but also require much more work to prepare the data for analysis. Typically, researchers retrieve these data from social media sites through application programming interfaces (APIs). Web sites and platforms use APIs to provide programmers with limited access to their servers and databases. Unfortunately, APIs are rarely designed with research in mind and are often inconvenient and limited for social scientists as a result. For example, Twitter's search API returns a small, non-random sample of tweets by default (what a user might want to read), rather than all of the tweets that match a given query (what a researcher building a sample would want). In addition, APIs typically limit how much data they will provide for each query and how many queries can be submitted within a given time period.

APIs provide raw data in formats like XML or JSON, which are poorly suited to most data analysis tasks. As a result, researchers must take the intermediate step of converting data into more appropriate formats and structures. Typically, researchers must also construct measures from the raw data, such as user-level statistics (e.g., number of retweets) or metadata (e.g., post length). A number of tools, such as NodeXL (Hansen, Shneiderman, & Smith, 2010), exist to make the process of obtaining and preparing digital trace data easier. However, off-the-shelf tools tend to come with their own limitations and, in our experience, gathering data amenable to computational analysis usually involves some programming work.

Compared to some traditional forms of data collection, obtaining and preparing social media data has high initial costs. It frequently involves writing and debugging custom software, reading documentation about APIs, learning new software libraries, and testing datasets for completeness and accuracy. However, computational methods scale very well and gathering additional data often simply means expanding the date range in a program . Contrast this with interviews, surveys, or experiments, where recruitment is often labor-intensive, expensive, and slow. Such scalability, paired with the massive participation on many social media platforms, can support the collection of very large samples.

Our application: The Scopus Bibliographic Database

We used a series of Scopus Bibliographic Database APIs to retrieve data about all of the publications in their database that contained the phrase 'social media' in their abstract, title, or keywords. We used the Python programming language to write custom software to download this data. First, we wrote a program to query the Scopus Search API to retrieve a list of the articles that matched our criteria. We stored the resulting list of 23,131 articles in a file. We used this list of articles as input to a second program, which used the Scopus Citations Overview API to retrieve metadata about all of the articles that cited these 23,131 articles. Finally, we wrote a third program that used the Scopus Abstract Retrieval API to download abstracts and additional metadata about the original 23,131 articles. Due to rate limits and the process of trial and error involved in writing, testing, and debugging these custom programs, it took a few weeks to obtain the complete dataset.

Like many social media APIs, the Scopus APIs returns data in JSON format. Although not suitable for analysis without processing, we stored this JSON data in the form it was given to us. Retaining the 'raw' data as it was provided by APIs allows researchers to construct new measures they might not have believed were relevant in the early stages of their research and to fix any bugs that they find in their data processing and reduction code without having to re-download raw data. Once we obtained the raw data, we wrote additional Python scripts to turn the downloaded JSON files into CSV tables which could be imported into Python and R, the programming languages we used to complete our analyses.

Results

The Scopus dataset contains a wide variety of data, and many different descriptive statistics could speak to various research questions. Here we present a sample of the sorts of summary data that computational researchers might explore. We begin by looking at where social media research is produced. Table 6.1 shows the number of papers produced by authors located in each of the six most frequently seen countries in our dataset.[1] We can immediately see that the English-language world produces much of the research on social media (which is perhaps unsurprising given that our search term was in English), but that a large amount of research comes from authors in China and Germany.

Country	Number of Papers
United States	7812
United Kingdom	1711
Australia	1096
China	926
Germany	787
Canada	771

Table 6.1: Top author countries by number of social media papers.

Next we look at the disciplines that publish social media research. Figure 6.1 shows the number of papers containing the term 'social media' over time. The plot illustrates that the quantity of published research on social media has increased rapidly over time. The growth appears to slow down more recently, but this may be due to the speed at which the Scopus database imports data about new articles.

Figure 6.1 shows the top ten disciplines, as categorized by Scopus. We see that the field started off dominated by computer science publications, with additional disciplines increasing their activity in recent years. This story is also reflected in the top venues, listed in Table 6.2, where we see that computer science venues have published social media research most frequently.

[1]Technically, each paper is assigned to the modal (i.e., most frequent) country among its authors. For example, if a paper has three authors with two authors located in Canada and one in Japan, the modal country for the paper would be Canada. Any ties (i.e., if more than one country is tied for most frequent location among a paper's authors) were broken by randomly selecting among the tied countries.

Publication Venue	Papers
Lecture Notes in Computer Science	935
ACM International Conference Proceeding Series	288
Computers in Human Behavior	257
CEUR Workshop Proceedings	227
Proceedings of the Hawaii International Conference on System Sciences	179
Journal of Medical Internet Research	170

Table 6.2: Venues with the most social media papers.

Title	Publication Venue	Cited by
Users of the world, unite! The challenges and opportunities of Social Media	Business Horizons	1876
Why we twitter: Understanding microblogging usage and communities	Proceedings of WebKDD / SNA-KDD 2007	645
Social media? Get serious! Understanding the functional building blocks of social media	Business Horizons	468
Social media: The new hybrid element of the promotion mix	Business Horizons	450
Role of social media in online travel information search	Tourism Management	389
Networked narratives: Understanding word-of-mouth marketing in online communities	Journal of Marketing	335

Table 6.3: Most cited social media papers.

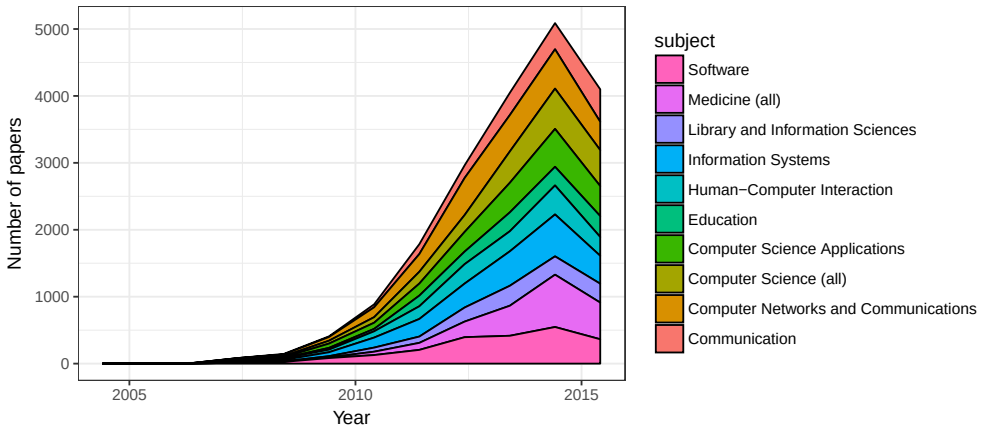


Figure 6.1: Social media papers published in the top ten disciplines (as categorized by Scopus), over time.

We then consider the impact of this set of papers as measured by the citations they have received. Like many phenomena in social systems, citation counts follow a highly skewed distribution with a few papers receiving many citations and most papers receiving very few. Table 6.3 provides a list of the most cited papers. These sorts of distributions suggest the presence of 'preferential attachment' (Barabási & Albert, 1999) or the 'Matthew effect' (Merton, 1968), where success leads to greater success.

Discussion

The summary statistics and exploratory visualizations presented above provide an overview of the scope and trajectory of social media research. We find that social media research is growing – both overall and within many disciplines. We find evidence that computer scientists laid the groundwork for the study of social media, but that social scientists, learning scientists, and medical researchers have increasingly been referring to social media in their published work. We also find several business and marketing papers among the most cited pieces of social media research even though neither these disciplines nor their journals appear among the most prevalent in the dataset.

These results are interesting and believable because they come from a comprehensive database of academic work. In most social science contexts, researchers have to sample from a population and that sampling is often biased. For example, the people willing to come to a lab to participate in a study or take a phone survey may have different attributes from those unwilling to par-

ticipate. This makes generalizing to the entire population problematic. When using trace data, on the other hand, we often have data from all members of a community including those who would not have chosen to participate. One of the primary benefits of collecting data from a comprehensive resource like Scopus is that it can reduce some types of bias in the data collection process. For example, we do not have backgrounds in education or medical research; had we tried to summarize the state of social media research by identifying articles and journals manually, we might have overlooked these disciplines.

That said, this apparent benefit can also become a liability when we seek to generalize our results beyond the community that we have data for. The large *N* of big data studies using social media traces may make results appear more valid, precise, or certain, but a biased sample does not become less biased just because it is larger (Hargittai, 2015). For example, a sample of 100 million Twitter users might be a worse predictor of election results than a truly random sample of only 1,000 likely voters because Twitter users likely have different attributes and opinions than the voting public. Another risk comes from the danger that data providers collect or filter data in ways that aren't apparent. Researchers should think carefully about the relationship of their data to the population they wish to study and find ways to estimate bias empirically.

Overall, we view the ease of obtaining and analyzing digital traces as one of the most exciting developments in social science. Although the hurdles involved represent a real challenge to many scholars of social media today, learning the technical skills required to obtain online trace data is no more challenging than the statistics training that is part of many PhD programs. Below, we present examples of a few computational analyses that can be done with this sort of data.

Network analysis

Social network analysis encompasses the most established set of computational methods in the social sciences (Wasserman & Faust, 1994). At its core, network analysis revolves around a 'graph' representation of data that tries to capture relationships (called edges) between discrete objects (called nodes). Graphs can represent any type of object and relationship, such as roads connecting a group of cities or shared ingredients across a set of recipes. Graph representations of data, and the network analytic methods built to reason using these data, are widely used across the social sciences as well as other fields

including physics, genomics, computer science, and philosophy. 'Social network analysis' constitutes a specialized branch of network analysis in which nodes represent people (or other social entities) and edges represent social relationships like friendship, interaction, or communication.

The power of network analysis stems from its capacity to reduce a very large and complex dataset to a relatively simple set of relations that possess enormous explanatory power. For example, Hausmann et al. (2014) use network data on the presence or absence of trading relationships between countries to build a series of extremely accurate predictions about countries' relative wealth and economic performance over time. By reasoning over a set of relationships in a network, Hausmann and his colleagues show that details of the nature or amount of goods exchanged are not necessary to arrive at accurate economic conclusions.

Network analysis has flourished in studies of citation patterns within scholarly literature, called 'bibliometrics' or 'scientometrics.' Bibliometric scholars have developed and applied network analytic tools for more than a half-century (Kessler, 1963; Hood & Wilson, 2001). As a result, bibliometric analysis provides an obvious jumping-off point for our tour of computational methods. Because network methods reflect a whole family of statistics, algorithms, and applications, we focus on approaches that are both well-suited to bibliometric analysis and representative of network analyses used in computational social science more broadly.

Our application: Citation networks

Our network analysis begins by representing citation information we collected from the Scopus APIs as a graph. In our representation, each node represents a paper and each edge represents a citation. Scopus provides data on incoming citations for each article. Our full dataset includes 35,620 incoming citations to the 23,131 articles in Scopus with 'social media' in their titles, abstracts, or keywords. 19,267 of these articles (83%) have not been cited even once by another article in Scopus and 18,324 (79%) do not cite any other article in our sample. The recent development of social media and the rapid growth of the field depicted in Figure 6.1 might help explain the sparseness (i.e. lack of connections) of the graph. As a result, and as is often the case in network analysis, a majority of our dataset plays no role in our analysis described in the rest of this section.

Once we create our citation graph, there are many potential ways to an-

alyze it. One important application, common to bibliometrics, is the computational identification of communities or clusters within networks. In network studies, the term 'community' is used to refer to groups of nodes that are densely connected to each other but relatively less connected to other groups. In bibliometric analyses, communities can describe fields or sub-fields of articles which cite each other, but are much less likely to cite or be cited by papers in other groups. Although there are many statistical approaches to community detection in network science, we use a technique from Rosvall and Bergstrom (2008) that has been identified as appropriate for the study of bibliometric networks (Šubelj, Eck, & Waltman, 2016). By looking at the most frequently occurring journals and publication venues in each community, we are able to identify and name sub-fields of social media research as distinct communities.

A citation graph is only one possible network representation of the relationships between articles. For example, the use of common topics or terminology might constitute another type of edge. Alternatively, journals or individual authors (rather than articles) might constitute an alternative source of nodes. In bibliometric analyses, for example, it is common for edges to represent 'co-citations' between articles or authors. Using this approach, papers are said to be tied together by a co-citation if they have both been cited in a third document (Small, 1973). Due to limited space, we only present the simplest case of direct citations.

Results

As is common in social networks, the large majority of articles with any citations connect to each other in one large 'component' or sub-network. Figure 6.2 shows a visualization of this large component. The optimal way to represent network data in two-dimensional space is a topic of research and debate. Figure 6.2 uses a force-directed drawing technique (Fruchterman & Reingold, 1991), the most widely used algorithm in network visualization, using the free/open source software package Gephi (Bastian, Heymann, Jacomy, et al., 2009). The basic idea behind the algorithm is that nodes naturally push away from each other, but are pulled together by the edges between them. Shades in each graph in this section reflect the communities of documents identified by Rosvall and colleagues' 'map' algorithm (Martin Rosvall & Bergstrom, 2008; M. Rosvall, Axelsson, & Bergstrom, 2010). Although the algorithm identified several dozen communities, most are extremely small, so we have shown



Figure 6.2: Network visualization of the citation network in our dataset. The layout is 'force directed' meaning that nodes (papers) with more edges (citations) appear closer to the center of the figure.

only the largest 6 communities in Figure 6.2. Each of these communities are summarized in Table 6.4 where the right-most column lists the three most common journals for the articles included in each community.

At this point, we could look in more depth at the attributes of the different communities. For example, in a bibliometric analysis published in the journal *Scientometrics*, Kovács, Looy, and Cassiman (2015) reported summary statistics for articles in each of the major communities identified (e.g., the average number of citations) as well as qualitative descriptions of the nodes in each community. We can see from looking at Table 6.4 that the communities point to the existence of coherent thematic groups. For example, Community 1 includes biomedical research while Community 3 contains papers published in communication journals. Earlier, we relied on an existing category scheme applied to journals to create Figure 6.1; all articles published in particular journals were treated as being within one field. Network analysis, however, can

Community	Description	Journals
Community 1	biomedicine; bioinformatics	Journal of Medical Internet Research; PLoS ONE; Studies in Health Technology and Informatics
Community 2	information technology; management	Computers in Human Behavior; Business Horizons; Journal of Interactive Marketing
Community 3	communication	Information Communication and Society; New Media and Society; Journal of Communication
Community 4	computer science; network science	Lecture Notes in Computer Science; PLoS ONE; WWW; KDD
Community 5	psychology; psychometrics	Computers in Human Behavior; Cyberpsychology, Behavior, and Social Networking; Computers and Education
Community 6	multimedia	IEEE Transactions on Multimedia; Lecture Notes in Computer Science; ACM Multimedia

Table 6.4: Description of each of the citation network clusters identified by the community detection algorithm, together with a list of the three most common journals in each community.

identify groups and categories of articles in terms of who is citing whom and, as a result, can reveal groups that cross journal boundaries. PLoS ONE, for example, is a 'megajournal' that publishes articles from all scientific fields (Binfield, 2012). As a result, PLoS ONE is one of the most frequently included journals in both Community 1 and Community 4. In a journal-based categorization system, articles may be misclassified or not classified at all.

Network analysis can also reveal details about the connections between fields. Figure 6.3 shows a second network we have created in which our communities are represented as nodes and citations from articles in one community to articles in the other communities are represented as edges. The thickness of each edge represents the number of citations and the graph shows the directional strength of the relative connections between communities. For example, the graph suggests that the communication studies community (Community 3) cites many papers in information technology and management (Community 2) but that this relationship is not reciprocated.

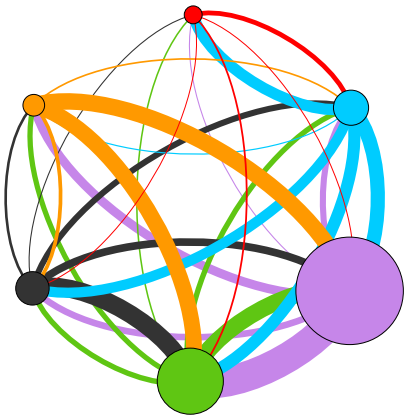


Figure 6.3: Graphical representation of citations between communities using the same grayscale mapping described in Table 6.4. The size of the nodes reflects the total number of papers in each community. The thickness of each edge reflects the number of outgoing citations. Edges are directional, and share the color of their source (i.e., citing) community.

Discussion

Like many computational methods, the power of network techniques comes from representing complex relationships in simplified forms. Although elegant and powerful, the network analysis approach is inherently reductive in nature and limited in many ways. What we gain in our ability to analyze millions or billions of individuals comes at the cost of speaking about particular individuals and sub-groups. A second limitation stems from the huge number of relationships that can be represented in graphs. A citation network and a co-citation network, for example, represent different types of connections and these differences might lead an algorithm to identify different communities. As a result, choices about the way that edges and nodes are defined can lead to very different conclusions about the structure of a network or the influence of particular nodes. Network analyses often treat all connections and all nodes as similar in ways that mask important variation.

Network analysis is built on the assumption that knowing about the relationships between individuals in a system is often as important, and sometimes more important, than knowing about the individuals themselves. It inherently recognizes interdependence and the importance of social structures. This perspective comes with a cost, however. The relational structure and interdependence of social networks make it impossible to use traditional statistical methods. SNA practitioners have had to move to more complex modeling

strategies and simulations to test hypotheses.

Text analysis

Social media produces an incredible amount of text, and social media researchers often analyze the content of this text. For example, researchers use ethnographic approaches (Kozinets, 2002) or content analysis (Chew & Eysenbach, 2010) to study the texts of interactions online. Because the amount of text available for analysis is far beyond the ability of any set of researchers to analyze by hand, scholars increasingly turn to computational approaches. Some of these analyses are fairly simple, such as tracking the occurrence of terms related to a topic or psychological construct (Tausczik & Pennebaker, 2010). Others are more complicated, using tools from natural language processing (NLP). NLP includes a range of approaches in which algorithms are applied to texts, such as machine translation, optical character recognition, and part-of-speech tagging. Perhaps the most common use in the social sciences is sentiment analysis, in which the affect of a piece of text is intuited based on the words that are used (Asur & Huberman, 2010). Many of these techniques have applications for social media research.

One natural language processing technique—topic modeling—is used increasingly often in computational social science research. Topic modeling seeks to identify topics automatically within a set of documents. In this sense, topic modeling is analogous to content analysis or other manual forms of document coding and labeling. However, topic models are a completely automated, unsupervised computational method—i.e., topic modeling algorithms do not require any sort of human intervention, such as hand-coded training data or dictionaries of terms. Topic modeling scales well to even very large datasets, and is most usefully applied to large corpora of text where labor-intensive methods like manual coding are simply not an option.

When using the technique, a researcher begins by feeding topic modeling software the texts that she would like to find topics for and by specifying the number of topics to be returned. There are multiple algorithms for identifying topics, but we focus on the most common: *latent Dirichlet allocation* or LDA (Blei, Ng, & Jordan, 2003). The nuts and bolts of how LDA works are complex and beyond the scope of this chapter, but the basic goal is fairly simple: LDA identifies sets of words that are likely to be used together and calls these sets 'topics.' For example, a computer science paper is likely to use words like 'algorithm', 'memory', and 'network.' While a communication ar-

ticle might also use 'network,' it would be much less likely to use 'algorithm' and more likely to use words like 'media' and 'influence.' The other key feature of LDA is that it does not treat documents as belonging to only one topic, but as consisting of a mixture of multiple topics with different degrees of emphasis. For example, an LDA analysis might characterize this chapter as a mixture of computer science and communication (among other topics).

LDA identifies topics inductively from the observed distributions of words in documents. The LDA algorithm looks at all of the words that co-occur within a corpus of documents and assumes that words used in the same document are more likely to be from the same topic. The algorithm then looks across all of the documents and finds the set of topics and topic distributions that would be, in a statistical sense, most likely to produce the observed documents. LDA's output is the set of topics: ranked lists of words likely to be used in documents about each topic, as well as the distribution of topics in each document. DiMaggio, Nag, and Blei (2013) argue that while many aspects of topic modeling are simplistic, many of the assumptions have parallels in sociological and communication theory. Perhaps more importantly, the topics created by LDA frequently correspond to human intuition about how documents should be grouped or classified.

The results of topic models can be used many ways. Our dataset includes 73 publications with the term 'LDA' in their abstracts. Some of these papers use topic models to conduct large-scale content analysis, such as looking at the topics used around health on Twitter (Prier, Smith, Giraud-Carrier, & Hanson, 2011; Ghosh & Guha, 2013). Researchers commonly use topic modeling for prediction and machine learning tasks, such as predicting a user's gender or personality type (Schwartz et al., 2013). Papers in the dataset also use LDA to predict transitions between topics (Wang, Agichtein, & Benzi, 2012), to recommend friends based on similar topic use (Pennacchiotti & Gurumurthy, 2011), and to identify interesting tweets on Twitter (Yang & Rim, 2014).

Our application: Identifying topics in social media research

We apply LDA to the texts of abstracts in our dataset in order to identify topics in social media research. We show how topics are extracted and labeled and then use data on topic distributions to show how the focus of social media research has changed over time. We begin by collecting each of the abstracts for the papers in our sample. Scopus does not include abstract text for 2,801 of the 23,131 articles in our sample. We examined a random sample of the en-

tries with missing abstracts by hand, and found that abstracts for many simply never existed (e.g., articles published in trade journals or books). Other articles had published abstracts, but the text of these abstracts, for reasons that are not clear, were not available through Scopus.[2] We proceed with the 20,330 articles in our sample for which abstract data was available. The average abstract in this dataset is 177 words long, with a max of 1,353 words and a minimum of 5 ("The proceedings contain 15 papers.").

We then remove 'stop words' (common words like 'the,''of,' etc.) and tokenize the documents by breaking them into unigrams and bigrams (one-word and two-word terms). We analyze the data using the Python *LatentDirichletAllocation* module from the *scikit-learn* library (Pedregosa et al., 2011). Choosing the appropriate number of topics to be returned (typically referred to as k) is a matter of some debate and research (e.g., Arun, Suresh, Madhavan, & Murthy, 2010). After experimenting with different values of k, plotting the distribution of topics each time in a way similar to the graphs shown in Figure 6.4, we ultimately set k as twelve. At higher values of k, additional topics only rarely appeared in the abstracts.

Results

Table 6.5 shows the top words for each of the topics discovered by the LDA model, sorted by how common each topic is in our dataset. At this point, researchers typically evaluate the lists of words for coherence and give names to each of the topics. For example, after looking at the words associated with Topic 1 we gave it the name 'Media Use.' Of course, many other names for this topic could be chosen. We might call it 'Facebook research' because it is the only topic which includes the term 'facebook.' Researchers often validate these names by looking at some of the texts which score highest for each topic and subjectively evaluating the appropriateness of the chosen name as a label for those texts. For example, we examined the abstracts of the five papers with the highest value for the 'Media Use' topic and confirmed that we were comfortable claiming that they were examples of research about media use. In this way, topic modeling requires a mixture of both quantitative and qualitative interpretation. The computer provides results, but making sense of those results requires familiarty with the data.

[2]This provides one example of how the details of missing data can be invisible or opaque. It is easy to see how missing data like this could impact research results. For example, if certain disciplines or topics are systematically less likely to include abstracts in Scopus, we will have a skewed representation of the field.

Media Use	Social Network Analysis	Consumer Analsyis	Education	Quantitative Analysis	Information Spread
social	social	media	students	based	twitter
media	data	social	learning	approach	tweets
"social media"	media	"social media"	knowledge	method	time
use	"social media"	new	research	proposed	information
study	information	marketing	education	data	messages
online	users	2015	technology	model	events
facebook	network	business	social	text	public
research	networks	brand	use	images	videos
communication	user	communication	media	results	crisis
public	paper	information	"social media"	media	users
political	web	consumers	design	user	mobile
article	analysis	companies	tools	search	data
findings	based	organizations	technologies	using	event
"use social"	online	management	development	image	location
2014	"social networks"	consumer	digital	topic	used
people	different	customer	2015	propose	2014
new	research	services	student	paper	emergency
results	content	strategies	educational	algorithm	disaster
networking	"social network"	customers	paper	problem	real
using	people	service	project	detection	youtube

Health	Sentiment Analysis	News	HCI	Influence	Methodology
health	content	news	systems	2015	purpose
information	springer	women	information	model	value
use	sentiment	study	privacy	influence	implications
patients	analysis	facebook	papers	al	findings
medical	user	posts	based	intention	limited
care	results	articles	music	et	paper
methods	negative	sexual	security	"et al"	methodology
results	generated	participants	personality	factors	approach
patient	online	page	cloud	perceived	publishing
participants	positive	young	alcohol	smoking	"publishing limited"
using	study	men	model	tobacco	emerald
related	opinion	stories	online	"2015 elsevier"	"emerald group"
internet	reviews	gender	include	satisfaction	"group publishing"
reported	comments	journalists	using	theory	practical
conclusions	switzerland	online	software	structural	originality
support	opinions	significantly	management	variables	design
used	"sentiment analysis"	female	proceedings	intentions	research
clinical	quality	group	discussed	equation	group
healthcare	users	exposure	contain	study	"originality value"
risk	media	pages	analysis	addiction	"methodology approach"

Table 6.5: Top 20 terms for each topic. Topics are presented in the order of their frequency in the corpus of abstracts.

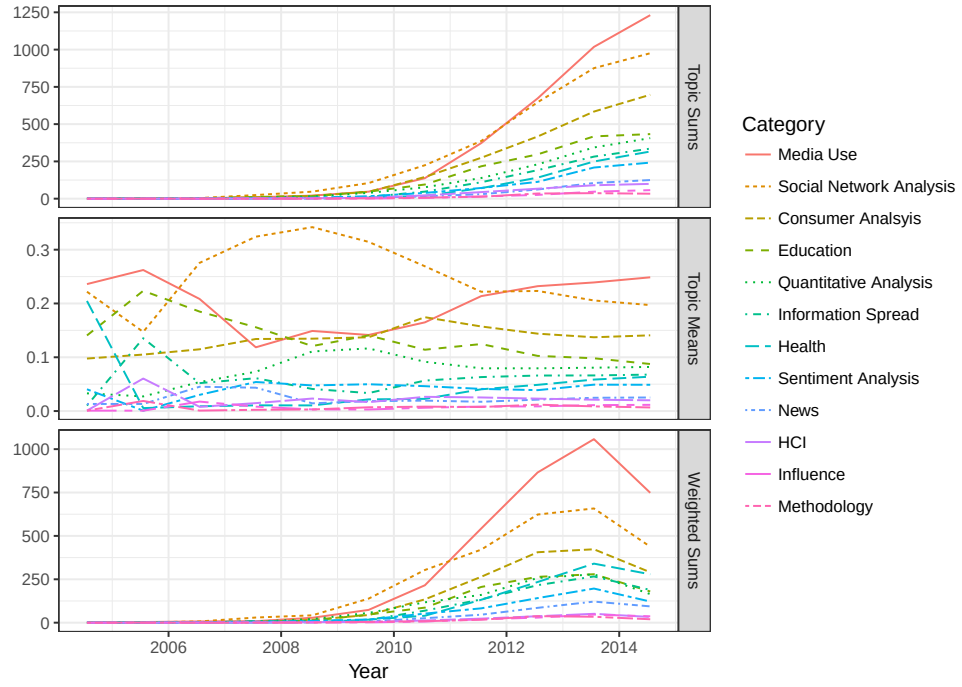


Figure 6.4: Statistics from our LDA analysis, over time. The top panel shows topic sums which capture the amount that each topic is used in abstracts, by year. The middle panel shows topic means which are the average amount that each topic is used in a given abstract. The bottom panel shows the amount that each topic is used in abstracts, by year, weighted by citation count.

The top panel of Figure 6.4 shows how the distribution of topics identified by LDA in our analysis has changed over time. The LDA algorithm gives each abstract a probability distribution over each of the topics, such that it sums to 1 (e.g., a given abstract may be 80% 'Social Network Analysis,' 20% 'Education,' and 0% everything else). To construct Figure 6.4, we sum these percentages for all of the documents published in each year and plot the resulting prevalence of each topic over time.[3]

The figures provide insight into the history and trajectory of social media research. Looking at the top figure, it appears that the 'Social Network Analysis' topic was the early leader in publishing on social media, but was overtaken by the 'Media Use' topic around 2012. This pattern is even more apparent when we look at the mean amount that each topic was used each year (the middle panel of Figure 6.4). In the bottom panel, we take a third

[3]More complex approaches such as dynamic LDA (Blei & Lafferty, 2006) are often better suited to identify the temporal evolution of topics.

look at this data by weighting the topics used in each paper by the log of the number of citations that the paper received. This final statistic characterizes how influential each topic has been. The overall story is similar, although we see that the 'Health' topic and the 'Media Use' topic are more influential than the non-weighted figures suggest.

Discussion

Some of the strengths of topic modeling become apparent when we compare these LDA-based analyses with the distribution of papers by discipline that we created earlier (Figure 6.1). In our earlier attempt, we relied on the categories that Scopus provided and found that early interest in social media was driven by computer science and information systems researchers. Through topic modeling, we learn that these researchers engaged in social network analysis (rather than interface design, for example). While some of our topics match up well with the disciplines identified by Scopus, a few are more broad (e.g., 'Media Use') and most are more narrow (e.g., 'Sentiment Analysis'). This analysis provides a richer sense of the topics of interest to social media researchers. Finally, these topics emerged inductively without any need for explicit coding, such as classifying journals into disciplines. This final feature is a major benefit in social media research where text is rarely categorized for researchers ahead of time.

Topic modeling provides an intuitive, approachable way of doing large-scale text analysis. Its outputs can be understandable and theory-generating. The inductive creation of topics has advantages over traditional content analysis or 'supervised' computational methods that require researchers to define labels or categories of interest ahead of time. While topic models clearly lack the nuance and depth of understanding that human coders bring to texts, the method allows researchers to analyze datasets at a scale and granularity that would take a huge amount of resources to code manually.

There are, of course, limitations to topic modeling. Many of LDA's limitations have analogues in manual coding. One we have already mentioned is that researchers must choose the number of topics without any clear rules about how to do so. Although a similar problem exists in content analysis, the merging and splitting of topics can be done more intuitively and intentionally when using traditional methods. An additional limitation is that topic modeling tends to work best with many long documents. This can represent a stumbling block for researchers with datasets of short social media posts or

comments; in these cases posts can be aggregated by user or by page to produce meaningful topics. The scope of documents can also affect the results of topic models. If, in addition to using abstracts about 'social media,' we had also included abstracts containing the term 'gene splicing,' our twelve topics would be divided between the two fields and each topic would be less granular. To recover topics similar to those we report here, we would have to increase the number of topics created.

As with network analysis, a goal of LDA is to distill large, messy, and noisy data down to much simpler representations in order to find patterns. Such simplification will always entail ignoring some part of what is going on. Luckily, human coders and LDA have complementary advantages and disadvantages in this regard. Computational methods do not understand which text is more or less important. Humans are good at seeing the meaning and importance of topics, but may suffer from cognitive biases and miss out on topics that are less salient (DiMaggio et al., 2013). Topic models work best when they are interpreted by researchers with a rich understanding of the texts and contexts under investigation.

Predicting citation

A final computational technique is statistical prediction. Statistical prediction can come in handy in situations where researchers have a great deal of data, including measures of an important, well-defined outcome they care about, but little in the way of prior literature or theory to guide analysis. Prediction has become a mainstream computational approach that encompasses a number of specific statistical techniques including classification, cross validation, and machine learning (also known as statistical learning) methods (Tibshirani, 1996). Arguably made most famous by Nate Silver (2015), who uses the technique to predict elections and sporting event outcomes, prediction increasingly colors public discourse about current events (Domingos, 2015).

There are many approaches to prediction. We focus on regression-based prediction because it offers a reasonably straightforward workflow. Begin by breaking a dataset into two random subsets: a large subset used as 'training' data and a small subset as 'holdout' or 'test' data. Next, use the training data to construct a regression model of a given outcome (dependent variable) that incorporates a set of features (independent variables) that might explain variations in the outcome. Apply statistical model selection techniques to determine the best weights (coefficients) to apply to the variables. Evaluate the per-

formance of the model by seeing how accurately it can predict the outcome on the test data. After selecting an appropriate model, assess and interpret the items that most strongly predict the outcome. One can even compare the performance of different or nested sets of features by repeating these steps with multiple groups of independent variables.

Interpreting the results of statistical prediction can be less clear-cut. The term 'prediction' suggests a deep knowledge of a complex social process and the factors that determine a particular outcome. However, statistical prediction often proves more suitable for exploratory analysis where causal mechanisms and processes are poorly understood. We demonstrate this in the following example that predicts whether or not papers in our dataset get cited during the period of data collection. In particular, we try to find out whether textual features of the abstracts can help explain citation outcomes. Our approach follows that used by Mitra and Gilbert (2014), who sought to understand what textual features of Kickstarter projects predicted whether or not projects were funded.

Our application: Predicting paper citation

We use multiple attributes of the papers in our dataset, including text of their abstracts, to predict citations. About 42% of the papers (9,713 out of 23,131) received one or more citations ($\mu = 3$; $\sigma = 19$). Can textual features of the abstracts explain which papers receive citations? What about other attributes, such as the publication venue or subject area? A prediction analysis can help evaluate these competing alternatives.

To begin, we generate a large set of features for each paper from the Scopus data. Our measures include the year, month, and language of publication as well as the number of citations each paper contains to prior work. We also include the modal country of origin of the authors as well as the affiliation of the first author. Finally, we include the publication venue and publication subject area as provided by Scopus. Then, we build the textual features by taking all of the abstracts and moving them through the following sequence of steps similar to those we took when performing LDA: we lowercase all the words; remove all stop words; and create uni-, bi-, and tri-grams.

We also apply some inclusion criteria to both papers and features. To avoid subject-specific jargon, we draw features only from those terms that appear across at least 30 different subject areas. To avoid spurious results, we also exclude papers that fall into unique categories. Specifically, we remove papers

which are the only publications in a given language, journal, or subject area. These sorts of unique cases can cause problems in the context of prediction tasks because they may predict certain outcomes perfectly. As a result, it is often better to focus on datasets and measures that are less 'sparse' (i.e., characterized by rare, one-off observations). Once we drop the 22,619 papers that do not meet these criteria, we are left with 1 papers.

We predict the dichotomous outcome variable *cited*, which indicates whether a paper received any citations during the period covered by our dataset (2004-2016). We use a method of *penalized logistic regression* called the least absolute shrinkage and selection operator (also known as the *Lasso*) to do the prediction work. Although, the technical details of Lasso models lie beyond the scope of this chapter, it, and other penalized regression models work well on data where many of the variables have nearly identical values (sometimes called collinear variables because they would sit right around the same line if you plotted them) and/or many zero values (this is also called 'sparse' data) (Friedman, Hastie, & Tibshirani, 2010; James et al., 2013). In both of these situations, some measures are redundant; the Lasso uses clever math to pick which of those measures should go into your final model and which ones should be, in effect, left out.[4] The results of a Lasso model are thus more computationally tractable and easier to interpret.

We use a common statistical technique called cross-validation to validate our models. Cross-validation helps solve another statistical problem that can undermine the results of predictive analysis. Imagine fitting an ordinary least squares regression model on a dataset to generate a set of parameter estimates reflecting the relationships between a set of independent variables and some outcome. The model results provide a set of weights (the coefficients) that represent the strength of the relationships between each predictor and the outcome. Because of the way regression works (and because this is a hypothetical example and we can assume data that does not violate the assumptions of our model), the model weights are the best, linear, unbiased estimators of those relationships. In other words, the regression model fits the data as well as possible. However, nothing about fitting this one model ensures that the same regression weights will provide the best fit for some new data from the same population that the model has not seen. A model may be overfit if it excellently predicts the dataset it was fitted on but poorly predicts new

[4]To put things a little more technically, a fitted Lasso model *selects* the optimal set of variables that should have coefficient values greater than zero and *shrinks* the rest of the coefficients to zero without sacrificing goodness of fit (Tibshirani, 1996).

data. Overfitting in this way is a common concern in statistical prediction. Cross-validation addresses this overfitting problem. First, the training data is split into equal-sized groups (typically 10). Different model specifications are tested by iteratively training them on all but one of the groups, and testing how well they predict the final group. The specification that has the lowest average error is then used on the full training data to estimate coefficients.[5] This approach ensures that the resulting models not only fit the data that we have, but that they are likely to predict the outcomes for new, unobserved results. For each model, we report the mean error rate from the cross-validation run which produced the best fit.

Our analysis proceeds in multiple stages corresponding to the different types of measures we use to predict citation outcomes. We start by estimating a model that includes only the features that correspond to paper and author-level attributes (year, month, and language of publication, modal author country). We then add information about the first author's affiliation. Next, we include predictors that have more to do with research topic and field-level variations (publication venue and subject area). Finally, we include the textual features (terms) from the abstracts.

Results

Table 6.6 summarizes the results of our prediction models. We include goodness-of-fit statistics and prediction error rates for each model as we add more features. A 'better' model will fit the data more closely (i.e., it will explain a larger percentage of the deviance) and produce a lower error rate. We also include a supplementary error rate calculated against the 'holdout' data created from a random subset of 10% of the original dataset that was not used in any of our models. An intuitive way to think about the error rate is to imagine it as the percentage of unobserved papers for which the model will correctly predict whether or not it receives any citations. The two error rate statistics are just this same percentage calculated on different sets of unobserved papers. Unlike a normal regression analysis, we do not report or interpret the full battery of coefficients, standard errors, t-statistics, or p-values. In part, we do not report this information because the results of these models are unwieldy – each model has over 2,000 predictors and most of those predictors have coefficients of zero! Additionally, unlike traditional regression results,

[5]For our Lasso models, cross-validation was used to select λ, a parameter that tells the model how quickly to shrink variable coefficients. We include this information for those of you who want to try this on your own or figure out the details of our statistical code.

Model	N features	Deviance (%)	CV error (%)	Hold-back error (%)
Controls	98	7	38	37
+ Affiliation	1909	23	39	37
+ Subject	2096	28	37	34
+ Venue	3902	55	34	30
+ Terms	4411	72	29	27

Table 6.6: Summary of fitted models predicting citation. The 'Model' column describes which features were included. The N features column shows the number of features included in the prediction. 'Deviance' summarizes the goodness of fit as a percentage of the total deviance accounted for by the model. 'CV error' (cross-validation error) reports the prediction error rates of each model in the cross-validation procedure conducted as part of the parameter estimation process. 'Holdout error' shows the prediction error on a random 10% subset of the original dataset not included in any of the model estimation procedures.

coefficient interpretation and null hypothesis testing with predictive models remain challenging (for reasons that lie beyond the scope of this chapter). Instead, we focus on interpreting the relative performance of each set of features. After we have done this, we refer to the largest coefficients to help add nuance to our interpretation.

The results reveal that all of the features improve the goodness of fit, but not necessarily the predictive performance of the models. As a baseline, our controls-only model has a 37% classification error on the holdout sample. This level of precision barely improves with the addition of both the author affiliation and subject area features. We observe substantial improvements in the prediction performance when the models include the publication venue features and the abstract text terms. When it comes to research about social media, it appears that venue and textual content are the most informative features for predicting whether or not articles get cited.

To understand these results more deeply, we explore the non-zero coefficient estimates for the best-fitting iteration of the full model. Recall that the Lasso estimation procedure returns coefficients for a subset of the parameters that produce the best fit and shrinks the other coefficients to zero. While it does not make sense to interpret the coefficients in the same way as traditional regression, the non-zero coefficients indicate what features the model identified as the most important predictors of the outcome. First, we note that among the 1,482 features with non-zero coefficients, only 2% are control

Feature	Type	Coefficient
Multiple Sclerosis Journal	venue	-2.969
Nature Communications	venue	2.871
Journal of Information Technology	venue	2.762
CrossTalk	venue	2.543
21	term	-2.472
NICTA Victoria Research Laboratory	affiliation	-2.260
The Department of Education, Sookmyung Women's University	affiliation	-2.196
20th ITS World Congress Tokyo 2013	venue	-2.191
Electronics and Communications in Japan	venue	-2.085
British Journal of Nursing	venue	2.077

Table 6.7: Feature, variable type, and beta value for top 10 non-zero coefficients estimated by the best fitting model with all features included. Note that the outcome is coded such that positive coefficients indicate features that positively predict the observed outcome of interest (getting cited) while negative coefficients indicate features that negatively predict the outcome.

measures (country, language, month, and year of publication). Similarly, 3% are subject features. In contrast, 15% are affiliation features, 34% are venue features, and a whopping 44% are textual terms. Once again, we find that publication venue and textual terms do the most to explain which works receive citations.

Closer scrutiny of the features with the largest coefficients adds further nuance to this interpretation. Table 6.7 shows the ten features with the largest coefficients in terms of absolute value. The Lasso model identified these coefficients as the most informative predictors of whether or not papers in our dataset get cited. Here we see that the majority of these most predictive features are publication venues. The pattern holds across the 100 features with the largest coefficients, of which 75 are publication venues and only 2 are textual terms from the abstracts. In other words, variations in publication venue predict which work gets cited more than any other type of feature.

Discussion

The results of our prediction models suggest that two types of features – publication venue and textual terms – do the most to explain whether or not papers on social media get cited. Both types of features substantially improve model fit and reduce predictive error in ten-fold cross-validation as well as on a hold-out sub-sample of the original dataset. However, the venue features appear

to have a much stronger relationship to our outcome (citation), with the vast majority of the most influential features in the model coming from the venue data (75 of the 100 largest coefficients).

As we said at the outset of this section, statistical prediction offers an exploratory, data-driven, and inductive approach. Based on these findings, we conclude that the venue where research on social media gets published better predicts whether that work gets cited than the other features in our dataset. Textual terms used in abstracts help to explain citation outcomes across the dataset, but the relationship between textual terms and citation only becomes salient in aggregate. On their own, hardly any of the textual terms approach the predictive power of the venue features. Features such as author affiliation and paper-level features like language or authors' country provide less explanatory power overall.

The approach has several important limitations. Most important, statistical prediction only generates 'predictions' in a fairly narrow, statistical sense. Language of prediction often sounds like the language of causality and inferring process, but these methods do not guarantee that anything being studied is causal or explanatory in terms of mechanisms. We do not claim that a paper's publication venue or the phrases in its abstract *cause* people to cite it. Rather, we think these attributes of a paper likely index specific qualities of an article that are linked to citation outcomes. Just because something is predictive does not mean it is deterministic or causal. We also note that the sort of machine learning approach we demonstrate here does not support the types of inferences commonly made with frequentist null hypothesis tests (the sort that lead to p-values and stars next to 'significant' variables in a regression model). Instead, the interpretation of learning models rests on looking closely at model summary statistics, objective performance metrics (such as error rates), and qualitative exploration of model results.

Conclusion

In this chapter, we have described computational social scientific analysis of social media by walking through a series of example analyses. We began with the process of collecting a dataset of bibliographic information on social media scholarship using a web API similar to those provided by most social media platforms. We then subjected this dataset to three of the mostly widely used computational techniques: network analysis, topic modeling, and statistical prediction. Most empirical studies would employ a single, theoretically-

motivated analytic approach, but we compromised depth in order to illustrate the diversity of computational research methodologies available. As we have shown, each approach has distinct strengths and limitations.

We believe our examples paint a realistic picture of what is involved in typical computational social media research. However, these analyses remain limited in scope and idiosyncratic in nature. For example, there are popular computational methods we did not cover in this chapter. Obvious omissions include other forms of machine learning, such as decision trees and collaborative filtering (Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994), as well as simulation-based techniques such as agent-based modeling (Macy & Willer, 2002; Wilensky & Rand, 2015).

Despite our diffuse approach, we report interesting substantive findings about the history and state of social media research. We discovered a number of diverse communities studying social media. We used different tools to identify these communities, including the categories provided by Scopus, the results of a community detection algorithm applied to the citation network, and the topics identified by topic modeling. Each analysis provided a slightly different picture of social media research. We learned that the study of social media related to media use and medical research is on the rise. We also learned that social network research was influential at the early stages of social media research, but that it is not as influential in the citation network. All of these findings are complicated by our final finding that subject area is not as good a predictor of whether a paper will receive a citation as the publication venue and the terms used in the abstract.

In the process of describing our analyses, we tried to point to many of the limitations of computational research methods. Although computational methods and the promise of 'big data' elicit excitement, this hype can obscure the fact that large datasets and fast computers do nothing to obviate the fundamentals of high quality social science: researchers must understand their empirical settings, design studies with care, operationalize concepts in ways that are valid and honest, take steps to ensure that their findings generalize, and ask tough questions about the substantive impacts of observed relationships. These tenets extend to computational research as well.

Other challenges go beyond methodological limitations. Researchers working with passively collected data generated by social media can face complex issues around the ethics of privacy and consent as well as the technical and legal restrictions on automated data collection. Computational analyses of

social media often involve datasets gathered without the sort of active consent considered standard in other arenas of social scientific inquiry. In some cases, data is not public and researchers access it through private agreements or employment arrangements with companies that own platforms or proprietary databases. In others, researchers obtain social media data from public or semi-public sources, but the individuals creating the data may not consider their words or actions public and may not even be aware that their participation generates durable digital traces (boyd danah & Crawford, 2012). A number of studies have been criticized for releasing information that researchers considered public, but which users did not (Zimmer, 2016). In other cases, researchers pursuing legitimate social inquiry have become the target of companies or state prosecutors who selectively seek to enforce terms of service agreements or invoke broad laws such as the federal Computer Fraud and Abuse Act (CFAA).[6]

We advise computational researchers to take a cautious and adaptive approach to these issues. Existing mechanisms such as Institutional Review Boards and federal laws have been slow to adjust to the realities of online research. In many cases, the authority and resources to anticipate, monitor, or police irresponsible behaviors threaten to impose unduly cumbersome restrictions. In other cases, review boards' policies greenlight research that seems deeply problematic. We believe researchers must think carefully about the specific implications of releasing specific datasets. In particular, we encourage abundant caution and public consultation before disseminating anything resembling personal information about individual social media system users. Irresponsible scholarship harms both subjects and reviewers and undermines the public trust scholars need to pursue their work.

At the same time, we remain excited and optimistic about the future of computational studies of social media. As we have shown, the potential benefits of computational methods are numerous. Trace data can capture behaviors that are often difficult to observe in labs and that went unrecorded in offline interactions. Large datasets allow researchers to measure real effects obscured by large amounts of variation, and to make excellent predictions using relatively simple models. These new tools and new datasets provide a real opportunity to advance our understanding of the world. Such opportunities

[6]See Sandvig's (2016) blogpost, "Why I am Suing the Government," for a thoughtful argument against the incredibly vague and broad scope of the CFAA as well as a cautionary tale for those who write software to conduct bulk downloads of public website data for research purposes.

should not be undermined by overly-broad laws or alarmist concerns.

Finally, much of computational social science, including this chapter, is data-focused rather than theory-focused. We would encourage others to do as we say, and not as we do. The great promise of computational social science is the opportunity to test and advance social science theory. We hope that readers of this chapter will think about whether there are theories they are interested in which might benefit from a computational approach. We urge readers with a stronger background in theory to consider learning the tools to conduct these types of analyses and to collaborate with technically minded colleagues.

Reproducible research

Computational research methods also have the important benefit of being extraordinarily reproducible and replicable (Stodden, Guo, & Ma, 2013). Unlike many other forms of social research, a computational researcher can theoretically use web APIs to collect a dataset identical to one used in a previous study. Even when API limits or other factors prohibit creating an identical dataset, researchers can work to make data available alongside the code they use for their analysis, allowing others to re-run the study and assess the results directly. Making code and data available also means that others can analyze and critique it. This can create uncomfortable situations, but we feel that such situations serve the long-term interests of society and scholarly integrity. Although not every computational researcher shares their code (Stodden et al., 2013) there are movements to encourage or require this (LeVeque, Mitchell, & Stodden, 2012; Stodden et al., 2013; Bollen et al., 2015).

We have tried to follow emerging best practices with regards to reproducibility in this chapter. We have released an online copy of all of the code that we used to create this chapter. By making our code available, we hope to make our unstated assumptions and decisions visible. By looking at our code, you might find errors or omissions which can be corrected in subsequent work. By releasing our code and data, we also hope that others can learn from and build on our work. For example, a reader with access to a server and some knowledge of the Python and R programming languages should be able to build a more up-to-date version of our dataset years from now. Another reader might create a similar bibliographic analysis of another field. By using our code, this reader should able to produce results, tables, and figures like those in this chapter. Data repositories, such as the Harvard Dataverse, make

storing and sharing data simple and inexpensive. When thinking of the opportunities for openness, transparency, and collaboration, we are inspired by the potential of computational research methods for social media. We hope that our overview, data, and code can facilitate more of this type of work.

Online supplements

All of the code used to generate our dataset, to complete our analyses, and even to produce the text of this chapter, is available for download on the following public website: https://communitydata.cc/social-media-chapter/

Because the Scopus dataset is constantly being updated and changed, reproducing the precise numbers and graphs in this chapter requires access to a copy of the dataset we collected from Scopus in 2016. Unfortunately, like many social media websites, the terms of use for the Scopus APIs prohibit the re-publication of data collected from their database. However, they did allow us to create a private, access-controlled, replication dataset in the Harvard Dataverse archive at the following URL: http://dx.doi.org/10.7910/DVN/W31PH5. Upon request, we will grant access to this dataset to any researchers interested in reproducing our analyses.

References

Arun, R., Suresh, V., Madhavan, C. E. V., & Murthy, M. N. N. (2010, June 21). On finding the natural number of topics with latent dirichlet allocation: Some observations. In *Advances in knowledge discovery and data mining* (pp. 391–402). Lecture Notes in Computer Science. Pacific-asia conference on knowledge discovery and data mining. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-13657-3_43

Asur, S. & Huberman, B. A. (2010, August). Predicting the future with social media. In *2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology (WI-IAT)* (Vol. 1, pp. 492–499). 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology (WI-IAT). doi:10.1109/WI-IAT.2010.63

Barabási, A.-L. & Albert, R. (1999, October 15). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512. doi:10.1126/science.286.5439.509

Bastian, M., Heymann, S., Jacomy, M., et al. (2009). Gephi: An open source software for exploring and manipulating networks. *ICWSM*, *8*, 361–

362. Retrieved July 20, 2016, from http://www.aaai.org/ocs/index.php/ICWSM/09/paper/viewFile/154/1009/

Binfield, P. (2012, February 29). *PLoS ONE and the rise of the open access Mega-Journal*. The 5th SPARC Japan Seminar 2011, National Institute for Informatics. Retrieved July 20, 2016, from http://www.nii.ac.jp/sparc/en/event/2011/pdf/20120229_doc3_binfield.pdf

Blei, D. M. (2012, April). Probabilistic topic models. *Commun. ACM*, *55*(4), 77–84. doi:10.1145/2133806.2133826

Blei, D. M. & Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on machine learning* (pp. 113–120). ACM. Retrieved April 21, 2016, from http://dl.acm.org/citation.cfm?id=1143859

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022. Retrieved December 3, 2015, from http://dl.acm.org/citation.cfm?id=944937

Bollen, K., Cacioppo, J. T., Kaplan, R. M., Krosnick, J. A., Olds, J. L., & Dean, H. (2015, May). *Social, behavioral, and economic sciences perspectives on robust and reliable science*. National Science Foundation. Retrieved from http://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf

boyd danah, d. & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679. doi:10.1080/1369118X.2012.678878

Chew, C. & Eysenbach, G. (2010, November 29). Pandemics in the age of twitter: Content analysis of tweets during the 2009 h1n1 outbreak. *PLOS ONE*, *5*(11), e14118. doi:10.1371/journal.pone.0014118

DiMaggio, P., Nag, M., & Blei, D. (2013, December). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of u.s. government arts funding. *Poetics*, *41*(6), 570–606. doi:10.1016/j.poetic.2013.08.004

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. New York, New York: Basic Books.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, *33*(1), 1–22. Retrieved July 20, 2016, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2929880/

Fruchterman, T. M. J. & Reingold, E. M. (1991, November 1). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164. doi:10.1002/spe.4380211102

Ghosh, D. (& Guha, R. (2013, March 1). What are we 'tweeting' about obesity? mapping tweets with topic modeling and geographic information

system. *Cartography and Geographic Information Science*, *40*(2), 90–102. doi:10.1080/15230406.2013.776210

Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Burlington, Massachusetts: Morgan Kaufmann.

Hargittai, E. (2015, May 1). Is bigger always better? potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 63–76. doi:10.1177/0002716215570866

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A. (2014, January 17). *The atlas of economic complexity: Mapping paths to prosperity*. MIT Press.

Hood, W. W. & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, *52*(2), 291–314. doi:10.1023/A:1017919924342

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in r*. New York: Springer.

Kessler, M. M. (1963, January 1). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10–25. doi:10.1002/asi.5090140103

Kovács, A., Looy, B. V., & Cassiman, B. (2015, June 20). Exploring the scope of open innovation: A bibliometric review of a decade of research. *Scientometrics*, *104*(3), 951–983. doi:10.1007/s11192-015-1628-0

Kozinets, R. V. (2002, February 1). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research*, *39*(1), 61–72. doi:10.1509/jmkr.39.1.61.18935

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., … Van Alstyne, M. (2009, February 6). Computational social science. *Science*, *323*(5915), 721–723. doi:10.1126/science.1167742

Leskovec, J. & Krevl, A. (2014, June). *SNAP datasets: Stanford large network dataset collection*. Retrieved from http://snap.stanford.edu/data

LeVeque, R. J., Mitchell, I. M., & Stodden, V. (2012). Reproducible research for scientific computing: Tools and strategies for changing the culture. *Computing in Science and Engineering*, *14*(4), 13–17.

Macy, M. W. & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, *28*, 143–166. Retrieved July 20, 2016, from http://www.jstor.org/stable/3069238

Merton, R. K. (1968). The matthew effect in science. *Science*, *159*(3810), 56–63. Retrieved September 27, 2014, from http://www.unc.edu/ ~ fbaum/teaching/PLSC541_Fall06/Merton_Science_1968.pdf

Mitra, T. & Gilbert, E. (2014). The language that gets people to give: Phrases that predict success on kickstarter. In *Proceedings of the 17th ACM conference on computer supported cooperative work & social computing* (pp. 49–

61). CSCW '14. New York, NY, USA: ACM. doi:10.1145/2531602.2531656

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011, October). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830. bibtex: pedregosa_scikit-learn:_2011. Retrieved June 7, 2016, from http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Pennacchiotti, M. & Gurumurthy, S. (2011). Investigating topic models for social media user recommendation. In *Proceedings of the 20th international conference companion on world wide web* (pp. 101–102). WWW '11. New York, NY, USA: ACM. doi:10.1145/1963192.1963244

Prier, K. W., Smith, M. S., Giraud-Carrier, C., & Hanson, C. L. (2011). Identifying health-related topics on twitter: An exploration of tobacco-related tweets as a test topic. In *Proceedings of the 4th international conference on social computing, behavioral-cultural modeling and prediction* (pp. 18–25). SBP'11. Berlin, Heidelberg: Springer-Verlag. Retrieved July 19, 2016, from http://dl.acm.org/citation.cfm?id=1964698.1964702

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on computer supported cooperative work* (pp. 175–186). CSCW '94. New York, NY, USA: ACM. doi:10.1145/192844.192905

Rosvall, M. [M.], Axelsson, D., & Bergstrom, C. T. (2010, April 17). The map equation. *The European Physical Journal Special Topics*, *178*(1), 13–23. doi:10.1140/epjst/e2010-01179-1

Rosvall, M. [Martin] & Bergstrom, C. T. (2008, January 29). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, *105*(4), 1118–1123. doi:10.1073/pnas.0706851105

Sandvig, C. (2016, July 1). Why i am suing the government [Social media collective research blog]. Retrieved October 23, 2016, from https://socialmediacollective.org/2016/07/01/why-i-am-suing-the-government/

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., … Ungar, L. H. (2013, September 25). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, *8*(9), e73791. doi:10.1371/journal.pone.0073791

Silver, N. (2015). *The signal and the noise: Why so many predictions fail–but some don't*. New York, New York: Penguin Books.

Small, H. (1973, July 1). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269. doi:10.1002/asi.4630240406

Stodden, V., Guo, P., & Ma, Z. (2013, June 21). Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. *PLOS ONE*, *8*(6), e67111. doi:10.1371/journal.pone.0067111

Šubelj, L., Eck, N. J. v., & Waltman, L. (2016, April 28). Clustering scientific publications based on citation relations: A systematic comparison of different methods. *PLOS ONE*, *11*(4), e0154404. doi:10.1371/journal.pone.0154404

Tausczik, Y. R. & Pennebaker, J. W. (2010, March 1). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. doi:10.1177/0261927X09351676

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. Retrieved July 20, 2016, from http://www.jstor.org/stable/2346178

Wang, Y., Agichtein, E., & Benzi, M. (2012). TM-LDA: Efficient online modeling of latent topic transitions in social media. (p. 123). ACM Press. doi:10.1145/2339530.2339552

Wasserman, S. & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

Wilensky, U. & Rand, W. (2015). *An introduction to agent-based modeling: Modeling natural, social, and engineered complex systems with NetLogo*. Cambridge, Massachusetts: MIT Press.

Yang, M.-C. & Rim, H.-C. (2014, July). Identifying interesting twitter contents using topical analysis. *Expert Syst. Appl. 41*(9), 4330–4336. doi:10.1016/j.eswa.2013.12.051

Zimmer, M. (2016, May 14). OkCupid study reveals the perils of big-data science. *WIRED*. Retrieved August 31, 2016, from https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/

Understanding Trust amid Delays in Crowdfunding

Yongsung Kim, Aaron Shaw, Haoqi Zhang, Elizabeth Gerber
Northwestern University
Evanston, IL
yk@u.northwestern.edu, {aaronshaw, hq, egerber}@northwestern.edu

ABSTRACT
Trust is essential for beginning and maintaining relationships online where assessing uncertainties and risks is difficult. While product delays have been shown to reduce trust in e-commerce settings, we understand little about the effect of delays on trust in the increasingly popular context of crowdfunding. In a mixed method study, we examine what factors influence backers' trust in crowdfunding when rewards are delayed. Based on in-depth interviews with crowdfunding participants, we found that a rich set of factors influenced backers' trust including backers' role identity and domain knowledge, backer's research on a creator's background, creators' communication during delays, and duration of delays. To better understand the factors affecting delays, we conducted a regression analysis with 4,089 delayed projects and found that the funding goal, number of backers, percent raised, number of reward levels, and creator's previous crowdfunding experience are associated with the duration of delay. We discuss design implications for managing delays and maintaining trust in crowdfunding.

Author Keywords
Crowdfunding; Trust; Expectations; Entrepreneurship; Communication; Delays

ACM Classification Keywords
H.5.3 Group and Organization Interfaces: Computer-supported cooperative work

INTRODUCTION
Trust is essential in social interaction especially when beginning a new relationship or maintaining an existing one [9]. Trust issues are more prominent in online social interactions than in offline settings due to the lack of social cues that help us detect uncertainty and risk in the physical world [8]. Within HCI and CSCW, trust has been widely studied in different online contexts [27, 7, 24, 40, 25] and trust attitudes differ across domains such as online communities, e-commerce, and peer production [9].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CSCW '17, February 25-March 01, 2017, Portland, OR, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-4335-0/17/03...$15.00
DOI: http://dx.doi.org/10.1145/2998181.2998207

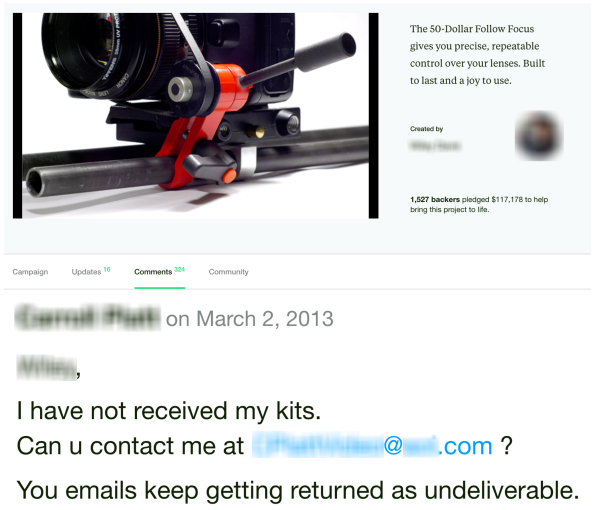


Figure 1. A screenshot of a project that was delayed on Kickstarter. Backers usually post their concerns and complaints about delays in the comment page.

Although crowdfunding has drawn a lot of attention from researchers across fields [49, 28, 35, 37, 4] few have studied issues of trust on crowdfunding platforms. Compared to traditional online marketplaces, the uncertainties and risks are greater in crowdfunding since there is (a) no guarantee that product will be made; (b) no risk assurance when a product is not made or delivered; and (c) far less information provided about the product to be made. Compared to donation platforms, motivations for participating in crowdfunding go beyond supporting a good cause to include collecting rewards and being part of a community [15]. These differences among platforms may lead to significant differences in people's trust attitudes, and by extension, the functioning of the platform for supporting its goals.

A prevalent issue on crowdfunding platforms is *reward delays* [36]. Reward delays refer to situations where project creators do not deliver their rewards on the promised delivery date that they had set when launching their campaigns (as shown in Figure 1, backers tend to communicate via comments page in the campaign website or email when there are delays). Research finds that more than 75% of rewards are delayed [36]. The popular press offers anecdotal reasons for delays including manufacturing, shipping problems, technical and design hurdles, and scaling when overfunded [10]. To our knowledge, however despite the increasing popularity of

crowdfunding, we lack empirical understanding of how delays affect trust, and if they do, what factors influence delays.

As HCI/CSCW researchers, our goal is to design a support tool for creators and backers to avoid or alleviate seemingly prevalent problems, delays, and as a first step we conducted two presented studies to provide insights into designing such support tools. First, we sought to understand whether or not backers even perceive delays as a problem in the crowdfunding settings due to a large difference in the context comparing to others such as traditional marketplace and e-commerce. If so, we tried to explore factors influence backers' trust in situations of delays. Second, since we found that delay is indeed a problem in crowdfunding from backer perspective, we sought to better understand project attributes that explain variation in delay duration so that a support tool with identified estimators can help novice creators to reason about and plan their project not just with regard to funding goals but also potential delays.

We first interviewed 14 backers who have supported crowdfunding projects and experienced delays in the past. We found that tolerance of delays differs depending on a backer's perceived role as an *investor*, *supporter*, or *consumer*. Regardless of backer role, backers reported that communication of delays is important to assure them from uncertainties and provide transparency. Further, we found that backers' domain knowledge, prior crowdfunding backing experience, and creators' credentials all influenced delay expectations. The duration of delays also played a crucial role in influencing backers' trust but the tolerance of delays varies across backer roles.

We also ran a regression analysis with 4,089 delayed projects on Kickstarter, the popular reward-based crowdfunding platform, to identify the project attributes associated with longer delays. Our results show that the number of reward levels, the creator's previous crowdfunding experience, the number of backers, the funding goal, and the percent raised are predictors for the duration of delays.

In summary, *our research contributes an empirical understanding of factors influencing delays and trust in crowdfunding platforms.* Based on these findings, we discuss design implications for a) helping creators to plan crowdfunding campaigns to minimize their delays, and supporting creators to take necessary actions when delays occur, and b) helping backers decision making process and adjust their expectations for potential delays.

RELATED WORK

Trust and Reputation

Trust is an essential part of social relationships [9] and has been widely studied in disciplines ranging from sociology and psychology to economics and political science [14, 46, 31, 41]. In this paper, we focus on trust in social computing. We use the definition of trust often cited by prior HCI and CSCW researchers [7, 47, 51, 9]; trust is the "general expectations held by an individual that the word of another can be relied on [41]". Also, trust "presupposes a situation of risk and the possibility of disappointment [29]". In crowdfunding context, we describe backer trust as backer's expectations that the word of a creator can be relied on, where there exists possible risks and disappointment (i.e. products not being made, being delayed, or not working as described, money not being spent on what the creator said to spend). Below, we compare and contrast previous research on trust and reputation in different context with respect to crowdfunding.

Online Communities

Within HCI and CSCW, trust has been widely studied in different online contexts and previous research argues that trust attitudes differ across domains [9]. For example, researchers have studied individual's attitudes about the trustworthiness of websites [9], relationships between trust and commitment in online communities [27], trust in computer-mediated communication settings [7], and trust on online peer production websites such as Wikipedia [24]. Many online communities use reputation systems to help encourage good behavior and deter norm violations [27].

Although trust has been studied in many domains, trust is overlooked in crowdfunding platforms by HCI and CSCW researchers, where trust plays a critical role in backing behavior due to great uncertainties and risks in getting rewards. Our in-depth interview expands our understanding of trust in the crowdfunding context especially when there are delays.

Marketplaces

Understanding of trust is essential for understanding interpersonal behavior in economic exchanges [12], particularly in online marketplaces where there are higher risks and uncertainties [5]. Pavlou and Gefen find that buyers have less trust in both individual sellers and the community of the sellers if they experience product delivery delay, thus influencing actual transactions [39]. It has also been shown that while trust can greatly improve the effectiveness of the market [1], lack of trust in a market, particularly in one characterized by dishonesty and cheating, can lead to market failure [17].

Unlike online marketplaces that provide assurances to consumers for delivering products and attempt to provide on-time delivery, crowdfunding platforms lack risk assurance for *project failure* (i.e. creators fail to deliver products after raising funds from backers) or *reward delays* (i.e. creators do not meet the promised delivery date). Lack of risk assurance might prevent backers who are less likely to take risks [11] from participating. Also, experience of delays might provide a bad experience to backers who had put their trust in projects; this might lead to loss of trust and attrition. Our work seeks to understand influential factors on trust in situations of delay and predictors of the duration of delays, so that we might help creators minimize delays in their crowdfunding campaigns and take necessary action when delays occur. Further, we hope to inform design decisions that help backers in the process of decision making online and to adjust their expectations for potential delays.

To mitigate their risks and uncertainties, existing online marketplaces use reputation systems to enhance trust [40, 25]. Customers write and rate about their past transactions with sellers and future customers use this data to measure the trustworthiness of the sellers. In the absence of such repu-

tation systems in crowdfunding platforms, we are interested in exploring what backers expectations are before backing a project, and how their risk-taking behavior and reaction to delays differ among backers with varying experience.

Donation platforms

In the non-profit donation context, reputation is often an important factor in initial donations [44]; trust and positive interactions are important factors for donor retention. In offline charities, trust, relationship building, expressing gratitude, communicating impact, commitment, satisfaction, and involvement are deemed important [42]. In online settings, like nonprofit donation platforms,completing a project leads to a larger donation [48], and having a positive experience makes donors more likely to return [3].

Unlike donation platforms on which the main motivation for donors is to support a good cause, crowdfunding platforms consist of other motivations for backer participation (e.g. collecting rewards, being part of a community) [15]. Also, backers in crowdfunding platforms have to understand reputations at an individual level rather than at an organizational level, which might make it harder for backers to gauge reputation.

Organizations

Researchers have studied trust among distributed workers [38, 52] and temporary groups [34] at organizational level as well. A case study of Hudson's Bay Company, a large distributed company, suggests that socialization, communication (e.g. establishing, tightening, negotiating communication norms), and participation (e.g. visiting local offices to generate situated knowledge and relationships) were important to establish and maintain trust. Another study of trust among distributed workers suggests that people rely on their first impression of perceived trustworthiness to evaluate co-worker's work since they lack reliable information about their co-worker's progress [52]. In temporary groups, where people with different skillsets work on a complex task until the task is completed, people use varying trusting behaviors to manage issues of uncertainty, risk, and expectations [34]; such as reducing their dependence on others or sharpening the perception of risks in projected actions. As trust is "required as an input condition to stimulate supportive activities in situations of uncertainty and risk [29]," we are interested in understanding how people establish initial trust under uncertainties and maintain trust in situations of delays in crowdfunding platforms.

Crowdfunding

Crowdfunding platforms allow many novice entrepreneurs to realize their dreams. They are not merely technical fundraising platforms but also places for social interaction and community building. Entrepreneurs learn from previous campaigns and seek help from experienced creators [22]. People back others' projects to be a part of a community [15], and creators leverage their social networks to raise funding [21]. Yet, despite the very social nature of crowdfunding, trust, one of the most important aspect in social relationship, has been overlooked by researchers. By understanding trust in crowdfunding, we hope to foster trustworthy behaviors in crowdfunding platform which can help maintaining existing creator-backer relationship, and also introduce new backers to contribute in crowdfunding who were hesitant to join due to mistrust.

Researchers have sought to understand how crowdfunding is changing the work of an entrepreneur. Previous works identified the characteristics of successfully funded campaigns, such as their communication strategies [35], project attributes that predict success [13, 19, 50], in hopes to better help novice entrepreneurs to successfully fund their campaigns.

One chronic problem in crowdfunding platforms is reward delays. Previous research show that more than 75% of successfully funded projects failed to deliver their products on the promised delivery date [36]. Interviews from top 50 funded creators in the popular press showed that reasons for the delays include manufacturing problems, shipping problems, overfunding, and technical and design changes [10]. Building on this preliminary understand of causes of delays, we seek more deeply to understand how backers feel about delays and the factors that might influence their trust.

Previous research showed that funding goal and percent raised can be used to predict delays; however, the analysis was limited in only the Design and Technology project type (N=314) and only used percent funded, number of backers, and funding goals for predicting delays [36]. We conduct an analysis with a larger and more diverse dataset (N=4089) which takes into account all project types and additional project attributes to better understand the attributes that are predictive of the duration of delays. We found that number of reward levels, creators' previous launching experience, number of backers, percent raised, and funding goals are associated with the duration of delays.

Most recently, Kim et al. studied factors that potential donors use to assess credibility of donation-based medical crowdfunding and found that communicative and emotional factors are unique in medical crowdfunding settings [23]. Our work is different from this work in that 1) we study trust in the context of reward-based crowdfunding, which differs from donation-based medical crowdfunding [23], and 2) our focus is on trust in situations of delays, while their study focused on assessment of credibility in the pre-donation stage.

INTERVIEW STUDY

We conducted an interview study to understand factors influencing backer trust in situations of delays in crowdfunding platforms. Specifically, we were interested in factors such as backers' expectations, delay experience, and their trust toward creators. We chose to do an interview study because we sought to have deeper understanding of backers' experience and expectations throughout the project, from before backing to receiving rewards. In the rest of the section, we will discuss the details of the study and the results.

Methods

Participants

We interviewed 14 crowdfunding project backers who have experienced delays on Kickstarter and Indiegogo—two of the most popular reward-based crowdfunding platforms in the

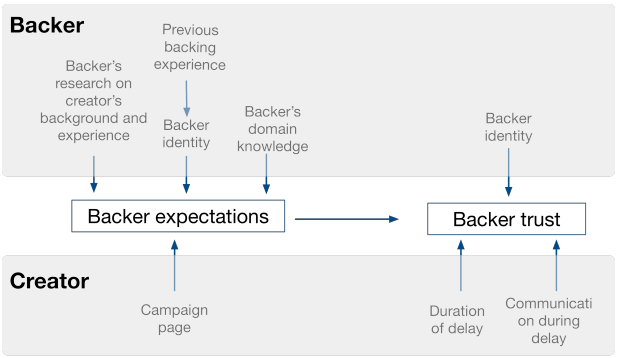


Figure 2. Factors including backer's previous backing experience, backer identity, domain knowledge, research on creator, and creator's campaign page influences backer expectations in terms of delays. Backer's trust is influenced by their expectations, their identity, duration of delays, as well as creator's communications during delays.

US. Project types, as defined by the crowdfunding platforms, included Art, Comics, Crafts, Design, Fashion, Film & Video, Food, Games, Music, Photography, Publishing, Technology, and Theater.

The amount of money the participants pledged ranged from $5 to $250 (M: 70.63, SD: 66.25). The participants backed 7.71 projects on average (SD: 17.42, Max: 68, Min: 1) with eleven participants backing multiple projects while 3 participants only backed one project. The participants' occupations include quality assurance engineer, program director, undergraduate students, graduate students, and professors. Participants' ages ranged from 18 to 41 with 8 females and 6 males. The participants experienced 6.07 months of delay in average (SD: 5.97) ranging from a few weeks to almost two years.

Procedure

We recruited participants through snowball sampling, which allowed us to find participants who have experienced delays. Through semi-structured interviews we asked participants about their profession, three most recent projects they have backed, and their expectations when backing the projects. We asked trust related questions derived from [39], including reliability, honesty, competency, and trustworthiness, which are often cited as base concepts for trust [38, 31, 26, 32]. We also asked about delay experience, their tolerance of delays, as well as importance of on-time delivery. Average interview length was 30 minutes. Interviews took place in person, over the phone, and through video call.

Analysis

We used selective coding and analysis to understand factors influencing trust when delays happen [45]. First, we flagged each instance related to delays, expectations, trust, and creators. After identifying all of the instances, we clustered instances into conceptual categories. Initial coding began after the collection of seven interviews. The emergent themes and the evidence that supports these themes are described in the Interview Results section.

INTERVIEW RESULTS

From our interviews with backers who have experienced delays, we found that factors such as (1) perceived backer role identity; (2) creator's communication strategies during delays; (3) duration of delay; (4) backers' domain knowledge; (5) backers' previous backing experience and purpose of using crowdfunding; and (6) creator's background and experience, all influence backers' trust when delays occurred (Figure 2). In the remainder of this section, we will discuss each of these factors in detail.

Variable	Mean	St. Dev.	Min	Max	Median
# Backed					
Supporter	4.25	1.5	2	5	5
Investor	15.2	29.52	1	68	2
Consumer	3	1.87	1	6	3
$ Pledged					
Supporter	70.09	66.56	10	199	40
Investor	66.55	75.37	5	250	35
Consumer	75.7	61.94	10	169	57.5
Delays					
Supporter	4.07	2.78	0	7	5
Investor	5.25	5.75	0	12	3
Consumer	8.63	7.78	0	24	7.5
Year					
Supporter	-	-	2012	2013	2012
Investor	-	-	2012	2015	2014
Consumer	-	-	2013	2015	2014

Table 1. Summary statistics for number of project backed (# Backed), amount of money pledged ($ Pledged), months of delays experienced (Delays), as well as in which year they first backed a project (Year).

Backers' role identities affected their perception of delays

Our analysis revealed three main role identities for backers, namely *supporters*, *investors*, and *consumers*. Table 1 shows the summary statistics for number of project backed, amount of money pledged, duration of delays (in month) experienced, as well as in which year they first backed a project for different backer roles. Backers identifying with these roles responded differently to delays. Below we discuss the characteristics of each role:

- *Supporters*. Supporters identified as community members on crowdfunding platforms who are willing to help and support other campaigns. Most of them began backing projects in the early development of crowdfunding (around 2012-2013) when crowdfunding platforms were not as popular as they are now. Supporters understood crowdfunding platforms as a place where creators can pursue their innovative ideas. They used the words "contribute", "support", and "help" to describe their backing behaviors. Supporters backed more projects than other type of backers and experienced shorter duration of delays. This is probably due to the fact that most of them have been in crowdfunding platforms longer than other type of backers and gained crowdfunding knowledge over time.
- *Investors*. Investors were those who see crowdfunding campaigns as start-up projects and were willing to invest their money and take risks. Investors mostly are tech-savvy people, either studying or having background in science or

engineering. They went through more rigorous processes when deciding which projects to back. They frequently used the word "invest" to describe their backing behavior. Similar to supporters, investors experienced relatively short duration of delays. Perhaps, due to their rigorous process in selecting campaigns, their median number of project backed was less than supporters and consumers.

- *Consumers*. Consumers saw crowdfunding as an online marketplace where product and delivery were guaranteed. Most consumers were new to the concept of crowdfunding and they were less aware of the risks such as project failure or product delays in crowdfunding platforms. They often used the word "buy" and "purchase" to describe their backing behavior. Consumers experienced longest delays and that may have be caused by unawareness of risks and lack of crowdfunding knowledge. This may have led them to spend more money on campaigns than other backer types.

Investors and supporters were generally more tolerant of delays and retained trust in the platforms despite delays. They didn't express the loss of trust even after experiencing delays since they understood the difficulties of new product development. A supporter who funded a weather app described his expectations: *"It's not a buyer-seller relationship, it's framed as backer-creator relationship. And I think that's emotionally really different. Because one it tells you that the creator is creating these things, it doesn't exist yet. And you know when someone is creating something that doesn't exist, there is no guarantee it can or will exist."* -P2. Investors and supporters also seem to understand the unexpected roadblocks that creators might encounter along their campaigns as one participant who understands the difficulty of tech project said: *"I understand that sometimes there will be a little hiccups"* -P3.

Investors' decision-making process for which projects to back was more rigorous than that of supporters. They would check the creator's previous experience through online professional networks and assess the feasibility of the project. For example, they would Google a creator's name to check their background, such as current or previous employment and job, and their prior experience in crowdfunding projects, before deciding to back a project. In contrast, supporters would just "support" a product if it seemed "cool" and relevant to them.

Some participants changed their perceived role from supporter to investor due to long delays. A backer who had seen himself as a supporter explained how the delay affected his role change:

> "....I really should've looked up what else they've done or look at the people who were leading the campaign. I wish I had done that. That would have been like a nice one little step of research before than just blindly throwing money at it. So definitely has made me a lot more cynical...I have to think a lot more about who is doing it, what are their qualifications, I basically would do research on them as if I was an investor. Which I kind of don't like. I think with crowdfunding, like I do wanna to be more casual, I don't wanna feel like I have to work and think like an investor here. I like to just be able to say, I think this is really cool, I like what you are doing, I assume that you are doing it, here is something to help about." -P3

With a better understanding of the inherent uncertainty of design and manufacturing, supporters and investors typically backed crowdfunding projects for non-essential products, such as microwave removable notebook or laser cutting arm, that they didn't need immediately. A supporter who backed an indoor sensor projects described his decision-making process for backing creators:

> "You know I am trying to think of a scenario where I would back a project on Kickstarter because they were creating something that I had an imminent need for on schedule. If there was something like, 'oh, this could really help my [current project on a deadline], well then I don't want a delay'. But I can't really think of a scenario where any reasonable person would use Kickstarter project in that way." -P2

Supporters or investors would back projects or products for entertainment instead of essential goods in their daily life, since they were aware of the risks of delays and the lack of guarantee of receiving the products.

Consumers, on the other hand, were less aware of risks and uncertainties of products being delayed or even made. A backer who pledged $30 to a cardboard tech project explained: *"I guess I thought it was just like more like buy a normal product, right? It's like, you sign up for it, if enough people do it, then they have the capital to make the product, and you get the product. I didn't really realize there's like continual [delays], maybe, I didn't realize how much risk and uncertainty was it."* -P6. Prior experiences in crowdfunding informed expectations for interactions, and consumers often lacked prior crowdfunding experiences.

Many consumers were also newcomer to crowdfunding who perceived crowdfunding platforms as online marketplaces such as Amazon or Ebay, where they would order products and the products would be shipped to them within a promised delivery date. Some newcomers ordered the products for close friends and family for special occasions, such as Christmas or birthdays and failed to receive the product by the date of the special occasion. A backer who backed a cardboard programmable kit described his disappointment in delays

> "The [project] really bugged me. Cause it was like, it was the Christmas present for the kids. So originally it was gonna be Halloween, we couldn't do that, then it's gonna be Christmas, we couldn't do that. So it's kind of like I told my kids that I will give them this thing and it's just sitting there we can't play with it. So that's like really frustrating. It was a gift, so it made it bad." -P6

Not only was the backer personally disappointed, but he perceived the delay as negatively impacting his plans and promises.

When consumers were asked about whether or not they would back some projects in the future, participants reported being

less likely to back something in the future, at least in the same project type of the delayed project (P1, P4, P6). A backer who was not aware of the risks and uncertainties about delays described how unlikely he would back other projects in the future: *"If it's someone I knew or it's like such a great product that I feel like this thing has to get made, then I would [back other projects], probably not otherwise. Cause it's such a pain."* -P6. Delays influenced intention to engage in crowdfunding in the future; however, personal relations or the desirability of a product may still overcome prior negative experiences.

In summary, backer role identity influenced reaction to delay and the varying degree of loss of trust when there were delays. Those who were more understanding of the risks and uncertainties of crowdfunding platform tended to be more understanding about delays and saw creators as trustworthy even after delays. People who regarded crowdfunding platform as an online marketplace were more worried and annoyed about delays.

Communication during delays were important for maintaining backer's trust

Regardless of the backers' perceived role, they all expressed the importance of communication during delays. Most of the backers said keeping them informed about what was going on during delays increased their trust because updates made them less worried about the possible risks and uncertainties (e.g. further delays, creators bailing out). Also, legitimate causes of delays made them more understanding of the delays.

The frequency of the communication was deemed important to the backers. A backer who backed a laser cutting arm product explained his expectations for the updates:*"It [on-time delivery] is important but I care more that I am always updated if something is going on...Not hearing about something for a year, and getting it randomly, is not acceptable."* -P3. Keeping backers informed whenever the creators achieved major milestones or encountered unexpected roadblocks helped reassure backers regarding the process of the project that they backed.

Although some backers were annoyed by the excessive amount of updates sent from the creators, updating backers during delays and being transparent about the current status were important for backers who trusted the creators and pledged money to them.

Not only was regular communication critical for fostering trust, but rich media of communication also helped backers to build trust. Backers gained trust in creators when creators sent behind-the-scene videos or photographs in their updates during delays to show that they were making progress. A backer who backed an indoor sensor project explained how nice it was to receive a video showing bits and pieces of the product: *"They were very good about keeping the backers informed. They made a really good [behind-the-scene] videos and had relatively frequent updates. I thought that was really good, was a nice way to at least communicate that they care and they were working towards something, even it is not coming together as quickly as they originally hoped."* -P2. Videos and pictures of products or trips to manufacturing sites seemed to help backers gain more trust during the delays, since those could be seen as evidence of progress.

Backers valued not only the frequency and the medium of the communication, they also valued understanding the cause of the delay (e.g. manufacturing and shipping). Backers wanted to know what was going on with the project and what was causing delays. A backer who backed a hardware project explained how he wanted to know about the cause of the delays: *"If something is wrong, I'd like to know what it is. And just let me know that this is gonna take another three months because of this. Then I am fine with it."* -P3. Explanations for the causes of the delays provided backers with an understanding of the reasons for the delays and helped backers set expectations for the extended delivery date for the products.

Backers perceived some explanations as more legitimate than others. For example, backers were more tolerant of external factors that were out of creator's control and saw such delays as inevitable (P2, P6, P12, P14). A backer who waited 4 months to receive his hardware product explained:

> "...but it also has to do with what are [the creator is] communicating with me. If [the creator] says, we ran into some hardware production issues, and we didn't like the quality that we are getting from the manufacturer, and we are working with them to deliver the best product possible. I say ok, I understand that...." -P2

Another backer also explained how one project, which was perfectly on time in manufacturing and shipping, got delayed because the day after the products shipped to the Seattle port from overseas, the port employees were on a strike and the products got stuck at the port until the employees returned to work (P12). Most backers understood the delays and even sympathized with the creators when the causes of the delays could not have been predicted or were beyond creators' control.

Backers did not perceive explanations as legitimate when the reasons provided were in the creator's locus of control. For example, a backer explained his possible reaction to a creator who would take a long vacation prior to the delayed product launch. *"...um, so it is really all about what your excuse is. Because some excuses are legitimate and I can understand that... and if you say 'I am gonna try to have this thing for you in August', and in July you take a month long vacation, and then I am not gonna be thrilled."* -P2.

Sometimes, legitimacy of the explanations was also related to the timeline of the project. One backer who backed a card game explained how the delay was acceptable because the shipping company made a mistake with shipping labels: *"for the [project], that one was interesting. So, they had super well run campaign, and then at the very last minute, um, they had bought from certain manufacturer gone to Amazon to send out, and then Amazon has screwed up labels on, so stuff got delayed"* -P6. Backers were more understanding and forgiving for the delays since the delays were not the creator's fault and also gave credit to the creator who had been running the

campaign very well and had everything kept within the timeline of the initial plan.

In summary, during delays, backers need assurance that the products they have backed are still in progress. Creators need to effectively communicate causes of delays and report their current status to maintain backers' trust. Backers felt creators were more trustworthy when they showed evidence of progress using pictures and videos than with mere text updates. They were also more understanding and tolerant of delays when the causes of delays were out of the creator's control and were clearly communicated.

Duration of delay influenced backer's experience and trust

Backers' thresholds for what they considered to be tolerable delays depended on their perceived role. Delays lasting up to a few weeks were deemed acceptable by all backer types. Consumers were the least tolerant of delays, while supporters and investors were much more tolerant. Although we couldn't quantify based on our interview results what the maximum length of delay deemed tolerable was to each backer, we found that even supporters and investors were frustrated and annoyed when delays lasted more than a year. They viewed long delays as evidence of poor planning and execution on the creator's part: *"You definitely don't want it to be like a year delayed...but I think if as a whole they are pushing everything, for a little bit, for a few months, I don't really mind that cause I won't notice."* -P4.

Long delays could make backers gave up on the creator and the project. One backer of a hardware project experienced a two-year delay described how he almost gave up on the project: *"so the [project] has been delayed for I think like two years now. Which is really really frustrating. Like I don't even expect to get it almost. I think it's coming, but it's just been really obnoxious."* -P6.

Some backers did not care about the delays as much since the estimated delivery date was usually far from when they backed the project (P4, P5, P10). A backer who supported a fashion project explained how long period of time between backing project and receiving products would make her care less about the delays: *"I don't think that [delay] was ever a big thing for me. Cause it's kind of like it's in six months or nine months [from when I backed till I get the product], that just feels far away. Plus you are not gonna be thinking [about the project]..."* -P4.

Backers used their domain knowledge to estimate delays

Backers used their own expertise and domain knowledge to assess the feasibility and difficulty of the project, which also allowed them to assess creator's estimation for delivery. Backers were more tolerant of delays if they perceived the projects as challenging. Further, backers used sources such as peer's opinion or media coverage to understand the difficulty of completing the project.

Backers were more understanding of the delays if the projects appeared technically challenging or required inventing something new that was not currently out in the market. A backer who backed a solar power bike lock project said that he was more tolerant of the delays because of the difficulty of the project: *"for [project], I mean I am not that annoyed [by delays] because I think that one is like really challenging piece of engineering that they are doing. So, in some sense, delays are more understandable. They are actually trying to invent a piece of hardware, electronics, that hasn't been created before."* -P6. The annoyance of the delays was mitigated by backers' understanding of the technological difficulties of the project.

Backers used different ways to estimate the level of difficulty of the project, which could be a mixture of their own domain knowledge, news coverage, comparison to current market product, as well as peer's opinion. A backer who backed a kinetic energy power charger project mentioned how she measured novelty and complexity of the project: *" [I assess novelty and complexity of the project by looking at] what's out there right now, my knowledge on how hard that thing they are trying to build, and maybe a little word of mouth like if people around me in that fields say 'woo, yeah, that project is a tough one'."* -P7.

Some backers also had difficulties in estimating the feasibility of delivering products on time due to the lack of expertise in that domain. One backer who had no hardware knowledge said why it was difficult for him to assess the estimated delivery date given by the creator: *"...Whereas, I don't have any hardware manufacturing experience, neither did the people who did the [project], so..I don't have a good way of assessing their time estimate."* -P2.

To sum up, backers use various sources to assess the achievability of the project as well as the promised delivery date set by the creators. The more domain knowledge that backers have in the project type seems to help backers to foresee how feasible it is for the creators to deliver their products within the given promised delivery date. Backers also adjust their expectations on on-time delivery based on such assessment, high expectations for the delays would influence less of their trust when there are delays.

Research on creator's background and experience informed backers' estimation of delays and trust

Backers not only used their own domain knowledge to estimate delays, but also used creator's credentials to estimate delays. Backers stated that they used different sources to measure a creator's credentials, including the creator's background and previous experience, whether or not the creator's projects were featured by the crowdfunding platform, and believing in wisdom of crowd by looking at how many other backers had backed the creator's project. Backers thought the more trustworthy the creators were, the higher chances the creators would deliver products on time.

Backers had lower expectations towards on-time delivery when they deemed the creator as lacking in experience or expertise. A backer who backed a weather sensor project stated: *"Whereas, I don't have any hardware manufacturing experience, neither did the people who did the [project name],*

so..Even before they said they were gonna make it in August, it was pretty clear that they weren't gonna make it." -P2.

In contrast, backers were more trusting of creators who they felt had the relevant experience and expertise. Similar to other context such as online dating [16], backers also searched creator's name on the web to verify their credibility. A backer who backed a weather software explained how the creator's credentials and background increased his trust:

> "... if you have some people make software and they say 'one of us used to work at Google' [and] 'one of us used to work at Apple' as designers and developers we really know what we are doing, and we can deliver'. And I would say 'ok.' And if someone says 'I am an independent singer song-writer, I've been recording videos on my computer and posting them on Youtube and people seemed to like it and I wanna make a studio album'. I am gonna say 'I don't know if I wanna be involved in that.'" -P2

Creator's credentials influenced backers' estimation of delivery as well as decision making.

In addition to expertise, some backers perceived creators as more trustworthy and capable of delivering product on-time if the crowdfunding platforms featured their projects. A backer explained his preference to back projects that were featured on the crowdfunding platform: *"...Usually the bigger projects have been vetted [by Indiegogo]. Like if you have large campaign in Indiegogo, Indiegogo will reach out to you [creators]."* -P5. Backers believed that featured projects were more trustworthy and they also seemed to believe popularity as a signal for trustworthiness.

Trust towards creators was also influenced by how many people had backed their projects, the backers took crowd's judgement as a way to measure creator's capability. A backer who backed several hardware projects said: *"I will just browse through things, I usually go to more popular things, just because it's more trustworthy, and it's easier to find. I don't really [have] criteria necessarily. It's just my subjective evaluation."* -P5.

In summary, backers use different sources to measure creator's credentials and that influence their estimation of the delays. Creator's credentials are inferred by their previous experience (e.g. job and crowdfunding campaigns), endorsement from the crowdfunding platforms, and other backer's backing behavior.

QUANTITATIVE STUDY

From our qualitative study, we found that backers, depending on their perceived role identity, are influenced by varying degree of reward delays. Also, they use different measures to estimate delay time, but it seemed that less experienced backers or backers who lack domain knowledge have difficulties in estimating delays. Motivated by our previous findings, we seek to explore project attributes that are associated with duration of delays so that such attributes can be used to help not only creators to prevent delays when they are planning their campaigns, but also help those backers who lack domain knowledge to better estimate potential delays.

Dataset

We combined the dataset originally collected by [13] with the dataset that was provided from the owner of Kickspy.com and TheKickbackMachine.com, websites that capture projects as they are launched. The dataset from [13] has 59,233 successfully funded projects between May 2009 to October 2015, and after matching with the Kickspy and TheKickbackMachine dataset (from July 2009 to October 2013) we were able to get 17,457 successfully funded projects with at least one update or one comment. We combined the datasets for two reasons: 1) Kickstarter changed their user interface so that we are no longer able to see list of backers for each campaign and the dataset provided by [13] has them, 2) The dataset from Kickspy and TheKickbackMachine has campaigns' updates and comments from backers and we needed the updates and comments to identify whether or not campaigns have been delayed. We periodically scraped updates and comments for all the projects and the scraping lasted until May 5, 2015.

Unit of Analysis: Delayed projects

The unit of analysis in this paper is all projects which have been delayed. We used the words "sorry", "apologiz(s)e", and "delay" in updates and the words "delay", " have not receive", "sorry", "apologiz(s)e" in comments to identify delayed projects. After this process, we were left with a dataset of 4,089 projects that were identified as delayed. In terms of the duration of delays, we calculated days passed since the latest estimated delivery date (often times projects have different reward tiers with different delivery dates) given by creators to the latest date either from the matched updates or the matched comments.

With 4,089 projects, we randomly selected 50 projects and manually checked whether or not the projects were actually delayed, as well as the duration of the delays. All the projects that we identified as delayed were indeed delayed. In the time of our visit to the campaign websites, 9 projects still did not deliver the products. Among 41 projects that have delivered products, the mean time difference between our delay length and actual delay length was 3.6 months (SD: 6.72). We found that some of the larger discrepancies were due to the campaigns for which creators delivered much later than May 5, 2015, on which our dataset did not capture comments nor updates.

Measures

Number of reward levels specifies the number of different rewards a project creator provides. Each level is associated with a pledge amount. For example, in a smart tumbler project, the reward for a $30 pledge might be a tumbler that recognizes caffeine while the reward for $99 pledge might be a tumbler that recognizes any liquid. Manufacturing is one of the issues causing delays [10] and having more reward levels leads to different batches for manufacturing which may cause more delays.

Funding goal is the amount of money that a project creator has to raise in order to get the money for their project. The

Variable	N	Mean	St. Dev.	Min	Max
Dependent Variable					
Duration of Delays (*delays*)	4,089	224	314	0.004	16,095
Independent Variable					
Funding goal (*goal*)	4,089	21,999	54,411	1	1,100,000
Project duration (*projectDuration*)	4,089	125.4	112.7	0	793.2
Percent raised (*percentRaised*)	4,089	13.8	650	1	41,535
Number of backers (*backers*)	4,089	826	2,759	2	87,142
Number of reward levels (*rewards*)	4,089	13.8	8.12	2	126
Number of projects backed (*backed*)	4,089	15.87	31.12	0	488
Number of projects created (*created*)	4,089	2.37	4.53	1	65

Table 2. Summary statistics for variables used in our analysis.

funding goal can be as small as a few hundreds dollars to upwards of millions of dollars. Kickstarter platform uses an all-or-nothing crowdfunding model where the project creators either get all the money that they raised if the amount is above the funding goal after the campaign period, or nothing if the amount is below the funding goal. Since many creators are novice entrepreneurs [22] who typically face difficulties in dealing with scaling issues, larger funding goals means having to produce more rewards which may lead to longer delays.

Project duration is time between a campaign end date and an estimated delivery date. It can be used as a proxy measure for the complexity of a project. This suggests an additional hypothesis that long project duration may yield longer delays.

Number of backers is the number of people who have backed the project. Having a large number of backers will require creators to ship larger numbers of rewards, which we hypothesize will lead to more delays.

Percent raised is the ratio between how much the creator has already raised and the funding goal. Because most project creators are novice entrepreneurs who lack effective planning strategies [22], we believe that largely exceeding their initial funding goals will require them to adjust their planning, which may lead to longer delays.

Number of projects created indicates how many projects the creator has launched before the current campaign. We use the number of projects launched as a measure of the creator's level of crowdfunding experience. Previous studies showed that creators learn crowdfunding skills over time [22] and learn from failures [18]; therefore, we hypothesize that more experienced creators will cause less delays.

Number of projects backed indicates how many projects the creator has backed before the current campaign. We hypothesize that creators who have backed other projects and have experienced delays may better understand the unpleasant experience with delays and thus seek to reduce delays in their own projects.

Project types are categories for projects. By default, Kickstarter has 15 project types, including Art, Comics, Crafts, Dance, Design, Fashion, Film & Video, Food, Games, Journalism, Music, Photography, Publishing, Technology, Theater.

Project type	N	% within DP	% DP within type
Art	156	3.82%	11.02%
Comics	266	6.51%	39.88%
Crafts	25	0.61%	15.63 %
Dance	6	0.15%	1.9%
Design	632	15.46%	55.54%
Fashion	164	4.01%	27.38%
Film & Video	588	14.38%	16.12%
Food	169	4.13%	22.18%
Games	876	21.42%	64.27%
Journalism	12	0.29%	16.22%
Music	423	10.34%	10.82%
Photography	57	1.39%	15.24%
Publishing	310	7.58%	18.66%
Technology	356	8.71%	60.03%
Theater	40	0.99%	5.24%

Table 3. The number (N) and distribution of the delayed projects (DP) across types and the percentage of delayed projects within each type.

Hypotheses

We hypothesize that an increase in *Number of reward levels*, *Funding goal*, *Project duration*, *Number of backers*, and *Percent raised* will lead to an increased duration of delays, while an increase in *Number of projects created* and *Number of projects backed* will lead to a decrease in duration of delays.

Analysis

Our analysis strategy involves the estimation of regression models on duration of delays. Our formal model is shown below and includes the full set of variables:

$$delays = \beta_0 + \beta_1 rewards + \beta_2 goal + \beta_3 projectDuration$$
$$+ \beta_4 backers + \beta_5 percentRaised + \beta_6 backed$$
$$+ \beta_7 created + \beta_8 type$$

We used negative binomial regression because duration of delays is an over-dispersed count with conditional variance that exceeds the conditional mean [43].

QUANTITATIVE STUDY RESULTS

Descriptive analysis results

Our descriptive analysis results are shown in Table 2. The mean duration of delay is 7.4 months and the mean funding goal was $21,999. The mean number of reward levels in delayed projects was 13. On average, projects raised 13.8 times of their initial funding goals and had 826 backers. Creators had backed 15 projects and created 2 projects on average. The distribution of the delayed project types are shown in Table 3.

Regression analysis results

Our regression results in Table 4 show that the number of reward levels and the funding goal help explain variations in project delays. The number of backers, percent raised, the number of projects created have a small, but statistically significant association with delay duration as well. Below, we present these results in greater detail using marginal effects for a typical project to facilitate interpretation. For the marginal effects, we hold continuous covariates at their median values and then report the impact of shifting between the 10^{th} and 90^{th} percentiles for statistically significant independent variables. We choose Design as the project type to interpret the regression results because Design projects often face most of the delay problems, such as manufacturing, shipping, design changes, that are mentioned in the previous findings [10].

In our model, the coefficients for funding goal and the number of reward levels are positive and statistically significant. In other words, larger funding goals and larger number of reward levels are associated with more delays in the project. This finding is in line with previous qualitative findings that creators would have difficulties managing and scaling their projects [22, 10]. Figure 3 visualizes predicted duration of delays across different funding goals. For example, with other variables set at the sample median, our model estimates that a design project with $2,176 as their funding goal (10^{th} percentile) would be expected to have 200.38 days of delays (about 6.5 months); a project with $55,000 as its funding goals (90^{th} percentile) would be expected to have 216.42 days of delays (about 7 months).

Figure 4 illustrates the positive relationship estimated by our model between duration of delays and the number of reward levels. For example, setting other variables at the sample median, our model estimates that a design project with 6 reward levels (10^{th} percentile) would be expected to have 197.94 days of delays (about 6.5 months); a similar project with 18 reward levels (90^{th} percentile) would be expected to have 214.38 days of delays (about 7 months).

The model also estimates that duration of delays increases as the number of backers and percent raised increases, but the point estimates are very small and the relationship are not substantively important. For example, setting other variables at the sample median, we expect that a Design project with 103 backers (10^{th} percentile) would have 202.23 days of delays (about 7 months) whereas a similar project with 2186 backers (90^{th} percentile) would have 207.78 days of delays (also about 7 months). In other words, more backers may help explain variation in delay duration, but the observed association is so small as to be almost meaningless.

	Dependent variable
	days delayed
rewards	0.0067^{***}
	(0.0025)
goal (in thousands)	0.0015^{***}
	(0.00028)
projectDuration	−0.00011
	(0.00014)
backers	0.000013^{**}
	(0.000005)
percentRaised	0.0000079^{***}
	(0.0000022)
backed	−0.00018
	(0.00049)
created	-0.0073^{*}
	(0.0043)
project type fixed effects	Yes
(Intercept)	5.239^{***}
	(0.047)
N	4,089
Log Likelihood	−26,096.923
θ	1.101^{***} (0.022)
AIC	52,240
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4. Results of negative binomial regression model estimating days delayed.

The coefficient for number of projects created is negative and marginally significant in our model, indicating that the experience of creating projects is associated with shorter delays. Our model estimates that a design project with a creator who had created one prior project (10^{th} percentile) would be expected to have 203.27 days of delays (about 7 months) whereas an identical project initiated by someone who had already created four prior projects (90^{th} percentile) would be expected to have 198.90 days of delays (about 6.5 months).

DISCUSSION

Greater uncertainties and risks in inventing or creating a new product set crowdfunding platforms apart from other online marketplaces and donation platforms [22]; thus backers' expectations towards delay and trust in situations of delays also differ from other platforms. Unlike in online marketplaces where shipping is one of the main problems for delays, there are other issues such as design, development, manufacturing, and shipping in crowdfunding. As a result, backers who are less aware of such uncertainties and risks are generally less aware of the possibility of delays. Those who are more aware of such issues use their own domain knowledge to expect and estimate potential delays. In situations of delays, factors such as 1) backer role identity, 2) communication strategies during delays, 3) duration of the delay, 4) backers' do-

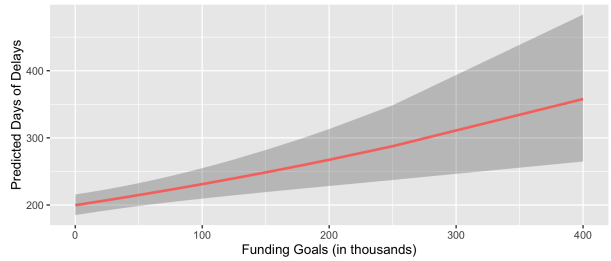


Figure 3. Estimated values for design projects showing the predicted days of delays using our model. X-axis is the funding goal while holding all the other variables at the sample median.

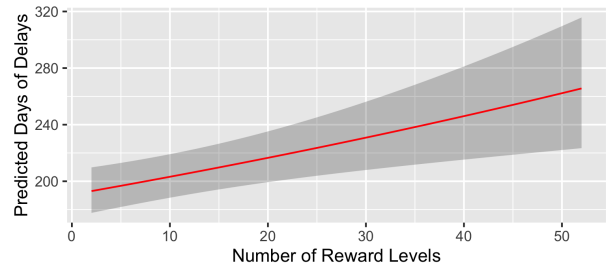


Figure 4. Estimated values for design projects showing the predicted days of delays using our model. X-axis is the number of reward levels while holding all the other variables at the sample median.

main knowledge, and 5) research on creator's credentials influenced backers' expectation and trust.

Although previous research on crowdfunding sought to understand predictive factors for successfully funded projects [19, 13, 50], few studies have considered delivery delays. Being able to estimate delays a priori can help both creators manage campaigns and backers manage expectations. Our regression results show that the number of reward levels, funding goal, creator's previous launching experience, project types, and number of backers help explain the duration of delays.

In the rest of the section, we discuss design considerations that may help backers in the process of decision making and adjust their expectations for potential delays. We then discuss implications for helping creators to plan their campaigns to minimize their delays, and supporting creators to take necessary actions when delays occur. Finally, we discuss how we can improve existing interface designs or policies on crowdfunding platforms to provide backers a better experience.

Implications

Our core findings around backer role identity, communication during delays, crowdfunding policies and accountability have implications for understanding and designing crowdfunding platforms and online interactions more generally.

Backer Role Identity

Backers have different perceived roles that informed their views on delays. Both *investor* and *supporter* type of backers understand the risks and uncertainties in crowdfunding platform and are much more tolerant of delays. In contrast, *consumer* type of backers often see crowdfunding platforms as typical online marketplaces such as Amazon or eBay, and they are less tolerant of delays and might yield high attrition rate.

As crowdfunding is becoming more and more popular [30], crowdfunding platforms should be sensitive to backers' *crowdfunding literacy*. While platforms are beginning to include notices about the underlying risks (i.e. creators fail to deliver product, late delivery) or the primary intention of their platforms (i.e. helping creators to realize their dream; but there is no guarantee they would succeed), literacy for many platform participants remains low.

Consumers in particular need help to understand the risks and uncertainties based on which to adjust their expectations accordingly. For example, a backer support tool which shows potential delays and delay estimations based on project attributes might help a backer's decision making process. Also, reputation systems based on creators' previous delays and failures can be useful for helping backers find "more trustworthy" creators.

Although crowdfunding platforms emphasize "crowd due diligence" [2], which expects backers to research the project and the creator before backing, we argue that there should be more information and previous crowdfunding histories shared within the platforms so that the backers could do less additional work (e.g., Google a creator's name) outside the crowdfunding platforms.

Our findings suggest that different role identities have varying trust attitudes even within the same context or domain, and socio-technical solutions need to meet different user group's needs. This builds directly on prior work on the importance of not only context [9] but also on different user groups in the same context. Designers often overlook the implications of the extra work required of other user groups in multi-user applications [20]. To foster trustworthiness in the system, we need to design interfaces that provide information tailored to different user groups to help them best make informed decisions without having to perform additional work.

Communications during delays

Effective communications during delays was deemed important for backers as it: (a) helped them resolve information asymmetry; (b) provided reassurances given uncertainties; (c) allowed them to set their expectations for possible upcoming roadblocks of the projects. Our findings show the richness of the communication medium, the frequency of the updates, and the legitimacy of reasons for the delays influenced backers trust.

Crowdfunding platforms can provide ways to remind creators to upload videos or pictures when cases of delays are detected through campaign updates or comments. Backers found creators to be more trustworthy when the creators showed behind-the-scene videos or pictures as evidence that they were making progress. However, some backers did mention that sometimes excessive updates felt like spam. Creators, therefore, should send updates sparingly by sending updates whenever they achieve a certain milestone or hit a roadblock but have detailed information to provide.

In-process communication can mitigate information asymmetry problems among people who have mistrust or potential loss of trust. Reputation systems are often used for prediction about future behavior during pre-transaction or pre-interaction stage [25]; we believe that in-process communications can be used as another risk management technique during the interaction process.

Duration of delay

Duration of delay also influences backers trust especially when there are long delays. Short delays such as a few weeks or a few months were considered by backers as within the margin of errors of the creators' estimated delivery dates, but long delays made backers question the competency of the creators or the preparedness of the project itself.

Similar to previous research tools that help creators predict their success in raising fund [19, 13, 50], crowdfunding platforms should help creators to estimate potential delays and the duration of delays with their campaign attributes. Minimizing delays can help creators who want to have subsequent projects launched on the platform avoid losing potential recurrent backers with long delays they inadvertently create.

However, we also have to consider the creator's needs when designing a support tool for creators to prevent delays or mitigate consequences of delays. For example, creators may not follow advice to reduce the funding goal to minimize delays, since they may prioritize additional funds over some delays caused to backers. Without compromising creators' needs, one suggestion is to create support tools to assist creators to set more conservative estimated delivery dates for reward types preferred by consumers and conservative number of rewards based on manufacturing capabilities. Further, since the tolerance of delays differ across backer roles, in cases of delays a support tool may assist creators to prioritize shipping to consumers over investors or supporters. However, creators should be very transparent and honest about adopting such practices to all the backers.

Crowdfunding Policies and Accountability

Crowdfunding platforms specifically claim that they are not liable of projects' performance or punctuality in their terms of use, but who should be held accountable when there are delays or even failures to deliver products? Some backers mentioned that they are concerned about accountability, and suggest that crowdfunding platforms should give more assurances to backers by overseeing projects.[1] As one backer expressed: *"I am mad if something does go wrong, I am kind of mad at them [creators], but it's also like you put your trust in them, it's not they guaranteed to do, they are supposed to, but they are not, they could fail...I think, my only issue is that Kickstarter and Indiegogo does nothing to help [such as business plan or progress check] with them, when they should."*

[1]A week before the submission, Indiegogo announced to team up with company called Arrow to help with design tools, prototyping services, manufacturing support, supply chain management assistance for hardware projects to deliver products in a timely fashion. See http://techcrunch.com/2016/05/20/indiegogo-teams-up-with-arrow/

Due to the lack of liability in crowdfunding platforms, some people would prefer to wait until a crowdfunding campaign is successfully funded and the creators launch their own website to purchase the product. They feel much more comfortable purchasing from the website since they are protected by consumer rights. As one participant mentioned: *"I think that you want people to continue using crowdfunding, but I think that I wanna be the person who buys it from the website after it's funded [because it felt more legitimate]..."*

If many backers start following such strategies, however, less and less projects will be able to reach their funding goals since people are waiting for others to take risks. As a result, less creative and innovative projects may be pursued and realized. Crowdfunding platforms, instead of merely stating they are not responsible for any disputes between creator and backer, can develop reputation systems with reward or penalties with regard to delay and failure that can help to mitigate potential dispute by providing backers more information about creator's crowdfunding history as well as hold creators more accountable for their actions.

Platforms' policy should benefit both sides of agents and especially provide risk management mechanisms for parties that are more prone to risks. This applies not only to crowdfunding platforms, but also to other platforms such as e-commerce, crowdsourcing marketplaces [33]. One starting place may be to design solutions that help both sides assess the risks and benefits involved, so that outcomes on the platform more realistically represent the wishes of parties to avoid market inefficiencies and failures due to the lack of helpful information for making informed decisions.

Limitations and future work

In this work, our study of trust on crowdfunding platforms is limited to situations of reward delays. Future work will attempt to build a deeper understanding of issues of trust on crowdfunding platforms more generally.

Another limitation of our dataset is the estimation of the duration of delays. Estimation errors were either due to the lack of updates or comments or nuanced comments that were difficult to detect with text matching. Therefore, in addition to our negative binomial regression analysis to estimate days of delays, we also used logistic regression, which is more robust to such errors, to check predictors for high delays in the duration of delays. Predictors for delays in both regression results were identical except *backers* and *created* were not significant in logistic regression. We need further investigation to understand this difference.

Lastly, within reward-based crowdfunding platforms, there likely exists backers who may support their friends, family, or weak-ties solely due to altruism. However, understanding altruism, the practice of acting generously towards others who have been generous to them, or anyone, in the past, was out of the scope of this paper. As people are more willing to trust an altruistic person than a non-altruistic person [6], future research will explore how reputation of 'altruistic behavior' might influence trust of creators.

CONCLUSION

In this paper, we studied backer's trust in situations of delays and predictors of delays in crowdfunding platforms. We found that a rich set of crowdfunding specific factors that influenced backers' trust, including backers' role identity as well as domain knowledge, backer's research on a creator's background, creator's communication during delays, and the duration of delay. Our regression analysis with 4089 delayed projects showed that the funding goal, number of backers, number of reward levels, creators previous crowdfunding experience, and project types are predictors of for the duration of delays.

This research contributes to our knowledge about trust in situations of delays in crowdfunding and provides initial insights for how we should design future computer supported cooperative systems for crowdfunding. These insights are important for expanding the number and breadth of people who participate and the new ventures realized. This work sets the stage for new CSCW research in trust in emerging CSCW contexts.

By helping creators to reduce delays and building trust among creators and backers, more backers may contribute to crowdfunding projects to help novice entrepreneurs, engineers, designers to bring innovative and creative ideas to the world.

ACKNOWLEDGEMENT

We thank members of the Delta Lab for valuable feedback and discussions throughout the project. We thank Julie Hui and Mike Greenberg for their help in shaping and pursuing this project.

REFERENCES

1. Paul S. Adler. 2001. Market, Hierarchy, and Trust: The Knowledge Economy and the Future of Capitalism. *Organization Science* 12, 2 (April 2001), 215–234.
2. Ajay Agrawal, Christian Catalini, and Avi Goldfarb. 2013. *Some Simple Economics of Crowdfunding*. University of Chicago Press, 63–97.
3. Tim Althoff and Jure Leskovec. 2015. Donor Retention in Online Crowdfunding Communities: A Case Study of DonorsChoose.Org. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. ACM, New York, NY, USA, 34–44.
4. Jisun An, Daniele Quercia, and Jon Crowcroft. 2014. Recommending Investors for Crowdfunding Projects. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. ACM, New York, NY, USA, 261–270.
5. Sulin Ba and Paul A. Pavlou. 2002. Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly* 26, 3 (2002), 243–268.
6. Pat Barclay. 2004. Trustworthiness and competitive altruism can also solve the tragedy of the commons. *Evolution and Human Behavior* 25, 4 (2004), 209–220.
7. Nathan Bos, Judy Olson, Darren Gergle, Gary Olson, and Zach Wright. 2002. Effects of four computer-mediated communications channels on trust development. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 135–140.
8. Coye Cheshire. 2011. Online trust, trustworthiness, or assurance? *Daedalus* 140, 4 (2011), 49–58.
9. Coye Cheshire, Judd Antin, Karen S. Cook, and Elizabeth Churchill. 2010. General and Familiar Trust in Websites. *Knowledge, Technology & Policy* 23, 3-4 (Sept. 2010), 311–331.
10. CNN. 2012. Why 84% of Kickstarter's top projects shipped late. (2012). http://money.cnn.com/interactive/technology/kickstarter-projects-shipping/
11. Karen S. Cook, Toshio Yamagishi, Coye Cheshire, Robin Cooper, Masafumi Matsuda, and Rie Mashima. 2005. Trust Building via Risk Taking: A Cross-Societal Experiment. *Social Psychology Quarterly* 68, 2 (June 2005), 121–142.
12. Patricia M. Doney and Joseph P. Cannon. 1997. An Examination of the Nature of Trust in Buyer-Seller Relationships. *Journal of Marketing* 61, 2 (1997), 35–51.
13. Vincent Etter, Matthias Grossglauser, and Patrick Thiran. 2013. Launch Hard or Go Home!: Predicting the Success of Kickstarter Campaigns. In *Proceedings of the First ACM Conference on Online Social Networks (COSN '13)*. ACM, New York, NY, USA, 177–182.
14. Diego Gambetta. 1988. *Trust: Making and Breaking Cooperative Relations*. Blackwell.
15. Elizabeth M. Gerber and Julie Hui. 2013. Crowdfunding: Motivations and Deterrents for Participation. *ACM Trans. Comput.-Hum. Interact.* 20, 6 (Dec. 2013), 34:1–34:32.
16. Jennifer L Gibbs, Nicole B Ellison, and Chih-Hui Lai. 2010. First comes love, then comes Google: An investigation of uncertainty reduction strategies and self-disclosure in online dating. *Communication Research* (2010).
17. Mark Granovetter. 1985. Economic action and social structure: The problem of embeddedness. *American journal of sociology* (1985), 481–510.
18. Michael D. Greenberg and Elizabeth M. Gerber. 2014. Learning to Fail: Experiencing Public Failure Online Through Crowdfunding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 581–590.
19. Michael D. Greenberg, Bryan Pardo, Karthic Hariharan, and Elizabeth Gerber. 2013. Crowdfunding Support Tools: Predicting Success & Failure. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 1815–1820.

20. Jonathan Grudin. 1988. Why CSCW Applications Fail: Problems in the Design and Evaluationof Organizational Interfaces. In *Proceedings of the 1988 ACM Conference on Computer-supported Cooperative Work (CSCW '88)*. ACM, New York, NY, USA, 85–93.

21. Julie S. Hui, Elizabeth M. Gerber, and Darren Gergle. 2014a. Understanding and Leveraging Social Networks for Crowdfunding: Opportunities and Challenges. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, USA, 677–680.

22. Julie S. Hui, Michael D. Greenberg, and Elizabeth M. Gerber. 2014b. Understanding the Role of Community in Crowdfunding Work. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 62–74.

23. Jennifer G Kim, Ha Kyung Kong, Karrie Karahalios, Wai-Tat Fu, and Hwajung Hong. 2016. The Power of Collective Endorsements: Credibility Factors in Medical Crowdfunding Campaigns. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4538–4549.

24. Aniket Kittur, Bongwon Suh, and Ed H. Chi. 2008. Can You Ever Trust a Wiki?: Impacting Perceived Trustworthiness in Wikipedia. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, New York, NY, USA, 477–480.

25. Peter Kollock. 1999. The production of trust in online markets. *Advances in group processes* 16, 1 (1999), 99–123.

26. Roderick M Kramer and Tom R Tyler. 1995. *Trust in organizations: Frontiers of theory and research*. Sage Publications.

27. Robert E Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph Konstan, Yuqing Ren, and John Riedl. 2012. *Building successful online communities: Evidence-based social design*. Mit Press.

28. Venkat Kuppuswamy and Barry L. Bayus. 2014. *Crowdfunding Creative Ideas: The Dynamics of Project Backers in Kickstarter*. Technical Report ID 2234765. Social Science Research Network, Rochester, NY.

29. Niklas Luhmann. 2000. Familiarity, confidence, trust: Problems and alternatives. *Trust: Making and breaking cooperative relations* 6 (2000), 94–107.

30. Massolution. 2015. 2015CF Crowdfunding Industry Report. (2015).

31. Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.

32. Daniel J McAllister. 1995. Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of management journal* 38, 1 (1995), 24–59.

33. Brian McInnis, Dan Cosley, Chaebong Nam, and Gilly Leshed. 2016. Taking a HIT: Designing around Rejection, Mistrust, Risk, and Workers Experiences in Amazon Mechanical Turk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2271–2282.

34. Debra Meyerson, Karl E Weick, and Roderick M Kramer. 1996. Swift trust and temporary groups. *Trust in organizations: Frontiers of theory and research* 166 (1996), 195.

35. Tanushree Mitra and Eric Gilbert. 2014. The Language That Gets People to Give: Phrases That Predict Success on Kickstarter. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 49–61.

36. Ethan R. Mollick. 2013. *The Dynamics of Crowdfunding: An Exploratory Study*. SSRN Scholarly Paper ID 2088298. Social Science Research Network, Rochester, NY.

37. Michael Muller, Werner Geyer, Todd Soule, Steven Daniels, and Li-Te Cheng. 2013. Crowdfunding Inside the Enterprise: Employee-initiatives for Innovation and Collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 503–512.

38. Michael O'Leary, Wanda Orlikowski, and J Yates. 2002. Distributed work over the centuries: Trust and control in the Hudson's Bay Company, 1670-1826. *Distributed work* (2002), 27–54.

39. Paul A. Pavlou and David Gefen. 2005. Psychological Contract Violation in Online Marketplaces: Antecedents, Consequences, and Moderating Role. *Information Systems Research* 16, 4 (Dec. 2005), 372–399.

40. Paul Resnick. 2002. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. *Advances in Applied Microeconomics* 11 (2002).

41. Julian B Rotter. 1967. A new scale for the measurement of interpersonal trust. *Journal of personality* 35, 4 (1967), 651–665.

42. Adrian Sargeant, John B. Ford, and Douglas C. West. 2006. Perceptual determinants of nonprofit giving behavior. *Journal of Business Research* 59, 2 (Feb. 2006), 155–165.

43. J Scott Long. 1997. Regression models for categorical and limited dependent variables. *Advanced quantitative techniques in the social sciences* 7 (1997).

44. Robin L Snipes and Sharon L Oswald. 2010. Charitable giving to not-for-profit organizations: factors affecting donations to non-profit organizations. *Innovative Marketing* 6, 1 (2010), 73–80.

45. James P. Spradley. 1980. *Participant Observation*. Holt, Rinehart and Winston.

46. Piotr Sztompka. 1999. *Trust: A Sociological Theory*. Cambridge University Press.

47. Catalina L. Toma. 2010. Perceptions of Trustworthiness Online: The Role of Visual and Textual Information. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, NY, USA, 13–22.

48. Rick Wash. 2013. The Value of Completing Crowdfunding Projects. In *ICWSM*.

49. Rick Wash and Jacob Solomon. 2014. Coordinating Donors on Crowdfunding Websites. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 38–48.

50. Anbang Xu, Xiao Yang, Huaming Rao, Wai-Tat Fu, Shih-Wen Huang, and Brian P. Bailey. 2014. Show Me the Money!: An Analysis of Project Updates During Crowdfunding Campaigns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 591–600.

51. Jun Zheng, Elizabeth Veinott, Nathan Bos, Judith S. Olson, and Gary M. Olson. 2002. Trust Without Touch: Jumpstarting Long-distance Trust with Initial Social Activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 141–146.

52. Roxanne Zolin, Pamela J Hinds, Renate Fruchter, and Raymond E Levitt. 2004. Interpersonal trust in cross-functional, geographically distributed work: A longitudinal study. *Information and organization* 14, 1 (2004), 1–26.

Black Lives Matter in Wikipedia: Collaboration and Collective Memory around Online Social Movements

Marlon Twyman
Technology & Social Behavior
Northwestern University
Evanston, IL, USA
marlont2@u.northwestern.edu

Brian C. Keegan
Information Science
Univ. of Colorado, Boulder
Boulder, CO, USA
brian.keegan@colorado.edu

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

ABSTRACT

Social movements use social computing systems to complement offline mobilizations, but prior literature has focused almost exclusively on movement actors' use of social media. In this paper, we analyze participation and attention to topics connected with the *Black Lives Matter* movement in the English language version of Wikipedia between 2014 and 2016. Our results point to the use of Wikipedia to (1) intensively document and connect historical and contemporary events, (2) collaboratively migrate activity to support coverage of new events, and (3) dynamically re-appraise pre-existing knowledge in the aftermath of new events. These findings reveal patterns of behavior that complement theories of collective memory and collective action and help explain how social computing systems can encode and retrieve knowledge about social movements as they unfold.

Author Keywords

Social movements; civil rights; computer-supported collective action; collaboration network; social computing; user behavior modeling

ACM Classification Keywords

H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces (collaborative computing, computer supported cooperative work); K.4.2 Computers and Society: Social Issues

INTRODUCTION

Contemporary social movements use social computing and social media platforms to mobilize supporters, negotiate meaning, alter their relationship to media gatekeepers, and reframe issues [3, 35]. How does knowledge about a social movement come into being while the movement unfolds? What dynamics characterize participation in social computing systems around ongoing and contentious political events as a movement coalesces?

This paper analyzes how the *Black Lives Matter* movement (*BLM*) has manifested itself within the English language Wikipedia. The English Wikipedia is the largest language edition and includes articles on topics relevant to the current study: breaking news, social movements, protests, and social justice issues. Wikipedia is also a place for collective memory building around historical events, such as the Arab Spring and the Vietnam War [7, 33, 36], as well as unfolding coverage of breaking news [25, 27, 22]. *BLM* provides a case study of a social movement generating breaking news around social justice issues across the United States.

We use trace data from *BLM* relevant articles' revision histories and pageviews on Wikipedia to understand dynamics of attention, collective memory, and knowledge production in response to a movement. Prior work on social movements and social computing has focused disproportionately on Twitter and Facebook use, emphasizing interpersonal interactions, event coordination, and information diffusion dynamics among movement participants. In contrast, Wikipedia exemplifies a collaborative knowledge and memory production system that aims to remain independent of movement influence. Analyzing the temporal changes around movement-related articles offers an ideal perspective on how knowledge about a social movement shifts while it is still in action.

Our results complement and extend existing analyses of social computing activity around social movements as well as prior work on Wikipedia participation dynamics. We find that movement events and external media coverage appear to drive attention and editing activity to *BLM*-related topics on Wikipedia. This activity and attention intensified as the movement gained prominence and momentum, leading to bigger spikes of attention to movement events, faster coverage of new movement topics, expanded discussion, and the emergence of networks of attention and collaboration around *BLM*-related articles. Specifically, we describe three key processes that we observe in Wikipedia around topics related to *BLM*: intensified documentation, collaborative migration, and dynamic re-appraisal.

Our findings show novel patterns of collective memory that occur in social computing in conjunction with social movement events. In addition, we describe novel dynamics of attention and collaboration within Wikipedia while also replicating patterns from prior Wikipedia research in other topic domains. We look only at events connected to a single movement, but the processes we observe may help explain general patterns of collective memory making within social computing systems.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
CSCW '17, February 25 - March 01, 2017, Portland, OR, USA
ACM 978-1-4503-4335-0/17/03.
http://dx.doi.org/10.1145/2998181.2998232

BACKGROUND AND PRIOR WORK

The "Black Lives Matter" movement

BLM emerged in the United States during the summer of 2014 in response to instances of police officers shooting or killing unarmed African Americans. The roots of the movement stretch much further back in time and, for some supporters, encompass the history of racial violence, exclusion, inequality, mass incarceration, and slavery in the United States. However, most accounts of the movement place its immediate origins in Ferguson, Missouri with the death of Michael Brown on August 9, 2014.

Michael Brown, an 18 year-old African American man, was killed in Ferguson, Missouri by Darren Wilson, a 28 year-old white Ferguson police officer. In response to Brown's death, protesters and activist groups adopted the phrase "Black Lives Matter" to metonymize their grievances. This phrase originated in response to the shooting death of another unarmed African American teenager, Trayvon Martin, in Sanford, Florida early in 2012 [13]. As it became clear that the events in Ferguson had provoked nationwide concern, the phrase spread on social media and in the press [10]. Subsequent events from 2014 through 2016 involving the deaths of African Americans associated with police broadened the attention around "Black Lives Matter" as a movement focused on racial injustice and police violence. The political climate and ambiguity surrounding *BLM*'s goals has since generated numerous supporters and critics of the movement.

From the beginning, social media systems played a critical role in the rise and popularization of *BLM* [13]. Twitter has generated a great deal of activity as a place where content and issues specific to the African American community have gained attention [40]. Following the deaths of Martin, Brown, and others, the dissemination of images, videos, and hashtags on Twitter and other social media served a pivotal role leading to the emergence of *BLM* [10, 19]. However, the impact of the *BLM* movement in social computing extends far beyond social media. We seek to understand this broader impact by studying activity in Wikipedia around topics related to *BLM*.

Social movements & social computing systems

Social media platforms have played influential roles in mobilizing participation in social movements. These platforms are designed to reduce coordination costs as well as enhance information sharing, potentially encouraging more people to engage in protest politics than would have otherwise [41]. "Networked counterpublics" deploy novel tactics such as "threadjacking" to garner attention, recruit newcomers, and mobilize supporters [20]. Broadly speaking, these are all instances of *computer-supported collective action* [45]. We investigate a set of related concerns about how computer-supported collective action unfolds in the context of commons-based peer production system: what patterns of attention, engagement, and collaboration emerge in social computing systems engaged in knowledge collaboration around contentious movements?

We extend prior work by investigating participation and attention in a peer production system that was not a primary site of mobilization for the movement under study. Recent studies of *BLM* have relied heavily on data from social media to examine how movement participants curate information and mobilize for action [5, 10, 19, 20]. In contrast, the contributors and readers of articles related to *BLM* on Wikipedia do not necessarily participate in the movement or share the perspectives of movement actors. Thus, rather than focus on social media as a mode of movement participation, we investigate whether and how the patterns of knowledge collaboration and information consumption in an online peer production community shape the representation of a social movement.

Breaking news & current events in Wikipedia

The Wikipedia collaborations around breaking news display unique patterns of activity and engagement within emergent topic spaces. Editors demonstrate remarkable ability to respond to current events: they create articles within minutes of major disasters and catastrophes [25], alter collaboration practices to support high-tempo interactions [26], rapidly regenerate prior organizational structures [27], and adopt differentiated social roles [23]. A broader literature has shown that editors act interdependently assuming nuanced social roles [1, 43], managing conflict dynamics [30, 44], and coordinating their actions to produce high quality content [29]. Other research has approached the study of the Wikipedia community itself through the focus of social movement theory (e.g., [21, 31]). This prior work has considered the dynamics of breaking news coverage on Wikipedia and has described Wikipedia editing as a form of social movement activity, but has not focused on editing activity around social movement events as they are occurring.

Articles related to *BLM* in the English language Wikipedia provide an ideal case to understand more about current social movements through the lens of social computing systems. The movement is a current event, the associated deaths and protests made breaking news, and the topic occupies a domain that bridges different content areas in the platform. Collaboratively producing knowledge about social movements could be dynamic and contentious in ways that other breaking news events are not. Prior research on breaking news in Wikipedia does not consider the need to situate new events with the context of older events. *BLM*, in particular, focuses on the deaths of a minority group in the United States, is embedded within historically contentious issues, and continues to generate new information. This case can extend our understanding of breaking news practices in Wikipedia to include a social movement and related events.

RQ1: How has Wikipedia editing activity and the coverage of *BLM* movement events changed over time?

RQ2: How have Wikipedians collaborated across articles about events and the *BLM* movement?



V·T·E	Black Lives Matter	[hide]
People	Patrisse Cullors · Alicia Garza · Opal Tometi · DeRay Mckesson · Shaun King · Marissa Johnson · Nekima Levy-Pounds · Johnetta Elzie	
Events	Ferguson unrest · 2015 Baltimore protests · 20th Anniversary of the Million Man March: Justice or Else · 2015–16 University of Missouri protests · 2016 Donald Trump Chicago rally protest	
Cases protested	2009–2011	Oscar Grant · Aiyana Jones · James Craig Anderson · Kenneth Chamberlain, Sr.
	2012	Ramarley Graham · Trayvon Martin · Rekia Boyd · Jordan Davis · Timothy Russell and Malissa Williams
	2013	Jonathan Ferrell · Renisha McBride
	2014	Dontre Hamilton · Eric Garner · John Crawford III · Michael Brown · Ezell Ford · Laquan McDonald · Akai Gurley · Tamir Rice · Antonio Martin · Jerame Reid
	2015	Charley Leundeu Keunang · Tony Robinson · Anthony Hill · Meagan Hockaday · Eric Harris · Walter Scott · Freddie Gray · The Emanuel Nine · Jonathan Sanders · Sandra Bland · Samuel DuBose · Jeremy McDole · Corey Jones · Jamar Clark
	2016	Bruce Kelley Jr. · Abdullahi Omar Mohamed
Related	Police brutality in the United States · Hands up, don't shoot · Trayvon Martin could have been me 35 years ago · Blue Lives Matter · Campaign Zero · Super Bowl 50 halftime show controversy	

Figure 1. The Wikipedia Template:Black Lives Matter containing relevant articles for analysis.

Collective memory in social computing

BLM also lets us explore how social computing systems serve as sites of collective memory. The movement centers around a series of recent deaths. Social movements often focus on transforming individual events such as these into collective memories [16, 39]. A groups' shared sense of the past contributes to collective action by providing common frames to articulate grievances, develop shared identity, regenerate networks and resources, and increase incentives for cooperation [12, 34]. Halbwachs's theory of collective memory argues that memories shape and are shaped by present concerns and developed through social interactions [14]. Observing memory building around a social movement with respect to related events provides new insights into collective memory in Wikipedia.

Prior work by Pentzold [36] describes the role of Wikipedia in constructing collective memories. Wikipedia's openness, popularity, and detailed archival records enables investigations into the processes of translating a topic's archival materials, historical scholarship, and other memories into an article. Subsequent research followed this perspective with a number of case studies. Memory-building processes for the Egyptian revolution across languages [7], the deaths of notable people [24], and the Vietnam War [33] have all extended the understanding of how editors collaborate, manage conflicts, and produce knowledge in a commemorative mode. The current study extends the application of collective memory to events and memories encoded into individual articles that in turn compose a larger network documenting evidence relevant to a social movement.

However, there are opposing perspectives regarding whether collective memory threatens or supports collective action. Each perspective implies different dynamics between activity and attention across current and historical events. On one hand, tragic events in the past act as proof that transgressing the status quo is dangerous, which reinforces shame, fear, and apathy [38]. Such events can be collectively self-censored and more favorable events privileged in order to preserve a positive self-image or dominant social identity. On the other hand, tragic events can be translated and sustained in oral, written, or visual histories as well as commemorated through rituals (*e.g.*, anniversaries) and artifacts (*e.g.*, literature). These collective memory repertoires provide interpretive material for people to make sense of current events, which in turn support solidarity and collective action [16]. In this view, activity around current events should be positively correlated with historical events as they are re-surfaced and sustained to support information seeking and mobilization.

> **RQ3**: How are events on Wikipedia re-appraised following new events?

DATA AND METHODS

This study uses the revision histories of English language Wikipedia pages related to *BLM* collected on April 21, 2016. A custom Python script built with the Wikitools library[1] retrieved data from the English Wikipedia's MediaWiki API.[2] The creation of the analysis page set required multiple steps. First, we built a list of potential pages by identifying the 59 articles belonging to the *Black Lives Matter* Wikipedia template (Figure 1) and the 74 articles in the *Black Lives Matter* Wikipedia category. From the union of these unique pages, we excluded pages if they were not focused on the movement, a death, or a protest event. These pages were excluded because of the current study's focus on coverage and attention related to the movement, motivating deaths, and the protests in response. The resulting articles covered the social movement itself (1 page), deaths of African Americans (37 pages), and protest events (4 pages) totaling 42 pages.

We included the redirects and talk pages as well. Redirects give insights into editors who were interested in contributing, but may have edited a similar page before the current version became the main article for a topic [17]. We map activity on the redirected page to the target article of the redirect. The overall editing on redirect pages represents a small amount of activity to the sample (1.7% of total), but is included to provide article creation data. Talk pages are a history of discussions that occur throughout the lifespan of articles. In the complete set, there were 86,940 revisions made by 6,795 editors to 141 articles and talk pages from January 6, 2009 to April 21, 2016.

To supplement the revision data, we also analyze the pageviews for the sample from January 1, 2014 to January 1, 2016. Pageviews are a count of the number of visitors to each Wikipedia article. The data was extracted using Python from a third-party database of daily pageviews data.[3] While this time range truncates some of the information that can be gained for some of the article sample, the two years of

[1] https://github.com/alexz-enwp/wikitools
[2] https://en.wikipedia.org/w/api.php
[3] http://stats.grok.se/

Article	Revisions	Editors	Talk Revisions	Talk Editors	Pageviews
Shooting of Michael Brown	7,141	904	19,454	391	4,254,371
Shooting of Trayvon Martin	6,639	805	19,707	782	1,145,483
Shooting of Oscar Grant	2,728	756	985	113	1,066,511
Charleston church shooting	2,544	575	2,118	162	903,191
Ferguson unrest	1,847	556	1,640	139	1,315,598
Black Lives Matter	1,626	408	534	74	450,516
Death of Eric Garner	1,526	486	1,178	122	1,873,056
Death of Freddie Gray	1,513	354	1,358	115	1,375,498
2015 Baltimore protests	1,034	298	634	248	611,371
Death of Sandra Bland	950	284	1,012	65	492,589
Total	27,548	4,372*	48,646	1,743*	13,488,184

* Indicates a unique count for each measure. Overall, 5,449 unique editors contributed.

Table 1. List of articles with the most revisions in the sample related to the movement. The number of editors, talk page activity, and pageviews are also included.

pageviews shows information consumption patterns about articles in the sample and daily changes in attention, especially after the movement began.

We focus on the three aforementioned research questions in order to extend prior research on social computing, social movements, and collective memory as well as studies of participation around breaking news and current events in Wikipedia. *BLM* provides a unique and exceptional opportunity to advance research at the intersection of these topics. Because prior work addressing these concerns and related phenomena does not support clear predictions or hypotheses to test, we adopt an exploratory and descriptive approach. We gain insights into the editing activity, attention, and collaboration surrounding *BLM* on Wikipedia. Temporal dynamics, correlations, and networks uncover different behavioral patterns showing the movement's impact on Wikipedia. Additionally, we qualitatively review discussions from the movement's talk page to further investigate collaborations surrounding *BLM*.

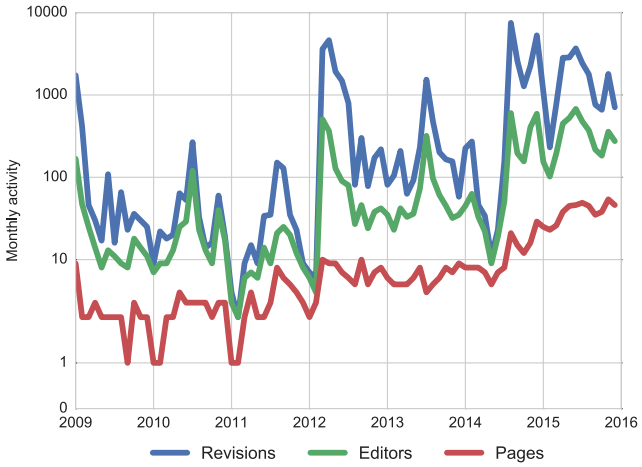


Figure 2. Aggregated monthly activity over time for all the pages included in analysis. The number of revisions (blue line), editors (green), and pages (red) are plotted.

RESULTS

RQ1: Intensified Documentation

Research Question 1 asked, "How has Wikipedia editing activity and coverage of the *BLM* movement events changed over time?" We examine this question in two ways. First, we describe the revision and pageview activity to the top articles and identify variation in monthly activity due to current events. Second, we observe changes in the latency between the occurrence of an event and the creation of its article. There is wide variance, but latencies decline for events since the beginning of *BLM*. These results suggest Wikipedia editors engage in an emergent practice we call *intensified documentation* as they create more content, more quickly, within the normative constraints of Wikipedia policies like neutral point of view or notability. As *BLM* gained notoriety, Wikipedians covered new events more rapidly and also created articles for older events.

Revision histories document several dimensions of knowledge creation in Wikipedia: the editor, time stamp, and content changed in each revision. Table 1 shows the number of revisions, editors, talk page revisions, talk page editors, and pageviews to the ten articles with the most revisions. The amount of activity on these ten articles and talk pages dominates all of the activity on the other 121 pages in the sample combined. The Top 10 pages and their corresponding talk pages account for 76,194 (87.6%) of all revisions, and 5,449 (80.2%) editors made these revisions. Excluding *BLM*, these top articles are about events that generated a large amount of media attention outside of Wikipedia [10, 13, 19]. The activity for the "Shooting of Michael Brown" and the "Shooting of Trayvon Martin" contribute 52,941 (60.9%) of all revisions by themselves. The "Shooting of Michael Brown" article also had the most pageviews, accounting for 31.5% of all views in the Top 10, reflecting its influence as a major news event.

Figure 2 visualizes the monthly activity for all articles in our corpus by number of revisions made (blue), unique editors contributing (green), and pages edited (red). The peaks in activity correspond closely to the death of Oscar Grant (January

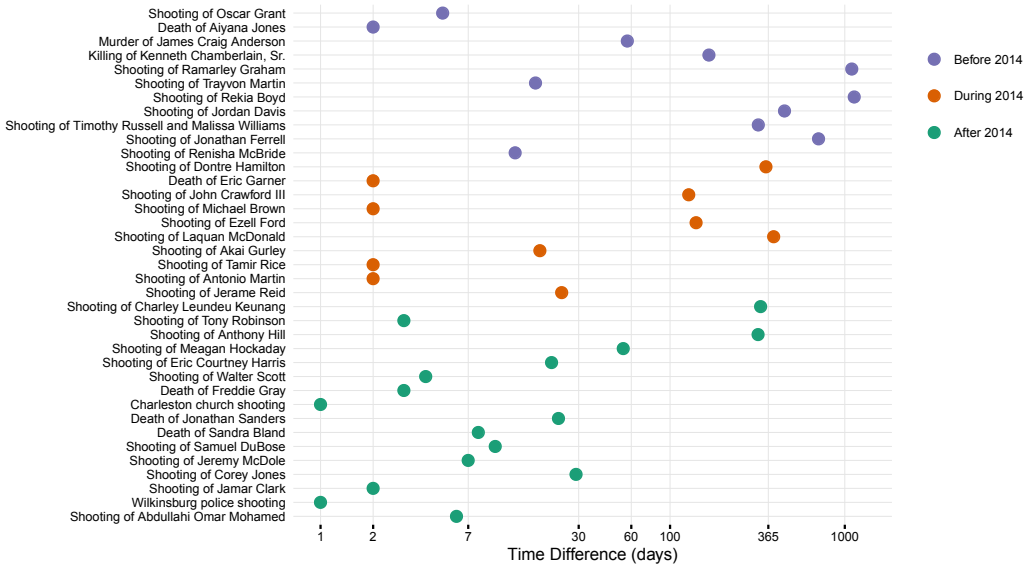


Figure 3. The time difference (lag) between *BLM*-related deaths and the date that the corresponding Wikipedia page is created. The *y*-axis is ordered chronologically from oldest death (top) to most recent. Points for each death are grouped by color corresponding to when they occurred: "Before 2014", "During 2014", or "After 2014".

1, 2009), the death of Trayvon Martin (February 26, 2012), the death of Michael Brown (August 9, 2014), the death of Tamir Rice (November 22, 2014), the protests in Baltimore (April 2015), and the Charleston Church shooting (June 17, 2015). We note one peak in July 2013 corresponds to the acquittal of George Zimmerman, the man who killed Martin, and consists almost entirely (98%) of edits to pages related to the shooting of Trayvon Martin and a few edits to pages about the shooting of Oscar Grant. The monthly activity plots illustrate that prominent events drive periods of high activity and reshape activity in aggregate. While they do not sustain the levels of peak activity, we observe a general trend towards an increasing level of activity across all three metrics as additional articles are created and added to the sample. However, the large peaks seem to suggest that activity in the *BLM* topic space is focused on individual events and does not necessarily imply sustained writing about the *BLM*-related topics.

The activity analysis above does not discriminate between different types of editors who may contribute to articles. The editing environment of Wikipedia is dynamic and will be different for different types of users (*e.g.*, registered, administrators, bots, and unregistered). Page protection statuses affect editing activity from different sets of users, especially unregistered ones [18]. There are 374 log events collected for the sample, including redirected pages. 21 of the 374 events are autoconfirmed page protections to the 42 main articles, which prevent unregistered users from making revisions. In the entire 141-page sample, there were 3,275 unregistered editors who made 7,703 revisions. These people would not have been able to edit anything when pages were under protection. It is possible that these protection spells suppressed activity from unregistered users who were vandalizing pages, and also users interested in contributing, but unfamiliar with the policies of Wikipedia to navigate the protections.

Temporal dynamics of article creation

The template for *BLM* includes 37 incidents that resulted in the deaths of African Americans (Figure 1), but how long after these incidents did it take for Wikipedia articles to be generated? The time between an event and the creation of its Wikipedia article event can signal its significance [27]. Unlike the "wiki-bituary" activity following the deaths of people who already had Wikipedia articles [24], the deceased subjects in our sample did not have articles until after their deaths generated media attention. Figure 3 plots the 37 events and the time elapsed between the event and article creation.

We note the wide variation in article creation latency. Only two of the deaths have articles created within 24 hours of the event ("Charleston Church Shooting" and "Wilkinsburg police shooting"); across the rest of events, it takes at least 48 hours. Even the deaths that were major news stories in the United States, such as "Shooting of Michael Brown," "Shooting of Trayvon Martin," and "Death of Eric Garner" took at least two days for their articles to be created. Many of the articles appear months or even years after the deaths occurred. The "Shooting of Rekia Boyd" (1,133 days) and the "Shooting of Ramarley Grant" (1,097 days) had the longest article creation lag. In total, 12 of the events had article creation lags of at least 100 days. None of these events were among the most active pages in the sample by number of revisions.

Also in Figure 3, we separate the deaths into three categories that correspond to different phases of the *BLM* movement: "Before 2014," "During 2014," and "After 2014." There are 11 death events that took place before 2014, 10 events during 2014, and 16 events after 2014. Before 2014 and the emergence of *BLM*, the 11 articles about deaths took on average 361 days to be created after each death. During 2014, when the *BLM* movement gained attention, the 10 pages about deaths took on average 107 days to be created. After 2014, the

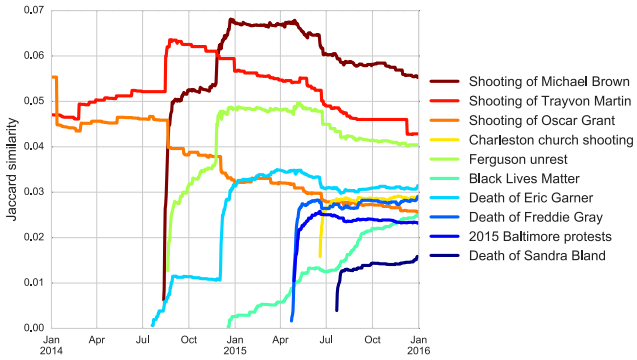


Figure 4. Changes in Jaccard similarity scores for top 10 articles. Similarity is computed daily for the cumulative editor sets between the focal article and editors to all articles in the corpus.

16 pages about deaths took on average 51 days to be created after each death. The deaths of most of these African Americans did not become articles until after widespread attention started to be given to *BLM* and these types of incidents.

The *BLM* movement has led to both reduced latency in article creation and increased coverage of relevant deaths, demonstrating an overall pattern of *intensified documentation*. The rise of *BLM* has coincided with Wikipedia reducing its average response time from 361 to 51 days (86% reduction) when creating new articles about such related deaths. In addition, there have been more articles created about related deaths in the 16 months after 2014 than in the 60 months before 2014. The *BLM* movement and the curation of a Wikipedia template and category organized the topic space and appears to have helped editors create knowledge about the events related to the movement.

RQ2: Collaborative migration

Research Question 2 asked, "How have Wikipedians collaborated across articles about events and the *BLM* movement?" To what extent did Wikipedia editors contribute across related articles? How does this activity change over time and in response to new events? We explore these questions in two ways. First, we examine the similarities between the sets of editors and changes over time across articles. Then we qualitatively investigate discussions on the "Black Lives Matter" talk page. Experienced editors were diligent in maintaining a neutral point of view, improving the sourcing, and making decisions about inclusion of related events in the article. These findings illustrate a distinctive mechanism we term *collaborative migration* in which groups of Wikipedia editors move among related event articles in the absence of traditional coordination structures to support co-authorship, such as templates or WikiProjects. In the case of *BLM*, many of these structures were not created until mid-2015.

Editor similarities over time

We analyze editor similarity across articles using Jaccard coefficients. The Jaccard coefficient captures the similarity between two sets of entities by computing the size of the intersection between entities in both sets, divided by the union of the sets. We computed daily Jaccard coefficients between the cumulative set of editors to a given article and the cumulative set of editors to all other articles in the corpus between January 2014 and January 2016. Substantively, this metric describes what fraction of users on an article in a given day had edited any other article in the corpus at any time leading up to and including that day. Small values indicate a small percentage of the editors on the article have engaged with other *BLM*-related articles while large values indicate there is substantial overlap between the editors of the article and the editors contributing across all other *BLM* articles. Assessing the similarity in editors between articles reveals whether the collaborators converge or diverge over time.

Figure 4 plots these daily Jaccard similarity coefficients for the Top 10 articles (Table 1). First, the daily editor similarities between articles are relatively low in absolute terms: the "Shooting of Michael Brown" article shows the greatest similarity in editor membership with the rest of the corpus but never rises above 7% similarity. Given the number of articles and the tendency for most editors to contribute only once, ~7% similarity over a set of 6,795 unique users is quite high.

Many newly-created articles like "Shooting of Michael Brown" follow a "rotated L"-shaped pattern of sharply rising similarity immediately following their creation followed by a plateau of relatively constant similarity in the cumulative editor sets. The immediate increase in similarity suggests a set of "early responder" editors rapidly come together to contribute to and frame these articles. However, the tendency for editor similarity to decrease over time across these articles suggests that new contributors to these articles in the weeks and months after the events themselves tend to edit only that article rather than working across articles. Figure 4 also shows discontinuities in articles' similarities around the time that *new* articles are created. For some events and some articles, this results in increased similarity scores. For others, the same event appears to drive down similarity, indicating that the article's editors overlap less with the full set.

The creation of the "Black Lives Matter" article (Fig. 4, cyan line) in late 2014 corresponds to noticeable increases in the similarity coefficients for the articles about Michael Brown, Ferguson, and Eric Garner, but minor decreases in the similarity coefficients for Trayvon Martin and Oscar Grant. Editors on the "Black Lives Matter" article also started editing the former articles, but did not edit the latter articles. All of the related event articles show a general tendency towards slowly-declining similarity over time. The *BLM* article's similarity coefficient, while relatively low compared to major event articles in absolute terms, has followed a different and constantly increasing trajectory. The trend suggests editors of event articles increasingly make contributions to the movement article as well.

Figure 5 measures editor similarity in an alternative way. It shows the extent to which editors of Top 10 articles (Table 1) collaborated in the rest of the articles over time. We compute the daily fraction of users who edited any of the Top 10 articles and the rest of the articles by *relative dates* since the non-Top 10 article was created. While there is wide variation in articles, the averaged values (in red) show the frac-

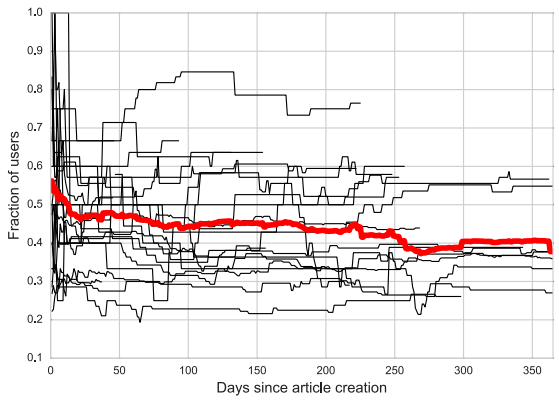


Figure 5. Fraction of editors from Top 10 *BLM* articles in Table 1 collaborating on non-Top 10 *BLM* articles by day since article creation. Average value across articles in red.

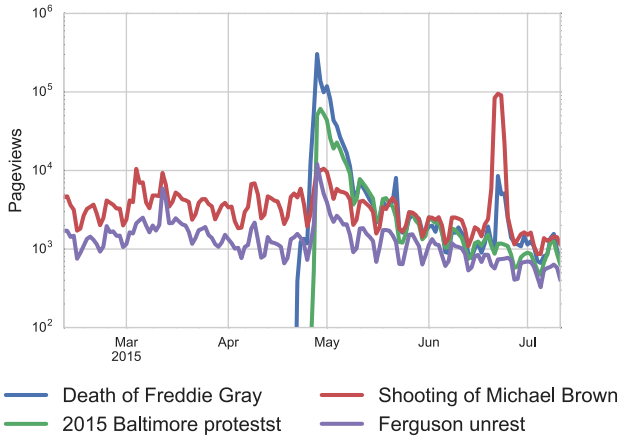


Figure 6. Example: Pageviews among four articles in the aftermath of Freddie Gray's death. Correlations were calculated from the daily attention dynamics.

tion is initially high (~50%) but declines only modestly. Editors migrate throughout the corpus across the lifespans of pages. The tendency for non-Top 10 articles to draw large fractions of their editorship from Top 10 articles, even when many of these articles are created weeks or months after the events themselves, shows that editors collaborate on high-profile events as well as less popular or overlooked events.

Black Lives Matter talk page discussion

The "Black Lives Matter" article in Figure 4 has an increasing trend of similarity that differs from other Top 10 articles. The article is the only one in the sample dedicated to the social movement and it is reliant upon the actual events that motivate it. We conducted an interpretive analysis of the "Black Lives Matter" talk page to better understand how the events are organized into knowledge on *BLM*. There were 74 editors who made 534 revisions (See Table 1), and 41 of these talk page editors also edited the *BLM* article itself. This indicates that only 10% of editors to the article participated in discussions on the talk page. Reviewing the posts that were made from December 21, 2014 to April 3, 2016 reveals deliberations about how Wikipedia rules should be applied to the movement's article.

A recurring debate surrounded *BLM*'s connection to associated events and deaths. Editors deliberated how to situate the *BLM* article with respect to other related events while maintaining Wikipedia's standards, such as keeping a neutral point of view[4], preventing it from becoming a "coat rack" article[5] for enumerating all related events, and identifying reliable sources[6]. During the time some of these discussions took place, neither the template nor category for *BLM* existed, and editors discussed whether and how to organize and situate individual events with the movement. One particularly active editor in the sample with 1,040 revisions, "MrX," was adamant that clear connections between related events and *BLM* needed documentation before inclusion:

> *Many of the articles linked make no mention of BLM. Again, this is an original research issue and a WP:NPOV issue... My remaining concern is that a list like this may be WP:UNDUE, and turns the article into somewhat of a WP:COATRACK. For example, did Oscar Grant III's death in 2009 actually inspire the movement that started four years later, or is this some revisionist history on the part of the organization?*

Editors' diligence against such inclusions suggests that Wikipedia remains a place for documenting events and revising accounts of the movement consistent with community standards such as NPOV. This included a debate about identifying a *BLM* protest. The most active editor in the sample, Mandruss (6,248 revisions), expressed concerns about what constitutes a related protest:

> *If I then protest the Killing of John Doe and invoke the BLM name, and some [reliable source] reports that I did so (without necessarily endorsing the connection), does that constitute a BLM protest?*

The talk page illustrates challenges editors faced when trying to produce a neutral and consistent record about a social movement without explicitly supporting it. Identifying reliable sources is challenging with breaking movement events that lack authoritative, centralized sources of information. Editors also needed to decide when to focus content on *BLM* and to what extent to incorporate pages about individual events into *BLM* for background. In the context of *collaborative migration*, these discussions show the choices made in organizing knowledge among events before infrastructures (WikiProjects, categories, and templates) exist for topics and are essential for collaboration.

RQ3: Dynamic re-appraisal

Research Question 3 asked, "How are events on Wikipedia reappraised following new events?" The declining event-article

[4]https://en.wikipedia.org/wiki/WP:NPOV
[5]https://en.wikipedia.org/wiki/WP:Coatrack
[6]https://en.wikipedia.org/wiki/WP:IRS

latency in Figure 3 suggests relationships between different incidents. Some articles may drive attention or editing activity to other articles. We examine the extent to which revision and pageview behavior about different events are correlated with each other. Unlike other research on social movements, we use these Wikipedia measures to observe how people rely on prior events as part of the sensemaking process associated with a social movement. These relationships highlight a general pattern we term *dynamic re-appraisal* in which Wikipedia users engaged in information seeking return to previous event articles in response to current events.

Figure 6 is an example of daily pageview activity for four articles in the *BLM* corpus in the aftermath of Freddie Gray's death in April 2015. The pageviews for all four appear to follow similar trends. There is a spike of activity on the "Death of Freddie Gray" and "2015 Baltimore protests" articles since these events were major news events at the time. The pageview activity for the "Shooting of Michael Brown" and "Ferguson unrest" articles also show attention spikes at the same time, despite these events happening nine months earlier. A little over a month later, in late June, we observe a spike in views to the "Death of Freddie Gray" and "Shooting of Michael Brown" articles without any corresponding increase in views to the "2015 Baltimore protests" or "Ferguson unrest" articles. These patterns highlight that attention to Wikipedia articles can spillover to adjacent articles [32]. We can use such highly-correlated behaviors to reveal topically similar or contextually relevant article relationships.

Tables 2–4 are the most positive correlations for daily revision, pageview, and revision-pageview activity (respectively). Revision activity (Table 2) has consistently lower correlations than the pageview (Table 3) or revision-pageview relationships (Table 4). The cost of editing articles is higher than simply viewing them, which is an intuitive explanation of the observed differences in correlation. The strongest correlations exist between articles about similar events, such as Freddie Gray and Baltimore protests or Michael Brown and Ferguson protests. These correlations also reveal temporal and spatial relationships. The article for Rekia Boyd was created in the same week as the articles for Freddie Gray and 2015 Baltimore protests. The Walter Scott shooting and Charleston church massacre both happened in Charleston, South Carolina in 2015 and the officer's indictment in the former happened days after the latter.

Article 1	Article 2	Corr.
Freddie Gray	Baltimore protests	0.672
Rekia Boyd	Baltimore protests	0.601
Jeremy McDole	James Craig Anderson	0.546
Tamir Rice	Ferguson unrest	0.516
Michael Brown	Baltimore protests	0.504

Table 2. Five article pairs with strongest daily revision correlations.

Article 1	Article 2	Corr.
Michael Brown	Ferguson unrest	0.918
Walter Scott	Charleston shooting	0.915
Freddie Gray	Baltimore protests	0.863
Ferguson unrest	Freddie Gray	0.842
Eric Courtney Harris	Freddie Gray	0.840

Table 3. Five article pairs with strongest daily pageview correlations.

Pageviews	Revisions	Corr.
Freddie Gray	Baltimore protests	0.942
Eric Courtney Harris	Baltimore protests	0.856
Akai Gurley	Baltimore protests	0.839
Walter Scott	Charleston shooting	0.836
Ferguson unrest	Baltimore protests	0.829

Table 4. Five article pairs with strongest revision-pageview correlations.

Clustering of article correlations

To make sense of these three different correlation relationships, we use hierarchical clustering to identify article pairs that have similar correlation values across all three article interactions. We compute the distances between clusters using the Ward variance minimization algorithm and manually tuned the distance parameter to return four easily-interpretable clusters. The results are clusters where the three correlation values for each article pair in the cluster are more similar to other article pairs in the cluster than outside the cluster. Each cluster of article relationships can be qualitatively interpreted as having distinctive information consumption and production patterns. We focus on the cluster where pages' pageview and revisions activity are most strongly and positively correlated with each other. Substantively, these article pairs are viewed and edited in similar ways on similar days—indicating that they have the most well-aligned production and consumption among those in our sample [42].

The set of article relationships in this highly-correlated cluster can also be visualized as a network to identify clusters of topics that behaved most similarly with each other. In Figure 7, each node in the network is an article and the articles are connected by directed links if these articles' pageview and revision time series behavior appeared in the most strongly correlated cluster. Correlations between articles in this network indicates that both articles' revision and pageview behaviors respond similarly to each other. This network perspective helps to surface latent relationships between topics based on the dynamics of users' activity editing or viewing these pages at the same time.

This correlation network is not completely connected. Articles about the shootings of Trayvon Martin, Michael Brown, and Oscar Grant, the most prominent and actively-edited events in our corpus, are found in the green-colored community at 3 o'clock in Figure 7. Despite the tremendous amount of activity on these articles, they are not the most central articles in the correlation network. Articles about the deaths of Freddie Gray and the 2015 Baltimore protests occupy the most central locations in this interaction network (in the pink-colored community at 10 o'clock in Figure 7).

Activity on these articles drove pageviews and revisions to each other, but not to other articles like "Black Lives Matter"

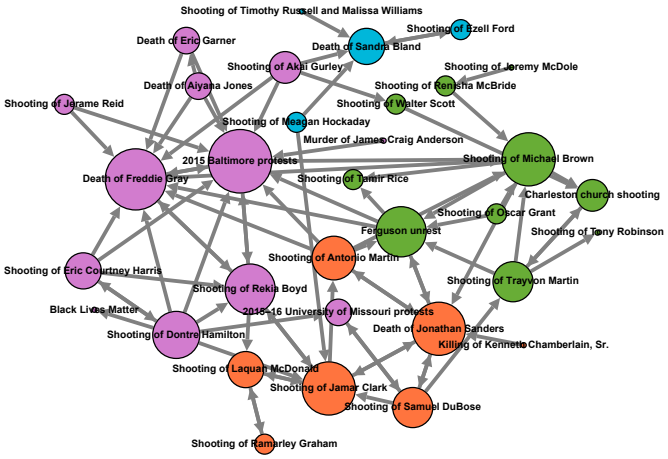


Figure 7. Network of article-article relationships based on correlations among pageview and revision activity. Articles are linked together if these articles occur in the same hierarchical cluster. Nodes are sized by number of connections and colored by a community detection algorithm.

(at 9 o'clock in Figure 7), implying there was little correlation between the "Black Lives Matter" article and these prominent event pages. *BLM* is only a peripheral node in this correlation network, having a single connection to the shooting of Dontre Hamilton. The lack of connectivity for *BLM* in this cluster implies that the activity on articles about individual events is not robustly correlated with the activity on the *BLM* movement article. In other words, while articles about different events exhibit strong correlations in pageview and editing activity, this event-related activity is not also driving activity in the article about the movement itself.

The daily editing and pageviews correlations between related-event articles show that articles can remain active sites of attention even when they are not breaking news stories. These "dynamic re-appraisal" processes capture users' engagement with content about some — but not all — related events as part of information seeking and sensemaking routines in the aftermath of new events. These behaviors capture emergent knowledge collaboration relationships within social computing systems beyond existing affordances such as coauthorship, hyperlinking, or categorization.

DISCUSSION

This analysis of activity around Wikipedia articles related to the *Black Lives Matter* movement contributes to prior literature on social movements by describing knowledge production and collective memory in a social computing system as the movement and related events are happening. We identify three distinct mechanisms — intensified documentation, collaborative migration, and dynamic re-appraisal — that occur throughout the *BLM*-related pages on Wikipedia. Through these mechanisms, the articles and the editors who worked on them support collective memory and knowledge production. The activity drives knowledge creation around *BLM*, organizes the topic space, and generates a neutral accounting of the movement and related events. The patterns of participation and attention around *BLM* pages also demonstrate some novel dynamics when compared with earlier analysis of Wikipedia. We discuss the connections between these mechanisms, theory, the implications of our findings for practice and design, as well as the limitations of this work below.

Intensified documentation

In the context of the *BLM* movement, the growth in the number of articles, acceleration of editing activity, and reduction in article creation latency reflect a general pattern of behavior we term *intensified documentation*. We do not claim that editors sought to advance the agenda of movement supporters or opponents. Instead, Wikipedians generated a neutral accounting of related events and created topical structures to navigate *BLM* content. Indirectly, Wikipedia's coverage arguably lends support to *BLM*'s claims about police violence in the United States being a systematic problem rather than isolated cases. This affinity between coverage and movement aims of increased visibility, mourning, and commemoration derive from the specific character and tactics of *BLM*. In general, intensified documentation practices may support or discredit movement frames over time. In this way, intensified Wikipedia coverage of social movement concerns seem likely to capture the context of a movement more thoroughly than other social computing platforms like Twitter.

Collaborative migration

Similar to prior work on breaking news in Wikipedia, we observe *collaborative migrations* as editors work across topically-related articles. In general, effective coverage of new events implies sustained engagement from editors to coordinate in the aftermath, manage conflicts arising in response, and support collaboration by developing infrastructures (like templates and categories) [27]. The *collaborative migrations* around *BLM* extend the understanding of breaking news in one key way. Wikipedia's coverage of *BLM* requires the integration of the social movement with events across a longer time horizon. Some events are breaking news with respect to the movement, but others are older events. Some events had articles well *before BLM* existed, while other events occurred years ago, but did not have an article until *after BLM*. New current event articles also prompted noticeable changes in the similarities of existing articles, sometimes increasing similarity as prior editors came in or decreasing similarity as new editors joined the editing population. The *BLM* article itself did not have high co-authorship similarity, but exhibited sustained growth in similarity unlike the event articles. This suggests that it played an increasingly central role facilitating collaboration and connections across events as its editor population became more similar to the other articles. There are likely other types of current events and topic domains that fit this pattern of a central reference article surrounded by a set of less connected articles.

Dynamic re-appraisal

The re-appraisal of past events in response to new events highlights how the distinct affordances of Wikipedia translate into collective memory practices. We find evidence of increased traffic to articles about past events even when there was little new information on them. We interpret this as evidence of Wikipedia readers making sense of new events by

reading about related prior events. We are not aware of prior work documenting this pattern on Wikipedia or other social computing systems.

We also observe correlations between knowledge production (revisions) and consumption (pageviews) within the same article and also between different articles in our study. In the context of *BLM* these correlations potentially reflect collective sense-making attempts to constrain uncertain, anomalous, or unjust events within narratives about events having similar familiar contexts and actors [37]. They also reflect patterns consistent with past Wikipedia research finding correlations between viewership and editorship [17].

Intentionally or not, the process of dynamic re-appraisal around events of public mourning and commemoration aligns with some of the goals and tactics embraced by *BLM* movement participants. Without breaking with community norms like NPOV, Wikipedia became a site of collective memory documenting mourning practices as well as tracing how memories were encoded and re-interpreted. Editors and readers who curated, visited, and re-visited information surrounding movement-related events contributed to this process.

Implications for practice and design

Movement participants or organizations may look to the *BLM* coverage on Wikipedia as an example of how to coopt a social computing system, but this reflects a profound misunderstanding of how Wikipedia works. Wikipedia's open nature and encyclopedic scope allow social movements to be neutrally documented, resulting in framing and narrative generation processes unlike those found in other social computing systems. Typically, a movement's framing describes the grievances at stake, determines how the movement is understood, and specifies the scope of issues to be addressed by fostering a "sense of severity, urgency, efficacy, and propriety" [2]. Movement narratives attempt to link past, present, and future events together and align individual and collective identities into stories that shared and referenced [37]. While the use of social media has transformed mobilization strategies [3, 6], movement actors should also consider the role of online communities like Wikipedia in framing movements for a broader audience. Distinct from settings like Twitter or Facebook, movements may struggle to exert control over Wikipedia narratives since the information must meet criteria like neutrality and notability.

Designers could better support the sorts of collaboration, collective sense-making, and attention dynamics we observed around *BLM* articles in Wikipedia. For example, the evidence we uncover of dynamic re-appraisal suggests that editors working on Wikipedia articles about a particular current event would likely benefit from understanding the coverage of related prior events. Likewise, readership patterns, category structures, and templates could be used to design a more active content suggestion system for Wikipedia readers. Previous work indicates that surfacing past activities or interests for editors can mobilize them to make more contributions [4]. Our findings here imply that similar interventions could help readers navigate a topically-linked set of articles related to unfolding events or movements.

Limitations and future work

Our analysis looked at only a single, prominent, contemporary social movement case in the United States and activity related to it on the largest language edition of Wikipedia. All of these features potentially limit the generalizability of these findings across other cases, movements, or contexts. However, social movements are ubiquitous and social computing systems increasingly function as spaces of collective knowledge production, sense-making, and commemoration. The dynamics we have described here can help guide future investigations into these phenomena.

A richer mixed-methods accounting of the framing contests, value conflicts, editor identities, and motivations would also complement the results we report. The logged revisions provide a rich, but convenient sample of data that censors contributions during page protections [18], overlooks the choice of information sources [8, 9] and the role of automated agents in supporting collaboration [11]. We do not have data on the demographic attributes of editors or readers in our dataset. While prior work has documented disparities in participation [15], we do not know whether or how the readers and editors of pages related to *BLM* reflect these broader trends. In addition, while the representation of *BLM* in Wikipedia may advance some movement goals, we cannot make claims regarding whether editors support the movement or not. Future research might address this in a variety of ways, including analyzing the content of revisions and revision histories of editors, administering surveys to collect demographic or motivational data, and performing interviews and qualitative analyses of editors' actions.

The descriptive and exploratory quantitative analyses we employed could be augmented as well. Regression, network, or time series modeling could be used to estimate the likelihood of existing editors contributing to movement articles or editors migrating across articles. Text analysis techniques could be used in conjunction with news data to examine the gaps or biases in coverage on Wikipedia compared to popular media narratives and frames. Sequence analysis approaches could be used to better disambiguate the direction of editor flows among these articles [28]. Observational causal inference approaches could also be used to identify the influence of dynamic re-appraisals on other collaborative proceses [32].

CONCLUSION

Wikipedia's openness, popularity, and legibility provides a unique opportunity to track the evolution of social movements' activity over time. Our analysis of the English Wikipedia's response to the "Black Lives Matter" movement and articles about related events extends understanding of the role social computing systems have in online collective action. Our findings point to the interplay between collective action and collective memory by highlighting three distinctive mechanisms: intensified documentation, collaborative migration, and dynamic re-appraisal.

ACKNOWLEDGEMENTS

The authors thank the BYOR workshop at Northwestern for reviewing early drafts of this work. We thank the anony-

mous reviewers and Cliff Lampe for their detailed comments, which have strengthened this work.

REFERENCES

1. Ofer Arazy, Felipe Ortega, Oded Nov, Lisa Yeo, and Adam Balila. 2015. Functional Roles and Career Paths in Wikipedia. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 1092–1105. DOI: http://dx.doi.org/10.1145/2675133.2675257
2. Robert D Benford. 1993. "You Could Be the Hundredth Monkey": Collective Action Frames and Vocabularies of Motive Within the Nuclear Disarmament Movement. *The Sociological Quarterly* 34, 2 (1993), 195–216.
3. W. Lance Bennett and Alexandra Segerberg. 2013. *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. Cambridge University Press, New York.
4. Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. 2007. SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI '07)*. ACM, New York, NY, USA, 32–41. DOI:http://dx.doi.org/10.1145/1216295.1216309
5. Munmun De Choudhury, Shagun Jhaver, Benjamin Sugar, and Ingmar Weber. 2016. Social Media Participation in an Activist Movement for Racial Equality. In *Tenth International AAAI Conference on Web and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13168
6. Jennifer Earl and Katrina Kimport. 2011. *Digitally Enabled Social Change: Activism in the Internet Age*. MIT Press.
7. Michela Ferron and Paolo Massa. 2011. Collective memory building in Wikipedia: the case of North African uprisings. *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (2011), 114–123.
8. Heather Ford. 2012. *Wikipedia Sources: Managing Sources in Rapidly Evolving Global News Articles on the English Wikipedia*. SSRN Scholarly Paper ID 2127204. Social Science Research Network, Rochester, NY. http://papers.ssrn.com/abstract=2127204
9. Heather Ford, Shilad Sen, David R. Musicant, and Nathaniel Miller. 2013. Getting to the Source: Where Does Wikipedia Get Its Information from?. In *Proceedings of the 9th International Symposium on Open Collaboration (WikiSym '13)*. ACM, New York, NY, USA, 9:1–9:10. DOI: http://dx.doi.org/10.1145/2491055.2491064
10. Deen Goodwin Freelon, Charlton D. McIlwain, and Meredith D. Clark. 2016. Beyond the Hashtags:# Ferguson,# Blacklivesmatter, and the Online Struggle for Offline Justice. *SSRN eLibrary* (2016). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747066
11. R. Stuart Geiger and Aaron Halfaker. 2013. When the Levee Breaks: Without Bots, What Happens to Wikipedia's Quality Control Processes?. In *Proceedings of the 9th International Symposium on Open Collaboration (WikiSym '13)*. ACM, New York, NY, USA, 6:1–6:6. DOI: http://dx.doi.org/10.1145/2491055.2491061 00010.
12. Timothy B Gongaware. 2010. Collective memory anchors: collective identity and continuity in social movements. *Sociological Focus* 43, 3 (2010), 214–239.
13. Erhardt Graeff, Matt Stempeck, and Ethan Zuckerman. 2014. The battle for 'Trayvon Martin': Mapping a media controversy online and off-line. *First Monday* 19, 2 (Jan. 2014). http://firstmonday.org/ojs/index.php/fm/article/view/4947
14. Maurice Halbwachs and Lewis A Coser. 1992. *On collective memory*. University of Chicago Press.
15. Eszter Hargittai and Aaron Shaw. 2015. Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information, Communication & Society* 18, 4 (April 2015), 424–442. DOI: http://dx.doi.org/10.1080/1369118X.2014.957711
16. Fredrick C Harris. 2006. It takes a tragedy to arouse them: Collective memory and collective action during the civil rights movement. *Social movement studies* 5, 1 (2006), 19–43.
17. Benjamin Mako Hill and Aaron Shaw. 2014. Consider the Redirect: A Missing Dimension of Wikipedia Research. In *Proceedings of The International Symposium on Open Collaboration (OpenSym '14)*. ACM, New York, NY, USA, 28:1–28:4. DOI: http://dx.doi.org/10.1145/2641580.2641616 00002.
18. Benjamin Mako Hill and Aaron Shaw. 2015. Page Protection: Another Missing Dimension of Wikipedia Research. In *Proceedings of the 11th International Symposium on Open Collaboration (OpenSym '15)*. ACM, New York, NY, USA, 15:1–15:4. DOI: http://dx.doi.org/10.1145/2788993.2789846
19. Sarah J Jackson and Brooke Foucault Welles. 2015a. # Ferguson is everywhere: initiators in emerging counterpublic networks. *Information, Communication & Society* (2015), 1–22.
20. Sarah J. Jackson and Brooke Foucault Welles. 2015b. Hijacking #myNYPD: Social Media Dissent and Networked Counterpublics. *Journal of Communication* 65, 6 (Dec. 2015), 932–952. DOI: http://dx.doi.org/10.1111/jcom.12185
21. Dariusz Jemielniak. 2014. *Common Knowledge?: An Ethnography of Wikipedia*. Stanford University Press, Stanford, California.

22. Brian C. Keegan. 2013. A History of Newswork on Wikipedia. In *Proceedings of the 9th International Symposium on Open Collaboration (WikiSym '13)*. ACM, New York, NY, USA, 7:1–7:10. DOI: http://dx.doi.org/10.1145/2491055.2491062

23. Brian C Keegan. 2015. Emergent Social Roles in Wikipedia's Breaking News Collaborations. In *Roles, Trust, and Reputation in Social Media Knowledge Markets*. Springer, 57–79.

24. Brian C. Keegan and Jed R. Brubaker. 2015. 'Is' to 'Was': Coordination and Commemoration in Posthumous Activity on Wikipedia Biographies. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 533–546. DOI: http://dx.doi.org/10.1145/2675133.2675238

25. Brian C. Keegan, Darren Gergle, and Noshir Contractor. 2011. Hot off the wiki: dynamics, practices, and structures in Wikipedia's coverage of the Tōhoku catastrophes. *Proceedings of the 7th international symposium on Wikis and open collaboration* (2011), 105–113.

26. Brian C. Keegan, Darren Gergle, and Noshir Contractor. 2012. Staying in the Loop: Structure and Dynamics of Wikipedia's Breaking News Collaborations. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration (WikiSym '12)*. ACM, New York, NY, USA, 1:1–1:10. DOI: http://dx.doi.org/10.1145/2462932.2462934

27. Brian C. Keegan, Darren Gergle, and Noshir Contractor. 2013. Hot Off the Wiki Structures and Dynamics of Wikipedia's Coverage of Breaking News Events. *American Behavioral Scientist* 57, 5 (May 2013), 595–622. DOI: http://dx.doi.org/10.1177/0002764212469367

28. Brian C. Keegan, Shakked Lev, and Ofer Arazy. 2016. Analyzing Organizational Routines in Online Knowledge Collaborations: A Case for Sequence Analysis in CSCW. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1065–1079. DOI: http://dx.doi.org/10.1145/2818048.2819962

29. Aniket Kittur and Robert E. Kraut. 2008. Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, New York, NY, USA, 37–46. DOI: http://dx.doi.org/10.1145/1460563.1460572

30. Aniket Kittur, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. 2007. He Says, She Says: Conflict and Coordination in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 453–462. DOI: http://dx.doi.org/10.1145/1240624.1240698

31. 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

32. Michael E. Kummer. 2014. *Spillovers in Networks of User Generated Content: Pseudo-Experimental Evidence on Wikipedia*. SSRN Scholarly Paper ID 2567179. Social Science Research Network, Rochester, NY. http://papers.ssrn.com/abstract=2567179

33. Brendan Luyt. 2015. Wikipedia, collective memory, and the Vietnam war. *Journal of the Association for Information Science and Technology* (2015).

34. Doug McAdam. 1986. Recruitment to high-risk activism: The case of freedom summer. *American journal of sociology* (1986), 64–90.

35. Sharon Meraz and Zizi Papacharissi. 2013. Networked Gatekeeping and Networked Framing on #Egypt. *The International Journal of Press/Politics* (Jan. 2013), 138–166. DOI: http://dx.doi.org/10.1177/1940161212474472

36. Christian Pentzold. 2009. Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Memory Studies* 2, 2 (2009), 255–272.

37. Francesca Polletta. 1998. Contending stories: Narrative in social movements. *Qualitative Sociology* 21, 4 (1998), 419–446.

38. Lester M Salamon and Stephen Van Evera. 1973. Fear, apathy, and discrimination: A test of three explanations of political participation. *American Political Science Review* 67, 04 (1973), 1288–1306.

39. Howard Schuman and Jacqueline Scott. 1989. Generations and collective memories. *American sociological review* (1989), 359–381.

40. Sanjay Sharma. 2013. Black Twitter?: Racial hashtags, networks and contagion. *New formations: a journal of culture/theory/politics* 78, 1 (2013), 46–64.

41. Zeynep Tufekci and Christopher Wilson. 2012. Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square. *Journal of Communication* 62, 2 (2012), 363–379. DOI: http://dx.doi.org/10.1111/j.1460-2466.2012.01629.x

42. Morten Warncke-Wang, Vivek Ranjan, Loren Terveen, and Brent Hecht. 2015. Misalignment Between Supply and Demand of Quality Content in Peer Production Communities. *ICWSM 2015: Ninth International AAAI Conference on Web and Social Media* (2015). http://www-users.cs.umn.edu/~morten/publications/icwsm2015-popularity-quality-misalignment.pdf

43. Howard T. Welser, Dan Cosley, Gueorgi Kossinets, Austin Lin, Fedor Dokshin, Geri Gay, and Marc Smith. 2011. Finding social roles in Wikipedia. In *Proceedings of the 2011 iConference (iConference '11)*. ACM, New York, NY, USA, 122–129. DOI: http://dx.doi.org/10.1145/1940761.1940778

44. Taha Yasseri, Robert Sumi, Andras Rung, Andras Kornai, and János Kertesz. 2012. Dynamics of Conflicts in Wikipedia. *PLoS ONE* 7, 6 (June 2012), e38869. DOI: http://dx.doi.org/10.1371/journal.pone.0038869

45. Haoqi Zhang, Andrés Monroy-Hernández, Aaron Shaw, Sean A. Munson, Elizabeth Gerber, Benjamin Mako Hill, Peter Kinnaird, Shelly D. Farnham, and Patrick Minder. 2014. WeDo: End-To-End Computer Supported Collective Action. In *Eighth International AAAI Conference on Weblogs and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8041

THIS DOCUMENT IS AN UNPUBLISHED DRAFT.
PLEASE DO NOT DISTRIBUTE OR CITE WITHOUT PERMISSION.

*The Hidden Costs of Requiring Accounts: Quasi-Experimental Evidence from Peer Production**

Benjamin Mako Hill
makohill@uw.edu

Aaron Shaw
aaronshaw@northwestern.edu

Abstract Online communities like Wikipedia produce valuable public information goods. While some of these communities require would-be contributors to create accounts, many do not. Does this requirement catalyze cooperation or inhibit participation? Prior research provides divergent predictions but little causal evidence. We conduct an empirical test using longitudinal data from 136 natural experiments where would-be contributors to wikis were suddenly required to log in to contribute. Requiring accounts leads to a small increase in account creation, but reduces both high and low quality contributions from registered and unregistered participants. Although the change deters a large portion of low quality participation, the vast majority of deterred contributions are higher quality. We conclude that requiring accounts imposes introduce an under-theorized tradeoff for public goods production in interactive communication systems.

INTRODUCTION

As a core principle, Wikipedia has always welcomed what it calls "unregistered editing" by people who have not created an account on the site.[1] Because vandalism and malicious editing persists on Wikipedia—as much as 97%

*Both authors contributed equally to this work and determined the order of names randomly.

[1] https://en.wikipedia.org/wiki/Wikipedia:Welcome_unregistered_editing

of it from unregistered editors[2]—Wikipedia participants have repeatedly suggested requiring contributors to create accounts.[3] Even during a series of crises around defamation, Wikipedia's leaders refused to require Wikipedia contributors to reveal their identity or use consistent pseudonyms (Jemielniak, 2014). Wikipedia's commitment to unregistered editing speaks to a longstanding concern in the study of interactive communication systems: do stable identifiers catalyze the creation of public information goods (Benkler, 2006; Fulk, Flanagin, Kalman, Monge, & Ryan, 1996; Monge & Contractor, 2003; Yuan et al., 2005)?

According to some scholars, stable identifiers support social connections that enable public goods production and prosocial cooperation in exchange networks (e.g., Axelrod, 1984; Faraj & Johnson, 2011; Ostrom, 1990; Yamagishi, Matsuda, Yoshikai, Takahashi, & Usui, 2009). Within online communities, identifiers and reputations increase social transparency, group attachment, trust, accountability, and commitment among members, leading to increased participation (Cheshire & Antin, 2008; Kraut, Resnick, & Kiesler, 2012; Ren et al., 2012). Even requiring low cost, disposable account names prevents negative disinhibited behavior, non-cooperation, and strategic gaming of reputation systems in computer-mediated communication without deterring good-faith participants (Friedman & Resnick, 2001; Joinson, 1998; Kraut et al., 2012; Walther, 1996).

Other research argues that minimized contribution costs facilitate large-scale participation (Benkler, 2002, 2006; Bennett & Segerberg, 2012; Cheshire, 2007; Kollock, 1999). In this view, the requirement to create a stable identifier, even a disposable pseudonym, imposes a transaction cost that inhibits cooperation that would otherwise occur. Unregistered contributions can also introduce valuable perspectives (Anthony, Smith, & Williamson, 2009; Kane, Johnson, & Majchrzak, 2014; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010) and can stimulate activity from experienced community members (Gorbatai, 2011).

Synthesizing this prior work, we anticipate that requiring peer production contributors to use stable identifiers will decrease both high and low quality contributions. We report the results of an empirical, observational test of

[2]See https://en.wikipedia.org/wiki/Wikipedia:Counter-Vandalism_Unit/Vandalism_studies The specific estimate depends on the measurement method used.

[3]Proposals to "prohibit anonymous users from editing" are the first item on Wikipedia's list of perennial proposal topics related to editing: https://en.wikipedia.org/wiki/WP:PEREN

these claims. The study exploits a series of 136 natural experiments that occurred in collectives engaged in peer production. Across all of these groups, would-be contributors went from being able to edit without an account to being obligated to create an account and log in in order to participate. In every case, the feature change was sudden and resulted in the sort of institutional shock that supports the use of quasi-experimental methods to identify causal effects (Murnane & Willett, 2011). Using exhaustive longitudinal data gathered from public database records, we employ a regression discontinuity design (RDD) to estimate the impact of the feature change.

Although requiring accounts causes the proportion of low quality contributions to decrease enormously, the large majority of deterred edits are of higher quality. These effects hold even among participants who had accounts prior to the intervention, indicating that unregistered contributions stimulate activity among established contributors. Overall, we find no support for the argument that user accounts actively catalyze short term cooperation. We conclude that stable identifiers like user accounts introduce previously undertheorized tradeoffs in public information goods production in open collaboration systems.

Public Information Goods Production and Stable Identifiers Online

Networked communication and interaction in online communities has facilitated large scale public information goods production (Benkler, 2002; Fulk et al., 1996; Kollock, 1999; Monge & Contractor, 2003). Many of the central elements of classical theories of public goods production and commons (Olson, 1965; Ostrom, 1990) do not change with the move to online environments and digital resources. However, several key aspects of the communicative practices involved undergo fundamental shifts. Two of the most frequently referenced shifts include (1) radically reduced transaction costs associated with participation (e.g., Benkler, 2006; Bennett & Segerberg, 2012; Fulk et al., 1996) and (2) technological mediation that allows individuals to participate with varying degrees of anonymity (e.g., Anonymous, 1998; Joinson, 1998; Walther, 1996).[4] To combat the negative effects of disinhibition due to anonymity, many communities require would-be contributors to create an account and log in (e.g. Reddit). On the other hand, requiring stable identifiers—even disposable pseudonyms—constitutes a cost. Communities

[4] Although many describe unregistered users of online communities as "anonymous," there are distinct types of anonymity (Anonymous, 1998).

like Wikipedia argue that the time required to create an account imposes a prohibitive barrier to would-be contributors.

Our analysis focuses on online collectives engaged in *commons-based peer production*, the form of digitally-networked, open public goods production that has created valuable resources such as Free/Libre Open Source Software, Wikipedia, and OpenStreetMap (Benkler, 2002, 2006). Benkler, Shaw, and Hill (2015) define peer production as "a form of open creation and sharing performed by groups online that: (1) set and execute goals in a decentralized manner; (2) harness a diverse range of participant motivations, particularly non-monetary motivations; and (3) separate governance and management relations from exclusive forms of property and relational contracts." Peer production encompasses some of the largest sustained instances of open collaboration and collective action achieved through Internet-mediated communication (Benkler et al., 2015; Cheshire, 2007; Fulk et al., 1996; Schweik & English, 2012). As a result, peer production communities provide ideal environments to analyze the effects of different sociotechnical arrangements on public information good production.

Existing research makes divergent predictions about the effects of requiring user accounts on public goods production in online communication. One body of prior studies argue that stable identifiers work as catalysts of cooperation by facilitating accountability, group identification, boundaries, and commitment that lead to higher quality participation. Others contend that requiring accounts imposes costly obstacles that deter participation. A related set of studies suggests that unregistered contributions may introduce diverse perspectives and stimulate activity among experienced community members. Thus, reductions in low quality content may lead to overall declines in high quality contribution as well. Our hypotheses, developed below, borrow from all of these approaches and reflect a synthesis of this earlier work.

Stable Identifiers as Catalysts of Cooperation

The value of stable identifiers in public information goods production and interactive communication networks derives from the importance of boundaries for cooperation and common pool resource production (Ostrom, 1990). At small scale, stable identifiers facilitate sustained communication in repeated interactions leading to the development of reputations and norms (Axelrod, 1984). Relying on identifiers, reputation systems can promote accountability and community boundaries in computer-mediated communication. Without

identifiers, anonymity online can lead to disinhibition and toxic behavior like spam and hate speech (Joinson, 1998) which tend to be counter-productive to the creation of large-scale public information goods (Kraut et al., 2012). Even light-weight identifiers can facilitate norm enforcement and more supportive interpersonal communication dynamics (Anonymous, 1998; Walther, 1996).

Stable identifiers can also support computer-mediated information sharing by reducing anonymity and enhancing trust and opportunities for social approval (Cheshire, 2007; Kollock, 1999). The low costs of reproducing and consuming digital information makes would-be contributors particularly susceptible to social incentives like feedback and status (Cheshire, 2007; Cheshire & Antin, 2008; Fulk et al., 1996; Kollock, 1999). Reduced anonymity between participants thus enhances the social feedback that sustains cooperation (Cheshire, 2007; Cheshire & Antin, 2008; Faraj & Johnson, 2011; Kollock, 1999) and reduces non-cooperative behavior (Yamagishi et al., 2009). Stable identifiers can also enhance identification and commitment online, leading to more robust, sustained participation (Kraut et al., 2012; Ren et al., 2012). Visible reputations support the formation of trust and bonding among participants, who develop a sense of collective identity (Ren et al., 2012), reinforcing group boundaries in positive ways.

In digital settings, low quality contributions can be easily undone. However, the process of filtering contributions requires resources and effort that could be allocated to more productive ends. In theory, the requirement that contributors create accounts imposes a barrier analogous to the use of "CAPTCHA" insofar as it selectively keeps out vandals and spammers who would not have made productive contributions anyway (Kraut et al., 2012). Appropriately "cheap" pseudonyms would thus lead to an increase in registered contributors and enhance the overall quality of participation so long as the system cannot be gamed too easily (Friedman & Resnick, 2001). Barriers to entry can, in this view, serve informal boundary-maintenance functions, ensuring that would-be contributors have sufficient motivation to overcome small obstacles while individuals unwilling to abide by minimal community norms will not get involved in the first place.

At minimum, the "catalysts" perspective leads us to expect that account requirements should not deter contributions from good-faith participants. In a stronger form, this perspective would predict that contributions from more socially integrated participants would lead to an overall increase in quality. Based on the approaches described above, a requirement to register an account

before contributing to a peer production project should drive away low quality contributions like spam while prompting an increase in the creation of new accounts and either stable or increasing rates of high quality contributions.

Barriers to Entry as Transaction Costs and Social Inhibitors

The speed and ease with which individuals can make small contributions to digitally-networked communication systems differentiates them from earlier knowledge sharing and management (Benkler, 2006; Bennett & Segerberg, 2012; Cheshire, 2007; Fulk et al., 1996; Kollock, 1999). For example, in Wikipedia, it takes a few seconds and very little specialized knowledge to add a missing comma to an article. Cost-oriented explanations of online collaboration argue that such radically low barriers to entry enhance the quantity of participation. Further arguments extend this to propose that even low quality participation stimulates further activity leading to enhanced content quality.

Classical theories of public goods and collective action treat barriers to entry as a resource constraint likely to deter participation (Olson, 1965; Oberschall, 1973). Low participation costs help explain why some collectives achieve critical mass while others do not (Marwell & Oliver, 1993). Reduced participation costs caused by new communication technology also explain the rise of online forms of collective action and group generalized exchange (Benkler, 2006; Bennett & Segerberg, 2012; Kollock, 1999). In online environments, low participation costs often translate into fluid or open community boundaries. Classical theories predict that these open boundaries might facilitate free riding (Olson, 1965) or undermine community stability and cohesion (Ostrom, 1990). However, open communities and crowds pursuing information sharing can benefit from larger group size as well as greater diversity of perspectives (Benkler, 2006; Bennett & Segerberg, 2012; Kane et al., 2014; Woolley et al., 2010; Zhang & Zhu, 2011). Open boundaries also make it easier for organizations to build informal, lightweight relationships with volunteers and then leverage opportunities to develop deeper identification and commitment (Ren et al., 2012).

The absence of a registration requirement may also facilitate easier onboarding of new participants and greater diversity of contributions. For example, newcomers to online communities often pursue low-commitment, marginal tasks before they create accounts (Antin & Cheshire, 2010). The opportunity to contribute to a project without a stable identifier allows par-

ticipants to experiment without risk or cost to their reputations (Anonymous, 1998; Walther, 1996), potentially yielding more innovative and novel contributions (Bernstein et al., 2011). Unregistered contributions can bring new, diverse knowledge to the table that attracts the attention of more experienced community members (Anthony et al., 2009) and provokes those experienced community members to make further improvements (Kane et al., 2014). The interactions that result can become self-sustaining (Bennett & Segerberg, 2012; Margolin & Liao, 2018) lead to higher quality information goods overall even when unregistered participants add relatively low quality content (Gorbatai, 2011).

Together these approaches contradict the "catalysts of cooperation" perspective on stable identifiers. The work emphasizing barriers to entry suggests that a shift to requiring accounts should decrease participation activity in communities, including contributions of both low and high quality. In terms of new account creation, these perspectives do not advance a clear prediction. Finally, research on the productive role of unregistered contributors suggests that requiring identifiers may inhibit overall quality improvements by reducing group size and reducing a source of stimulus for follow-on contributions, leading existing community members to participate less as well.

Tradeoffs of Stable Identifiers in Peer Production

Synthesizing the bodies of research summarized above, we suggest that stable identifiers in the organization of online public information goods entail tradeoffs. On the one hand, stable identifiers can enhance group boundaries, accountability, cohesion, and dynamics of reputation and status. Imposing stable identifiers can drive away spammers, reduce the burden of norm enforcement, and lead to virtuous cycles of engagement. On the other hand, stable identifiers impose a barrier to participation that should reduce the likelihood of contribution among some would-be participants who are not vandals. In losing these would-be participants, communities will also shrink in terms of good faith contributions and members. These reductions in contributions from outsiders, both good and bad, will undermine the activity of established contributors indirectly.

The tradeoffs we anticipate underscore the importance of several questions that prior studies have failed to resolve. On the catalysts side, studies on the beneficial impacts of stable identifiers tend to overlook whether the identifiers deter good faith participation. Meanwhile, studies in the barriers camp

have approached transaction costs as a relatively homogeneous concept, failing to make specific predictions about the effects of common types of costs. In the context of peer production and other forms of online collective action, this leads us to question whether requiring accounts for participation would constitute a sufficiently costly barrier to entry that it would drive away otherwise motivated and interested contributors? Alternatively, requiring account creation might, as many project administrators and systems designers believe, keep spammers out while having little or no effect on good faith contributors.

We derive our hypotheses from this tradeoffs perspective. First, drawing from the view that stable identifiers are not sufficiently burdensome to drive away all contributors, we anticipate that *requiring account creation for participation in peer production communities will cause an increase in the number of newly registered accounts* (*H1*).

We also expect that requiring accounts will depress the rate of contributions. Drawing from the theoretical perspective derived above, we divide this into two hypotheses. Based on both the catalysts and costs perspectives, we expect that *requiring account creation for participation in peer production communities will cause a reduction in the number of subsequently removed contributions* (*H2*). Communities require accounts because they believe the effect of doing so will primarily be to provide a costly barrier biased toward deterring low quality and unwanted forms of participation like vandalism and spam. Because vandals and spammers are not motivated through collaboration, their contributions are unlikely to be catalyzed by an account requirement.

Our third hypothesis anticipates that the imposition of the costs associated with creating and signing-in to a user account will also deter would-be contributors who would have added something of value and that *requiring account creation for participation in peer production communities will cause a reduction in high quality contributions* (*H3*). This reflects the view that requiring accounts is a broadly-felt costly barrier whose effect on would-be contributors of high quality knowledge will not be outweighed by the benefits of accounts as catalysts of cooperation.

Finally, building on the argument that unregistered contributors stimulate activity by existing community members by Gorbatai (2011), Kane et al. (2014) and others, we hypothesize that *requiring account creation for participation in peer production communities will cause a reduction in contributions from participants who edited with accounts prior to the requirement* (*H4*). Directly contradicting the catalysts point of view, this hypothesis also extends

the transaction costs argument by proposing an indirect impact of the design change on contributors who already hold accounts.

METHODS

We test these hypotheses by combining observational digital trace data from a sample of 136 peer production projects with an analytic strategy that allows us to identify causal effects. Prior research on the impact of stable identifiers and small transaction costs on public information goods does not compare across multiple communities or incorporate evidence capable of supporting causal claims. Other studies that address causal concerns have incorporated findings from small samples with limited external validity. Additionally, previous work has focused on the effect of transaction costs and stable identifiers on the aggregate *quantity* of participation without differentiating or evaluating its quality (Jay, Dunne, Gelsthorpe, & Vigo, 2016).

The communities in our study underwent a policy change after which contributing without an account became unavailable. We estimate the causal effect of the policy change on communities that previously had unregistered activity by comparing trace data on contributions immediately before and after the change using a quasi-experimental design described below. This approach follows a small number of studies that use changes to communication systems in peer production as a way of drawing causal inference (Zhang & Zhu, 2011; Geiger & Halfaker, 2013; Slivko, Kummer, & Saam, 2016).

Empirical Setting: Wikis and Wikia Inc.

Our empirical setting is a large population of peer production communities engaged in the collaborative creation of *wikis*. The term "wiki" refers to a type of software designed to facilitate the collaborative, asynchronous writing and distribution of textual content; the communities that use wiki software; and the products created by these groups (Leuf & Cunningham, 2001). Wikipedia is the most famous example of a wiki, but there are hundreds of thousands of other wikis with different goals, topics, and scopes. Through the enormous success of Wikipedia—one of the five most popular websites in the world—wikis figure prominently in Benkler's (2002) original conceptualization of commons-based peer production and are among the most visible and impactful public information goods online.

We analyze a population of wikis hosted by the for-profit firm *Wikia*.

Wikia, started in 2004, sought to apply Wikipedia's production model to a broad range of topics. Wikia was founded by Jimmy Wales, Wikipedia's founder, and Angela Beesley, an active and respected early Wikipedia contributor. Wikia's policies, structures, and technologies have been heavily influenced by Wikipedia. In terms of the content and scope, Wikia wikis vary enormously, addressing popular culture and "fan culture" topics as well as things like software, food, and fashion. Some of the largest wikis host information about massive multiplayer online video games like Halo, television shows like Breaking Bad, or cultural phenomena like puppetry. Although many firms host wikis (e.g., PBWiki, WikiSpaces, and SocialText), Wikia is unique in that it only hosts publicly accessible, volunteer-produced content.

Wikia wikis constitute a large, inclusive population of peer production communities. Wikia relies on peer production to create the content of its websites. All of its wikis invite contributions from the general public and all Wikia content is distributed freely and openly. The vast majority of Wikia communities allow contributions without an account.

Data

In September 2012, Wikia staff shared with us a list of 181 wikis that, according to their records, had changed a configuration setting that required all would-be contributors to create accounts before making an edit. The staff also provided information and documentation that the intervention happened at the request (or at least with the consent) of administrators from the affected communities. They explained that on Wikia, configuration of the wiki was generally left to administrators and that it was the staff's belief that only a small proportion of non-administrators—and no unregistered users—would have been involved in the decision to make the change. We conducted text-based searches on a randomly selected 10% of wikis in our sample and did not find any evidence that the interventions were announced ahead of time. The fact that the change was implemented in software meant that would-be contributors had no obvious way of avoiding exposure to it.

We collected public datasets on the full contribution histories of each of the affected wikis with the exception of 4 wikis whose database records has been deleted by Wikia. For each wiki, Wikia either publishes the data as an XML database dump or exposes the relevant information to construct an XML dump through their API. The XML dumps contain the full history of

every revision made by every user to every page in each wiki.[5]

The data we received from Wikia staff included only the most recent time that the option for unregistered editing had been updated in the site's configuration database. We found that the dates shared by Wikia staff corresponded to a complete cessation of editing from users without accounts in the vast majority of cases. For 8 wikis, we found compelling evidence that the configuration change had occurred earlier and adjusted the date accordingly. Details of this process, visualizations of the edit histories of these wikis, and additional models in which we do not adjust for the dates appear in the "Adjusted Cutoffs" section of our online supplement. Making this adjustment does not alter the pattern of our results.

Because our hypotheses concern the effect of requiring accounts on active communities with both registered and unregistered contributors, we restrict our sample to wikis with minimal activity and where at least some contributions were made by unregistered users before they are blocked. We use the following inclusion criteria: wikis must (1) be at least 4 weeks old when the intervention occurred; (2) have received at least 1 contribution from an anonymous contributor in the 8-week period before the change; (3) have at least one contribution in at least 70% of the weeks during the 16-week period before and after the change; and (4) experience at least a 90% decrease in the number of contributions from unregistered users in 8-week periods before and after the change indicating that the block actually occurred. We exclude at least one wiki from the main analysis for each of these reasons leaving us with a total of 136 wikis. We provide details on our sample selection, descriptive statistics for the excluded wikis, and models fit on a dataset that does not exclude any of these wikis in the "Sample Selection" section of our supplement. Although returning excluded wikis to our analysis reduces the size of our effects slightly, it does not change the substance of our results.

Measures

Using the XML database dumps, we construct an analytic dataset where the unit of analysis for the study is the "wiki week." The dataset includes 2,171 observations from a total of 136 wikis. We use four measures of our outcome variables—each of which varies within wiki over time. Tables of descriptive

[5]The dumps do not include data from pages whose entire history of contribution has been deleted. Since pages consisting only of low quality content can be deleted by administrators in this way, the measures of damage described below should be interpreted as incomplete and conservative.

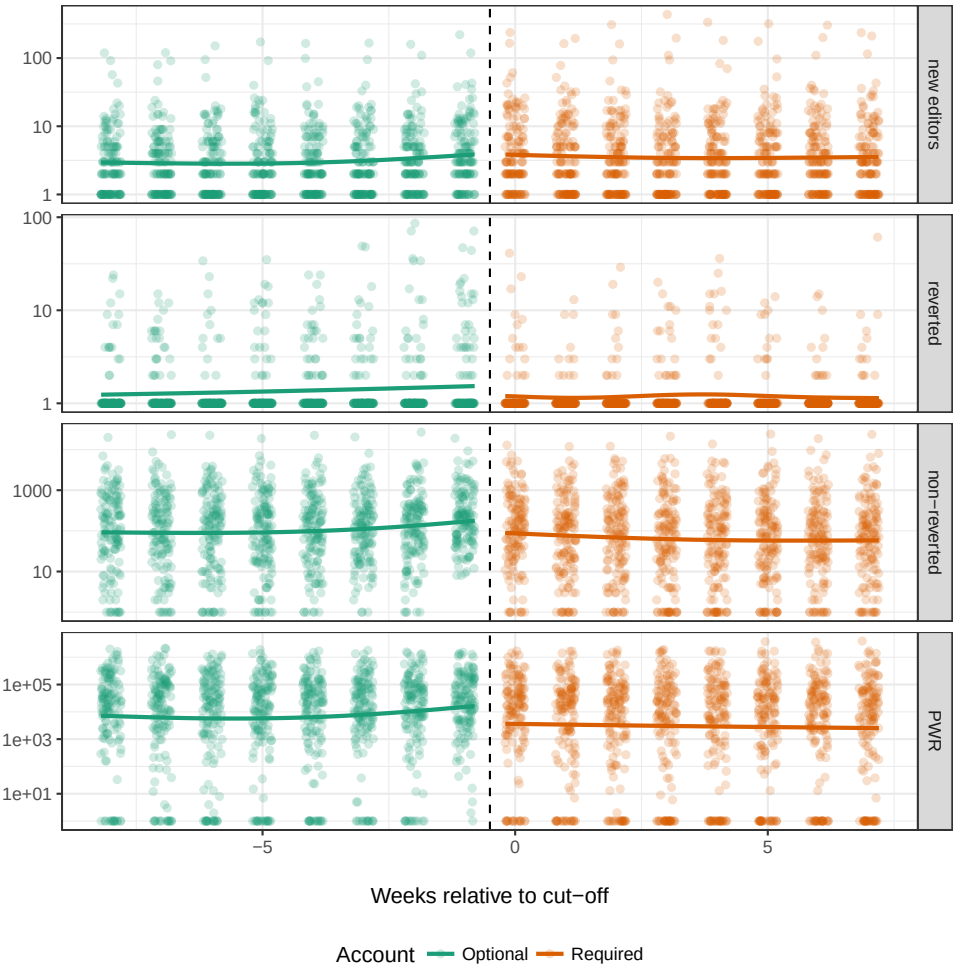


Figure 1: Plot of all dependent variable values for all wikis over all weeks in the dataset. Points are jittered horizontally within weeks, one outlier is dropped from the reverted edits plot (the second panel from the top), and Y-axes are all log-transformed to facilitate interpretation. Central trend lines before and after the cut-off generated with a generalized additive model (GAM) smoother.

statistics summarizing the size and activity of the wikis in the sample at the point of the intervention as well as the four dependent variables in our analytic dataset appear in our online supplement. Figure 1 presents scatterplots with GAM-smoothed trend lines before and after the cut-off for all four dependent variables across the full 16-week analytic window. In each panel, we observe a large amount of variation as well as a series of discontinuous shifts at the cut-off point.

For our first dependent variable (*H1*), we count the number of user ac-

counts on each wiki that have made a contribution for the first time in the history of that wiki during each week (*new accounts*). Because the XML databases only include data on the contributions to each wiki, this measure will not reflect accounts that were created but that never went on to make a contribution. If a user had created an account previously, but never made a contribution to the wiki in question, the account will be marked as new during the week it makes its first contribution.

Our second hypothesis (*H2*) looks at low quality edits which we measure by counting the number of edits that are subsequently removed (*reverted*) in their entirety. We adopt the simplest and most widely used approach to detecting reverts by focusing on what are called "identity reverts" (Priedhorsky et al., 2007; Halfaker, Kittur, & Riedl, 2011; Piskorski & Gorbatâi, 2017).[6] Using the identity revert approach, a contribution (*A*) is considered reverted if and only if a second user's contribution (*B*) returns the page to a state that is identical to its state before edit *A*. The affordances of MediaWiki make it easy for any user user who can edit (with or without an account) to create identity reverts and the software has additional affordances that allow administrators and users with special privileges to do so even more easily. Because large proportion of reverts involve vandalism and other intentionally disruptive contributions, identity reverts provide an efficient and conservative measure of damage (Flöck, Vrandečić, & Simperl, 2012). In Wikia, reverts are relatively rare, and only 1 out of every 340 edits in our dataset is reverted and half of the wikis in our sample (69 wikis) experience no reverts during our analytic window.

We construct two measures to test the effect of the intervention on higher quality contributions (*H3*). First, we count the number of contributions made to a wiki in a given week that are not subsequently reverted (*non-reverted*). Because the effort that goes into an edit can vary greatly, a single non-reverted contribution might reflect a large amount of new high quality content, the removal of some content added previously by another user, or the fix of a simple typo.

Our second more conservative measure of quality addresses this threat by relying on the collective judgment of a wiki's contributors. Computer scientists Adler and de Alfaro (2007) developed an approach that defines higher quality wiki content as material that persists through more subsequent contri-

[6]We detect reverts using the *python-mwreverts* software library published by Aaron Halfaker at the Wikimedia Foundation. See: https://github.com/mediawiki-utilities/python-mwreverts

butions by others. The most widespread operationalization of this measure is *persistent word revisions (PWR)*. The PWR for a given contribution is the sum of its tokens (i.e., words and formatting elements) multiplied by the number of subsequent versions of the page (within some fixed window) in which each token persists.[7] To construct this measure, we use the *python-mwpersistence* software library published by the Wikimedia Foundation.[8] Because highly edited pages have more revisions which create more opportunities for token revisions, we follow Adler and de Alfaro (2007) and sum the number of persistent tokens over a 7 revision window.

We select the dependent variables described above because we believe they provide the best available indicators of overall wiki community health and well-being. Wikis that attract new contributors and generate more non-reverted contributions or contributions that last longer are thriving wikis. These measures are also consistent with the theoretical predictions we develop above as well as prior empirical approaches to modeling attempts to build public information goods in online communities (Cheshire & Antin, 2008; Halfaker et al., 2011; Schweik & English, 2012; Zhang & Zhu, 2011). Differentiating between contributions of low (*H2*) and high (*H3*) quality allows us to consider the impact of the intervention in a more holistic fashion than much of this earlier work.

Our key independent variable is a measure of time in weeks centered on the date that each wiki blocked anonymous contributions (*week*). We also include a dichotomous variable (*acct_req*) that captures whether a wiki had blocked editing by users without accounts during each week (i.e., when $week \geq 0$). This point occurred at different moments in calendar time because wikis instituted the requirement individually.

Analytic Strategy and Models

We model the effect of wikis' transition to require user accounts with a *regression discontinuity design* (RDD).[9] An RDD seeks to identify causal effects by comparing outcomes before and after a threshold value marking an exogenous change in a "forcing variable" (Lee & Lemieux, 2010). Valid inference relies

[7] A detailed description of the method and a bibliography of papers that developed and employ the technique is available at: https://meta.wikimedia.org/wiki/Research:Content_persistence . Flöck, Erdogan, and Acosta (2017) describe some limitations of the approach we use.

[8] https://github.com/mediawiki-utilities/python-mwpersistence

[9] Murnane and Willett (2011) provide an excellent introduction to RDD studies.

on an assumption that outcomes are equal in expectation immediately around the threshold. Of course, the dependent variables (number of newcomers, reverted edits, etc.) might tend to increase or decrease around the threshold. RDD models this underlying "secular" trend over the forcing variable to create a more precise estimate of the change attributable to the discontinuity.

In this case, we adapt RDD methods to a within-subjects analysis with repeated observations (panel data) for every subject (wiki) capturing each subject's individual secular trend. The forcing variable is a relative measure of time (*week*) centered at the discontinuity. Our estimand (τ) is the average effect of the intervention on the communities that underwent the change. The parameter estimate $\hat{\tau}$ associated with *acct_req* captures any discontinuous shift in the outcomes at the point the wikis required accounts.

The panel RDD approach provides insight into the instantaneous effects of requiring accounts on the treated communities. We adopt this "within subjects" RDD approach for two reasons. First, it lets us compare the treated wikis to themselves immediately before and after the treatment in the absence of control cases. We do so because true control cases do not exist given the quasi-experimental nature of the change. Although matching on observables is often used to construct control cases in such situations, our preliminary evaluations and experience of Wikia led us to believe that wikis which implemented the account requirement likely differed in unobserved ways from wikis that did not. Second, we estimate the instantaneous effects because differences in outcomes farther from the discontinuity are more likely due to sources of heterogeneity unrelated to the account requirement. We discuss these choices and assumptions as well as alternative approaches at greater length in the online supplement.

Our models also each include 408 fixed effects control variables. First, we include a vector of dummy variables ***wiki*** that constitute wiki-level fixed effects. To model trends within each wiki, we interact the ***wiki*** vector with *week* and $week^2$.[10] Fixed effects in repeated measure models are typically considered preferable to random effects because their inclusion ensures that estimates of $\hat{\tau}$ reflect only within group variation (Angrist & Pischke, 2008; Murnane & Willett, 2011). In this way, our three vectors of fixed effects capture all observed and unobserved characteristics of wikis that have a consistent effect across all weeks (such as each project's start date or initial audience size) as

[10]Because of visual and statistical evidence of curvilinear secular trends within many of our wikis, we use a quadratic specification. Although a cubic model had improved goodness of fit, there was no substantive change in $\hat{\tau}$, so we report the more parsimonious model.

well as a separate quadratic trend for each wiki (such as growth or decline in the popularity of a topic covered by a wiki). As our hypotheses focus exclusively on the estimate of $\hat{\tau}$, the fixed effects serve exclusively as controls and we do not report the parameter estimates for them. The models thus take the form:

$$Y = \tau acct_req + \beta \boldsymbol{wiki} + \beta(\boldsymbol{wiki} \times week) + \beta(\boldsymbol{wiki} \times week^2) + \varepsilon \quad (1)$$

We set Y equal to each outcome measure in turn: *new users* (M1), *reverted* (M2), *non-reverted* (M3a), and *PWR* (M3b). To evaluate the effect of the intervention on contributions from experienced community members (*H4*), we re-estimate models 2-4 restricting the sample to editors who had made at least one contribution under a registered account prior to the cut-off.

Since all the dependent variables are skewed counts, we use negative binomial regression.[11] We calculate heteroskedasticity-robust standard errors using the HC_1 estimator (MacKinnon & White, 1985) and follow Angrist and Pischke's (2008) suggestion to report the larger of robust or conventional standard errors. Finally, we conduct a supplementary analysis that decomposes $\hat{\tau}$ across each wiki in our sample. To do this, we re-estimate our models adding a term interacting ***wiki*** with *acct_req* to recover individual wiki-level estimates for $\hat{\tau}$. Although some wikis in our analysis had contributions from unregistered editors after the cut-off, we estimate the impact of the intervention irrespective of compliance. Our results thus correspond to an *Intent-to-treat (ITT)* estimate (Murnane & Willett, 2011) in that we measure the impact of a change even though it may not have been perfectly effective in every case.

RESULTS

We find substantial discontinuous shifts in each of our dependent variables when wikis began requiring accounts: an increase in the number of new editors and decreases in the numbers of reverted edits, non-reverted edits, and PWR. These outcomes support hypotheses *H1-H3*. We also find similar decreases among editors who contributed with accounts prior to the intervention, supporting hypothesis *H4*.

Table 1 summarizes the results of *H1-H3* and includes regression models fit using the full sample. We report the estimated parameter ($\hat{\tau}$) associated

[11] Given the relatively small amount of variation in *reverts* in our sample, we report a logistic regression model of whether wikis experience any reverted edits in a section titled "Logistic Specification for Reverts" in our online supplement. The results suggest a similar takeaway.

	new editors (M1)	*reverted* (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	0.201***	−1.124***	−0.491***	−0.586***
	(0.057)	(0.332)	(0.094)	(0.118)
Deviance	2157.498	549.198	2540.722	2599.352
Num. obs.	2171	2171	2171	2171

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Summary of regression models. We omit estimates for our 408 control variables in the form of fixed effects for ***wiki*** and its interaction with *week* and *week*^{2}.

with *acct_req* from each model. Because these are non-linear models, interpreting the coefficients directly can be challenging. In order to facilitate interpretation, we also present the model-predicted values for prototypical wikis in Figure 2.[12]

Model M1 supports *H1* that requiring accounts leads to a higher number of new user accounts. We estimate that the number of accounts making their first contribution rose from 6.64 to 8.12 new users per week in a prototypical community at the point that they required accounts ($\beta = 0.201, \sigma = 0.057, p < 0.001$). This change constitutes a 22% increase. However, the standard errors are relatively large and the 95% confidence interval for the parameter estimate corresponds to an increase of between 0.62 and 2.44 new editors (9% and 37% of the total in a prototypical wiki). Over the course of eight weeks following the change, we predict a total number of new editors that is 122% what we would have predicted in the absence of the shift.

Model M2 provides strong support for *H2*, that requiring accounts drives down levels of low quality edits. M2 estimates a decline from 6.64 to 2.16 reverted edits per week (68%) in a prototypical community ($\beta = -1.124, \sigma = 0.332, p = 0.001$). Our 95% confidence interval corresponds to a decrease between 5.5 and 2.5 reverted contributions (83% and 38% of the total in a prototypical community). For less active wikis in our sample this would involve a complete elimination of reverted edits. Over the course of eight weeks, we predict a total number of reverts that is 32% what we would predict with unregistered contribution permitted.

In Models M3a and M3b we find that the block also caused a decrease in quality edits (supporting *H3*). Using our first measure of quality, M3a shows

[12]Figure 2 summarizes an "average" trend by plotting predicted values from a negative binomial model that regresses the fitted values generated by the original models on a quadratic specification of week and the indicator for the cut-off.

	reverted (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	—0.785*	—0.478***	—0.473***
	(0.320)	(0.093)	(0.121)
Deviance	481.858	2518.216	2508.188
Num. obs.	2164	2164	2164

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Summary of regression models estimated on sample restricted only to editors whose accounts were created prior to the cut-off. Estimates for our 408 fixed effects for ***wiki*** and its interaction with *week* and $week^2$.

that the cut-off led to a decrease from 658 to 403 non-reverted contributions each week (39%) in a prototypical community ($\beta = -0.491, \sigma = 0.094, p < 0.001$). Our 95% confidence interval for this parameter estimate encompasses between 323 and 174 contributions per week (i.e., decreases of 49% and 27% of the total). Over eight weeks, we predict a total number of non-reverted edits that is 61% what we would have predicted otherwise. The estimates in M3b are very similar to those in M3a. We estimate that the intervention led to a decrease of 44% from 126,000 to 70,200 PWR per week at the cut-off in a prototypical community ($\beta = -0.586, \sigma = 0.118, p < 0.001$). with 95% confidence intervals between 70,400 and 37,900 word revisions per week (i.e., decreases of 56% and 30% of the total). Over eight weeks, we predict a total number of PWR that is 56% the predicted total in the absence of the shift.

The model results on the restricted sample of editors with accounts prior to the cut-off (corresponding to *H4*) is shown in Table 2. M1 is not included because all new accounts among editors active before the cut-off must have been created before as well. The results for Models M2, M3a, and M3b on the restricted sample suggest very similar pattern to those from the full sample above. M2 finds that the number of reverted edits among experienced editors undergoes a discontinuous decrease after the intervention and the magnitude is smaller than in the full sample. Models 3a and 3b also show a discontinuous decrease in the number of non-reverted edits and PWR respectively. The magnitudes of each estimate are nearly identical to those reported in the main analysis. This follows the pattern observed in both Gorbatai (2011) and Kane et al. (2014), where contributions from unregistered editors stimulate contributions from existing editors.

In order to understand the heterogeneity of the effect among wikis, we fit a series of models incorporating an interaction between ***wiki*** and *acct_req*.

These models generate a separate point estimate of $\hat{\tau}$ for each wiki. The results, described in our online supplement, show substantial heterogeneity. For M1, we find that *more* wikis show a decrease in new editors (73) than an increase (60) suggesting that our conclusion in terms of H1 may be driven by large increases among a subset of wikis. The results for M2, M3a, and M3b each suggest that about 1 wiki in our sample experiences a shift that is opposite in sign to the average effect for every 2 wikis that shift in the average direction. In the case of M2, absence of reverts in many wikis means our estimates are are very close to 0 for about half of our sample (72 wikis). Overall, this suggests that aggregate estimates accurately characterize the trends among wikis while demonstrating that heterogeneous effects also occurred.

Threats to Validity

Several threats to the validity of our findings stem from challenges common to RDDs and the panel RDD approach we adopt (Lee & Lemieux, 2010; Murnane & Willett, 2011). Below, we briefly discuss the most salient threats and, when possible, the sensitivity analyses used to assess them. We also revisit the assumptions that underpin the validity and generalizability of our findings. In our online supplement, we present additional analysis, technical threats, and the full results of all sensitivity analyses and alternative model specifications.

Administrators' knowledge of the intervention poses arguably the most critical threat to our assumption of the exogeneity of the precise timing of the intervention in relation to the dependent variables. This threat, a specific case of what is called "crossover" in quasi-experimental methods, arises from the fact that some community administrators would have known of the intervention ahead of time and could have manipulated their pre/post-treatment behavior. Although it is impossible to prove the absence of crossover, we test the sensitivity of the findings to the inclusion of the administrators in the analysis by dropping edits from administrators. The results are similar but the effects are slightly larger in magnitude.

In other sensitivity analyses, the effects of the intervention on low and high quality edits are robust to alternative specifications. Under all approaches explored in the supplement, our main analysis estimates for M2, M3a, and M3b retain the same sign. Differences in the magnitude of the point estimates under these alternative specifications remain smaller than the original standard errors. In contrast, the effects on new account creation (M1) are more fragile. Under several specifications, the point estimate shrinks substantially

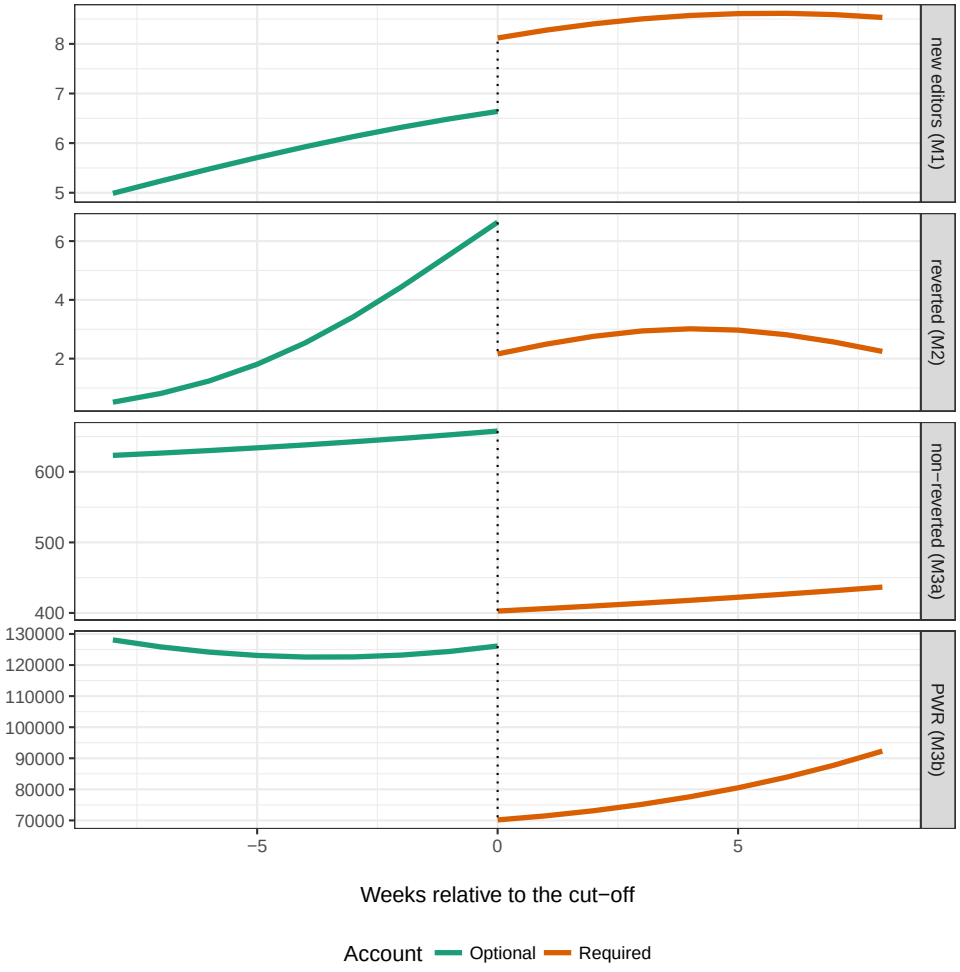


Figure 2: Visualizations of our model predicted values for prototypical wikis for all four outcomes. To represent the effect of an "average" wiki in our sample, trend lines reflect a fitted regression model of window weeks and the cut-off on the predicted values from each of our models reported in Table 1. We do not plot predicted values corresponding to *H4* because the estimates align so closely with those for *H3a* and *H3b*.

or becomes indistinguishable from zero. We also find that the increase estimated in the main analysis is not robust to dropping the largest wikis from the analysis (see the subsection on "Influential Cases" in the online supplement). The estimates of *acc_req* in these models are consistently positive but smaller in magnitude and are not statistically significant in models that drop the largest 5% and 10% of wikis. Although we can imagine explanations the effect of requiring accounts on new editors might vary with project size or other factors, we know of no theoretical reasons to anticipate this. It is also possible that the fragility of the estimate is an artifact of fact that the effect on new editors is relatively small and noisy and that much of the estimated effect is concentrated in large wikis. As a result of these tests, the results of M1 (see Table 1) should be interpreted with care.

This panel RDD approach rests on the assumption that individual wikis were "equal in expectation" to themselves immediately before and after the point of the intervention. As discussed earlier, we are concerned that administrators of the treated wikis played a role in the decision to require accounts. We have tried to evaluate this assumption through both qualitative investigation (i.e., conversations with wiki administrators and manual review of on-wiki discussions by participants) as well as empirical sensitivity analysis described above and reported in the online supplement. Although we find no likely causes of bias in this regard, we cannot eliminate the possibility of this threat. In the absence of an empirical matching strategy or credible control group with which to construct alternative baselines for estimating treatment effects, we do not claim that our results generalize to untreated wikis. Additional assumptions and further analysis would be necessary to extend our analysis in this way.[13]

Tests of the external validity of our estimates lie beyond the scope of the current paper. Our central finding that requiring accounts causes a decline in both low quality and high quality contributions resonates with related work in the contexts of citizen science (Jay et al., 2016) and discussion forums (Bernstein et al., 2011). In particular, Jay et al. (2016) estimate an increase of over 60% in the number classifications to a citizen science platform when accounts are not required. The Wikimedia Foundation has also released an unpublished study with comparable point estimates of a 25—30% decrease in overall productivity when accounts were requested of would-be contributors.[14] We believe that these data points provides reasons to be optimistic about the

[13] See our online supplement for further discussion of this issue.

[14] https://meta.wikimedia.org/wiki/Research:Asking_anonymous_editors_to_register

generalizability of our results.

Lastly, our estimates may not generalize to other types of account requirements. Although the design of the Wikia account creation and sign-in system resembles similar systems on a number of peer production projects, social media platforms, and online discussion communities, the system design could be adjusted to streamline the experience of account creation, emphasize the value of stable identifiers, or incorporate other elements. We anticipate that such alternative designs could alter the effects we estimate here.

Discussion

The results support our hypotheses and suggest that stable identifiers impose barriers to entry that introduce a tradeoff for online communities. While the account requirement almost completely eliminates reverted (low quality) contributions and induces some individuals to create new accounts, it also decreases high quality participation enormously. We find evidence for this tradeoff even within the editing activity of editors registered prior to the cutoff demonstrating how participation by unregistered editors stimulates activity across the board. Accumulated over many weeks, these effects would profoundly change the scope and character of participation within these projects.

The results diverge from earlier findings and theories consistent with the "catalysts" perspective which would have predicted stable or increased rates of high quality contributions. What explains this? Differences between experimental contexts (e.g., Yamagishi et al., 2009) and the wikis in our sample offer one potential explanation. Perhaps communities that require accounts for participation at the point they are initially constituted (like those in most experiments) experience different effects than existing communities (like those in our sample). It might also be the case that specific institutional arrangements support some types of public goods production or collective action better than others. In this sense, our findings may be interpreted as supporting the idea that particular types of community boundaries may promote or hinder particular sorts of common pool resource management (Cox, Arnold, & Villamayor Tomás, 2010). The account requirements in our study may also lead to subtle long-term effects on community culture that we do not estimate or observe due to the constraints of our measures and analytic approach.

Although our findings resonate more with the costly barriers perspective, they deviate from it as well. The decline we see among registered editors does not follow from a pure transaction costs approach which might

also predict that these registered editors would contribute more new content (i.e., PWRs) following the removal of low quality participation that may have required them to focus on vandal fighting instead. Instead, unregistered contributions—even the low quality ones—seem to induce community activity. This supports earlier claims about the value of group size and peripheral participation for collective action and public goods production in online groups or crowds (Anthony et al., 2009; Bennett & Segerberg, 2012; Gorbatai, 2011; Kane et al., 2014; Margolin & Liao, 2018; Olson, 1965; Woolley et al., 2010; Zhang & Zhu, 2011). Our analysis cannot unpack this finding more fully because our data does not include additional information about the individual contributors who had participated without accounts prior to the cut-off. Future work should analyze the content of unregistered contributions more deeply to understand the value they add to communities and use network analysis and related methods to explain the structural mechanisms by which unregistered contributions enhance group collaboration.

Our theory that community boundaries introduce tradeoffs by reducing both low and high quality contributions suggests some possible common ground between the catalysts and costs perspectives. Even a strong proponent of the catalysts point of view might not be surprised that an imprecise intervention like an account requirement would deter some high quality contributions. However, the lopsided empirical tradeoff we observe in our results strongly favors preserving opportunities for unregistered contribution. For example, in the prototypical wiki shown in Figure 2, our models estimate that the account requirement will deter 260 edits but that only 4.5 of those contributions (1.7%) would have been reverted. In order to justify requiring accounts in our empirical context, the damage prevented by eliminating any single reverted edit would need to outweigh the loss of 58 non-reverted contributions. Given that removing damage in peer production is typically easier than substantive content development, we find it hard to imagine how these findings could be used to support arguments in favor of requiring accounts. In that sense, our estimates will likely surprise both the administrators of the wikis in our sample who took steps to require accounts as well as "catalysts" theorists whose work supported the decision to do so.

At the same time, if we focus on the results in terms of relative proportions (instead of the raw numbers of low and high quality contributions), a sufficiently large decline in low quality contributions could provide benefits that offset a smaller decline in high quality contributions. Similarly, if each individual low quality act were incredibly damaging and costly—something

that does not seem to be the case in our setting—the tradeoff might be seen as worthwhile. Such scenarios extrapolates beyond the current study in a number of ways and introduces risks of generalizing outside our population.

In any case, the visible benefits of reduced vandalism or spam must be weighed against the hidden costs of reduced high quality contributions. Theories in this space need to more fully accommodate the tensions between open community boundaries and effective safeguards of contribution quality. The empirical tradeoff at the heart of our analysis suggests that requiring accounts will not benefit most communities in the long run. Instead, our findings suggest that most communities should preserve opportunities for unregistered contributions. While unidentified and anonymous participation may bring challenges around disinhibition, lack of accountability, and breakdowns of trust, our results indicate that requiring accounts erects a boundary around participation that undermines the capacity of open communities to produce valuable public goods. To fight low quality contributions, these communities might seek more precise barriers to entry or might attempt to lower the cost of dealing with undesirable forms of participation.

Conclusion

Our study contributes to communication research on public information goods, stable identifiers in computer-mediated communication, information systems, and online communities and crowds. We provide the first large-scale observational evidence of the causal effects of requiring accounts in peer production systems. The tradeoffs we theorize and observe synthesizes prior research that implied incongruous empirical predictions. Our findings reveal limitations of this earlier work and show that a specific implementation of community boundaries through imposing the use stable identifiers impacts participation in multiple ways. Overall, the account requirement decreases both low and high quality contributions to the wikis in our study. Although it deters a large portion of all low quality participation, the vast majority of deterred contributions are high quality. Most surprisingly, we observe a large decrease among contributors who hold accounts already and are not directly impacted by the design change.

The findings illustrate how the empirical tradeoffs play out across diverse communities. We conclude that the results support the preservation of opportunities for unregistered, low cost forms of contribution to public information goods online. However, these implications depend on contextual fac-

tors such as the costs of dealing with low quality participation. For example, automated tools to filter spam can scale well with minimal overhead. Large projects such as the English language Wikipedia have anti-vandalism bots that effectively minimize the cost of removing many low quality contributions. As a result, Wikipedia can afford to enable unregistered editing and draw benefits from good-faith contributors who would not participate otherwise. This illustrates how the nature of contribution costs likely also matter more than prior work had suggested. Completing a CAPTCHA is not the same as creating an (even pseudonymous) account on a site. Both the content and framing of interfaces shape how newcomers experience the boundaries of a community organization and interpret the significance of participation.

The results of this study suggest several directions for future research. A field experiment could support valid comparisons of communities with and without account requirements over longer time periods. Subgroup analyses might decompose the outcomes we observe here to identify particular types of wiki communities that experienced the effects of the intervention to varying degrees. Other approaches can support deeper investigation of the mechanisms by which unregistered contributions support overall group activity. Community managers and systems designers might also investigate alternative boundary management or quality control systems as means of promoting the sort of prosocial cooperation and collective action identified in previous research. For example, prior work has described system designs that support privacy as well as safety in the moderation of user-generated content (Frey, Bos, & Sumner, 2017). Our findings speak to the importance of continuing to test and refine both the theories and empirical approaches in these domains.

References

Adler, B. T. & de Alfaro, L. (2007). A content-driven reputation system for the Wikipedia. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 261–270). WWW '07. New York, NY, USA: ACM. doi:10.1145/1242572.1242608

Angrist, J. D. & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Anonymous. (1998). To Reveal or Not to Reveal: A Theoretical Model of Anonymous Communication. *Communication Theory*, *8*(4), 381–407. doi:10.1111/j.1468-2885.1998.tb00226.x

Anthony, D., Smith, S. W., & Williamson, T. (2009, August). Reputation and reliability in collective goods: The case of the online encyclope-

dia Wikipedia. *Rationality and Society*, *21*(3), 283–306. doi:10.1177/1043463109336804

Antin, J. & Cheshire, C. (2010). Readers are not free-riders: Reading as a form of participation on Wikipedia. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (pp. 127–130). CSCW '10. New York, NY, USA: ACM. doi:10.1145/1718918.1718942

Axelrod, R. M. (1984). *The evolution of cooperation*. New York, NY: Basic Books.

Benkler, Y. (2002). Coase's penguin, or, Linux and the nature of the firm. *Yale Law Journal*, *112*(3), 369–446.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Benkler, Y., Shaw, A., & Hill, B. M. (2015). Peer production: A form of collective intelligence. In T. W. Malone & M. S. Bernstein (Eds.), *Handbook of collective intelligence* (pp. 175–204). Cambridge, MA: MIT Press.

Bennett, W. L. & Segerberg, A. (2012). The logic of connective action. *Information, Communication & Society*, *15*(5), 739–768. doi:10.1080/1369118X.2012.670661

Bernstein, M. S., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., & Vargas, G. (2011). 4chan and /b/: An analysis of anonymity and ephemerality in a large online community. In *Fifth International AAAI Conference on Weblogs and Social Media*. Palo Alto, CA: AAAI Publications. Retrieved October 2, 2017, from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2873/4398

Cheshire, C. (2007). Selective incentives and generalized information exchange. *Social Psychology Quarterly*, *70*, 82–100.

Cheshire, C. & Antin, J. (2008). The social psychological effects of feedback on the production of Internet information pools. *Journal of Computer-Mediated Communication*, *13*(3), 705–727. doi:10.1111/j.1083-6101.2008.00416.x

Cox, M., Arnold, G., & Villamayor Tomás, S. (2010, December). A review of design principles for community-based natural resource management. *Ecology and Society*, *15*(4). doi:10.5751/ES-03704-150438

Faraj, S. & Johnson, S. L. (2011). Network exchange patterns in online communities. *Organization Science*, *22*(6), 1464–1480. doi:10.1287/orsc.1100.0600

Flöck, F., Erdogan, K., & Acosta, M. (2017). TokTrack: A complete token provenance and change tracking dataset for the English Wikipedia. In *Proceedings of the International Conference on Web and Social Media (ICWSM '17)*. Palo Alto, California: AAAI. Retrieved August 10, 2017, from https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15689/14890

Flöck, F., Vrandečić, D., & Simperl, E. (2012). Revisiting reverts: Accurate revert detection in Wikipedia. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media* (pp. 3–12). HT '12. New York, NY: ACM. doi:10.1145/2309996.2310000

Frey, S., Bos, M. W., & Sumner, R. W. (2017, July). Can you moderate an unreadable message? 'Blind' content moderation via human computation. *Human Computation*, *4*(1), 78–106. doi:10.15346/hc.v4i1.5

Friedman, E. J. & Resnick, P. (2001). The social cost of cheap pseudonyms. *Journal of Economics & Management Strategy*, *10*(2), 173–199. doi:10.1111/j.1430-9134.2001.00173.x

Fulk, J., Flanagin, A. J., Kalman, M. E., Monge, P. R., & Ryan, T. (1996). Connective and communal public goods in interactive communication systems. *Communication Theory*, *6*(1), 60–87. doi:10.1111/j.1468-2885.1996.tb00120.x

Geiger, R. S. & Halfaker, A. (2013). When the levee breaks: Without bots, what happens to Wikipedia's quality control processes? In *Proceedings of the 9th International Symposium on Open Collaboration (OpenSym '13)* (6:1–6:6). New York, NY: ACM. doi:10.1145/2491055.2491061

Gorbatai, A. (2011). *Aligning collective production with demand: Evidence from Wikipedia*. Cambridge, MA. Retrieved from http://faculty.chicagobooth.edu/workshops/orgs-markets/past/pdf/gorbatai.pdf

Halfaker, A., Kittur, A., & Riedl, J. (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)* (pp. 163–172). New York, New York: ACM. doi:10.1145/2038558.2038585

Jay, C., Dunne, R., Gelsthorpe, D., & Vigo, M. (2016). To sign up, or not to sign up?: maximizing citizen science contribution rates through optional registration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1827–1832). CHI '16. New York, NY: ACM. doi:10.1145/2858036.2858319

Jemielniak, D. (2014). *Common knowledge?: An ethnography of Wikipedia*. Stanford, CA: Stanford University Press.

Joinson, A. (1998). Causes and implications of disinhibited behavior on the Internet. In J. Gackenbach (Ed.), *Psychology and the Internet: Intrapersonal, interpersonal, and transpersonal implications* (pp. 43–60). San Diego, CA, US: Academic Press.

Kane, G. C., Johnson, J., & Majchrzak, A. (2014). Emergent life cycle: The tension between knowledge change and knowledge retention in open online coproduction communities. *Management Science*, *60*(12), 3026–3048. doi:10.1287/mnsc.2013.1855

Kollock, P. (1999). The economies of online cooperation: Gifts and public goods in cyberspace. In M. Smith & P. Kollock (Eds.), *Communities in cyberspace* (pp. 220–239). London, UK: Routledge.

Kraut, R. E., Resnick, P., & Kiesler, S. (2012). *Building successful online communities: evidence-based social design*. Cambridge, MA: MIT Press.

Lee, D. S. & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355. doi:10.1257/jel.48.2.281

Leuf, B. & Cunningham, W. (2001). *The Wiki way: Quick collaboration on the web*. Boston, MA: Addison-Wesley Longman.

MacKinnon, J. G. & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*(3), 305–325. doi:10.1016/0304-4076(85)90158-7

Margolin, D. & Liao, W. (2018). The emotional antecedents of solidarity in social media crowds. *New Media & Society*. doi:10.1177/1461444818758702

Marwell, G. & Oliver, P. (1993). *The critical mass in collective action: A micro-social theory*. Cambridge, MA: Cambridge University Press.

Monge, P. R. & Contractor, N. S. (2003). *Theories of communication networks*. Oxford, UK: Oxford University Press.

Murnane, R. J. & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. New York, NY: Oxford University Press.

Oberschall, A. (1973). *Social conflict and social movements*. Englewood Cliffs, NJ: Prentice-Hall.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. New York, NY: Cambridge University Press.

Piskorski, M. J. & Gorbatâi, A. (2017). Testing Coleman's social-norm enforcement mechanism: Evidence from Wikipedia. *American Journal of Sociology*, *122*(4), 1183–1222. doi:10.1086/689816

Priedhorsky, R., Chen, J., Lam, S. (K., Panciera, K., Terveen, L., & Riedl, J. (2007). Creating, destroying, and restoring value in wikipedia. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (pp. 259–268). GROUP '07. New York, NY: ACM. doi:10.1145/1316624.1316663

Ren, Y., Harper, F., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., & Kraut, R. (2012). Building member attachment in online communities: Applying theories of group identity and interpersonal bonds. *Management Information Systems Quarterly*, *36*(3), 841–864.

Schweik, C. M. & English, R. C. (2012). *Internet success: A study of open-source software commons*. Cambridge, MA: MIT Press.

Slivko, O., Kummer, M., & Saam, M. (2016). Identification of Causal Effects in the Context of Mass Collaboration. In *Mass Collaboration and Education* (pp. 391–411). Berlin, Germany: Springer.

Walther, J. B. (1996). Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research*, *23*(1), 3–43. doi:10.1177/009365096023001001

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*(6004), 686–688. doi:10.1126/science.1193147

Yamagishi, T., Matsuda, M., Yoshikai, N., Takahashi, H., & Usui, Y. (2009). Solving the lemons problem with reputation. In K. S. Cook, C. Snijders, V. Buskens, & C. Cheshire (Eds.), *eTrust: Forming relationships in the online world* (pp. 73–108). New York, NY: Russell Sage Foundation.

Yuan, Y., Fulk, J., Shumate, M., Monge, P. R., Bryant, J. A., & Matsaganis, M. (2005). Individual participation in organizational information commons. *Human Communication Research*, *31*(2), 212–240. doi:10.1111/j.1468-2958.2005.tb00870.x

Zhang, X. & Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at chinese Wikipedia. *The American Economic Review*, *101*(4), 1601–1615. doi:10.2307/23045913

THIS IS AN UNPUBLISHED DRAFT OF A CHAPTER FROM A FORTHCOMING BOOK. WHEN PUBLISHED, THE CITATION FOR THIS CHAPTER WILL BE: Hill, B. M., & Shaw, A. D. (2017) Studying Populations of Online Communities. In Foucault Welles, B. & González-Bailón, S. (Eds.), *The Handbook of Networked Communication*. New York, New York: Oxford University Press.
BEFORE CITING OR DISTRIBUTING, PLEASE CHECK FOR AN UPDATED VERSION AT:
https://mako.cc/academic/hill_shaw-populations_of_communities-DRAFT.pdf

Studying Populations of Online Communities

Benjamin Mako Hill*
makohill@uw.edu

Aaron Shaw*
aaronshaw@northwestern.edu

INTRODUCTION AND BACKGROUND

The discussion sections of empirical research on digital communities are littered with unsatisfying phrases. Networked communication researchers have read and, in many cases written, sentences like the following: our study is only of one web forum and we cannot know how our work generalizes to others; because we have gathered data only from Wikipedia, we only propose that our findings might be antecedents of that community's success; messages in our dataset were collected from one Facebook group and we cannot speak to all the conversation that may have happened elsewhere. These phrases are acknowledgements of real problems. We think we can do better.

The limitations identified in the sentences above each stem, in large part, from the fact that most empirical online community research looks within individual communities. Although there have been literally thousands of papers published about individual peer production projects like Wikipedia and Linux, there has been vanishingly little research that compares peer production projects to each other. Likewise, studies of communication in social net-

*Authors contributed equally and author order was determined randomly.

working sites and discussion groups have nearly always focused on interactions within a single network like Facebook or an individual discussion forum. Such research does not always generalize well. Furthermore, we foreclose many types of research questions by selecting research sites for their size, longevity, or the engagement level of their participants.

In this chapter we argue that by studying groups of communities, we can enhance the quality of online communities research in multiple ways. First, we can mitigate many common threats to the validity of online community research designs. Moreover, by studying groups of online communities, we open the door to answering types of questions that are unanswerable when our datasets begin and end at communities' borders.

A few key points about population-level analysis and online communities will help contextualize the rest of our argument. First, we are not advocating for anything especially radical by arguing that research on networked communication in online communities ought to do more to adopt population-level approaches. Empirical communication research into online communities has existed for several decades[1] and we cite many examples of population-level approaches. Moreover, population-level analysis exemplifies well-established traditions in organization science, which has sought a turn away from models of organizations as self-contained entities to models that treat them as actors in complex social environments; a shift that inspired a proliferation of comparative, population-level approaches to organizational analysis (Scott & Davis, 2006). We draw inspiration from these approaches and seek to describe how they might be applied in the context of research on online communities. We also believe that the empirical contexts of online communities present an extraordinary opportunity to extend these concepts and frameworks.

A second point concerns what exactly we mean when we talk about populations of online communities. In ecology, the term "population" is used to describe all interbreeding organisms that share a geographical area. In demography, the term is used to refer to groups of humans. In organizational sociology, it is used to refer to similar types of organizations that might compete for resources like customers or suppliers. Although others have defined the term more narrowly (e.g. Weber, Fulk, & Monge, 2016), we use the term "population" broadly and inclusively to refer to any groups of online communities whose membership is defined through similarity, competition, or

[1] We do not attempt to provide a comprehensive accounting of this prior work. Interested readers might seek out Kraut and Resnick (2012), which is an excellent book-length synthesis of much empirical online community research.

interaction. As is the case in ecology, demography, and organizational sociology, there are many ways to define populations of online communities and a single online community might belong to a large number and variety of distinct populations.

Most of our discussion will focus on populations of online communities defined through their use of a common technical platform. Examples of populations of this type might include discussion forums that use the same bulletin board software, wikis that use the same hosting service, or software development projects using the same tools to host their code and coordinate their work. For example, SourceForge or Github each provide a common technological platform that hosts millions of different online communities dedicated to creating software projects, LinkedIn and Facebook each act as platforms to host online communities associated with many different offline organizations and interest groups, and Reddit hosts hundreds of thousands of "sub-reddit" communities around different discussion topics. Because communities within this type of population are mediated by a common technology, data is often more readily available and consistent as a single host or platform provider may provide access via a single Application Programming Interface (API) or database. The resulting datasets support the advantages of computational social science (Lazer et al., 2009) and can often facilitate direct comparative analysis.

Of course, there are other ways to define populations. For example, a population of online communities might include all communities hosting conversations about a set of topics or beliefs or serving a single geographic area. Examples might be as narrow as the group of all Star Wars fan culture communities or as broad as all music sharing networks. The analogy to an industry of firms or a denomination of churches maps more or less directly in cases like these. A population might also be described in terms of communities whose membership overlaps or through which messages diffuse. An exciting body of research, including studies of diffusion which we discuss below, has constructed large platform-spanning populations of communities.

In the rest of this chapter, we use examples from recent empirical research to highlight five benefits of studying populations of online communities. First, we argue that studies of populations can lead to increased generalizability. That said, community-spanning research designs not only improve the type of research already done within online communities, but also bring entirely different questions into the realm of answerability. In particular, we highlight the ability to study community-level variables like group size, inter-

community behavior like knowledge diffusion, and the ways that communities affect each other through dynamics like competition. We also argue that data from populations of online communities makes it possible to combine many of these benefits with the benefits of intra-community analyses. Finally, we discuss a series of limitations before concluding.

Benefit 1: Generalizability

The first benefit of studying populations of communities is simple and straightforward: studies of a single community – no matter how exhaustive, granular, and expertly designed – may produce findings that hinge on the idiosyncrasies of the community being studied. While this is intuitive and widely-acknowledged, analyses of networked behavior and group processes frequently focus on a single platform or online community (e.g., Facebook, Twitter, Wikipedia) and result in findings that do not apply more broadly. Analyzing data drawn from multiple sites, communities, or platforms can lead to greater generalizability.

Early scholars of online communities gathered and compared evidence from case studies of the discussion-based systems Usenet, text-based role-playing games like "multi-user dungeons" (MUDs), the pioneering community "the WELL," and others (e.g., Kollock & Smith, 1999). Over time, the forces that brought about the rise of computational social science (Lazer et al., 2009), including widespread availability of digital trace data, and reduced costs of computing and storage, have made more direct comparisons across communities possible. However, despite the existence of opportunities for inter-community empirical work, the vast majority of studies in online community research still focus on a single site. Generalizability is hurt because communities are heterogeneous in ways that are frequently poorly understood.

This point is illustrated by Hargittai (2015) who uses data from offline surveys to describe the user bases of large online communities like Facebook and Twitter in traditional demographic terms. Although Hargittai's point is that participants in these communities are far from a cross-section of society, she also suggests that the demographics of online communities' user bases are distinct from each other. Of course, decades of social science have shown that demographic characteristic like age, race, gender, and skills are correlated with many of the attitudes and behaviors that social scientists study. Because online communities are different from the general population in demographic terms, Hargittai suggests that generalizability to society is often unjustified. An implication of her findings is that, to the extent that communities also

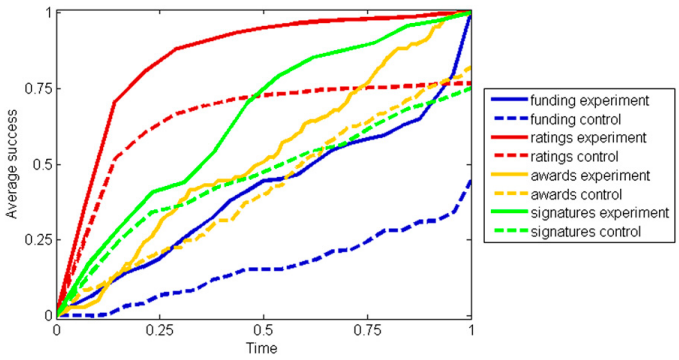


Figure 1: The success-breeds-success effect over time. The curves represent running numbers of donations (blue), positive ratings (red), awards (yellow), and campaign signatures (green) in the experimental conditions (solid lines) and the control conditions (dashed lines). The horizontal axis is normalized so that 0 marks the time of experimental intervention, and 1 marks the end of the observation period. The vertical axis is normalized so that for each system a value of 1 equals the maximum across time and conditions. (Figure and caption adapted from van de Rijt, Kang, Restivo, and Patil (2014)).

differ systematically from each other, generalization between communities will be fraught as well.

Studies across populations of online communities can help increase confidence in generalizability. For example, two studies published by Michael Restivo and Arnout van de Rijt look at the effects of social awards in Wikipedia called "barnstars" (Restivo & van de Rijt, 2012, 2014). In the first of these projects, the authors select very active Wikipedians and award barnstars to random subsets. Award recipients go on to edit the encyclopedia more than users not given an award. They also go on to receive more awards in the future. The authors suggest that such peer-to-peer forms of public recognition may produce a "success-breeds-success" dynamic. In a follow-up paper, van de Rijt, Kang, Restivo, and Patil (2014) compare the evidence of this effect in Wikipedia to very similar field experiments in three different online communities and show that the dynamic they identified in Wikipedia is also present in donations to Kickstarter, positive ratings of products on Epinions.com, and signatures on the e-petition site Change.org. The effects in all four experiments are shown together in Figure 1.

Many of the most studied online communities – even those that appear extremely similar – are unusual in ways that impact findings and limit generalizability. For example, in an extensive comparative analysis of the ten largest language editions of Wikipedia, Ortega (2009) details many instances where

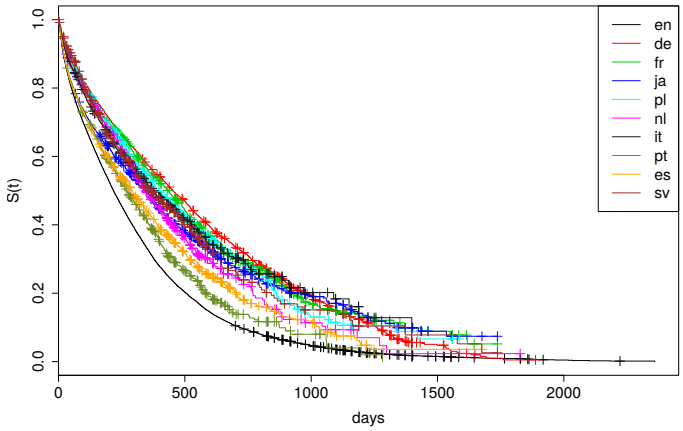


Figure 2: Survival functions that show the likelihood that editors participating in the core group of contributors continue to participate in each project once they cease to be a member of that core group. English Wikipedia is labeled "en" and is represented by the black line closest to the *x*-axis. (Figure adapted from Ortega (2009) page 125.)

the English edition is very different from Wikipedias in other languages. One example is editor retention. English Wikipedia's low editor retention has been an extremely widely studied problem (see Halfaker, Kittur, & Riedl, 2011; Halfaker, Geiger, Morgan, & Riedl, 2013), but most analyses of it can only speculate as to whether editor retention represents an issue endemic to all other large wikis. Ortega shows that the English-language community has extremely low retention among the most committed editors compared to other language Wikipedias (Figure 2). His results suggest that something particular about the English language community should explain at least part of its retention issues.

Generalization unsupported by direct comparison is risky business. This concern is akin to the challenge of scientists using "model organisms" to understand basic biological processes (Fields & Johnston, 2005). For example, much of what scientists know about genetics stems from research conducted with *Drosophila melanogaster*. That said, the question of when we should, and should not, generalize from *Drosophila* to other species is often not clear. Moreover, model organisms – like the most widely studied online communities – are selected precisely because of their extreme or idiosyncratic characteristics. Geneticists study *Drosophila* because it reproduces more quickly than other species. Online community researchers study Wikipedia precisely because it has attracted so many readers, contributions, and contributors and because it has generated articles of exceptional quality. We do not always understand the relationship between these unusual features and the basic social

and communicative processes we seek to study. In such situations, a combination of careful theorizing combined with healthy doses of analytic modesty and skepticism become a researcher's most important assets.

Benefit 2: Studying Community-level Variables

Although Wikipedia's enormous size and high quality articles motivate its starring role in online community research, understanding *why* Wikipedia became so big and high quality is remarkably difficult. The question "Why did Wikipedia succeed?" implies treating Wikipedia itself as the unit of analysis and measuring its success (however you define it) at the level of the community. As introductory research methods textbooks explain (e.g., King, Keohane, & Verba, 1994, pg. 139–142), answering this type question requires variation across multiple projects, some of which succeed and some of which fail. Ideally, to study why Wikipedia succeeded, you would compare it to failed Wikipedias or, at least, to projects that were trying to be something very similar to what Wikipedia became (Hill, 2013). This sort of comparison and inference represents an especially compelling opportunity created by studying populations of online communities.

Organization-level variables have a long tradition in organizational research where scholars have used data on variation in firm performance to understand which kinds of firms thrive. For example, a body of work has established and tested a theory of the "liability of newness" that suggests that newer firms are more likely to fail (Schoonhoven, 2015). Other research has examined how the structure of communication affects the performance of teams and work groups engaged in collaboration of various kinds (e.g., Crowston & Howison, 2006; Cummings & Cross, 2003). Because structure exists within groups, studying multiple groups provides variation that makes it possible to understand how structure can affect group-level outcomes. As a result, a large body of research in organizational communication has focused on group, team, or even firm-level variables.

Both the scholarly and popular literature about the power of online communities makes claims about project-level performance and outcomes. As a result, there have been many prior calls to pursue studies across projects and communities (e.g., Crowston, Wei, Howison, & Wiggins, 2008; Kraut & Resnick, 2012; Benkler, Shaw, & Hill, 2015). Schweik and English's (2012) book *Internet Success* provides a compelling example of the power of this approach. Deeply influenced by Ostrom's (1990) work on common pool resource management, Schweik and English frame their analysis of free/libre

open source software (FLOSS) projects in terms of questions about the successful provision of information commons. They create a stage model of when FLOSS projects will be abandoned and when they will develop into effective communities. They seek to explain these outcomes using dozens of project-level variables like leadership style and the copyright license of the project. They find that measures of clarity in leadership communication are among the best predictors of successful FLOSS commons and that other potential antecedents of FLOSS success cited in earlier literature, like license choice, do little to explain project outcomes.

Schweik and English's variables are nearly all at the level of the FLOSS project or community. Although previous case studies of leadership and governance reveal quite a bit about how FLOSS communities function, there is a sense in which they are limited to describing *how* these communities have operated. Because the sample of FLOSS projects analyzed by Schweik and English (2012) in *Internet Success* varies in terms of success (as the authors define it) and its theoretical antecedents, their analysis can begin to answer the question of *why* communities succeed. In this way, population-level research can engage with substantively and theoretically important questions in a manner that smaller scale analyses or comparisons struggle to do.

Perhaps the most important organization-level variables in networked communication research are related to technology. However, individual online communities tend to run on a single piece of software that consistently mediates the experience of every participant and the vast majority of research on the effect of technology on online communities uses data in which the technology itself is held constant. In the best cases, researchers are able to take advantage of a technological change or shock within a community to understand the impact of technology (e.g., Geiger & Halfaker, 2013). By looking across numerous communities, researchers can also build samples where technology varies as a way of testing these claims systematically (e.g., Shaw & Benkler, 2012). Such platform-level comparisons exemplify the sorts of analyses that become possible with the inclusion of community-level variables.

BENEFIT 3: STUDYING DIFFUSION BETWEEN COMMUNITIES

The ability to answer questions related to the way that information and practices flow between communities is another benefit of population-level research. For example, we know that a huge proportion of messages on social media systems like Twitter include links to sources outside Twitter and to other online communities (Bruns & Stieglitz, 2012, 3/4). As a result, studies that at-

tempt to characterize processes of information transmission without looking beyond the boundaries of an individual community or platform are necessarily incomplete. Although it can be difficult, research across communities can paint more accurate pictures of the diffusion.

Studies of diffusion have played a central role in communication scholarship since Rogers's (1962) seminal work on the diffusion of innovation. Subsequent studies by Valente (1995) and others have situated these diffusion processes within networks. Most of this research has sought to understand how information flows between individuals. For example, Bennett and Segerberg (2013) used data on message diffusion on Twitter to understand processes related to power, politics, and collective action. Of course, Twitter makes up only a part of Twitter users' media diets and many tweets point to further engagement in other media (Bruns & Stieglitz, 2012, 3/4). This speaks to the importance of research designs that can encompass a larger set of communities and thereby offer a more comprehensive explanation of diffusion processes.

The power of looking across communities is illustrated by a recent study by Graeff, Stempeck, and Zuckerman (2014) that traces the diffusion and development of conversations about the shooting of Trayvon Martin in February 2012. In addition to traditional media sources including television transcripts and newspapers, the analysis uses data from Media Cloud, an enormous dataset of many thousands of online media sources (Benkler, Roberts, Faris, Solow-Niederman, & Etling, 2013); search volume data from Google; messages on Twitter; signature data on the e-petitioning site Change.org; and a unique dataset from bit.ly that provided measures of how often people viewed social media material in a variety of communities including Twitter and Facebook. The paper traces how, after initially being reported only in local news, Martin's killing faded entirely from the media for weeks. A publicist hired by Martin's family and their lawyer then thrust the event back into the local media, bringing national media coverage that led to an online petition on the website Change.org. Then, a staff member at Change.org took notice of the petition and successfully engaged celebrities like Talib Kweli and Spike Lee to tweet messages about it, drawing increased engagement and attention. This increased attention led to further media coverage.

Graeff and colleagues' accounting provides a compelling answer to the question of how Trayvon Martin's killing became one of the most important news stories of 2012 precisely because they follow the story as it diffuses through different media sources and across different online communities, networks, and platforms. By looking across communities, the authors

Date	Change.org petition traffic	Traffic referred by social media
12 March	11,141	7,486
13 March	34,345	20,712
14 March	305,672	45,952
15 March	190,354	72,165
16 March	80,268	50,798

Table 1: Total traffic to the Trayvon Martin Change.org petition page and traffic referred to by social media, several days when the petition was active. (Table from Graeff, Stempeck, and Zuckerman (2014).)

discover that activity in social media communities like Twitter and Facebook was driven by activity in Tumblr and Blogger, and that the activity in these communities was driven by earlier activity in Twitter, which in turn was driven by activity around an e-petition. Table 1 reproduces some of Graeff et al.'s (2014) data summarizing traffic to the Change.org petition as well as traffic referred through social media sites. The multiple data sources combine to illustrate how the controversy developed across many types of media.

Although the dataset on the Trayvon Martin controversy used by Graeff and colleagues remains far from comprehensive, it offers the opportunity to go much further than an analysis of any single source could. The study also builds on previous work that had looked at network gatekeeping within Twitter (Meraz & Papacharissi, 2013) and is able to extend this theory beyond a single communication network. As is always the case with new methodologies, this approach also introduces new limitations. In particular, there are important challenges in terms of source selection, new opportunities for sample bias, and significant technical difficulties caused by the need to collect and integrate heterogeneous datasets.

Other studies consider other types of information transmission processes across community boundaries. A highly-cited study by Leskovec, Backstrom, and Kleinberg (2009) used an enormous dataset of 1.6 million different mainstream media sites and blogs to theorize a set of temporal patterns through which short distinctive phrases diffuse from news sites to blogs. A literature on code reuse in FLOSS has shown how code written for one project can be used in another (e.g., Mockus, 2007) and research in computer science has looked at how images are copied and moved between websites (Seneviratne et al., 2009). These lines of inquiry illustrate how studies of information flow across communities are increasingly possible and can lead to theoretically

salient insights. By considering a larger part of the communicative and collaborative landscape, these approaches enable the analysis of diffusion processes that may not exist at all within communities.

Diffusion processes can also involve the transmission of organizational practices. In organization science, sociologists have studied dynamics of organizational "isomorphism" whereby organizations adopt routines and structures used in other organizations (Meyer & Rowan, 1977; DiMaggio & Powell, 1983, e.g.). Although studies of this type of diffusion remain much less common in online communities, they are possible. In one recent study, Zhu, Kraut, and Kittur (2016) considers the diffusion of a model for coordinating through topic-focused task forces within Wikipedia called "Collaboration of the Week." Because they look at both a group-level process and group-level outcomes of productivity across many groups, Zhu and her colleagues' can evaluate the effectiveness of the practice at a more general level. Their work shows how a population-level research design also makes it possible to understand the ways that the task force model of collaboration spread and, together with the Graeff et al. study, illustrates how such research can alter and enhance our understanding of multiple kinds of diffusion processes.

Benefit 4: Studying Ecological Dynamics

Research on diffusion and information transmission underscores how online communities do not exist in a vacuum. By studying individual communities, we tend to study them in isolation from each other and from the external forces that shape or threaten them. This point leads us to another line of inquiry that population-level datasets of online communities can address: the way that online communities interact with their environments and the surrounding ecology of similar communities. We use the terms "environment" and "ecology" here in the same way that scholars in organizational studies have applied them to describe a collection of factors and pressures outside organizations' boundaries.[2] In studies of firms, an environment consists of competitors, suppliers, and customers. For example, one could describe the studies of diffusion discussed above as a type of research on environments. In this section, we highlight the opportunity to learn about online communities by studying how they experience ecological forces through their environments. Questions of this type remain almost entirely unexplored in online

[2]See Scott and Davis (2006) for both an excellent general overview of organizational theory and a detailed account of organization scientists' definitions of organizational environments.

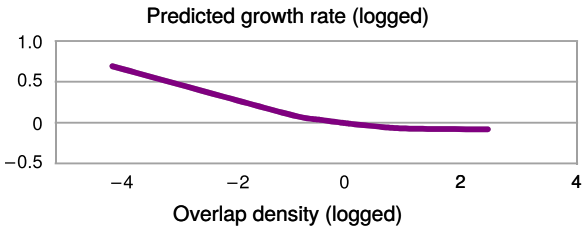


Figure 3: Relationship between member overlap density and membership growth as Usenet groups as predicted by an empirical model. Figure adapted from Wang, Butler, and Ren (2013), page 426.

community research, but present promising opportunities.

Ecological studies of organizations are well established within organizational research. Seminal work in organizational science borrowed the metaphor of ecologies from biology to assert a set of theories around resource competition among firms (Hannan & Freeman, 1993; Carroll & Hannan, 2000). A central insight drives this body of work: firms succeed or fail because of conditions in their environment like the the number of competing firms. It turns out that the much of the likelihood of a firm's failure can be predicted based on what other firms in the same "niche" are doing. To pick just one example, firms in either very empty or very competitive niches (i.e., brand new sectors versus markets that are crowded with other firms) fail more frequently that firms in moderately competitive markets (Carroll & Hannan, 2000). Although originally focused only on for-profit firms, this work has been extended to non-profit and social movement organizations (McPherson, 1983; Soule & King, 2008).

Several pieces of recent research on online communities adopt this type of ecological approach. One important example is Wang, Butler, and Ren (2013) which uses 64 months of longitudinal data from a stratified random sample of 240 Usenet discussion groups to understand how competition for participants' attention affects communities' growth. For each discussion group, the authors created a measure of membership overlap by identifying participants and then measuring the number of other Usenet discussion groups that each user participates in during every month. Wang and colleagues show that groups whose members participate in more groups grow more slowly. The relationships predicted by their model are shown in Figure 3. They use group-level variables like community size to unpack this finding and conclude that larger groups experience more difficulty in growth and are more vulnerable to the deleterious effects of competition.

Population-level datasets make it possible to imagine research like that the Wang et al. study. Additionally, taking an ecological perspective allows them to identify substantively important relationships. Their models of community growth rate improve substantially when they take membership overlap into account, explaining up to 14% of the variance overall and returning significantly better fit than alternative specifications without the overlap measure. Although their strongest controls are group-level factors like total membership, Wang et al.'s paper shows that we can make important predictions about how much an online community will grow by considering activities outside communities' boundaries.

An ecological approach to studying online communities remains rare, but other examples exist. Gu, Konana, Rajagopalan, and Chen (2007) considered competition between online investment communities. Two other recent pieces have extended the findings and approach of Wang et al. (2013): Zhu, Chen, et al. (2014) studying a group of 9,495 IBM discussion groups and Zhu, Kraut, and Kittur (2014) in a community of 5,673 wikis. Remarkably, the latter finds that membership overlap improves the survival rate for new wikis. Future work may confirm or diverge from these findings as well as the earlier findings of organizational research that focused exclusively on firms.

Benefit 5: Studying Multilevel Processes

A final benefit of population-level analyses of online communities is the ability to analyze complex, multilevel social processes. This benefit derives partly from the extraordinary granularity and breadth of online trace data, with which there is often no need to sacrifice micro-level detail in the analysis of meso- and macro-level dynamics. However, while multilevel analysis may have an affinity with exhaustive digital trace data, it need not proceed through data-intensive or computational approaches. Insightful qualitative inquiry can also account for multilevel social processes. We believe that research on populations of online communities present an exceptional opportunity in this regard.

Scholars of organizational communication and group processes have implemented multilevel approaches in a variety of domains. Prior work by Faraj and Johnson (2011) advocates the use of multilevel network analysis to examine complex social processes at the individual-level. Indeed, the rise of massive scale online datasets has coincided with the creation of novel multilevel methods, including the use of multilevel network analysis to study (among other things) multi-team systems (Contractor, 2013; Zaccaro, Marks,

& DeChurch, 2012). Howard's (2002) concept of "network ethnography" provides a mixed-methods example of this approach that utilizes exhaustive knowledge of macro- and meso-level network structures both to identify appropriate cases for in-depth analysis as well as to guide the process of interpreting and generalizing findings. To date, relatively little research capitalizes on the opportunities to apply multilevel analysis across many online communities. Several studies from organizational research use multilevel approaches to study individual and group-level relational processes across multiple online communities (Faraj & Johnson, 2011), but such work remains exceptional. We think there should be more.

A study of social capital and relationships in World of Warcraft (WoW) guilds by Williams et al. (2006) provides an excellent example of multilevel work that combines the advantages of large-scale trace data with in-depth qualitative evidence. WoW guilds are arbitrarily sized clans of players which vary in size, scope, formality, and strategic focus. A player might, for instance, want to join up with a "raid-oriented family friendly guild" of marauding orcs who meet on weekends. The authors use an exhaustive observational dataset collected with automated programs connecting to different types of WoW game servers over several months to build a representative sample of individual players and guilds.[3] The authors created a stratified sample of these guilds and a representative sample of their members. They then conducted 48 interviews with individuals from the sample over the game's chat platform to understand how members participate in guilds.

Williams and colleagues' elaborate their account of both group-level and individual-level patterns of activity using quotations contextualized through their knowledge of player and guild characteristics. For example, they show how players articulate group-level identities and find very legitimate forms of social support through participation in guilds. One interviewee explained the experience of playing in a large guild with everything but a citation to Putnam's (2000) influential argument about the putative decline of civic associations in America:

> It's kind of like a bowling team or a softball league...it's just as social event in here, probably more often then [*sic*] bowling since I talk to these people several nights a week (Williams et al., 2006, pg. 347).

The quote illustrates the existence of large-scale solidarity and associationism

[3]Note that additional details of the data collection are published in a companion article by Ducheneaut, Yee, Nickell, and Moore (2006).

in the supposedly alienated world of video games. The authors can generalize from this kind of evidence because they used guild-level data to ensure that their findings are representative of experiences in guilds of varying sizes and characteristics. The findings integrate the richness and depth of insightful interviewing with the analytic leverage of a research design that can generalize across multiple communities.

An important takeaway is that individuals' experiences online often occur within the context of sub-communities like guilds. As a result, understanding whether and how WoW players build sustained social relationships and social capital entails accounting for a complex interplay of both factors at the game, guild, and individual levels. In this way, Williams and colleagues' study illustrates the value of multilevel modeling strategies as a tool for incorporating measures nested across levels of analysis and social interaction in online communities that are, themselves, part of larger communities. Although their study is qualitative in nature, Williams has collaborated with others on quantitative studies using the same dataset in similar ways (e.g., Shen, Monge, & Williams, 2014).

Multilevel analyses can also account for ecological dynamics in communities' environments and look across non-nested communities. For instance, Faraj and Johnson (2011) test claims about the prevalence of patterns of interpersonal exchange using exhaustive micro-level data collected from a random sample of discussion groups. They draw conclusions at the individual and group levels while accounting for community-level variation as well. Studying interpersonal communication processes in multiple online communities by simultaneously analyzing multiple levels of social organization and behavior offers a particularly promising path for future work.

LIMITATIONS

Alongside these benefits, population-level research on online communities introduces a number of limitations. Some of these limitations are inherent to research on populations and others derive from particular characteristics of populations of online communities. First, multiple populations of communities can interact in ways that suggest an even higher level of analysis that may bring additional insights. Second, identifying distinct communities poses difficulties in many situations. Third, research in this area exists in the interstices of multiple fields, leading to challenges related to integrating disparate literatures and concepts. Finally, it is important to acknowledge that studying populations of communities can require more effort and can be more skill

and resource-intensive than studying a community in isolation.

While studies on populations of communities may provide insights and generalizability that exceed smaller scale comparisons or case studies, it is important to note that a population of communities (as we described it above) may be distinct from still larger sets of populations, known as a metapopulation in ecology research (Hanski, 1999). Along these lines, organizational sociologists have suggested that studying organizations within industries ignores important high-level effects across industries and that it is also important to study populations of industries – literally, populations of populations (e.g. Ruef, 2000). In the context of online communities, examples of populations and corresponding metapopulations might look like all of the communities on a particular platform or hosting service (e.g., Wikia, GitHub, or EdX) and the full set of potentially comparable communities (e.g., all wikis, FLOSS projects, or MOOCs). There may also be other dimensions along which to define metapopulations, such as the sets of communities encompassed by particular language groups, legal regimes, or infrastructural features.

The distinction between populations and metapopulations implies new types of limitations of generalizability analogous to those confronted by single case studies. While a study may include all of the software development communities hosted by GitHub or SourceForge, or even both, there may be underlying biases that make these two groups systematically different from other otherwise similar communities. There may also be ecological forces that reflect competition between the different populations. As usual, unknown bias has no empirical remedy other than an effort to encompass more cases – in this case more populations. Likewise, the identification of ecological forces at a metapopulation level remain inaccessible to studies within a single population. These possibilities suggest future opportunities for research.

A second issue relates to the ways in which the nature and character of online communities resist easy classification or definition. A single community, such as the English language Wikipedia, may contain hundreds or even thousands of distinct sub-communities. In Wikipedia, these include WikiProjects – groups of individuals who work together on topics of shared interest (Morgan, Gilbert, McDonald, & Zachry, 2013). Like guilds in WoW, WikiProjects are online communities nested in English Wikipedia (for example), which is nested within the broader Wikipedia project, which is nested within the even broader Wikimedia movement. Such a fractal structure is not unique to Wikipedia and may represent a broader characteristic of large-scale, open organizations on the Internet.

Given these examples, language and theories developed to analyze more clearly bounded formal organizations such as firms, non-profits, and even more open volunteer projects or social movements may prove inadequate. Even the term "community" is contested and is used in a wide variety of ways by researchers studying online interaction (Bruckman, 2006). This disagreement about what constitutes a "community" makes the definition of a population (or metapopulation) of online communities analytically slippery as well. Research on organizations and collective action has, historically, elided this concern by adopting a very broad set of inclusion criteria (e.g., Marwell & Oliver, 1993). Ultimately though, if every sort of remotely collective endeavor can be also called a "community," the term ceases to hold meaning.

The murky empirical and conceptual boundaries around online communities make defining populations difficult. We advocate a pragmatic, but ultimately somewhat unsatisfying, approach to this issue: researchers should define populations by a combination of objectively observable criteria along with folk and scientific wisdom. Sometimes, as in studies examining all of the communities hosted through a common software platform or company, this may be straightforward and the results may correspond to both scientific intuition as well as the understanding of participants within these populations. Under other circumstances, it may be necessary to choose between different possible analytic boundaries. For instance, in analyses of emergent collaboration networks such as those engaging in breaking news coverage on microblogging platforms, Wikipedia, or Facebook, groups of participants may act in ways that are consistent with researchers' understandings of what it means to be a community even though they do not perceive themselves as such. In these situations, it becomes critical to acknowledge such divergent perspectives and proceed with caution. It may not make sense to address certain theoretically interesting questions with data drawn from unsuitable cases. Just because we can model something as an community does not mean we should. Relatedly, just because a platform describes something as a group (a Facebook group, a guild within a game, a WikiProject) doesn't mean it is the correct unit of analysis for a particular question

A third challenge is that existing research conducting population-level analyses of online communities spans disciplinary boundaries. Our own work has been influenced by organizational studies, social psychology, political sociology, interpersonal communication, and human computer interaction. Such diverse intellectual orientations may create opportunities for academic brokerage, but these opportunities come at a cost. As in finance (Zuckerman, 1999), academic endeavors that fail to fit into existing categorical schema may

pay an illegitimacy premium. There are also opportunity costs that come with learning multiple research fields. The institutionalization of novel schema takes time and proceeds unevenly.

We believe the best response to this problem is to draw strategically from the most compelling domains of research. For example, the papers we have highlighted in this essay have drawn on the literature on organizational behavior, work groups, computer science, and teams. We have also turned to the sociological literature focused on comparative analysis of movements. Others have emphasized links to studies of emergent networks and collective intelligence. All of these approaches have borne fruitful results. Of course, this type of pluralistic approach introduces challenges as well. While reading broadly, it is important to ground one's work in established traditions of scholarship. A broad set of influences is not a license to cherry-pick theories to fit data. An open challenge lies in the synthesis and integration of findings across disparate approaches. For the time being, many findings remain mutually contradictory or unintelligible as they speak of similar phenomena in different ways.

A final limitation is purely practical but still important to note. Simply put, studying multiple communities can require more time, more effort, more skills, or more resources than a study of a single community. For one, population-level datasets can be large and unwieldy and building and analyzing them may be outside the abilities of many communication researchers. The dataset used by Graeff et al. (2014) was created by professional engineering staff. Many population-level analyses, like Leskovec et al. (2009), were carried out by computer scientists. When using these large and varied datasets, there are often trade-offs between scope and scale. In the short term, collaborations with computer scientists and engineers is a popular strategy. In the longer term, population-level studies will become more widespread as the technology necessary to complete them becomes more established, easier-to-use, and more reliable. For example, the deployment of standardized APIs and large data releases from large community-hosting platforms has already made some types of population-level studies much more accessible. We expect that this will only increase with time.

Discussion

In summary, we have argued that online communities research can benefit enormously from studying populations of communities. We have tried to show how population-level research designs that span communities can of-

fer increased generalizability and can open the door to new kinds of questions including those that focus on theoretically important community-level variables, processes of diffusion across communities, and the way that communities interact with their environments through ecological competition. Finally, we've argued that digital trace datasets and a variety of analyses that cross multiple levels can allow research to realize many of these benefits without tossing aside the benefits of intra-community studies. By narrating this argument with a series of in-depth examples of exemplary work from communication and beyond, we have also pointed to concrete examples of how this can be done. Although these studies still reflect a small proportion of research on online communities, similar approaches are increasingly common.

Additional benefits from population-level research about online communities may emerge as researchers experiment and explore this type of work. One benefit we find particularly exciting concerns the degree to which analysis across many online communities may provide stronger empirical support for policy-making and design decisions. These decisions can benefit from insights into the effects of specific interventions. This sort of evidentiary basis holds particular relevance for computer-mediated communication systems where code may shape *de facto* institutional arrangements, norms, and behavioral patterns (Lessig, 1999). For example, in our own work we used a population of wikis to explore the effects of a requirement to create an account on subsequent community-level activity. Estimates of an average effects of a widely debated policy decision like this, across many projects, can give designers and community leaders greater confidence in the generalizability of a finding in a way that can inform subsequent policy decisions. Exploration of heterogeneous effects can provide an understanding of how a design might succeed or fail.

In ways that we discussed earlier, many studies of online communities and networked communication seek to understand the mediating effects of specific technologies. Since the design of technical platforms and tools, in most cases, operate at the level of communities, a shift to empirical analysis and theorizing at the level of communities can make this work more directly usable. There are already significant overlaps in the scholarly communities studying networked communication and the designers creating social computing systems. Crafting research capable of speaking across these boundaries will enhance the impact of communication research and will provide communication researchers with greater opportunities to test, extend, and refine their theories in conversation with designers and computer scientists.

Of course, while we see many opportunities in population-level research, the approach is far from a silver bullet. The limitations we have sketched out are real, significant, and a subset of the challenges that confront population-level analysis. Just as we hope that continued growth of population-level studies will help establish benefits of the approach, we also hope it will paint a better picture of the limitations of population-based approaches and support the development of better ones. We believe that an increased attention to populations marks one step toward better online community research and toward a deeper understanding of networked communication.

Acknowledgements

We thank the Community Data Science Collective for helping us develop these ideas. We thank the editors, Brooke Foucault Welles and Sandra Gonzalez Bailon, and reviewers of this chapter, for their helpful feedback. Portions of the project were completed at the Helen R. Whitely Center at the University of Washington's Friday Harbor Laboratories. Financial support for this work came from the National Science Foundation (grants IIS-1617129 and IIS-1617468).

References

Benkler, Y., Roberts, H., Faris, R., Solow-Niederman, A., & Etling, B. (2013). *Social Mobilization and the Networked Public Sphere: Mapping the SOPA-PIPA Debate* (SSRN Scholarly Paper No. ID 2295953). Social Science Research Network. Rochester, NY. Retrieved September 22, 2013, from http://papers.ssrn.com/abstract=2295953

Benkler, Y., Shaw, A., & Hill, B. M. (2015). Peer Production: A Form of Collective Intelligence. In M. Bernstein & T. Malone (Eds.), *The Handbook of Collective Intelligence*. Cambridge, MA: MIT Press.

Bennett, W. L. & Segerberg, A. (2013). *The logic of connective action: Digital media and the personalization of contentious politics*. New York, New York: Cambridge University Press.

Bruckman, A. (2006). A new perspective on "community" and its implications for computer-mediated communication systems. In *CHI '06 extended abstracts on Human factors in computing systems* (pp. 616–621). Montréal, Québec, Canada: ACM. doi:10.1145/1125451.1125579

Bruns, A. & Stieglitz, S. (2012, July). Quantitative Approaches to Comparing Communication Patterns on Twitter. *Journal of Technology in Human Services*, *30*, 160–185. 00083. doi:10.1080/15228835.2012.744249

Carroll, G. R. & Hannan, M. T. (2000). *The demography of corporations and industries*. Princeton, New Jersey: Princeton University Press.

Contractor, N. (2013, March 28). Some assembly required: Leveraging Web science to understand and enable team assembly. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *371*(1987), 20120385. doi:10.1098/rsta.2012.0385. pmid: 23419854

Crowston, K. & Howison, J. (2006, December). Hierarchy and centralization in free and open source software team communications. *Knowledge, Technology & Policy*, *18*(4), 65–85. doi:10.1007/s12130-006-1004-8

Crowston, K., Wei, K., Howison, J., & Wiggins, A. (2008, March). Free/Libre Open-source Software Development: What We Know and What We Do Not Know. *ACM Comput. Surv.* *44*(2), 7:1–7:35. 00212. doi:10.1145/2089125.2089127

Cummings, J. N. & Cross, R. (2003). Structural properties of work groups and their consequences for performance. *Social Networks*, *25*(3), 197–210. Retrieved June 8, 2015, from http://www.sciencedirect.com/science/article/pii/S0378873302000497

DiMaggio, P. J. & Powell, W. W. (1983, April). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160. JSTOR: 2095101

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006, January 10). Building an MMO With Mass Appeal A Look at Gameplay in World of Warcraft. *Games and Culture*, *1*(4), 281–317. doi:10.1177/1555412006292613

Faraj, S. & Johnson, S. L. (2011). Network Exchange Patterns in Online Communities. *Organization Science*, *22*(6), 1464–1480. doi:10.1287/orsc.1100.0600

Fields, S. & Johnston, M. (2005, March 25). Whither Model Organism Research? *Science*, *307*(5717), 1885–1886. doi:10.1126/science.1108872. pmid: 15790833

Geiger, R. S. & Halfaker, A. (2013). When the Levee Breaks: Without Bots, What Happens to Wikipedia's Quality Control Processes? In *Proceedings of the 9th International Symposium on Open Collaboration* (6:1–6:6). WikiSym '13. 00010. New York, NY, USA: ACM. doi:10.1145/2491055.2491061

Graeff, E., Stempeck, M., & Zuckerman, E. (2014, January). The battle for 'Trayvon Martin': Mapping a media controversy online and off-line. *First Monday*, *19*(2). Retrieved January 27, 2016, from http://firstmonday.org/ojs/index.php/fm/article/view/4947

Gu, B., Konana, P., Rajagopalan, B., & Chen, H.-W. M. (2007). Competition among virtual communities and user valuation: The case of investing-related communities. *Information Systems Research*, *18*(1), 68–85. JSTOR: 23211832

Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013, May 1). The Rise and Decline of an Open Collaboration System How Wikipedia's Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist*, *57*(5), 664–688. doi:10.1177/0002764212469365

Halfaker, A., Kittur, A., & Riedl, J. (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* (pp. 163–172). WikiSym '11. New York, NY, USA: ACM. doi:10.1145/2038558.2038585

Hannan, M. T. & Freeman, J. (1993). *Organizational Ecology*. Cambridge, MA: Harvard University Press.

Hanski, I. (1999, March 18). *Metapopulation Ecology*. 03605. Oxford, UK: Oxford University Press.

Hargittai, E. (2015, May 1). Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 63–76. doi:10.1177/0002716215570866

Hill, B. M. (2013). Almost Wikipedia: What eight early online collaborative encyclopedia projects reveal about the mechanisms of collective action.

In *Essays on volunteer mobilization in peer production*. PhD Dissertation. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Howard, P. N. (2002, January 12). Network Ethnography and the Hypermedia Organization: New Media, New Organizations, New Methods. *New Media & Society*, *4*(4), 550–574. doi:10.1177/146144402321466813

King, G., Keohane, R. O., & Verba, S. (1994, May 2). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Kollock, P. & Smith, M. (1999, February 10). *Communities in Cyberspace*. London: Routledge.

Kraut, R. E. & Resnick, P. (2012, March 23). *Building Successful Online Communities: Evidence-Based Social Design* (). Cambridge, MA: The MIT Press.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., … Van Alstyne, M. (2009, February 6). Computational Social Science. *Science*, *323*(5915), 721–723. doi:10.1126/science.1167742

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the Dynamics of the News Cycle. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 497–506). KDD '09. New York, NY, USA: ACM. doi:10.1145/1557019.1557077

Lessig, L. (1999). *Code and Other Laws of Cyberspace*. New York, NY: Basic Books.

Marwell, G. & Oliver, P. (1993). *The critical mass in collective action: A micro-social theory*. Cambridge, Massachusetts: Cambridge University Press.

McPherson, M. (1983). An ecology of affiliation. *American Sociological Review*, *48*(4), 519–532. doi:10.2307/2117719. JSTOR: 2117719

Meraz, S. & Papacharissi, Z. (2013, January 27). Networked Gatekeeping and Networked Framing on #Egypt. *The International Journal of Press/Politics*, 138–166. doi:10.1177/1940161212474472

Meyer, J. W. & Rowan, B. (1977, September). Institutionalized organizations: Formal structure as myth and ceremony. *The American Journal of Sociology*, *83*(2), 340–363. 20516. JSTOR: 2778293

Mockus, A. (2007, May). Large-Scale Code Reuse in Open Source Software. In *First International Workshop on Emerging Trends in FLOSS Research and Development, 2007. FLOSS '07* (pp. 7–7). First international workshop on emerging trends in floss research and development, 2007. floss '07. doi:10.1109/FLOSS.2007.10

Morgan, J. T., Gilbert, M., McDonald, D. W., & Zachry, M. (2013). Project Talk: Coordination Work and Group Membership in WikiProjects. In *Proceedings of the 9th International Symposium on Open Collaboration*

(3:1–3:10). WikiSym '13. New York, NY, USA: ACM. doi:10.1145/2491055.2491058

Ortega, F. (2009). *Wikipedia: A Quantitative Analysis* (Ph.D. Dissertation, Universidad Rey Juan Carlos, Madrid, Spain). Retrieved June 22, 2009, from http://libresoft.es/Members/jfelipe/phd-thesis

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. New York, NY: Cambridge University Press.

Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York, NY: Simon and Schuster.

Restivo, M. & van de Rijt, A. (2012). Experimental Study of Informal Rewards in Peer Production. *PLoS ONE*, *7*(3), e34358. doi:10.1371/journal.pone.0034358

Restivo, M. & van de Rijt, A. (2014). No praise without effort: Experimental evidence on how rewards affect Wikipedia's contributor community. *Information, Communication & Society*, *17*(4), 1–12. doi:10.1080/1369118X.2014.888459

Rogers, E. M. (1962). *Diffusion of Innovations*. New York, NY: The Free Press of Glencoe.

Ruef, M. (2000, November 1). The emergence of organizational forms: A community ecology approach. *American Journal of Sociology*, *106*(3), 658–714. doi:10.1086/318963

Schoonhoven, C. B. (2015). Liability of Newness. In *Wiley Encyclopedia of Management*. John Wiley & Sons, Ltd. Retrieved January 30, 2016, from http://onlinelibrary.wiley.com/doi/10.1002/9781118785317.weom030067/abstract

Schweik, C. M. & English, R. C. (2012). *Internet Success: A Study of Open-source Software Commons*. Cambridge, MA: MIT Press.

Scott, W. R. & Davis, G. F. (2006). *Organizations and organizing: Rational, natural and open systems perspectives*. Upper Saddle River, NJ: Pearson Prentice Hall.

Seneviratne, O., Kagal, L., Weitzner, D., Abelson, H., Berners-Lee, T., & Shadbolt, N. (2009). Detecting creative commons license violations on images on the world wide web. In *Proceedings of the 18th Internetional World Wide Web Conference*. Www2009. Madrid, Spain. Retrieved January 31, 2016, from http://dig.csail.mit.edu/2008/Papers/WWW2009/paper.pdf

Shaw, A. & Benkler, Y. (2012). A Tale of Two Blogospheres: Discursive Practices on the Left and Right. *American Behavioral Scientist*, *56*(4), 459–487. doi:10.1177/0002764211433793

Shen, C., Monge, P., & Williams, D. (2014, October 1). The evolution of social ties online: A longitudinal study in a massively multiplayer online

game. *Journal of the Association for Information Science and Technology*, *65*(10), 2127–2137. doi:10.1002/asi.23129

Soule, S. A. & King, B. G. (2008, May). Competition and resource partitioning in three social movement industries. *The American Journal of Sociology*, *113*(6), 1568–1610. doi:10.1086/587152. JSTOR: 25145846

Valente, T. W. (1995, January 10). *Network Models of the Diffusion of Innovations*. Cresskill, N.J: Hampton Press.

van de Rijt, A., Kang, S. M., Restivo, M., & Patil, A. (2014, May 13). Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences*, *111*(19), 6934–6939. doi:10.1073/pnas.1316836111. pmid: 24778230

Wang, X., Butler, B. S., & Ren, Y. (2013). The impact of membership overlap on growth: An ecological competition view of online groups. *Organization Science*, *24*(2), 414–431. doi:10.1287/orsc.1120.0756

Weber, M. S., Fulk, J., & Monge, P. (2016, February 11). The Emergence and Evolution of Social Networking Sites as an Organizational Form. *Management Communication Quarterly*, 0893318916629547. doi:10.1177/0893318916629547

Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N., & Nickell, E. (2006, January 10). From Tree House to Barracks The Social Life of Guilds in World of Warcraft. *Games and Culture*, *1*(4), 338–361. doi:10.1177/1555412006292616

Zaccaro, S. J., Marks, M. A., & DeChurch, L. (Eds.). (2012, May 22). *Multiteam Systems: An Organization Form for Dynamic and Complex Environments*. Routledge.

Zhu, H., Chen, J., Matthews, T., Pal, A., Badenes, H., & Kraut, R. E. (2014). Selecting an effective niche: An Ecological view of the success of online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 301–310). CHI '14. New York, NY, USA: ACM. doi:10.1145/2556288.2557348

Zhu, H., Kraut, R. E., & Kittur, A. (2014). The impact of membership overlap on the survival of online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 281–290). CHI '14. New York, NY, USA: ACM. doi:10.1145/2556288.2557213

Zhu, H., Kraut, R. E., & Kittur, A. (2016). A Contingency View of Transferring and Adapting Best Practices within Online Communities. In *Proceedings of the Conference on Computer Supported Cooperative Work and social Computing*. Cscw 2016. San Francisco, CA: ACM Press. Retrieved January 29, 2016, from http://haiyizhu.com/wp-content/uploads/2015/09/BestPracticeTransfer.pdf

Zuckerman, E. W. (1999, March). The categorical imperative: Securities analysts and the illegitimacy discount. *The American Journal of Sociology*, *104*(5), 1398–1438. JSTOR: 2990940

Journal of Communication ISSN 0021-9916

ORIGINAL ARTICLE

The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing

Aaron Shaw[1] & Eszter Hargittai[2]

1 Aaron Shaw, Department of Communication Studies, Northwestern University, Evanston, IL 60208, USA
2 Eszter Hargittai, Institute of Communication and Media Research, University of Zurich, Zurich 8050, Switzerland

Digital inequalities undermine the democratizing potential of the Internet. While many people engage in public discourse through participatory media, knowledge gaps limit engagement in the networked public sphere. Participatory web platforms have unique potential to facilitate a more equitable production of knowledge. This paper conceptualizes a pipeline of online participation and models the awareness and behaviors necessary to become a contributor to the networked public sphere. We test the theory with the case of Wikipedia editing, relying on survey data from a diverse, national sample of U.S. adults. Our findings underscore the multidimensionality of digital inequalities and suggest new pathways toward closing knowledge gaps by highlighting the importance of education and Internet skills for online stratification processes.

Keywords: Digital Inequality, Knowledge Gap, Online Participation, Internet Skills, Survey Research, Wikipedia.

doi:10.1093/joc/jqx003

Influential early accounts of the participatory Internet and networked media systems proclaimed their potential to democratize knowledge sharing (e.g., Benkler, 2006; Jenkins, 2006) and close long-standing "knowledge gaps" (Tichenor, Donohue, & Olien, 1970). These claims appear partially prescient, as online activities have become ever more ubiquitous spaces of social interaction that shape critical life-course trajectories and outcomes (DiMaggio & Bonikowski, 2008; Hampton, Lee, & Her, 2011; Hargittai & Shaw, 2013). However, scholars have also suggested that the Internet may expand knowledge gaps (Bonfadelli, 2002). Pervasive and durable digital inequalities mean that online environments often exacerbate existing patterns of social exclusion (Robinson et al., 2015). Stratification in Internet experiences, awareness, and skills compound and reinforce each other, contributing to a "Matthew effect" despite the increasing availability of new online activities and interfaces (van Deursen, Helsper, Eynon, & van Dijk, 2017; Pearce & Rice, 2013;

Corresponding author: Aaron Shaw; e-mail: aaronshaw@northwestern.edu

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Zillien & Hargittai, 2009). As a result, participatory digital spaces present critical domains for understanding whether communication in the digitally networked information environment perpetuates knowledge gaps.

Prior research has provided neither theories nor empirical analyses of specific processes by which participation divides in public communication of knowledge emerge, whether online or off. From writing newspaper editorials (Perrin & Vaisey, 2008) to posting photos and videos online (Schradie, 2011), scholarship has described how some people share their ideas publicly while others do not, but has not addressed why the divergent patterns exist. Widespread participation in public discourse underpins the effective functioning of the "networked public sphere" and democratic civil society (Benkler, 2006; Friedland, Hove, & Rojas, 2006). However, the exceptional scale and participatory affordances of digitally networked interactive communication platforms have not eliminated participation gaps in engagement with knowledge in society (Blank, 2013; Cacciatore, Scheufele, & Corley, 2014; Hargittai & Walejko, 2008; Robinson et al., 2015; Schradie, 2011). Understanding the mechanisms of inequalities in online participation around knowledge production is an important task for communication scholarship.

Explaining these phenomena requires new approaches to digital inequality. Many earlier studies either focus on inequalities of Internet access or the relationships between demographic backgrounds and what people do online in the forms of information seeking and interpersonal communication (see Robinson et al., 2015 for a review). While some work has explored inequalities in active engagement (see Brake [2014] and Hargittai & Jennrich [2016] for reviews of this research), the mechanisms whereby Internet users go from having access to the web to becoming more active participants in public online spaces have remained unexplored. Our paper contributes a more rigorous investigation of the factors and processes that shape online knowledge-production activities. Expanding prior approaches to digital inequality and knowledge gaps, we theorize and model a pipeline of online participation. This approach incorporates distinct steps heretofore unaddressed in the literature that are prior and necessary to becoming a contributor to the participatory web. We use a unique survey dataset collected in summer 2016 from a diverse, national sample of 1,512 U.S. adults to test our model empirically on the case of Wikipedia, arguably the most widely accessed and impactful example of participatory knowledge production online (comScore, 2016). While our empirical focus is on Wikipedia and the U.S., the theoretical approach can extend to other online participatory activities and contexts, such as the comment threads of major news media sites like the BBC or the Washington Post, social media platforms like Facebook or Twitter, as well as sites dedicated to particular engagement activities like online health communities or political volunteering.

Overall, the results illustrate that knowledge gaps help explain inequalities in knowledge production activities online. We find that the pipeline metaphor characterizes the data accurately and that different factors explain engagement at stages of the pipeline in ways that previous research had not considered. Participation on

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Wikipedia reflects inequalities along background attributes, including (but not limited to) education, Internet experiences and skills, and gender. The paper illuminates digital stratification processes and the knowledge gap in the digital age both by elaborating the metaphor of a pipeline of participation and by operationalizing it in relation to an important site of knowledge production in the networked public sphere. Our diverse, national sample supports more generalizable inferences about participation gaps. We conclude that future research and interventions to overcome digital participation gaps should not focus exclusively on gender or class differences in content creation, but expand their scope to multiple aspects of digital inequality across pipelines of participation. In particular, our findings support broader efforts to address knowledge- and skill-based barriers to entry among potential contributors to the networked public sphere, from online health communities to political discussions, from cultural conversations to job-search forums.

Knowledge gaps, digital inequalities, and online participation

According to the "knowledge gap hypothesis" (Tichenor et al., 1970), differential media access and skills compound differential levels of education, expanding inequalities of knowledge and engagement with important matters of public concern. Scholars have elaborated, tested, and debated this claim around print and broadcast media for decades, drawing mixed conclusions (e.g., Ettema & Kline, 1977; Hwang & Jeong, 2009).

Contemporary discussions of inequalities and stratification processes in digital media environments build on the knowledge gap debates. Inequalities in Internet use and online media engagement demonstrate a knowledge gap dynamic at work (e.g., Bonfadelli, 2002; Cacciatore et al., 2014; Pearce & Rice, 2013). Key determinants of digital inequalities also include, autonomy of use, experiences, and skills (Chen & Wellman, 2004; van Deursen et al., 2017; Dutton & Blank, 2014; Haight, Quan-Haase, & Corbett, 2014; Hargittai, 2002; Livingstone & Helsper, 2007; Martínez-Cantos, 2017; Robinson et al., 2015; Warschauer, 2003).

Knowledge- and content-production activities represent an arena where digital inequalities play an especially large role in shaping individual activities and attainment (Brake, 2014). Emancipatory accounts of networked communication emphasize the potential of the participatory Internet to empower individuals to become active producers of knowledge and cultural resources (rather than simply consumers; Benkler, 2006; Jenkins, 2006). However, digital inequality research has challenged such one-sided enthusiasm with evidence that participation in online content production remains highly differentiated across multiple sites, activities, and places (e.g., Blank, 2013; Correa, 2010; Hargittai & Walejko, 2008; Schradie, 2011, 2015).

Internet-use skills stand out among the most consistent predictors of differentiated online content production activities. For example, women and low–socio-economic status individuals tend to produce less content online than men and high-socioeconomic status

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

individuals (Schradie, 2011, 2015), but differences in Internet skills help explain the gendered variations in multivariate models (Correa, 2010; Hargittai & Shaw, 2015; Hargittai & Walejko, 2008; Martínez-Cantos, 2017). We seek to expand the theoretical and empirical literature on digital inequality and participation gaps by better understanding how factors such as skills may explain differentiated online content production.

Theorizing online content production gaps as pipelines

Stratification processes in the online media systems where knowledge and content production happen remain less understood than general patterns of Internet use. Prior theoretical work posits that even once people bridge the first-order digital divide and connect to the Internet, inequalities in their online behaviors will still exist (DiMaggio, Hargittai, Celeste, & Schafer, 2004; Selwyn, 2004) and that these inequalities reinforce each other in feedback loops (van Deursen et al., 2017). Differentiated Internet uses appear in a wide array of online activities, ranging from the types of content people seek and consume to material they produce and share (Blank, 2013; Hargittai & Walejko, 2008; Hargittai & Shaw, 2015; Schradie, 2011, 2015). Overcoming participation divides and knowledge gaps requires understanding mechanisms of stratification within these specific online activities more precisely.

To fill this gap in the literature, we adopt a pipeline metaphor to describe the multiple stages of Internet uses and skills necessary to engage in more active forms of knowledge production online. The pipeline metaphor builds on research into gender-based inequalities in science, technology, engineering, and math fields (e.g., Blank, 2013; Correa, 2010; van Deursen et al., 2017; Hargittai & Walejko, 2008; National Academy of Sciences, National Academy of Engineering, & Institute of Medicine, 2007; Schradie, 2011, 2015). We use the pipeline to represent a sequence of stages through which an individual must pass in order to become more actively engaged in knowledge consumption and production activities through digital media. First, a user must have heard of a site to be able to contribute to it. Second, a user must have visited the site to participate on it. Third, a user must understand that it is possible to make contributions to the site in order to add content. Only once these conditions are met can a user participate on a site. This model describes a general sequence applicable to diverse contexts of online engagement, such as discussion threads on major news media sites like *The Guardian* or *Slate*, social media platforms like Facebook or Instagram, as well as sites dedicated to particular engagement activities like support groups or learning. Figure 1 visualizes the pipeline in the context of online participation.

Research on online participation gaps tends to skip the first three stages and only collects data on who contributes online rather than establishing at which points people may have dropped out of the pool of potential contributors in the first place. That is, if a user does not know of a participatory site, she will not visit that

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

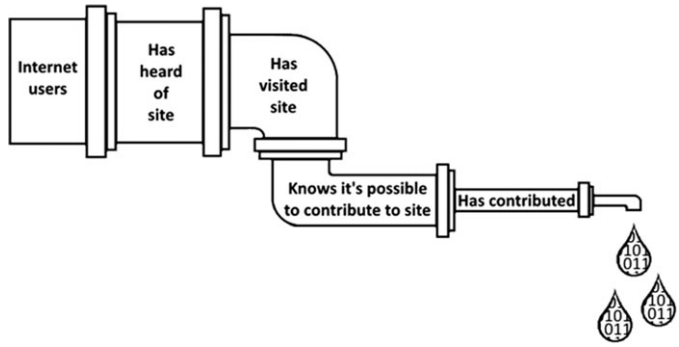


Figure 1 Schematic drawing of a pipeline of online participation.

site and will not interact on that site. Even if a user has heard of a participatory site, if she does not visit it, she will not participate on it. If a user visits a site, but does not know what types of behavior, i.e., active contributions, are possible on the site, then she will not engage in such behaviors. Conceptualizing online participation as a pipeline extends digital inequality literature to consider precursors to content creation and participation.

Prior work that examines broad categories of online participatory behavior cannot determine at which point in the process of using the Internet people drop out of the potential pool of online contributors. Approaching online participation as a pipeline examines these stages in the process in their own right. We anticipate that distinct factors explain variations and inequalities at the different stages of this pipeline, and explain our expectations below after some context on participation gaps in Wikipedia. Metaphorically speaking, we anticipate evidence of "leaks" at different points along the pipeline of participation, explained by underlying knowledge gaps. Our data set and methods allow us to to test these claims in a way that prior data could not.

Participation gaps in Wikipedia

The setting for our study is Wikipedia, the most widely-visited participatory web site with unique potential to overcome knowledge gaps. Inequalities in contributions to Wikipedia are a significant issue, given that the site is one of the most utilized knowledge resources on the Internet (Reagle & Rhue, 2011; Wagner, Graells-Garrido, Garcia, & Menczer, 2016). Accordingly, who edits the site may contribute to content biases in the coverage of the encyclopedia and exacerbate knowledge gaps. Wikipedia thus provides an ideal context in which to understand how stratification processes in online participation relate to knowledge gaps within the networked public sphere (Benkler, 2006; Friedland et al., 2006).

Gender gaps attract the bulk of scholarly and press attention about unequal contributions to Wikipedia. The population of U.S. adult contributors to the

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

English-language edition of Wikipedia appears to be at least 75% male (Hill & Shaw, 2013). Explanations of the Wikipedia gender gap tend to focus on existing contributors to the site and the culture of the community. While differences in Internet skills have been shown to matter for how people participate online, little of the work that looks at who contributes to Wikipedia has considered skills (for an exception, see Hargittai & Shaw, 2015). Recent studies investigate two main aspects of digital inequality in Wikipedia. One group of projects looks at inequalities in content coverage, with an emphasis on gendered coverage gaps (e.g., Adams & Brückner, 2015; Graells-Garrido, Lalmas, & Menczer, 2015; Johnson et al., 2016; Wagner et al., 2016). Generally speaking, this line of work emphasizes unequal participation rates across sub-populations (e.g., male vs. female editors; rural vs. urban editors). The authors point to structures of privilege and bias in knowledge production (e.g., men are considered more notable in news media and scholarly sources overall) as the forces shaping unequal coverage patterns.

A second group of studies focuses on inequalities in Wikipedia participation more directly, also with emphasis on gender divides (Bear & Collier, 2016; Ford & Wajcman, 2017; Hargittai & Shaw, 2015; Hill & Shaw, 2013; Shane-Simpson & Gillespie-Lynch, 2017). Some emphasize the culture of the existing Wikipedia community (e.g., Bear & Collier, 2016), where others emphasize skill-based barriers to entry (e.g., Hargittai & Shaw, 2015) as explanatory factors. On participation gaps by expanding prior models of how individuals become Wikipedia editors. We contribute a more precise conceptual model of a pipeline of participation. We also analyze a more diverse national dataset that includes many individuals who have never edited the site.

Thinking of engagement with Wikipedia in terms of a pipeline goes beyond earlier research. Studies of Wikipedia editing tend to take for granted that people know about Wikipedia, visit Wikipedia, and understand that Wikipedia can be edited. For example, most research on the Wikipedia gender gap focuses almost exclusively on current contributors or social and technical infrastructures of the Wikipedia community in order to determine why people edit Wikipedia and why fewer women edit than men (e.g., Bear & Collier, 2016; Ford & Wajcman, 2017; Lam et al., 2011). Conceptually and empirically, this work overlooks the processes by which such stratification happens. Previous research also risks selecting on key dependent variables by only analyzing data on people who have already become Wikipedia readers and contributors. Without collecting similar data on non-readers and non-contributors, it is not possible to determine what factors influence becoming an editor (Hargittai & Shaw, 2015).

What processes explain persistent participation gaps in content production on Wikipedia? We address this by testing the pipeline model described above. Our expectations specific to the pipeline for Wikipedia derive from earlier related findings. We anticipate that both individuals' backgrounds as well as their Internet-specific experiences and skills shape whether they understand that Wikipedia can be edited

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

and contribute to Wikipedia, and that gender and Internet skills have particularly strong relationships with these outcomes (Correa, 2010; Hargittai & Shaw, 2015; Schradie, 2015). We also expect that Internet experiences and background attributes explain who knows of the site and who has visited it, although we expect gender to play a less important role at these earlier points in the pipeline of digital participation given that the divide in general Internet use across men and women has all but disappeared over time (Martínez-Cantos, 2017; Ono & Zavodny, 2016).

We test our model on a diverse sample of adult Internet users in the U.S.—including many who had never visited or edited Wikipedia—so that we could draw credible inferences about the factors associated with Wikipedia editing and how those relate to prior work on the knowledge gap hypothesis. Our analysis expands on prior studies both theoretically and empirically. Theoretically, we test the pipeline model of participation to see how having heard of, having visited, and knowing that Wikipedia can be edited might help explain who edits Wikipedia. Empirically, we contribute to the literature by examining this question on a national sample that is diverse in terms of age, socioeconomic status, and geography, factors that prior research has not considered in this domain.

Data and methods

To answer our research questions, we collected survey data from a national sample of U.S. adults (18 years old or over) in summer 2016. The survey instrument incorporated detailed measures of individuals' background attributes and Internet experiences consistent with the theories and empirical findings of prior studies. Following the approach applied by Hargittai and Shaw (2015), we used logistic regression to model the likelihood that respondents had ever made an edit to Wikipedia. We also tested for differential outcomes at other points in the pipeline of participation by modeling who had heard of Wikipedia, who had visited Wikipedia, and who knew that Wikipedia can be edited. Details of our data collection and methods follow below.

Data collection

We contracted with an independent research organization, the NORC at the University of Chicago, to administer questions to their AmeriSpeak panel online. The panel is representative of the U.S. population using "area probability sampling and includes additional coverage of hard-to-survey population segments such as rural and low-income households that are underrepresented in surveys relying on address-based sampling" (NORC, n.d.). After pretesting the survey with 23 respondents and updating it based on the results in early May, 2016, we ran the survey from May 25 to July 5, 2016. We included an attention-check question and only analyzed responses from participants who passed this question. In total, we have valid responses from 1,512 American adults 18 and over, which constitutes a 37.8% survey response rate.[1]

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Measures: independent variables

Demographic and socioeconomic factors

Background variables about respondents, such as their age, gender, education, income, and race/ethnicity, were supplied by the NORC based on their earlier data collection about the AmeriSpeak panel. Here we describe what coding we used for these measures. We report age as a continuous variable. We created three education categories: high school or less, some college, and college degree or more. Income was reported in 18 categories, which we recoded to their midpoint values to make it a continuous variable. In the regression analyses, we used the square root of income as this transformation produces a distribution much closer to normal. Race and ethnicity are dummy variables for White, Hispanic, African American, Asian American, Native American, and "Other." We created a dichotomous "coupled" measure for those either married or living with a partner. We have a dummy variable for those employed either full time or part time. We also have a dummy variable signaling rural residence. Finally, we have a continuous measure of household size.

Internet experiences and skills

Following prior literature, we include measures for how long people have been Internet users, how much autonomy they have in accessing the Internet when and where they want, how much time they spend online, and their Internet skills. We asked respondents when they first started using the Internet by offering the following answer options with their recoded values in parentheses: "within the past year" (1), "1 to 5 years ago" (2.5), "more than 5, but less than 10 years ago" (7.5), and "10 or more years ago" (12.5). To measure autonomy of use, we asked "At which of these locations do you have access to the Internet, that is, if you wanted to you could use the Internet at which of these locations?" followed by nine options, including home, workplace, and friend's home. To assess frequency of use, we asked "On an average weekday, not counting time spent on email, chat and phone calls, about how many hours do you spend visiting Web sites?" and then asked the same question about "average Saturday or Sunday." The answer options ranged from "None" to "6 hours or more," with six options in between. We calculated weekly hours spent on the Web by multiplying the answers to the first question by five and the second question by two, and summing the two figures.

For measuring Internet skills, we used a validated, established index (Hargittai & Hsieh, 2012; Wasserman & Richmond-Abbott, 2005). Respondents were presented with 13 Internet-related terms (such as tagging, PDF, spyware) and were asked to rank their level of understanding of these items on a five-point scale ranging from "no understanding" to "full understanding." We then calculated the mean for all items as the Internet skills measure (Cronbach's alpha = .94). For our regression models we centered this measure around its mean so that all coefficients reflect relationships for an individual of average skills.

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Measures: dependent variables

Early in the survey, we asked respondents: "Have you ever heard of the following sites and services?" with Wikipedia among the sites on the list. The vast majority (96.5%) reported having heard of Wikipedia. We then asked the people who had heard of the site:

> Have you ever **visited** the following sites and services? For each site, indicate if no, you have never visited it; yes, you have visited it in the past, but do not visit it nowadays; yes, you currently visit it sometimes; yes, you currently visit it often.

Of those who had heard of Wikipedia, 82.5% said that they had ever visited the site (i.e., have done so in the past or do so currently).

Later in the survey, we presented the statement "Wikipedia is a site that:" with multiple options from among which respondents were asked to "select all that apply." Those who did not select the "can be edited by you" option were coded as not knowing that they could edit Wikipedia. We use these responses to construct a dichotomous measure for knowing that Wikipedia can be edited, for which 68.3% of the sample is coded as affirmative.

To assess who had ever made contributions to Wikipedia, we asked those who reported ever having visited the site the following question: "Have you ever edited a Wikipedia page by fixing a mistake or adding new material?" with "no" and "yes" as the possible answers. A small minority, 8.2%, of the full sample responded in the affirmative.

The sample

Table 1 reports summary statistics for all of the measures. We have close to equal representation of women (51%) and men. The average age was 48.7 years. The majority were White (71%), followed by Hispanics (12%), African Americans (11%), Asian Americans (3%), Native Americans (2%), and people who reported "Other" races (1%). The median income was $55,000; the mean income was $71,478. Just over a quarter of the group (26%) had no more than a high school education, 32% had some college education, and 43% had a college degree or more. Less than two-thirds (61%) lived with a partner (married or not), 62% were employed either full time or part time, and 13% lived in a rural area. The median household size was two people; the mean was 2.6. In sum, while a diverse sample, it was more educated and had a higher income than the average American, as is the case with Internet users generally (Pew Research Center, 2017).

Regarding online experiences, very few of our respondents were new to the Internet: this is not surprising given that Internet use statistics have plateaued in the United States in recent years (Pew Research Center, 2017). The average participant had been using the Internet for over ten years. The median number of access locations was five; the mean was 4.8. The median number of hours participants spent online weekly was 12 hours; the mean was 14.8. Our skill measure ranged from 1–5 and the mean was 3.4 (standard deviation: 1.1), showing that respondents varied considerably in their online know-how.

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Table 1 Descriptive Statistics for Independent Variables

	Percent	Mean	SD	N
Background				
Age (18–94)		48.74	16.87	1512
Income in U.S. $1,000s (2.5–225)		71.48	54.40	1512
Household size (1–6)		2.61	1.32	1512
Female	51			1512
Coupled	61			1512
Employed	62			1512
Rural resident	13			1512
Education				
High school or less	26			1512
Some college	32			1512
Bachelor's or higher	43			1512
Race & Ethnicity				
White	71			1511
Hispanic	12			1511
Black	11			1511
Asian	3			1511
Native American	2			1511
Other	1			1511
Internet Experiences				
Internet autonomy (0–9)		4.80	2.28	1512
Internet use frequency (0–42)		14.75	10.75	1491
Years of Internet use (1–12.5)		11.11	2.78	1512
Internet skills (1–5)		3.37	1.08	1511

Note: For interval measures the range (*min.—max.*) appears in parentheses.

Analyses

First, we present bivariate statistics to show how background characteristics and online skills relate to Wikipedia editing. Then we show results from logistic regression analyses to establish what factors explain each dependent variable when accounting for the independent variables together. To clarify and aid interpretation, we also present marginal effects plots visualizing key relationships in our data on the basis of prototypical values that we use to generate model-based predictions (King, Tomz, & Wittenberg, 2000).

For all models, we present findings for the full sample. That is, we include respondents who have never heard of Wikipedia and those who have never visited the site. Such people, by definition, have never made edits to the site and are centrally relevant to questions of inequality when it comes to understanding whose voices are represented on Wikipedia.

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Consistent with prior work, we also explored alternative model specifications incorporating multiplicative interaction terms (Hargittai & Shaw, 2015). With one exception, these specifications did not improve the overall model fit, based on the results of likelihood ratio tests. The only exception was an interaction between higher education and Internet skills in explaining who has edited Wikipedia. As a result, we report the corresponding model with this term included and do not include interaction terms for any of the other models.

Results

Descriptive analysis: participation gaps

Table 2 summarizes the four dependent variables in our analysis and supports the intuition behind the pipeline model in the context of Wikipedia editing. We see a steep dropoff in each successive outcome. At the extremes, nearly all (97%) of our respondents have heard of Wikipedia, whereas only 8% have ever edited it. To the best of our knowledge, this is the first such estimate ever collected in a national survey of U.S. adults. The figure on Wikipedia readership is consistent with other estimates (Hill & Shaw, 2013).

Table 3 presents bivariate analyses of background characteristics and Internet experiences with who has edited Wikipedia. Here, we report relationships that are significant at $p < 0.001$ (the table presents all relationships). Among women, 5.7% reported having ever changed something on a Wikipedia page compared to 10.9% of men. Only 4.1% of those in the highest quartile of age (62 and over) had edited the online encyclopedia, compared to 16.3% of those in the lowest quartile of age (34 and under). Among those with no more than a high school education, 3.7% had edited Wikipedia, compared to 11.8% of those with a college degree or more.

Regarding the relationship with Internet experiences, 3.7% of those in the bottom quartile of number of access locations had edited the site, compared to 14.2% of those in the top quartile. Three percent of the least frequent Internet users report editing, compared to 11.9% of the most frequent users. The skill differences are

Table 2 Descriptive Statistics for Dependent Variables

	Percent	N
Has heard of Wikipedia	97	1503
Has visited Wikipedia	83	1486
Knows Wikipedia is editable	68	1512
Has edited Wikipedia	8	1483

Note: All are dichotomous measures and percentages are calculated as a proportion of the number (*N*) of valid responses to each question.

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Table 3 Wikipedia Contribution by Respondent Attribute and Prior Internet Experiences

	Percentage	χ^2 statistic
Background		
Age LQ	16.25	39.62***
Age HQ	4.09	11.29***
Income in U.S. $1,000s LQ	9.41	.48
Income in U.S. $1,000s HQ	7.96	.03
Household size LQ	8.96	.14
Household size HQ	8.60	.12
Female	5.69	12.49***
Male	10.87	12.49***
Coupled	7.75	.54
Uncoupled	8.97	.54
Employed	9.30	3.37*
Unemployed	6.45	3.37*
Rural resident	7.96	.00
Urban resident	8.27	.00
Education		
High school or less	3.65	13.59***
Some college	7.10	.94
Bachelor's or higher	11.83	18.22***
Race & Ethnicity		
White	8.58	.44
Hispanic	7.60	.03
Black	7.83	.00
Asian	4.26	.55
Native American	8.33	.00
Other	7.69	.00
Internet experiences		
Internet autonomy LQ	3.72	8.13***
Internet autonomy HQ	14.18	23.57***
Internet use frequency LQ	3.02	14.68***
Internet use frequency HQ	11.89	8.40**
Years of Internet use LQ	3.33	12.64***
Years of Internet use HQ	9.63	12.64***
Internet skills LQ	.81	34.76***
Internet skills HQ	18.51	72.01***

Note: For continuous measures, we report the prevalence of the outcome within the lowest quartile (LQ) and the highest quartile (HQ) of the distribution. *$p < .05$, **$p < .01$, ***$p < .001$

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

especially notable. Among the least skilled quartile, less than one percent had edited Wikipedia, while among the most skilled quartile 18.5% had.

Modeling the pipeline of Wikipedia contribution

Table 4 provides a correlation matrix for all the measures in our study. While all correlations greater than zero meet conventional thresholds for statistical significance, none (except for some between-categories within the same measure) are greater than 0.4, and thus we have no multicollinearity concerns. Table 5 reports the results of our regression analyses. For each dependent variable, we estimate two logit models, regressing the same dichotomous outcome (e.g., whether a respondent has ever edited Wikipedia) on the block of background measures in the first model and then including the Internet experience and skill measures in the second model. We report coefficients as raw log-odds and include the standard error in parentheses. Asterisks indicate which coefficients meet conventional thresholds for statistical significance against a null-hypothesis of no association with the outcome.

Models 1 and 2 present the factors that explain whether respondents have heard of Wikipedia. Income, employment status, education at the bachelor's degree level or higher, and Internet skills predict this outcome. Note that these models do not include measures for the Asian or "Other" race categories due to sparse variation in the dependent variable within these categories (nearly all respondents in the two categories had heard of Wikipedia).

Models 3 and 4 regress the outcome "has visited Wikipedia" on the same blocks of predictors. Again, income, higher education, and Internet skills emerge as salient predictors. Internet autonomy (number of access points) and years of use also help explain this outcome.

In Models 5 and 6, we consider what explains whether individuals know that Wikipedia can be edited. Here age, gender, and several racial/ethnic identity categories (Black, Hispanic, Other) emerge as salient explanatory factors where they did not before. Income no longer explains the outcome. Education level associates strongly with knowing Wikipedia can be edited. Measures of Internet autonomy, years of use, and skills also all remain significant.

Finally, Models 7 and 8 regress "has edited Wikipedia" on the same blocks of predictors. The results in Model 7 reveal strong associations between editing Wikipedia and respondents' genders, ages, and education levels. Consistent with the bivariate statistics described earlier, being female associates strongly with never editing Wikipedia, even when controlling for other background factors. In addition, younger respondents are more likely to have edited, as are those who have an education level of some college or a bachelor's degree or higher, as compared to those with only a high school degree or less.

Incorporating Internet experience and skill measures in Model 8 alters the picture in several ways. A likelihood ratio test indicates that the overall model fit improves (test statistic = 61.9, 4 df, $p = 0$). All of the associations revealed in Model 7 persist, but are attenuated to some degree. Notably, we also observe that greater

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Table 4 Matrix of Pearson Correlation Coefficients for All Measures

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1. Age																							
2. Income in U.S. $1,000s	.1																						
3. Household size	−.3	.1																					
4. Female	−.1	−.1	.1																				
5. Coupled	.1	.3	.3	−.1																			
6. Employed	−.4	.1	.1	0	0																		
7. Rural resident	0	−.1	0	0	.1	−.1																	
8. High school or less	−.1	−.2	.1	.1	0	−.1	.1																
9. Some college	0	−.1	0	0	−.1	0	0	−.4															
10. Bachelor's or higher	.1	.3	−.1	−.1	.1	.1	−.1	−.5	−.6														
11. White	.2	.1	−.2	−.1	.1	−.1	.1	0	−.1	.1													
12. Hispanic	−.2	−.1	.2	.1	0	.1	−.1	0	.1	−.1	−.6												
13. Black	−.1	−.1	0	.1	−.2	0	−.1	0	0	0	−.6	−.1											
14. Asian	−.1	.1	0	0	0	.1	0	−.1	0	.1	−.3	−.1	−.1										
15. Native American	0	0	0	0	0	0	0	0	.1	−.1	−.2	0	0	0									
16. Other	0	0	0	0	0	0	0	0	0	0	−.2	0	0	0	0								
17. Internet autonomy	−.3	.1	.1	0	0	.3	−.1	−.2	0	.2	0	0	0	0	0	0							
18. Internet use frequency	−.3	−.1	.1	0	−.1	0	0	0	.1	−.1	−.2	.1	.1	0	0	0	.1						
19. Years of Internet use	0	.1	−.1	0	0	.1	0	−.3	0	.2	.1	−.1	0	.1	−.1	0	.1	.1					
20. Internet skills	−.4	.1	.1	−.1	0	.3	−.1	−.2	0	.2	0	0	0	0	0	0	.4	.2	.3				
21. Has heard of Wikipedia	−.1	.1	0	0	0	.1	0	−.1	0	.1	0	0	−.1	0	0	0	.1	0	.1	.2			
22. Has visited Wikipedia	−.2	.2	0	0	0	.2	−.1	−.2	0	.2	0	0	0	.1	0	0	.3	.1	.3	.4	.4		
23. Knows Wikipedia is editable	−.3	.1	.1	−.1	0	.2	0	−.2	0	.2	.1	0	−.1	.1	0	−.1	.3	.1	.2	.4	.2	.4	
24. Has edited Wikipedia	−.2	0	0	−.1	0	.1	0	−.1	0	.1	0	0	0	0	0	0	.1	.1	.1	.2	.1	.1	.2

Note: Non-zero values are statistically significant at $p \leq .05$ level.

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Table 5 Regression Models 1–8

	Has heard of WP		Has visited WP		Knows WP editable		Has edited WP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(Intercept)	3.91***	2.27*	1.92***	0.11	2.65***	1.47**	−0.30	−2.51**
	(0.93)	(1.14)	(0.46)	(0.63)	(0.40)	(0.52)	(0.54)	(0.91)
Background								
Age	−0.04**	−0.02	−0.04***	−0.01	−0.06***	−0.04***	−0.05***	−0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Income in U.S. $1,000s	0.15*	0.13*	0.10***	0.07*	0.04	0.01	−0.02	−0.05
	(0.06)	(0.07)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)
Household size	−0.08	−0.10	−0.04	−0.04	0.03	0.03	−0.08	−0.12
	(0.13)	(0.13)	(0.07)	(0.08)	(0.06)	(0.06)	(0.09)	(0.09)
Female	−0.33	−0.34	0.17	0.24	−0.32*	−0.27*	−0.76***	−0.55**
	(0.31)	(0.32)	(0.15)	(0.17)	(0.13)	(0.13)	(0.21)	(0.21)
Coupled	−0.19	−0.03	−0.19	−0.10	−0.09	−0.03	0.03	0.17
	(0.35)	(0.36)	(0.18)	(0.20)	(0.15)	(0.16)	(0.23)	(0.24)
Employed	0.87*	0.74*	0.60***	0.44*	0.11	−0.05	−0.27	−0.38
	(0.36)	(0.37)	(0.17)	(0.18)	(0.14)	(0.15)	(0.24)	(0.25)
Rural resident	−0.32	−0.36	−0.11	0.01	0.17	0.28	0.22	0.38
	(0.37)	(0.38)	(0.20)	(0.22)	(0.18)	(0.19)	(0.30)	(0.31)
Education (base = HS or less)								
Some college	0.74*	0.25	0.90***	0.33	0.84***	0.52**	0.70*	0.25
	(0.33)	(0.35)	(0.18)	(0.20)	(0.16)	(0.17)	(0.34)	(0.36)
Bachelor's or higher	2.08***	1.48**	1.82***	1.33***	1.75***	1.37***	1.51***	1.46***
	(0.51)	(0.53)	(0.21)	(0.23)	(0.17)	(0.18)	(0.33)	(0.41)
Race & Ethnicity (base = White)								
Hispanic	−0.11	0.26	−0.21	0.27	−0.66**	−0.43*	−0.37	−0.17
	(0.58)	(0.60)	(0.26)	(0.29)	(0.21)	(0.22)	(0.33)	(0.35)

(*Continued*)

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Table 5 Continued

	Has heard of WP		Has visited WP		Knows WP editable		Has edited WP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	−0.83*	−0.71	−0.38	−0.26	−1.07*	−1.03***	−0.22	−0.20
	(0.39)	(0.41)	(0.24)	(0.26)	(0.20)	(0.21)	(0.33)	(0.34)
Asian			0.48	0.54	−0.01	−0.03	−1.23	−1.06
			(0.66)	(0.70)	(0.44)	(0.45)	(0.74)	(0.75)
Native American	−0.25	−0.05	−0.83	−0.59	−0.11	0.13	0.29	0.57
	(1.08)	(1.10)	(0.51)	(0.54)	(0.48)	(0.49)	(0.79)	(0.86)
Other			0.68	1.14	−1.19	−1.38*	0.09	−0.13
			(0.83)	(0.90)	(0.64)	(0.70)	(1.08)	(1.10)
Internet Experiences								
Internet autonomy		0.09		0.10*		0.10**		0.08
		(0.08)		(0.04)		(0.03)		(0.05)
Internet use frequency		0.01		0.00		−0.00		0.02
		(0.02)		(0.01)		(0.01)		(0.01)
Years of Internet use		0.07		0.07*		0.05*		0.04
		(0.04)		(0.03)		(0.02)		(0.06)
Internet skills		0.58**		0.86*		0.51***		1.22***
		(0.18)		(0.10)		(0.07)		(0.24)
Edu. BA or higher × Internet skills								−0.58*
								(0.29)
AIC	398.20	380.37	1169.72	1010.52	1574.07	1471.32	777.06	715.11
BIC	467.29	470.49	1249.27	1111.03	1653.88	1572.14	856.58	820.86
Log Likelihood	−186.10	−173.19	−569.86	−486.26	−772.04	−716.66	−373.53	−337.56
Deviance	372.20	346.37	1139.72	972.52	1544.07	1433.32	747.06	675.11
Num. obs.	1502	1482	1485	1465	1511	1490	1482	1462

Note: Coefficients are log-odds with standard errors in parentheses. *$p < .05$, **$p < .01$, ***$p < .001$

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Internet use frequency and Internet use skills positively associate with ever editing Wikipedia.

To further illustrate the results from our regression models, we visualize predicted outcomes given prototypical values of the independent variables. We follow the approach developed by King et al. (2000) and used by Hargittai and Shaw (2015) on similar data. As noted earlier, considerable prior work has focused on gender differences in Wikipedia editing. Thus, we present a visualization of our results that maps the relationships between each outcome, gender, and education and Internet skills (our two strongest predictors). This helps put our visualizations in perspective in relation to earlier work (e.g., Hargittai & Shaw, 2015).

Figure 2 shows the predicted probabilities for each of our four dependent variables when we allow gender, education level, and Internet skills to vary using the empirically observed values from the data and hold other measures at typical values. In the top-left plot, we observe only small variations in the predicted probability of having heard of Wikipedia across the distribution of Internet skills and education levels. Differences primarily emerge at the very low end of the Internet skills spectrum and among those who have completed less than a bachelor's degree. The plot on the top-right shows the predicted probability of having visited Wikipedia and reveals how strongly this outcome relates to Internet skills and education level. The gap between the gray (BA or higher) and black (less than a BA) lines gets widest at the low end of the Internet skills spectrum. In contrast, at the high end of the skills distribution, educational divides in who visits Wikipedia nearly disappear.

The second row of plots in Figure 2 show the predicted probabilities for the less prevalent outcomes: knowing that Wikipedia can be edited (bottom left) and having contributed to Wikipedia (bottom right). Across all three of our key predictors (gender, education level, and Internet skills), we observe substantial differences in who knows that Wikipedia can be edited. At the extremes, an immense gap separates the predicted probability that a low-skill, low-educated female and a high-skill, high-educated male know that Wikipedia can be edited (about 28% versus 94%, respectively). Such vast inequalities underscore that this step in the pipeline of participation accounts for a huge decrease in the population of potential Wikipedia editors.

Finally, we visualize variations in the probability of editing Wikipedia in the bottom right plot of Figure 2. In comparison to the other three plots, we note the low probability of this outcome across the board. The predicted probability of editing Wikipedia remains very close to zero for anyone at the lower end of the Internet use skills distribution. At the upper end of the skills distribution, the predicted values again diverge along all three of the predictors. In other words, a gender gap in Wikipedia contribution does not emerge among low or average values of Internet skills. Instead, the gap exists almost exclusively among the most highly-skilled and highly-educated Internet users, all of whom remain much more likely to edit Wikipedia than individuals with lower levels of Internet skills or education. The

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Figure 2 Marginal effects plots.

Note: Predicted probabilities of having heard of Wikipedia (top left); having visited Wikipedia (top right); knowing that Wikipedia can be edited (bottom left); and having contributed to Wikipedia (bottom right).

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

model predicts that the lower-educated, highly-skilled male respondents are just as likely to edit Wikipedia as the highly-educated, highly-skilled female respondents.

Discussion

The results provide empirical support for the idea that knowledge gaps contribute to a pipeline of online participation in the networked public sphere. Individuals do not automatically engage in content production on even the most participatory sites, but must have the necessary knowledge to do so. Participation divides emerge through a sequence of necessary awareness and behaviors. Education, Internet skills, and age have robust associations with outcomes at every step in the pipeline. Other factors, such as income, employment status, and racial/ethnic background, help explain earlier stages in the pipeline even though they do not associate with who contributes content. Gender only matters at later stages in the pipeline, despite the important and valid emphasis of prior research and public debate on the Wikipedia gender gap. Distinct from prior studies, we provide evidence that participation divides in who creates content online are likely due to variations in who has visited a site and has the requisite knowledge that such contributions are possible (see Figure 2). This contributes a clearer understanding of stratification processes in knowledge production in the networked public sphere.

These findings make several other contributions to prior research. First, bivariate comparisons in our data reveal participation gaps across many dimensions of background and Internet experiences that had not been documented previously. In multivariate regression models, age, education level, and Internet skills are the only measures that consistently explain outcomes throughout the pipeline of participation. Individuals who are younger, more highly educated, and more skilled Internet users are more likely to have ever heard of, visited, known about editing, and edited Wikipedia.

Other background attributes and Internet experiences explain variations at different stages in the pipeline and to different degrees. Some characteristics (income, employment status), as well as Internet experiences and autonomy, help explain variations in having heard of Wikipedia or visiting Wikipedia. Other attributes (racial/ethnic background), as well as Internet experiences and autonomy, help explain intermediate stages (visiting Wikipedia and/or knowing Wikipedia can be edited). A gender gap emerges only at the latter stages in the pipeline, in terms of who knows the site can be edited and who has edited the site.

The results appear consistent with our expectations, but several patterns surprised us. Participation divides in the middle stages of the pipeline are vast compared to those at the pipeline's endpoints. In particular, we note that large differences emerge in our models of who has visited Wikipedia and who knows Wikipedia can be edited (see Figure 2). These gaps along dimensions of education, Internet skills, and (in the case of knowing Wikipedia can be edited) gender all dwarf inequalities in the predicted probability of ever having edited Wikipedia.

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Prior research has neither identified nor considered these gaps. Our findings suggest that the intermediate stages matter a great deal and should attract much closer scrutiny in the future. We speculate that reducing gaps at these intermediate stages of the participation pipeline would make more equitable participation patterns possible. Future research should investigate the effects of interventions to test this claim.

Limitations of this study include standard concerns about the validity of self-reported behavioral data, as well as unknown biases in our estimation related to the sampling techniques, measurement errors, or omitted variables. For example, we note that the study includes no measures of underlying interests or attitudes, which play a role in other kinds of online behavior (Brake, 2014). Furthermore, the study only includes U.S. adult Internet users, and the results may diverge from patterns in other populations. Engagement patterns with Wikipedia may also differ from engagement with other sites. We have no way of empirically testing or bounding the impact of these issues, but encourage readers to approach our results with appropriate care. Wherever possible, we have documented alternative estimation procedures and decisions in this paper, none of which have eliminated the patterns of results we report. We hope to see others pursue follow-up studies to evaluate any concerns empirically. We also hope that future studies will evaluate similar questions using panel data in order to assess whether the patterns we report here change over time. Future work may also elaborate alternative conceptual frameworks to emphasize different dimensions of the selection and stratification processes involved.

For all of these findings, we draw confidence from the fact that our analyses involved an extremely diverse national sample that includes respondents who had never heard of nor visited Wikipedia, one of the most popular participatory sites on the Internet. We also incorporated an additional measure of rural/urban location into our analysis based on prior work (Johnson et al., 2016) that earlier participation gap research had not used, but found no evidence of an association with who edits the online encyclopedia.

Conclusion

We observed a sequence of awareness and actions necessary to engage in online participation and show how knowledge gaps manifest at every step. These participation divides may contribute to the expansion of knowledge gaps along the lines proposed by the original knowledge gap hypothesis (Tichenor et al., 1970). Online platforms may offer more egalitarian and inclusive practices of knowledge production than prior media ecosystems (Benkler, 2006; Reagle & Rhue, 2011). At the same time, any biases encoded into Wikipedia may also propagate widely, as intelligent and algorithmic systems incorporate user-generated content such as Wikipedia's automatically into search results and other basic infrastructures of the networked public sphere (e.g., Johnson et al., 2016). The proliferation of downstream systems using content generated by participatory sites means that digital inequalities and skills

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

gaps may have far-reaching and unexpected consequences. These aspects of knowledge gaps in the networked public sphere also deserve deeper investigation.

Our results illuminate important dimensions of participation gaps that prior research and policy debates have either not emphasized or not addressed at all. The findings about education levels and Internet skills deserve particular emphasis. Research on Wikipedia participation has neither identified nor estimated an education divide, a concept particularly relevant to the digital inequality literature. In terms of the skills divide, our findings are consistent with Hargittai & Shaw's (2015), and we show further evidence that Wikipedia contribution remains vanishingly rare at the low end of the Internet skills spectrum. Transforming the culture of participation among existing Wikipedians—an area of intervention that receives considerable attention—will not overcome participation gaps. The fact that less educated, lower skilled, and older individuals are less likely to engage at every point in the participation pipeline suggests that interventions aimed at addressing these factors will be necessary to produce a more equitable and representative online encyclopedia. We believe these findings lend support to interventions that reduce technical and knowledge-based barriers to entry. We also suggest interventions aimed at including lower income, unemployed, and underrepresented racial and ethnic groups at earlier stages in the participation pipeline where gaps in Internet experience and autonomy may present distinct obstacles to subsequent engagement. When it comes to overcoming gender gaps in the case of Wikipedia, our results suggest that continued emphasis on recruiting female editors must include efforts to disseminate the knowledge that Wikipedia can be edited.

Future work should pursue causal identification and longitudinal tests of the patterns of results we have reported here. Evaluations of interventions aimed at redressing participation divides and pipelines of contribution in Wikipedia and beyond also offer a promising avenue of inquiry. Studies that focus on pathways of participation (e.g., Antin, Cheshire, & Nov, 2012) should seek to understand the role of factors such as Internet skills and education. We also encourage future studies to analyze stratification processes and pipelines of participation in online knowledge-production activities beyond Wikipedia. Such comparative analyses may reveal pathways towards greater equity and engagement among different demographic groups or people at different levels of Internet experiences and skills. Our focus on a single, highly-impactful site like Wikipedia has allowed us to elaborate an initial statement and evaluation of the pipeline framework. This extends prior work on digital inequality in the context of knowledge gaps. Better understanding these phenomena can inform design research and policy interventions aimed at realizing a more equitable and inclusive digital media environment.

Authors note

The order of author names does not reflect differences in contribution to the research.

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Aaron Shaw is Assistant Professor of Communication Studies at Northwestern University and a 2017–2018 Lenore Annenberg and Wallis Annenberg Fellow in Communication at the Center for Advanced Study in the Behavioral Sciences at Stanford University. Eszter Hargittai is Professor and Chair of Internet Use and Society at the Institute of Communication and Media Research at the University of Zurich. Shaw is Faculty Associate and Hargittai is on the Fellow Advisory Board of the Berkman Klein Center for Internet and Society at Harvard University.

Acknowledgements

The authors are grateful to Merck (Merck is known as MSD outside the United States and Canada) and the Robert and Kaye Hiatt Fund at Northwestern University for support. Hargittai also appreciates the time made available through the April McClain-Delaney and John Delaney Professorship at Northwestern University to conduct this work.

The authors appreciate input from the 2015–2016 cohort of research assistants of the Web Use Project at Northwestern University, especially Sam Mandlsohn, and acknowledge staff at the NORC at the University of Chicago for administering the survey discussed here. Audiences in the Research Seminar of the Institute for Communication and Media Research at the University of Zurich, the Bring Your Own Research Workshop at Northwestern, the Department of Media and Communications at the London School of Economics and Political Science, the Annual Conference of the International Communication Association, and the Annual Meeting of the American Sociological Association, as well as Darren Gergle, Ellen Helsper, and the editor and reviewers of the *Journal of Communication*, provided valuable feedback and direction on earlier versions of this work.

Note

1 We note that the NORC provides ranked survey weights for data from the AmeriSpeak panel which, in theory, enable an approximation of a representative sample of U.S. adults. We do not use the weights in our analyses here. Our rationale stems from standard concerns about sparsity and survey weighting, as well as empirical testing of our models with and without the weights (Bollen, Biemer, Karr, Tueller, & Berzofsky, 2016). Given that we are modeling several rare outcomes, sparsity poses a real threat. Following three widely used and analytically distinct procedures discussed in Bollen and colleagues' review, we find (1) weak evidence that the model residuals are correlated with the weights, (2) no evidence that adding the weights improves overall model fit, and (3) some evidence of meaningful overall differences in the coefficients with and without weights applied (see Bollen et al., 2016 for details of testing procedures). The overall pattern of findings does not change in weighted versus unweighted models.

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

References

Adams, J., & Brückner, H. (2015). Wikipedia, sociology, and the promise and pitfalls of Big Data. *Big Data & Society*, **2**(2), 1–5. doi:10.1177/2053951715614332

Antin, J., Cheshire, C., & Nov, O. (2012). Technology-mediated contributions: editing behaviors among new wikipedians. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 373–382). New York, NY: ACM. doi: 10.1145/2145204.2145264

Bear, J. B., & Collier, B. (2016). Where are the women in Wikipedia? Understanding the different psychological experiences of men and women in Wikipedia. *Sex Roles*, **74**(5), 254–265. doi:10.1007/s11199-015-0573-y

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Bollen, K. A., Biemer, P. P., Karr, A. F., Tueller, S., & Berzofsky, M. E. (2016). Are survey weights needed? A review of diagnostic tests in regression analysis. *Annual Review of Statistics and Its Application*, **3**(1), 375–392. doi:10.1146/annurev-statistics-011516-012958

Bonfadelli, H. (2002). The Internet and knowledge gaps: A theoretical and empirical investigation. *European Journal of Communication*, **17**(1), 65–84. doi:10.1177/0267323102017001607

Brake, D. R. (2014). Are we all online content creators now? Web 2.0 and digital divides. *Journal of Computer-Mediated Communication*, **19**(3), 591–609. doi:10.1111/jcc4.12042

Cacciatore, M. A., Scheufele, D. A., & Corley, E. A. (2014). Another (methodological) look at knowledge gaps and the Internet's potential for closing them. *Public Understanding of Science*, **23**(4), 376–394. doi:10.1177/0963662512447606

Chen, W., & Wellman, B. (2004). Charting digital divides: Comparing socioeconomic, gender, life stage, and rural-urban internet access and use in five countries. In W. Dutton, B. Kahin, R. O'Callaghan & A. Wyckoff (Eds.), *Transforming Enterprise* (pp. 467–497). Cambridge, MA: MIT Press.

comScore, Inc. (2016). comScore ranks the top 50 U.S. digital media properties for January 2016. *comScore.* Retrieved June 6, 2017 from http://www.comscore.com/Insights/Rankings/comScore-Ranks-the-Top-50-US-Digital-Media-Properties-for-January-2016 and archived at https://perma.cc/KRD4-YVZ9

Correa, T. (2010). The participation divide among "online experts": Experience, skills and psychological factors as predictors of college students' web content creation. *Journal of Computer-Mediated Communication*, **16**(1), 71–92. doi:10.1111/j.1083-6101.2010.01532.x

van Deursen, A., Helsper, E., Eynon, R., & van Dijk, J. (2017). The compoundness and sequentiality of digital inequality. *International Journal of Communication*, **11**, 22.

DiMaggio, P., & Bonikowski, B. (2008). Make money surfing the web? The impact of Internet use on the earnings of U.S. workers. *American Sociological Review*, **73**(2), 227–250. doi:10.1177/000312240807300203

DiMaggio, P., Hargittai, E., Celeste, C., & Schafer, S. (2004). Digital inequality: From unequal access to differentiated use. In K. Neckerman (Ed.), *Social Inequality* (pp. 355–400). New York, NY: Russell Sage Foundation. https://www.russellsage.org/research/reports/dimaggio

Dutton, W. H., & Blank, G. (2014). The emergence of next generation Internet users. *International Economics and Economic Policy*, **11**(1–2), 29–47. doi:10.1007/s10368-013-0245-8

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Ettema, J. S., & Kline, F. G. (1977). Deficits, differences, and ceilings: Contingent conditions for understanding the knowledge gap. *Communication Research*, **4**(2), 179–202. doi: 10.1177/009365027700400204

Ford, H., & Wajcman, J. (2017). 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science*, **47**(4), 511–527. doi: 10.1177/0306312717692172

Friedland, L. A., Hove, T., & Rojas, H. (2006). The networked public sphere. *Javnost—The Public*, **13**(4), 5–26. doi:10.1080/13183222.2006.11008922

Graells-Garrido, E., Lalmas, M., & Menczer, F. (2015). First women, second sex: Gender bias in Wikipedia. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media* (pp. 165–174). New York, NY: ACM. doi:10.1145/2700171.2791036

Haight, M., Quan-Haase, A., & Corbett, B. A. (2014). Revisiting the digital divide in Canada: The impact of demographic factors on access to the internet, level of online activity, and social networking site usage. *Information, Communication & Society*, **17**(4), 503–519. doi:10.1080/1369118X.2014.891633

Hampton, K. N., Lee, C., & Her, E. J. (2011). How new media affords network diversity: Direct and mediated access to social capital through participation in local social settings. *New Media & Society*, **13**(7), 1031–1049. doi:10.1177/1461444810390342

Hargittai, E. (2002). Second-level digital divide: differences in people's online skills. *First Monday*, **7**(4). Retrieved January 28, 2018 http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/942/864 and archived at https://perma.cc/996D-T6VL

Hargittai, E., & Hsieh, Y. P. (2012). Succinct survey measures of web-use skills. *Social Science Computer Review*, **30**(1), 95–107. doi:10.5210/fm.v7i4.942

Hargittai, E., & Jennrich, K. (2016). The online participation divide. In L. A. Friedland & M. Lloyd (Eds.), *The Communication Crisis in America, and How to Fix It* (pp. 199–213). Palgrave Macmillan. doi:10.1057/978-1-349-94925-0_13

Hargittai, E., & Shaw, A. (2013). Digitally savvy citizenship: The role of Internet skills and engagement in young adults' political participation around the 2008 presidential election. *Journal of Broadcasting & Electronic Media*, **57**(2), 115–134. doi:10.1080/08838151.2013.787079

Hargittai, E., & Shaw, A. (2015). Mind the skills gap: The role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information, Communication & Society*, **18**(4), 424–442. doi:10.1080/1369118x.2014.957711

Hargittai, E., & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication & Society*, **11**(2), 239–256. doi:10.1080/13691180801946150

Hill, B. M., & Shaw, A. (2013). The Wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PLoS One*, **8**(6), e65782. doi:10.1371/journal.pone.0065782

Hwang, Y., & Jeong, S.-H. (2009). Revisiting the knowledge gap hypothesis: A meta-analysis of thirty-five years of research. *Journalism & Mass Communication Quarterly*, **86**(3), 513–532. doi:10.1177/107769900908600304

Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York, NY: New York University Press.

Johnson, I. L., Lin, Y., Li, T. J.-J., Hall, A., Halfaker, A., Schöning, J., & Hecht, B. (2016). Not at home on the range: Peer production and the urban/rural divide. In *Proceedings of the*

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

2016 CHI Conference on Human Factors in Computing Systems (pp. 13–25). New York, NY: ACM. doi:10.1145/2858036.2858123

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, **44**(2), 347–361. doi:10.2307/2669316

Lam, S. K., Uduwage, A., Dong, Z., Sen, S., Musicant, D. R., Terveen, L., & Riedl, J. (2011). WP:clubhouse?: An exploration of Wikipedia's gender imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (pp. 1–10). New York, NY: ACM. doi:10.1145/2038558.2038560

Livingstone, S., & Helsper, E. (2007). Gradations in digital inclusion: Children, young people and the digital divide. *New Media & Society*, **9**(4), 671–696. doi:10.1177/1461444807080335

Martínez-Cantos, J. L. (2017). Digital skills gaps: A pending subject for gender digital inclusion in the European Union. *European Journal of Communication*, **32**(5), 419–438. doi:10.1177/0267323117718464

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine of the National Academies. (2007). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering*. Washington, DC: National Academies Press.

NORC. (n.d.). Amerispeak: NORC's breakthrough panel-based research platform. Retrieved February 5, 2018 from http://www.norc.org/Research/Capabilities/pages/amerispeak.aspx and archived at https://perma.cc/2TBM-KLUJ

Ono, H., & Zavodny, M. (2016). Internet and gender. In N. Naples (Ed.), *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies* (pp. 1–4). John Wiley & Sons, Ltd. http://onlinelibrary.wiley.com/doi/10.1002/9781118663219

Pearce, K. E., & Rice, R. E. (2013). Digital divides from access to activities: Comparing mobile and personal computer Internet users. *Journal of Communication*, **63**(4), 721–744. doi:10.1111/jcom.12045

Perrin, A. J., & Vaisey, S. (2008). Parallel public spheres: Distance and discourse in letters to the editor. *American Journal of Sociology*, **114**(3), 781–810. doi:10.1086/590647

Pew Research Center. (2017). Internet/broadband fact sheet. *The Pew Internet & American Life Project*. Retrieved February 5, 2018 from http://www.pewinternet.org/fact-sheet/internet-broadband/ and archived at https://perma.cc/LJ8G-FAVQ

Reagle, J., & Rhue, L. (2011). Gender bias in Wikipedia and Britannica. *International Journal of Communication*, **5**, 21.

Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., … Stern, M. J. (2015). Digital inequalities and why they matter. *Information, Communication & Society*, **18**(5), 569–582. doi:10.1080/1369118X.2015.1012532

Schradie, J. (2011). The digital production gap: The digital divide and Web 2.0 collide. *Poetics*, **39**(2), 145–168. https://doi.org/16/j.poetic.2011.02.003

Schradie, J. (2015). The gendered digital production gap: Inequalities of affluence. In L. Robinson, S. R. Cotten, & J. Schulz (Eds.), *Communication and information technologies annual* (**Vol. 9**, pp. 185–213). Bingley, UK: Emerald Group Publishing Limited. doi:10.1108/S2050-206020150000009008

Selwyn, N. (2004). Reconsidering political and popular understandings of the digital divide. *New Media & Society*, **6**(3), 341–362. doi:10.1177/1461444804042519

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Shane-Simpson, C., & Gillespie-Lynch, K. (2017). Examining potential mechanisms underlying the Wikipedia gender gap through a collaborative editing task. *Computers in Human Behavior*, **66**, 312–328. doi:10.1016/j.chb.2016.09.043

Tichenor, P. J., Donohue, G. A., & Olien, C. N. (1970). Mass media flow and differential growth in knowledge. *Public Opinion Quarterly*, **34**(2), 159–170. doi:10.1086/267786

Wagner, C., Graells-Garrido, E., Garcia, D., & Menczer, F. (2016). Women through the glass ceiling: Gender asymmetries in Wikipedia. *EPJ Data Science*, **5**(1). doi:10.1140/epjds/s13688-016-0066-4

Warschauer, M. (2003). *Technology and social inclusion: Rethinking the digital divide*. Cambridge, MA: MIT Press.

Wasserman, I. M., & Richmond-Abbott, M. (2005). Gender and the Internet: Causes of variation in access, level, and scope of use. *Social Science Quarterly*, **86**(1), 252–270. doi:10.1111/j.0038-4941.2005.00301

Zillien, N., & Hargittai, E. (2009). Digital distinction: Status-specific Internet uses. *Social Science Quarterly*, **90**(2), 274–291.

Downloaded from https://academic.oup.com/joc/article-abstract/68/1/143/4915319
by Stanford University Libraries, aaronshaw@northwestern.edu
on 27 April 2018

Managing Organizational Culture in Online Group Mergers

CHARLES KIENE, University of Washington, USA
AARON SHAW, Northwestern University, USA
BENJAMIN MAKO HILL, University of Washington, USA

Research in social computing has typically conceptualized community growth as a process through which a group welcomes newcomers individually. However, online communities also grow through formal and informal mergers, where groups of newcomers with shared experiences join in batches. To understand this process, we conducted a six month, comparative ethnography of two mergers of *World of Warcraft* raid guilds. While one merger led to a single, thriving community, the other led to the dissolution of both pre-merger groups. Analysis of our ethnographic data suggests that differences in managing organizational culture (a concept drawn from organization studies) led the successful and failed cases to diverge. The study contributes to our understanding of why some attempts to integrate members of different communities are more successful than others. We outline several ways that community leaders, researchers, and designers can effectively take organizational culture into account.

CCS Concepts: • **Human-centered computing → Ethnographic studies**; **Empirical studies in collaborative and social computing**; *Computer supported cooperative work*;

Keywords: organizational mergers; organizational culture; guilds; virtual worlds; newcomer socialization; online communities; social computing and social navigation; computer-mediated communication; World of Warcraft

ACM Reference Format:
Charles Kiene, Aaron Shaw, and Benjamin Mako Hill. 2018. Managing Organizational Culture in Online Group Mergers. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, CSCW, Article 89 (November 2018). ACM, New York, NY. 21 pages. https://doi.org/10.1145/3274358

1 INTRODUCTION

Influenced by industrial and organizational psychology, social computing research has typically approached the creation and growth of new communities as a task of attracting and socializing newcomers individually [22]. In practice, however, online community growth can often look quite different, as groups may draw in newcomers with deep experience from other communities who arrive together in batches. In mergers, both formal and informal ways, communities may grow by merging with other communities. These groups of newcomers share a history of experiences, familiar patterns of interaction with each other, and assumptions about how effective online communities should be managed, organized, and led. If unmanaged, differences between subgroups within a post-merger community can lead to instability, conflict, attrition, and even collapse.

Authors' addresses: Charles Kiene, University of Washington, Department of Communication, Seattle, WA, 98195, USA, ckiene@uw.edu; Aaron Shaw, Northwestern University, Communication Studies, Evanston, IL, 60208, USA, aaronshaw@northwestern.edu; Benjamin Mako Hill, University of Washington, Department of Communication, Seattle, WA, 98195, USA, makohill@uw.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
© 2018 Copyright held by the owner/author(s).
2573-0142/2018/11-ART89
https://doi.org/10.1145/3274358

In studies of the mergers of firms, differences in organizational culture between pre-merger groups can lead to conflict, discontent among members, and reductions in team performance [18, 30, 37]. Given the importance of culture and the existence of mergers in online groups and communities, we would expect culture to matter when two or more online groups come together as well. However, we know of no work that has studied group culture and mergers in online groups. Because firms and virtual groups differ in terms of incentives, obligations, resources to resolve problems, and more, we understand little about the process of online group mergers and the way that online group leaders should manage them.

Drawing from a six-month ethnography of two pairs of similar *World of Warcraft* raid guilds that were planning and undergoing mergers, we describe how one merger resulted in a single, thriving community, while another failed and led to the dissolution of both pre-merger groups. Supported by analysis of our field notes, interviews, other data, and theory, we show that organizational culture played a critical role in the individuals' ability to participate effectively within the merged groups. We show how the mergers destabilized organizational cultures and how the participants in the successful merger strategically selected each other, engaged in intentional socialization across the groups, and cultivated solidarity to produce a new, stabilized organizational culture.

Our results illustrate the importance of organizational culture for the design and management of online groups. In particular, we expand the existing literature on newcomer socialization in social computing by demonstrating how this concept entails collective dimensions that go beyond the emphasis of prior work on newcomers as individuals. Our findings also build on previous studies of *World of Warcraft* raid guilds by showing the consequences of guild cultures and cultural conflicts on guild survival and performance. We argue that leaders of online communities, as well as future social computing research and design, should attend to organizational culture and its role in shaping the outcomes of cooperative work and play.

2 BACKGROUND AND THEORY

2.1 Newcomer Socialization

Newcomer socialization has been one of the most important topics in research into online communities [9, 22]. This work has relied heavily on a distinction between collective, *institutionalized* practices of socialization and on-the-job, *individualized* socialization [36]. In particular, researchers studying online communities have sought to evaluate the relevant effectiveness of the two different socialization tactics in different online contexts [22]. With their roots in organizational studies, the goal of both institutionalized and individualized socialization techniques has been to support the effective integration of new employees at firms who, typically, are hired one-by-one. Although classic examples of institutionalized socialization might involve the simultaneous integration of many newcomers, the process is still fundamentally conceived of as integrating individuals who enter without strong connections to each other, the organization they are joining, or prior collective identities.

Newcomers to traditional organizations can arrive in groups that have been socialized together—for example, through an acquisition of one company by another. This also happens in social computing environments when members of one community merge with or join another group [34]. Unlike both institutionalized and individualized socialization, this process requires navigating two distinct *organizational cultures* within the merged group. Although this process has been studied in organization science [6, 18], this model of newcomer socialization has received very little attention in studies of online communities.

2.2 Organizational Mergers and Culture

Organizational researchers have acknowledged organizational culture as playing an important role in determining the success of mergers. In particular, scholars have suggested that the leaders' ability to consider issues of cultural compatibility when considering mergers and acquisitions, and their ability to manage cultural differences during integration, influences the newcomers' abilities to integrate into new groups as well as their post-merger performance [18, 29, 30, 37]. Studies of the role of culture in mergers have largely focused on firms and major corporations [6, 29, 30, 35, 37].

We build on this prior work and borrow the concept of *organizational culture* from Edgar Schein as the framework for our analysis of organizational mergers. Defined by Schein [30], organizational culture describes:

> A pattern of shared, basic assumptions that was learned by a group as it solved its problems of external adaptation and internal integration, that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems.

In Schein's framework, organizational culture has three layers: a top layer with tangible indicators of norms and values called *artifacts*, a middle layer of espoused beliefs that define the unwritten norms and expectations of behavior, and finally, an underlying layer of *common sense* basic assumptions for guiding the group's everyday operations [30]. Schein saw group cultures as a "...common language of shared assumptions about the basic logistical operations..." that groups used to move from missions to accomplished goals. In this sense, organizational culture is a mental model for *what has worked in the past* and *how things should be going forward*. A culture's stability within a group depends on its success in achieving the group's goals. Schein argues that analyzing a group's culture at the level of shared basic assumptions allows for a better, more precise understanding for how groups should manage conflict and change [29]. For leaders of merging firms, understanding this deeper level of one's organizational culture allows for greater insight into how to manage potential incompatibilities.

Prior social computing research has addressed many dimensions of culture within online communities. For example, Levina and Arriaga [23] used Pierre Bourdieu's theory of cultural production to explain how the different types of content producers and consumers in online fields create unique power relations over the Internet. Yang et al. [42] showed how cultural differences should be considered in the design of social search systems. Eagar et al. [16] used virtual ethnography in a comparative study of online fan communities to understand how inter-community conflict can construct social boundaries.

Scholars have also addressed dimensions of culture in online gaming communities. For example, several studies of massive, multiplayer online games (MMOGs) consider intra-group conflict and cohesion dynamics [8, 28] as well as the importance of culture for newcomer socialization and community norms [22]. That said, few studies have considered the impact of cultural variation across online groups. Although very recent work by Tan [34] has shown that online groups frequently include a large number of members with shared experience in antecedent groups, the dynamics or effects of mergers in online groups have never been studied. Organizational culture provides a theoretical frame to help us do so.

3 EMPIRICAL SETTING

3.1 *World of Warcraft* Raid Guilds

MMOGs have provided ideal settings for social computing researchers to study phenomena related to organization, human-computer interaction, and computer-supported cooperative work [2, 12, 15, 25, 38]. Of particular interest has been *World of Warcraft* (WoW). Launched in 2004 by Blizzard



Fig. 1. A screen capture of the lead author's WoW interface during a failed encounter with a raid boss, Fallen Avatar, in the *Tomb of Sargeras* raid on *Heroic* difficulty. The 13 rectangles at the center-bottom of the screen represented the status of all of the members on the raid team. Because all 12 of the members have the "Dead" status and the raid boss is still alive, this would be considered a *wipe*.

Entertainment, WoW continues to be one of the most widely-played MMOGs with millions of active players [41]. In WoW, players create and control avatars that exist in the game's virtual world. A player's character increases in power by collecting experience from activities, like slaying monsters and completing quests. The most powerful player characters can engage in *raiding*, the game's most challenging and rewarding activity.

Raids are private instances of the game world set within a dungeon, in which the game's most powerful monsters, called "raid bosses," defend against the player characters. Player characters must join formal groups ranging between 10 to 30 other players to successfully defeat raid bosses. Raid bosses hold the game's most sought after items, which are distributed to players as "loot" if the raid is successful. These digital items, like the *Robe of Aqueous Command* or the *Gauntlets of Spiteful Haunting*, increase the power of player characters who posses them by giving them statistical advantages in future combat. Raids "reset" weekly, restoring all raid bosses to life along with new random selections of loot. This arrangement provides an incentive for players to try again each week, even once they have been successful. WoW releases increasingly challenging game content over time, in a series of *tiers*. For example, in January 2017, WoW released a patch containing the *Nighthold* tier of raid content. In June 2017, the game added the *Tomb of Sargeras* raid tier that included a more difficult raid with a new series of raid bosses. Raids can be attempted at several difficulty levels: *Normal, Heroic,* or *Mythic,* as well as a special *Raid Finder* difficulty that can only be accessed using the game's Raid Finder[1] system.

Encounters with a raid boss typically last for 5 minutes, although more difficult bosses may last longer. A completely successful raid boss encounter results in the raid boss being killed while every player character in the group is still alive. A failure scenario results in a "wipe"—the defeat of the entire group of player characters before they have dealt enough damage to the raid boss to defeat it, as shown in Figure 1.

[1]Raid Finder is an in-game tool that matches players with others seeking to complete a specific raid.

As is common in many MMOGs, WoW players join together in formal groups called *guilds* to organize and schedule recurring activities [24, 38]. Guilds in WoW are formed by a single player, the *Guild Master*, who is afforded special privileges by the game to manage and organize the guild's bank (a shared digital space for storing virtual items) as well as a group calendar, in-game chat channel, and an informational dashboard. The Guild Master is afforded considerable power over their guild community: they may invite other members, assign ranks with corresponding privileges, and can even disband the guild entirely. Guilds can consist of up to 1000 members, although most guilds in WoW consist of 35 or fewer [15].

Most raiding guilds are focused on *progression*—a term used to describe the number of raid bosses that a guild has defeated in the game's most recent raid tier [3]. Raid guilds typically schedule raid activities geared toward progression for several hours at a time, multiple times per week. Because the game limits the size of raid groups to 30 players, raid guilds with more than 30 members select players for their raid team. This choice is usually made on the player's character's level, power as derived from their gear, and whether the character specializes in a role that fits within the raid team's composition.

Raid guilds are categorized by players in an unofficial typology: *hardcore, medium-core, and casual* [3, 10]. Hardcore raid guilds prioritize progression raiding over all other activities and compare their progress to other hardcore raid guilds [3]. Hardcore raid guilds primarily take on the game's hardest difficulty, *Mythic,* which requires exactly 20 members on a raid team. Progression in medium-core raid guilds usually takes place within *Heroic* difficulty raids, which require between 10 to 30 raid team members. Medium-core raid guilds are the most common type of raid guild and they are often more socially-oriented than hardcore guilds. Casual raid guilds usually consist of friends who take on the easier, *Normal* difficulty raids. These guilds are typically socially-oriented with very few in-game goals.

3.2 Studying Organizational Culture in WoW Raid Guilds

WoW raid guilds are a popular laboratory setting for studying teams and organizations in social computing research. A large body of research on guilds has studied member recruitment and retention, group stability, group dynamics, social dynamics, structural organization, and community [1, 3, 10, 14, 15, 24, 27, 32, 38]. Although none of this work has engaged with literature on organizational culture, some have described different patterns of group dynamics across guilds, hinting at cultural differences. For example, Choi et al. [10] differentiate guilds as either "task-oriented" or "social-oriented"; Bardzell et al. [3] categorize raid guilds as engaged in either "farm progression raiding" or "achievement progression raiding"; and Williams et al. [38] suggest that raid guilds are socially organized like military "barracks" or social "tree house" play spaces. Our study extends this line of research by using organizational culture to understand how guilds with similar social structures face challenges of integration and socialization.

Research has shown that many factors can feed attrition and lead to the collapse of even solidly established guilds [28]. Because raid guilds require some number of participants to continue, and because guilds may struggle to attract and retain members, guilds with too few members seek mergers as a way of avoiding collapse [38]. These mergers are frequently the source of conflict. For example, Williams observed that guild mergers were "...as contentious as any real world corporate-merger..." and usually resulted in conflicts over management style that were referred to by his informants as "culture shock" [38]. Although not focused on mergers, work by Chen [8] has shown how this type of intra-group conflict can also threaten the guilds' continued existence. However, neither Williams nor any other prior studies of online communities have attempted to explain why some mergers halt attrition, while other merged groups continue along a path of member attrition, often exacerbated by intra-group conflict.

4 METHODOLOGY

We pursued a six-month, comparative, ethnographic study of two pairs of WoW raiding guilds that were undergoing mergers. The two ethnographic studies were conducted separately and were sequential, but overlapping. Both cases were then analyzed together at the end of our time in the field, allowing us to develop inductive codes and reflect on how our observations related to prior literature on newcomer socialization and organizational culture in the context of online group mergers.

Although comparative or "parallel" ethnographies are rare in social computing research, they have played an important role in organizational studies [4, 20] and provide an ideal method for analyzing the impact of cultural differences across organizations. Culture scholars have emphasized the importance of ethnography for understanding the effect of culture, especially in new contexts [19, 29, 30, 38]. Ethnography is also seen as especially appropriate when investigating phenomena in virtual worlds, where engaging in the activities of games gives greater contextual sense of what is occurring [2, 5, 39]. Taking an ethnographic approach also allowed us to explore online community mergers, a previously unstudied phenomenon, in an open-ended way that captured the perspectives of participants as accurately as possible.

The lead author conducted all of the ethnographic fieldwork, attempting to follow the guidelines for ethnography of virtual worlds in Boellstorff et al.'s *Ethnography and Virtual Worlds: A Handbook of Method* as closely as possible [5]. Prior to the study, he had almost ten years of experience as a raid team member and guild leader in *WoW*, which provided him with proficient knowledge and skill for participating in the field site [5]. Guilds were recruited to our study via the official *WoW* forums from the middle of June to the first week of July in 2017. Because we sought to compare how typical guild mergers unfolded, we selected for active, medium-core progression raid guilds that had some degree of prior organization, consisted of experienced players, and expressed interest in merging with, or acquiring, other guilds. All of the guilds involved were engaged with the most difficult raid tier at the time, the *Tomb of Sargeras*. The lead author performed the recruitment and was transparent about the purpose and duration of the intended study, as well as the fact that he was willing to participate as a member of the raid team if requested. As per our IRB-approved research protocol, both guild and player names have been changed throughout to maintain the anonymity of our research subjects.

The lead author spent roughly 6 months in the field as part of this study. From June 20 through December 9, 2017, he participated in weekly raid activities with the two merged groups, using two separate characters, for an average of 10 hours per week, per group. This duration was selected to coincide with the June 20, 2017 release of the *Tomb of Sargeras*, which ended with the release of a new, more difficult raid on November 28. Additionally, he observed out-of-game interactions and discussions on the guilds' forums and community *Discord* servers (a third-party text and voice communication platform that is popular with online game players [40]). As part of his participation in raiding activities, he used screen recording software to capture his experience. He reviewed these recordings the following day, reflected on social interactions, and took comprehensive field notes. Over the six-month period, he produced 120 pages of field notes, capturing these reflections. He also collected and transcribed roughly 8 hours of semi-structured, open-ended interviews with guild leaders and members in the periods after their mergers. Finally, we collected quantitative data on raid group attendance from the third-party application *WarcraftLogs*[2] for all of the guilds involved in the study over the period of field work.

Our analytic methods involved iteratively collecting and coding data to identify theoretical concepts that are "...integrated, consistent, plausible, close to the data..." and which we could

[2] https://www.warcraftlogs.com/

	Merger A (→Sun Bros$_{SD+F}$)		Merger B (→Divinity$_{D+M}$)	
Empirical Attributes	**Salty Dogs$_{SD}$**	**Fragment$_{F}$**	**Divinity$_{D}$**	**Mirage$_{M}$**
Founded	2008	2016	2017	2016
Progression raid nights per week	2	2	2	2
Highest raid experience	Heroic	Heroic	Mythic	Mythic
Total active and inactive members	88	82	44	?
Raid team size before merger	11	15	16	13
Progression in January 2017 raid tier	10/10	10/10	10/10	10/10
Theoretical Attributes				
Progression style	Farm	Farm	Achievement	Achievement
Goal orientation	Task / Social	Task / Social	Task	Task / Social
Social organization	Tree House	Tree House	Barracks	Barracks

Table 1. Features of the four pre-merger guilds in our case studies along dimensions frequently used by players to describe raid guilds in WoW, as well as theoretical attributes based on past guild research: progression style [3]; goal orientation [10]; and social organization [38]. Because of the way it was disbanded, information on Mirage$_{M}$'s pre-merger membership was unavailable through WoW's public application programming interface (API).

operationalize for further testing [17]. Our analysis proceeded by first inductively coding the field notes and interviews. Most of our attention was drawn to the social interactions between the players and members of the guilds. This process revealed distinct social norms within the groups in both merger cases. As we iteratively coded our data, we incorporated sensitizing codes derived from theory [7] on organizational culture, while continuing to generate new inductive codes as well. Ultimately, our codes influenced a series of memos produced by the lead author that were discussed by the entire team and iterated on repeatedly. The findings below are derived from these memos and address the empirical puzzle that emerged through the course of the fieldwork: why did one of the two mergers result in a thriving, successful community while the other resulted in complete collapse?

5 ETHNOGRAPHIC CASE STUDIES

The two mergers and four guilds included in this analysis are described in the rest of this section. Features of the groups are listed in Table 1 and focus on *empirical* attributes used by players to describe raid guilds in WoW as well as *theoretical* attributes of guilds we identified as salient in previous research. Within each merger, the guilds were similar to each other in most of these ways. As a result, we had no reason to suspect that either guild merger would result in success or failure when we entered the field. For clarity, we use the following subscript and color-coding system to denote the guilds involved in our two merger case studies:

Merger A: **Salty Dogs$_{SD}$ + Fragment$_{F}$ → Sun Bros$_{SD+F}$**
Merger B: **Divinity$_{D}$ + Mirage$_{M}$ → Divinity$_{D+M}$**

5.1 Merger A: Salty Dogs$_{SD}$ + Fragment$_{F}$ → Sun Bros$_{SD+F}$

Salty Dogs$_{SD}$ was a medium-core progression raid guild founded in 2008. They were highly organized and used a community website for announcements as well as a *Discord* server for voice-based communication outside of WoW. They used formal organizational processes, such as a detailed application process for new members, and a system for dividing loot, in which players bid for loot

from raid bosses using a point system maintained by the guild leader in an external spreadsheet. Players in Salty Dogs$_{SD}$ fell into one of two categories: raid team members, who participated in the weekly scheduled raid activities, and social members who participated only in the guild's community but not in its formal raids. Salty Dogs$_{SD}$ members shared strong social bonds outside of the game, often playing other games together like *Elder Scrolls Online* and casually socializing about their daily lives on their *Discord* server and guild website. By September, 2017, attrition had reduced Salty Dogs$_{SD}$ to only 11 raid team members.

Fragment$_{F}$ was established in September 2016 as a progression raid guild to take on the *Heroic* difficulty raid bosses in the game's most recent expansion. Only a few weeks after its formation, Fragment$_{F}$'s initial Guild Master and several of his friends abandoned the guild as a result of poor performance in initial raids. Melvin, a Fragment$_{F}$ member with no prior Guild Master experience, took on the leadership role, becoming the group's new Guild Master. Fragment$_{F}$ found relative success in defeating raid bosses over time while also managing attrition and membership churn. By September 2017, however, Fragment$_{F}$ membership had dwindled to 15 players.

To preserve the group's ability to continue raiding, Salty Dogs$_{SD}$' leaders sought a merger with another guild. After advertising in the official WoW forums for a merger, the Salty Dogs'$_{SD}$ leaders explored three potential guilds and eventually decided to merge with Fragment$_{F}$. Fragment$_{F}$'s leader Melvin took interest in the merger request post from Salty Dogs$_{SD}$ as a way of shoring up his own group's membership. After deciding to merge, the members of Salty Dogs$_{SD}$ left their guild and joined Fragment$_{F}$ as a group. Although the merger was understood by all parties as Salty Dogs$_{SD}$ joining Fragment$_{F}$, the new combined guild was renamed *Sun Bros* to recognize the change.

The merger resulted in a single, thriving community. Before merging, Salty Dogs$_{SD}$ and Fragment$_{F}$ had progressed through 8 out of the 9 raid bosses in the June, 2017 raid tier at the *Heroic* level difficulty at a relatively similar pace. Both had struggled to defeat the 9th raid boss, with Salty Dogs$_{SD}$ attempting and failing 34 times and Fragment$_{F}$ attempting and failing 18 times. As a newly-merged raid team, they failed 37 more times but also succeeded in defeating the 9th raid boss 6 times in the period following the merger. In the succeeding months, attendance remained consistently at or above 20 raid team members per raid (see the top panel of Figure 2).

The lead author joined Salty Dogs$_{SD}$ first as a casual member of the community on July 3, 2017, but later took a more active role as a member of its raid team in August. As a casual member, the lead author was permitted access to Salty Dogs' voice channel in *Discord* during their weekly, scheduled raids as well as the ability to observe one raid team member's perspective of the raid using the Twitch livestream service. The lead author was an active raid team member in the newly merged raid team from September 30 to December 9, 2017, and has remained as a casual member of the community following the study.

5.2 Merger B: Divinity$_{D}$ + Mirage$_{M}$ → Divinity$_{D+M}$

Divinity$_{D}$ was a medium-core progression raid guild founded in 2017 by the group's Guild Master, Aiden, a former hardcore raider. While Divinity$_{D}$ lacked a community website, it used *Discord* for voice-based communication outside of the game as well as for making announcements and some informal socializing. Divinity$_{D}$ had few formal processes for applying for membership, dividing loot, or making group decisions. The players in Divinity$_{D}$ were categorized as either raid team or social members, but there were very few of the latter. By mid-June, 2017, Divinity$_{D}$'s raid team included only 16 members (see the bottom panel of Figure 2) prompting the need for a merger. The lead author joined Divinity$_{D}$ as a member of its raid team on June 20, 2017, immediately after recruiting the guild into this study.

Mirage$_{M}$ was a hardcore progression raid guild established in late 2016 to challenge the most difficult *Mythic* level raid bosses. Playing up to 3 or 4 nights per week, Mirage$_{M}$ had been very

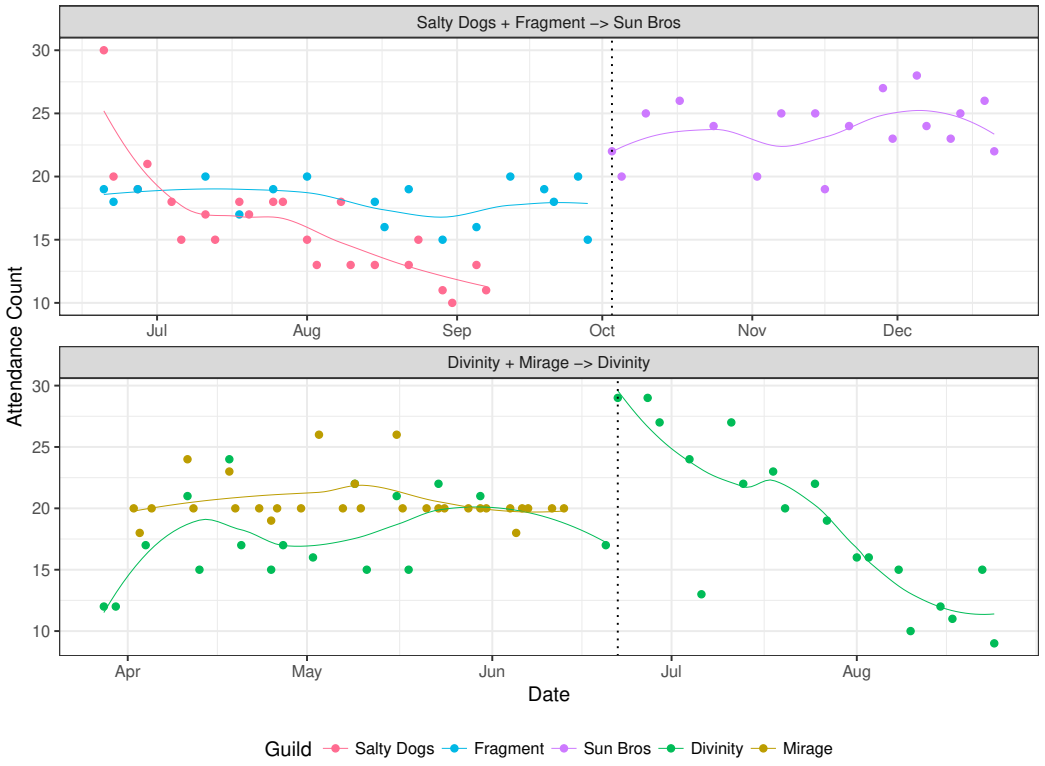


Fig. 2. Attendance in the observed guilds over six month periods around the mergers. Panels reflect the two pairs of merging guilds. Points reflect the observed attendance in each guild during raids. Colored lines reflect LOESS smoothed averages. Dashed, vertical lines reflect the point in time of the merger. Data was collected using the *Warcaft Logs* Interface Addon.

successful and defeated nearly all of the *Mythic* difficulty bosses in the January 2017 raid tier. However, after failing to defeat the last 3 *Mythic* raid bosses in the raid tier over several weeks in May, Mirage$_{M}$ underwent a schism in which its leaders kicked 13 of its 20 raid team members out of the guild, took all of the digital resources for themselves, and then disbanded the guild completely.

The 13 ejected members of Mirage$_{M}$ joined together under the temporary leadership of Oomkin who reached out to Divinity$_{D}$ in hopes that Aiden, Oomkin's in-game friend, would take in the group. Aiden agreed, and on June 22, the former Mirage$_{M}$ members were merged into Divinity$_{D}$. Divinity$_{D}$'s leadership viewed this move as a strategic decision to better manage player churn and to fill out its depleted roster with skilled players accustomed to raiding at a higher level of difficulty and intensity.

The post-merger Divinity$_{D+M}$ team progressed through the June, 2017 raid tier at a steady pace and had completed 8 out of the 9 *Heroic* difficulty raid bosses by July 24. Similar to Sun Bros$_{SD+F}$, the combined group encountered challenges on the final, most difficult raid boss, which they failed to defeat 63 times. This failure, combined with intra-group conflict, led to an attrition crisis (see the right half of the bottom panel in Figure 2). Over the following weeks, some Divinity$_{D+M}$ members fled to a new raid guild while others quit the game entirely. Recruiting efforts by Divinity$_{D+M}$'s

	Merger A (→Sun Bros$_{SD+F}$)		Merger B (→Divinity$_D$)	
Feature	**Salty Dogs$_{SD}$**	**Fragment$_F$**	**Divinity$_D$**	**Mirage$_M$**
Decision making	Democratic	Democratic	Top-down	Democratic
Leader-member distance	Low	Low	High	Medium
Delegation of leadership	Yes	Yes	No	Yes
Importance of community	High	High	Low	High

Table 2. Pre-existing differences and similarities of cultural assumptions that emerged from our coding process. The differences between the groups that were involved in Merger B proved consequential for the outcome of the merger.

leadership to balance the churn were unsuccessful. Divinity$_{D+M}$ membership dipped below the minimum of 10 players and the guild disbanded by the end of August.

6 FINDINGS

When we began our field work, we saw no *ex ante* reason to expect one merger to succeed and the other to fail. After all, the guilds in each merger were empirically and theoretically similar to each other, as shown in Table 1. Both the Mirage$_M$ and Divinity$_D$ members were competent WoW players with years of experience in having success in defeating the game's most difficult raid bosses. Some of the members from both groups knew each other before merging. Entering the field, we would have predicted the merger between Divinity$_D$ and Mirage$_M$ was more likely to be successful than the merger between Salty Dogs$_{SD}$ and Fragment$_F$ in large part because the players involved in the former seemed more committed to successful raiding. Given that our outcome surprised us, our analysis focused on understanding why Sun Bros$_{SD+F}$'s merger was more successful than Divinity$_{D+M}$'s.

Our analysis revealed two overarching findings. First, we found a greater cultural compatibility between the groups in Merger A (Sun Bros$_{SD+F}$). In the next subsection, we describe these differences in depth and explain how cultural incompatibilities led to sustained and repeated conflict in Merger B (Divinity$_{D+M}$). The second, more important finding focuses on the processes by which the groups managed cultural conflict. The teams involved in the successful merger used three strategies to effectively minimize and manage differences in organizational culture: (1) they took culture into account when planning their merger and strategically selected who to merge with; (2) they sought to intentionally socialize new members into their organizational culture through both formal and informal strategies; and (3) they cultivated collective solidarity by hosting informal socially-oriented group events in addition to their task-based group activities. Finally, we also found two additional factors that help to explain the outcomes of the mergers we observed: effective leadership and extensive tool use. We discuss these in a third subsection below.

6.1 Pre-existing Differences in Organizational Culture

Using Schein's concept of organizational culture as a framework for our analysis, we produced a set of key basic assumptions shared by members of the guilds in each case to understand what differences, if any, could have produced incompatibilities in these mergers. These core cultural features, listed in Table 2, were produced in our inductive coding process. Although both mergers involved bringing together groups with distinct assumptions, interactions between members revealed much more evidence of cultural compatibility between Salty Dogs$_{SD}$ and Fragment$_F$ than

between Divinity$_{D}$ and Mirage$_{M}$. Similarities and differences in culture reflected "common sense" assumptions about the nature of progression raid guilds and how they should be run.

Salty Dogs$_{SD}$ and Fragment$_{F}$ arrived at their merger with similar norms around group decision making. For example, when unsure about how to proceed during a scheduled raid night (such as choosing to end early or continue onto a new raid boss), raid team leaders in both guilds would issue the in-game command `/readycheck` to call a vote by everyone currently in the raid group. They would make decisions based on the majority's opinion. Additionally, both of these guilds demonstrated democratic strategies for decision making outside of the game. For example, Salty Dogs$_{SD}$' leaders held a group discussion on their *Discord* server with their guild members to decide which guild to merge with. Fragment$_{F}$ leaders would also hold verbal discussions about the guild's future in the guild's *Discord* server and would use Google Forms to survey its members. In both cases, the general assumption among guild members was that decisions would involve democratic participation of both leaders and members of the community.

Salty Dogs$_{SD}$ and Fragment$_{F}$ also handled the relationship between leaders and subordinates similarly. In both groups, subordinates who consistently attended progression raid nights and volunteered to help other members were rewarded with leadership roles. In Salty Dogs$_{SD}$, this was the case with Aysa and Jin who, despite being relatively new to the guild, were promoted to guild leaders in August as a result of helping out and consistently attending raid nights. In Fragment$_{F}$, Smalls was promoted to a leadership role after suggesting the organizational tool *WoW Audit* to the Guild Master and helping out other members during progression raid nights. Leaders of both guilds also formed social bonds with their subordinates by playing other games together outside of progression raid nights and having informal conversations over in-game chat channels or their guild *Discord* servers. Getting to know each other and developing friendships supported a strong sense of community in these guilds. Along all of these dimensions, approaches to government, leadership, decision-making, and community importance reflected similarities in assumptions between Salty Dogs$_{SD}$ and Fragment$_{F}$.

Although similar in many ways, the merger between Salty Dogs$_{SD}$ and Fragment$_{F}$ revealed some minor instances of resistance to change as well. For example, the different ways that Salty Dogs$_{SD}$ and Fragment$_{F}$ distributed loot during progression raids resulted in occasional moments of confusion for Salty Dogs$_{SD}$ members, who would need guidance on how to navigate Fragment$_{F}$'s loot system. Fragment$_{F}$'s leaders helped to resolve these relatively minor structural differences.

Although it was not clear before they merged, members of Divinity$_{D}$ and Mirage$_{M}$ held very different assumptions along the dimensions summarized in Table 2. These differences led to sustained conflict and frustration. For example, disagreements about the appropriate relationship between leaders and members became especially pointed. While nearly defeated by a difficult raid boss, the leaders of Divinity$_{D+M}$ called for the raid team to reset the boss encounter by giving up and letting the remaining living team members be killed. Several former Mirage$_{M}$ members (still new to Divinity$_{D+M}$) shouted back over the voice channel to ignore the leader's call and continue trying to defeat the boss. The remaining members successfully defeated the raid boss, but Divinity$_{D}$ leaders stopped the raid to call out the insubordination. Aiden, the Guild Master, added, "Last I checked, coaches make calls and their team still follows through, even if they don't agree." This conflict illustrates how Divinity$_{D}$'s assumptions about appropriate leader-subordinate interactions differed from Mirage$_{M}$'s.

Other conflicts arose over group decision-making. For example, during one night in which the raid team was struggling, once again, to defeat the same boss, Locke and Fizz, two Mirage$_{M}$ members, summarily quit. Locke left a departure message in the guild's *Discord* server lamenting the new guild's leadership style:

> I also feel like this guild is not built as most should be, there is a lot of awkward tension and voices are not being heard and even worse, opinions are not being openly voiced-makes for a very uncomfortable situation ...There is a lot of decisions that are made without a proper organized and democratic system which makes some people feel abandoned or useless for lack of better terms-I understand that may be how yall do things but regardless I no longer wish to be apart of it all.

In follow up interviews, Divinity$_{D+M}$'s Guild Master Aiden revealed that he felt that Locke's criticism was invalid. Aiden argued that making decisions as a group is inefficient, saying, "I've yet to see a guild that works democratically," and, "You don't want to slow the raid down by making it democratic." Other members of the Divinity$_{D}$ group expressed a preference for "militaristic" and "competitive" leadership styles, which they attributed to successful participation in hardcore guilds in the past. Aiden compared leading a raid guild to his experiences managing a business in the real world. Hector, another Divinity$_{D}$ leader, compared raid leading to his experience directing a music group at his university. The Divinity$_{D}$ members assumed that effective guild leadership involved investing leaders with power over their subordinates and authority to make decisions for the group.

On the other hand, Mirage$_{M}$ was used to progression raiding as both an organized team as well as a community of friends who all participated in group decision-making. In the moments before progression raid nights would begin, Mirage$_{M}$ members would often reminisce about their past experiences playing WoW together. After raids, Mirage$_{M}$ members would sometimes play other games together while socializing over the guild's *Discord* voice channels. During an interview, Oomkin, Mirage$_{M}$'s unofficial leader, stressed community-oriented values of feeling like a team in a raid guild and expressed an assumption that guild leaders should be "...making sure that everyone is equal and represented properly..." The misalignment between these expectations and the norms in Divinity$_{D+M}$ remained unresolved.

6.2 Strategies for Merging Organizational Cultures

Although pre-merger cultural (in)compatibility can partially explain Sun Bros$_{SD+F}$'s successful merger and Divinity$_{D+M}$'s failure, post-merger group actions shaped the merger outcomes at least as much. Sun Bros$_{SD+F}$ effectively employed several techniques to recognize, mitigate, and manage differences in organizational culture. These techniques included (1) strategic selection, (2) intentional socialization, and (3) cultivating solidarity to retain new and existing members.

6.2.1 Strategic Selection: Considering Culture When Deciding to Merge. Perhaps the most effective strategy that Sun Bros$_{SD+F}$ used to minimize cultural conflict was to pay careful attention to issues of organizational culture when selecting potential partners for their merger. This supports Schein's claim that leaders involved in successful mergers critically consider organizational culture in the process [30]. After publicly posting on WoW forums about the desire for a merger, Salty Dogs$_{SD}$' members engaged in a week-long back and forth process with three guilds: Next Week, Memory Loss, and Fragment$_{F}$. In each case, Salty Dogs$_{SD}$ engaged in raid activities with the group to assess potential incompatibilities and synergies. Once the Salty Dogs$_{SD}$ leaders finished these trial raids, they held a group-wide discussion on their *Discord* server to discuss the three potential merge partners. One prospective guild, Next Week, was reported to be behind Salty Dogs$_{SD}$ in their progression through the June, 2017 raid tier. Additionally, one of the Salty Dogs$_{SD}$ guild leaders, Aysa, reported a more social incompatibility with Next Week, saying, "...it's just that their attitude—they had a much more...a much more immature, much more crude, attitude..."

The second potential guild, Memory Loss, had progressed much further in the June, 2017 raid tier than Salty Dogs$_{SD}$ and was already restructuring into a hardcore raid guild to progress to *Mythic*

raid difficulty. Because succeeding at this higher level of difficulty requires intense discipline and commitment to the raid team, hardcore raid guilds typically have a more regimented organizational style [38]. Cheesy, a Salty Dogs$_{SD}$ member participating in the guild merger discussion, noted the incompatibilities they would face with both Next Week and Memory Loss, both of which were at different stages in their raid progression and difficulty level:

> We're not going half way back down [in progression to] like 5/9 [raid bosses] and we're not running with people who are like 5 bosses into Mythic either...Because we don't really belong to like either of those. There's a long way in-between those two worlds.

The third potential guild, Fragment$_{F}$, appeared to have the best potential synergy with Salty Dogs$_{SD}$. Outlaw, another Salty Dogs$_{SD}$ member who participated in the joint activity with Fragment$_{F}$, described the similarities in his report:

> It was almost like our alt-runs[3] on Friday. It was almost exactly like that. Where we were effin' around, having a good time. Messing with everybody, and then when it came to the fight, everybody calmed down.

The guild leader of Fragment$_{F}$, Melvin, joined the Salty Dogs$_{SD}$ *Discord* server to take questions from, and get to know, Salty Dogs$_{SD}$' members. Ultimately, the remaining Salty Dogs$_{SD}$ members agreed that the best guild to merge with would be the one that most closely fit their group culture. Another guild member summarized this:

> I was thinking we go to basically the closest to what we're doing. You know what I mean? So, there's not like culture shock for anyone or anything because I think if you want to keep what we have together, we want to make it as smooth as possible.

In a group interview after the merger, Fragment$_{F}$'s leaders explained that their members had also considered potential conflicts when deliberating on their decision to merge with Salty Dogs$_{SD}$. Although Melvin was apparently ready to take the Salty Dogs$_{SD}$ group with no questions asked, another leader in Fragment$_{F}$, Aramus, played devil's advocate to Melvin to address the potential for conflict. Aramus explained:

> We were only seeing the potential positive sides. I was just trying to bring up everything that could be negative about it, which isn't the best look, but I think it was important at the time.

Aramus was concerned that the members of Salty Dogs$_{SD}$ would require too much attention from Fragment$_{F}$'s leaders and reminded Melvin that Fragment$_{F}$ didn't "...super need this..." and that "...if it's going to be negative for our current members, it's not something that we should do". Melvin expressed appreciation for Aramus' perspective and agreed that they needed "...to look at our—at any issue we tackle—from all sides; it's not enough to say, well, 'let's just be blind to the negative consequences of something". In these ways, both Fragment$_{F}$'s and Salty Dogs$_{SD}$'s leaders considered and debated the potential merger carefully to assess the potential for cultural conflicts.

In contrast, the merger between Divinity$_{D}$ and the group of 13 raid team members from Mirage$_{M}$ was conducted in haste and without consideration of cultural compatibility between the merging groups. On June 16, the Mirage$_{M}$ group of 13 was removed suddenly from their old guild. In need of a new guild to participate in the upcoming June 20 release of a new raid tier, Oomkin organized the merger with Divinity$_{D}$'s leader Aiden. Aiden briefly considered merging with Oomkin's group of experienced raid team members as well as with two other possible guilds that happened to be on other game servers. Out of both the convenience of not having to pay to change servers, and because Aiden and Oomkin shared a social bond, Aiden quickly and unilaterally decided to take in the Mirage$_{M}$ group.

[3]This is another name for their optional *Normal* difficulty raid nights.

In Divinity$_{D+M}$, the process for merging was swift as well. There was no trial to allow the Mirage$_{M}$ members to take part in a Divinity$_{D}$ raid activity. Divinity$_{D}$'s leadership never even formally announced that a merger had occurred. Divinity$_{D}$'s members were not made aware of the merger until halfway through their June 22 raid activity when several of Divinity$_{D}$'s members privately messaged the guild leaders with confusion about the new players in their raid group. One of the Divinity$_{D}$ leaders, Hector, responded to the confusion over the voice channel: "Yes, we did merge. We took in a handful of people from Mirage$_{M}$ and now we're a bigger raid team. That's basically the long story short." This response underscored the extent to which Divinity$_{D+M}$'s leadership did not consider cultural compatibility in their merger.

6.2.2 Intentional Socialization: Planning Opportunities to Incorporate New Members. Sun Bros$_{SD+F}$ developed formal processes to ensure that members could learn or co-create a shared organizational culture. In Sun Bros$_{SD+F}$, socialization occurred in the context of learning rules and norms during the two weekly progression raid nights, as well as through access to explicit community rules. Additionally, the merged group held easier *Normal* raid nights to provide opportunities to integrating everyone in the community in a more social-oriented activity. Even the process of renaming the merged group to *Sun Bros* incorporated participatory brainstorming and voting by members of both sub-groups. Doing so was not costless; the Guild Master, Melvin, used a $10 game service to rename Fragment$_{F}$ to Sun Bros$_{SD+F}$.

On the other hand, Divinity$_{D+M}$ lacked any explicit community rules or mission statements. Instead, Divinity$_{D+M}$'s socialization process was highly individualized and occurred exclusively during its two weekly progression raid nights. This process primarily involved newcomers receiving negative feedback from Divinity$_{D}$ leaders for breaking unclear or unstated norms. For example, after the Mirage$_{M}$ members disobeyed Hector's order to reset, they were told never to ignore the orders of their raid team leader.

Divinity$_{D+M}$ did not hold intentional opportunities for socialization outside of progression raid nights in the way that Sun Bros$_{SD+F}$ did. When asked why Divinity$_{D+M}$ only played two nights a week together and never on easier levels of difficulty, its leader Aiden explained that they "...don't want to burn too many players out..." and that "...we also don't want to be chained to [raiding] like our actual job, which is the case for a lot of raiding guilds". Divinity$_{D+M}$'s leaders actively sought to minimize the amount of time spent playing together, assuming that progression raiding already took up enough of everyone's time. As a result, newcomers to Divinity$_{D+M}$'s group were given insufficient opportunities for socialization into its new organizational culture.

6.2.3 Cultivating Solidarity: Retaining New and Existing Members. Retention is a challenge for most new online communities and guilds [10, 12, 13]. Social computing scholars have found that the line between work and play is increasingly blurred, especially in the case of online games like WoW [25, 43]. As we have shown, progression raiding entails work-like features, such as subordination of team members to guild leadership, commitment to the team, and repetition in the sense that members come together every week to fight against the same raid bosses—frequently failing dozens of times. Play in the context of progression raiding is linked with work as a result of the work-like traits needed for raid teams to succeed. Commitment to a team in the face of failure and volunteer work can be difficult to sustain—especially in a setting intended for play.

The different ways in which guild leaders managed this balance between work and play contributed to the merger's divergent outcomes. Both merged guilds engaged in progression raids during their two scheduled raid activities. Members of the raid teams of both merged guilds were expected to show up on these progression raid nights and work by defeating the raid bosses that were defeated the week before, as well as repeatedly trying to defeat more difficult, undefeated raid

bosses. In both mergers, progression raid nights often ended in failure and left raid team members with no progress made and no new loot or achievements to justify the work put into playing.

While this feeling of unrewarding work is an inevitable aspect of progression raiding, Sun Bros$_{SD+F}$' leaders attempted to balance the work of progression raiding with more relaxed time for play. The guild leaders carved out a weekly, optional *Normal* difficulty raid activity on a night separate from the two more difficult progression raid activities. As a result, they cultivated solidarity by balancing their work with opportunities for play.

These *Normal* raid nights differed from the workplace of progression raiding, allowing newcomers to interact with each other in a relaxed environment that was less critical and performance contingent. Additionally, these activities were socially-oriented as non-raid team members from the guild, as well as friends and family of guild members, were encouraged to attend. In other words, *Normal* raid nights acted as a *third place*, a sociological term referring to a "home away from home", neutral space, like a bar or bowling alley that is separate from the "first" and "second" places of the home and workplace. Although video games are often referred to as *third places* [25, 31], task-oriented, action-based teams like progression raid guilds can transform the game into a workplace where members must follow attendance rules, meet performance standards, and respect a leader-subordinate hierarchy. The social bonds that the Sun Bros$_{SD+F}$ raid team members formed on *Normal* raid nights helped develop in-game friendships that built trust and motivated new members to commit to working with the guild's raid team.

Divinity$_{D+M}$ lacked comparable *third place* activities or environments. In fact, when Mirage$_{M}$ members tried to implement a *Normal* raid night during the first week of their merger, the idea was shelved by Divinity$_{D}$ leaders who weren't interested in playing more than the scheduled progression raid nights. Like colleagues in a workplace setting, the relationship among the raid team members in Divinity$_{D+M}$ was simply business. Without a *third place* oriented toward having fun and socializing, Divinity$_{D+M}$'s members failed to build social bonds that might have resulted in higher commitment to the guild in the face of struggles and misunderstandings.

6.3 Additional Factors: Effective Leadership and Technology Use

Two other factors shaped the outcome of the mergers, but do not fit neatly within the organizational culture framework: effective leadership and extensive use of communication tools. Because online activity is decentralized, virtual team leaders face the challenge of building trust among their members without direct interaction [21]. Scholars have shown that successful management of communication in online teams is a critical component of effective team leadership [26]. While our analysis suggested that organizational culture played the most important role in determining the success and failure of the mergers in our sample, extensive use of communication tools may have helped build trust between the leaders and members of the successful guild merger as well.

Both Divinity$_{D+M}$ and Sun Bros$_{SD+F}$ used the same communication tools to organize their teams, but Sun Bros$_{SD+F}$' more extensive use of their voice-based and asynchronous text-based channels in their *Discord* server made its leadership more effective in building trust among its members. The optional weekly *Normal* raid activity organized by Sun Bros$_{SD+F}$ allowed for an additional opportunity to build trust among its members by using their synchronous voice channel in Discord. Sun Bros$_{SD+F}$ text channels in their *Discord* server were also used extensively as an informal space for sharing personal stories, pictures, and banter. While Divinity$_{D+M}$ had a *Discord* server as well, its members mostly stuck to using their voice channel during its two weekly main progression raid activities and rarely during other days.

Although Sun Bros$_{SD+F}$ used communication tools to coordinate activities and manage guild work more effectively than Divinity$_{D+M}$, our strong sense was that this was a product (rather than a cause) of active guild social life and more effective guild leadership. Indeed, every technology

used by Sun Bros$_{SD+F}$ was freely available to Divinity$_{D+M}$ and known to its guild leadership. In other words, Sun Bros$_{SD+F}$' successful use of *Discord* and other organizational technologies was likely caused by differences in the way that they managed organizational culture and may have influenced but was not a direct cause of their merger's relative success. This is consistent with prior work showing that technology may influence organizational structures in some ways, but their influence is dependent on the cultures in which they're adopted and used [4, 42].

7 DISCUSSION

Applying Schein's concept of organizational culture, we find that cultural differences profoundly shaped the results of the mergers in our study. We find that the management of cultural compatibility and integration determined the outcomes for both Sun Bros$_{SD+F}$ and Divinity$_{D+M}$. Leadership and extensive use of communication tools also mattered, but these differences appear to have emerged as expressions of organizational culture as well. In contrast to prior social computing research that has considered socialization primarily as a problem that communities solve for individuals, our results underscore how merging existing groups entails a distinct consideration: collective cultural integration. This result is particularly important given recent work that suggests that many online communities grow through incorporating groups of users with shared experiences [34]. We find that organizations can manage mergers through tactics such as strategic selection, intentional socialization, and cultivating solidarity.

Our work builds heavily on a body of research in organization and management studies that has stressed the often overlooked importance of compatibility of cultures in organizational mergers [29, 30]. Schein [30] proposed that leaders considering organizational mergers must know their own group culture well enough to detect incompatibilities and synergies with the other group and engage in joint activities to understand each other's basic assumptions. Our work also builds on prior guild research in social computing: group mergers are contentious [38]; matching recruits with appropriately fitting groups plays an important role in newcomer retention [10]; forming social bonds in online gaming groups motivates member retention [13]; attrition and unresolved conflict between subgroups can lead to group collapse [8, 14, 28]; players form social experiences on platforms outside of the game [11]; and MMOGs can potentially serve as *third places* that encourage solidarity among peers [25, 31].

Extending these insights, our work illustrates how mergers elicit friction between different organizational cultures, resulting in some degree of instability. To understand the way that groups undergoing mergers attempt to stabilize, we borrow Ann Swidler's concepts of *settled* and *unsettled* cultures. Swidler conceived of culture as a toolkit and a "...repertoire from which actors select different pieces for constructing lines of action..." [33]. Swidler argues that culture has a role in both "...sustaining existing strategies of action" as well as developing new ones. She argues that *unsettled* cultures are those in which "...competing ways of organizing action are developed or contending for dominance". *Settled* cultures refers to those which are "...not in open competition with alternative models for organizing experience". According to Swidler, groups fighting for changes in cultural beliefs move from unsettled to settled times—a continuum from ideological movements to traditions to *common sense* assumptions about every day life.

In our synthesis of Swidler and Schein, we argue that groups and organizations with a shared history of experiences in accomplishing tasks have transitioned from an unsettled culture in which beliefs and norms were tested against each other to a settled culture in which beliefs about the basic operations become common sense. These surviving beliefs become the structure for the group's culture, providing stability and consistency. For example, a group formed by recruiting random strangers will not automatically come with clear, basic assumptions to which everyone agrees. Instead, through their history of competing beliefs for organizing action in their shared

activities, some beliefs will dominate and lay the foundation for establishing the basic underlying assumptions that structure the ways that the group reacts to its environment and accomplishes goals. In Schein's terms, this is how organizational culture comes to be. In Swidler's terms, mergers *unsettle* organizational cultures. In this sense, effective mergers minimize the degree to which unsettling destabilizes both groups while also minimizing the time and effort needed to return to a settled state. Effective mergers navigate the threats of instability while also reaping the benefits of an expanded group with a larger base of knowledge, skills, and other resources upon which to draw.

For Sun Bros$_{SD+F}$, assessing compatibility in organizational culture minimized the degree to which their two group cultures became unsettled by the merger. Strategically selecting merger partners also minimized the time and effort needed to return the group to a settled state. Although their post-merger raid performance was marked with considerably more failed attempts than their pre-merger performance, these failures did not result in instability and conflict. Failing, while frustrating, did not undermine their enjoyment or persistence. Instead, Sun Bros$_{SD+F}$' members developed new inside jokes and played other games together outside of WoW. The post-merger group eventually achieved a newly stable, settled state in which they completed new raids together.

Divinity$_D$ and Mirage$_M$'s merger failed precisely because the merged group did not achieve a settled state together. By neither assessing nor acting to overcome incompatibility in organizational cultures, both groups ignored potential conflicts that resulted in negative experiences during merged group activities. The shift from settled states of group stability as separate communities to an unsettled one, in which the belief systems of both groups were in competition, never resulted in a productive synthesis of the two communities. Instead, Divinity$_D$ leaders assumed that their belief systems for how progression raid guilds should be managed was *common sense* and that it was the Mirage$_M$ members' responsibility to assimilate. This expectation of assimilation was made more difficult in the absence of any formal socialization processes. Divinity$_D$ leaders ignored the possibility that the conflicts in *common sense* assumptions with Mirage$_M$ members would destabilize their organizational culture and result in the dissolution of both groups.

Both mergers resulted in an unsettled state immediately post-merger. However, instead of managing conflicts and minimizing the time and effort needed for the group to achieve a settled culture, Divinity$_D$'s leaders prolonged cultural conflicts with Mirage$_M$ members, resulting in negative group experiences. With no informal group activities oriented toward developing collective solidarity or social bonds, the members of the merged guild were not motivated to stay in the community. Because of the negative experiences that resulted, members from both Divinity$_{D+M}$ sub-groups either quit the game or found new guilds with which to continue raiding.

Our work has several limitations. For example, longer observation might reveal that the "successful" merger has other problems that we did not have the chance to watch develop. We chose a six-month period *ex ante* because it reflected a single cycle of the game. We also anticipated this period would be long enough to understand group culture because of the lead author's prior experience with WoW and understanding of raid guilds. A longer study of Divinity$_{D+M}$ would not have been possible given that the merger ended in collapse a few months after our study began. Also, while Sun Bros$_{SD+F}$ thrived, they scheduled a temporary hiatus in January 2018 in anticipation of the release of a new game expansion to WoW in August 2018.

Our study is also limited by its focus on organizational dimensions of culture within the guilds and their game world activities. Future work should compare organizational cultures of raid guilds across and between different regional settings, like the comparative study between Chinese and North American WoW players done by Nardi [25]. Other aspects of the cultural backgrounds of players are also likely to impact the organizing processes and cultural dynamics of raid guilds and other types of online groups.

7.1 Implications for Leadership, Design, and Research

Our study points to several paths forward for leadership, research, and design. First, virtual team leaders and online community managers can seek to better understand and accommodate the organizational culture of groups as they encounter potentially destabilizing intergroup conflicts. In that research has shown that these types of mergers are common on platforms like Reddit [34], unsettled and destabilized organizational cultures caused by informal and *ad hoc* mergers may also be common. We suggest that community leaders critically consider cultural incompatibilities between established members and groups of newcomers as their communities grow through these processes.

Second, our work illustrates the usefulness of Schein's idea of organizational culture for assessing group dynamics in social computing research. As we described in our findings, mergers of seemingly similar organized groups pursuing similar goals can have a variety of outcomes. The theoretical construct of organizational culture underscores how everyday processes and basic shared assumptions of these groups help explain differences in performance and survival. Prior approaches to culture in social computing have emphasized national, linguistic, or platform-specific (e.g., WoW vs. League of Legends vs. Reddit) cultural differences, but these can obscure the less immediately visible forms of collective understanding. Our work shows how comparative case studies can foreground these differences as well as some of their consequences.

Our results also underscore the importance of *third place* virtual activities as an effective strategy for newcomer socialization and for building solidarity to retain new and existing members. *Third places* provide a space in which hierarchical boundaries between group members and group leaders are blurred and cultural knowledge for the group can be tested in a setting that's more forgiving than their primary "work" activity. Socially oriented activities in virtual *third places*, like playing online games together or casually chatting over *Discord* servers, may result in social bonds and friendships that increase commitment to and retention in groups. Our work suggests that these are important considerations even in settings like games that are often described as *third places* themselves. When social computing involves the socialization of new individuals and collectives into focused, task-oriented, organized collaboration, we anticipate that online communities that incorporate *third place* activities will build stronger, deeper connections among participants than those that do not. In that *third place* activities span technical systems controlled by different interests (e.g., *Discord* servers are distinct from game infrastructure), supporting *third places* may pose a challenge to traditional socio-technical design processes.

8 CONCLUSION

We investigated the phenomenon of organizational mergers in online groups through a six-month, comparative ethnography of two pairs of raid guild mergers in *WoW*. One merger failed and led to the dissolution of both sub-groups, while the other merger resulted in a thriving community. We show that organizational culture shaped the outcomes in both cases, determining the ability of newcomers to join and contribute to both of the post-merger groups as well as the degree to which conflicts could be managed or resolved. The results expand existing notions of newcomer socialization in social computing by demonstrating how the process operates at a collective level. We also introduce Ann Swidler's concept of settled and unsettled organizational cultures to the social computing literature, synthesizing prior work from cultural sociology and management studies to capture the ways in which mergers can present both a threat and opportunity by destabilizing an existing cultural consensus of shared assumptions within a community. In order to sustain active, committed cooperation, the designers of social computing systems and leaders of communities must

navigate instability effectively. Our work demonstrates how mergers provide a unique opportunity to see how organizational culture plays a fundamental role in this process.

ACKNOWLEDGMENTS

The authors would thank the guilds that allowed us to play with them and participate in their community. We also also like to thank the Community Data Science Collective, Nathan TeBlunthuis, Sayamindu Dasgupta, Kirsten Foot, Nicole Crenshaw, Brian Keegan, for their support and/or feedback as part of this project. We are also thankful to our our CSCW reviewers and Associate Chairs for their thoughtful and detailed feedback, which greatly improved this paper. Lastly, CK wants to thank Hillary Rudolph for putting up with the months of late-night gaming and for being supportive during this project. Portions of the project were completed at the Helen R. Whitely Retreat Center at the University of Washington's Friday Harbor Laboratories. Financial support for this work came from the National Science Foundation (grants IIS-1617129 and IIS-1617468), Northwestern University, and the University of Washington.

REFERENCES

[1] Muhammad Aurangzeb Ahmad, Zoheb Borbora, Cuihua Shen, Jaideep Srivastava, and Dmitri Williams. 2011. Guild play in MMOGs: rethinking common group dynamics models. In *Proceedings of the Third International Conference on Social Informatics (SocInfo'11)*. Springer-Verlag, Berlin, Heidelberg, 145–152. http://dl.acm.org/citation.cfm?id=2050728.2050753

[2] William Sims Bainbridge. 2010. *The Warcraft civilization: social science in a virtual world.* MIT, Cambridge, Massachusetts.

[3] Jeffrey Bardzell, Jeffrey Nichols, Tyler Pace, and Shaowen Bardzell. 2012. Come meet me at ulduar: progression raiding in world of warcraft. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 603–612. https://doi.org/10.1145/2145204.2145296

[4] Stephen R. Barley. 1986. Technology as an occasion for structuring: evidence from observations of ct scanners and the social order of radiology departments. *Administrative Science Quarterly* 31, 1 (1986), 78–108. http://www.jstor.org/stable/2392767

[5] Tom Boellstorff, Bonnie Nardi, Celia Pearce, T. L. Taylor, and George E. Marcus. 2012. *Ethnography and virtual worlds: a handbook of method.* Princeton University Press, Princeton.

[6] Glenn R. Carroll and J. Richard Harrison. 2002. Come together? The organizational dynamics of post-merger cultural integration. *Simulation Modelling Practice and Theory* 10, 5 (Dec. 2002), 349–368. https://doi.org/10.1016/S1569-190X(02)00090-4

[7] Kathy Charmaz. 2015. *Constructing grounded theory: A practical guide through qualitative analysis* (2nd ed.). SAGE, Thousand Oaks, California.

[8] Mark Chen. 2011. *Leet noobs: the life and death of an expert player group in world of warcraft* (first printing edition ed.). Peter Lang Inc., International Academic Publishers, New York.

[9] Boreum Choi, Kira Alexander, Robert E. Kraut, and John M. Levine. 2010. Socialization tactics in Wikipedia and their effects. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, New York, 107–116. https://doi.org/10.1145/1718918.1718940

[10] Bo Choi, Robert Kraut, and Mark Fichman. 2008. Matching people and groups: Recruitment and selection in online games. (Jan. 2008). http://repository.cmu.edu/hcii/92

[11] Nicole Crenshaw and Bonnie Nardi. 2016. "It was more than just the game, it was the community": social affordances in online games. In *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE Computer Society, Koloa, Hawaii, 3781–3790. https://doi.org/10.1109/HICSS.2016.471 00006.

[12] Laura Dabbish, Robert Kraut, and Jordan Patton. 2012. Communication and commitment in an online game team. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 879–888. https://doi.org/10.1145/2207676.2208529

[13] Thomas Debeauvais, Bonnie Nardi, Diane J. Schiano, Nicolas Ducheneaut, and Nicholas Yee. 2011. If you build it they might stay: retention mechanisms in world of warcraft. In *Proceedings of the 6th International Conference on Foundations of Digital Games (FDG '11)*. ACM, New York, NY, USA, 180–187. https://doi.org/10.1145/2159365.2159390

[14] Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J. Moore. 2006. "Alone together?": exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 407–416. https://doi.org/10.1145/1124772.1124834

[15] Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J. Moore. 2007. The life and death of online gaming communities: a look at guilds in world of warcraft. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 839–848. https://doi.org/10.1145/1240624.1240750

[16] Toni Eagar, Jenine Beekhuyzen, and John Campbell. 2015. When online communities collide: Boundary identity construction and spanning. In *Proceeings of the Thirty Sixth International Conference on Information Systems (ICIS '15)*. Association for Information Systems, Atlanta, Georgia, 10. http://aisel.aisnet.org/icis2015/proceedings/HumanBehaviorIS/2/

[17] Barney G. Glaser and Anselm L. Strauss. 1967. *The discovery of grounded theory: Strategies for qualitative research*. Aldine Publishing Company, Chicago, Illinois.

[18] J. Richard Harrison and Glenn R. Carroll. 2006. *Culture and demography in organizations*. Princeton University Press, Princeton, N.J. 00125.

[19] Colin Jerolmack and Shamus Khan. 2014. Talk is cheap: ethnography and the attitudinal fallacy. *Sociological Methods & Research* 43, 2 (May 2014), 178–209. https://doi.org/10.1177/0049124114523396

[20] Katherine C Kellogg. 2011. *Challenging operations: medical reform and resistance in surgery*. University of Chicago Press, Chicago. OCLC: 780487085.

[21] M. Audrey Korsgaard, Arnold Picot, Rolf T. Wigand, Isabelle M. Welpe, and Jakob J. Assmann. 2010. Cooperation, coordination, and trust in virtual teams: Insights from virtual games. In *Online Worlds: Convergence of the Real and the Virtual*, William Sims Bainbridge (Ed.). Springer, London, UK, 253–264. https://doi.org/10.1007/978-1-84882-825-4_20

[22] Robert E Kraut, Paul Resnick, and Sara Kiesler. 2012. *Building successful online communities: evidence-based social design*. MIT Press, Cambridge, MA.

[23] Natalia Levina and Manuel Arriaga. 2014. Distinction and status production on user-generated content platforms: using bourdieu's theory of cultural production to understand social dynamics in online fields. *Information Systems Research* 25, 3 (Sept. 2014), 468–488. https://doi.org/10.1287/isre.2014.0535

[24] Bonnie Nardi and Justin Harris. 2006. Strangers and friends: collaborative play in world of warcraft. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW '06)*. ACM, New York, NY, USA, 149–158. https://doi.org/10.1145/1180875.1180898

[25] Bonnie A. Nardi. 2009. *My life as a night elf priest: an anthropological account of World of Warcraft*. University of Michigan, Ann Arbor, Michigan. http://hdl.handle.net/2027/spo.8008655.0001.001

[26] Arnold Picot, Jakob J Assmann, M Audrey Korsgaard, Isabell M Welpe, Julia V Gallenkamp, and Rolf T Wigand. 2009. A multi-level view of the antecedents and consequences of trust in virtual leaders. In *Proceedings of the 15th Americas Conference on Information Systems (AMCIS 2009)*. Association for Information Systems, Atlanta, Georgia, 1–11.

[27] N. Poor and M. Skoric. 2016. Play together, stay together? Community cohesion and stability in an mmo. In *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS '16)*. IEEE Computer Society, Koloa, Hawaii, 2266–2275. https://doi.org/10.1109/HICSS.2016.283

[28] Nathaniel Poor and Marko M. Skoric. 2014. Death of a guild, birth of a network: online community ties within and beyond code. *Games and Culture* 9, 3 (May 2014), 182–202. https://doi.org/10.1177/1555412014537401

[29] Edgar H. Schein. 1996. Culture: the missing concept in organization studies. *Administrative Science Quarterly* 41, 2 (June 1996), 229. https://doi.org/10.2307/2393715 02078.

[30] Edgar H. Schein. 2004. *Organizational culture and leadership* (3rd ed. ed.). Jossey-Bass, San Francisco. 34527.

[31] Steinkuehler Constance A. and Williams Dmitri. 2006. Where everybody knows your (screen) name: online games as "third places". *Journal of Computer-Mediated Communication* 11, 4 (Oct. 2006), 885–909. https://doi.org/10.1111/j.1083-6101.2006.00300.x

[32] Zahirah Suhaimi-Broder, David Kirschner, and J. Patrick Williams. 2014. Structural roles in massively multiplayer online games: a case study of guild and raid leaders in World of Warcraft. In *Symbolic Interaction and New Social Media*, Mark D. Johns, Shing-Ling S. Chen, and Laura A. Terlip (Eds.). Studies in Symbolic Interaction, Vol. 43. Emerald Group Publishing Limited, Bingley, UK, 121–142. https://www.emeraldinsight.com/doi/abs/10.1108/S0163-239620140000043016

[33] Ann Swidler. 1986. Culture in action: symbols and strategies. *American Sociological Review* 51, 2 (1986), 273–286. https://doi.org/10.2307/2095521

[34] Chenhao Tan. 2018. Tracing community genealogy: how new communities emerge from the old. In *Proceedings of the Twelfth International Conference on Web and Social Media (ICWSM '18)*. AAAI, Palo Alto, California, 395–404. https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17811

[35] Deborah J. Terry, Craig J. Carey, and Victor J. Callan. 2001. Employee adjustment to an organizational merger: an intergroup perspective. *Personality and Social Psychology Bulletin* 27, 3 (March 2001), 267–280. https://doi.org/10.1177/0146167201273001

[36] John Van Maanen and Schein, Edgar H. 1979. Toward a theory of organizational socialization. In *Research in organizational behavior*, Barry M. Staw (Ed.). JAI Press, Greenwich, CT, 209–264.

[37] Roberto A. Weber and Colin F. Camerer. 2003. Cultural conflict and merger failure: an experimental approach. *Management Science* 49, 4 (2003), 400–415. https://doi.org/10.1287/mnsc.49.4.400.14430 00619.

[38] Dmitri Williams, Nicolas Ducheneaut, Li Xiong, Yuanyuan Zhang, Nick Yee, and Eric Nickell. 2006. From tree house to barracks the social life of guilds in world of warcraft. *Games and Culture* 1, 4 (Oct. 2006), 338–361. https://doi.org/10.1177/1555412006292616

[39] Dmitri Williams and Marko Skoric. 2005. Internet fantasy violence: a test of aggression in an online game. *Communication Monographs* 72, 2 (June 2005), 217–233. https://doi.org/10.1080/03637750500111781

[40] Luke Winkie. 2017. Inside discord, the chat app that's changing how gamers communicate. (June 2017). https://www.rollingstone.com/glixel/news/inside-discord-the-chat-app-thats-changing-how-gamers-communicate-w489086

[41] Tessa Wong. 2016. China's love affair with world of warcraft. *BBC News* (May 2016). http://www.bbc.com/news/world-asia-china-36209913

[42] Jiang Yang, Meredith Ringel Morris, Jaime Teevan, Lada A. Adamic, and Mark S. Ackerman. 2011. Culture matters: a survey study of social Q&A behavior. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*. AAAI Press, Palo Alto, California, 409–416. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2755

[43] Nick Yee. 2006. The labor of fun: how video games blur the boundaries of work and play. *Games and Culture* 1, 1 (Jan. 2006), 68–71. https://doi.org/10.1177/1555412005281819

Received April 2018; revised July 2018; accepted September 2018

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.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
CHI 2018, April 21–26, 2018, Montréal, QC, Canada.
ACM ISBN 978-1-4503-5620-6/18/04.
https://doi.org/10.1145/3173574.3173929

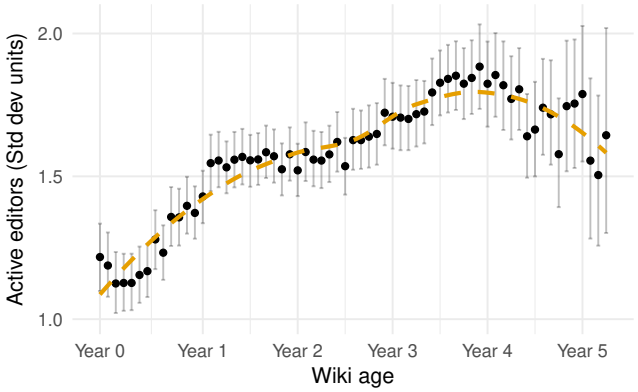


Figure 1. Mean of the number of editors with at least 5 edits per month in standard deviation units for wikis in our sample. The dashed lines represent the results of a LOESS regression. The error bars represent bootstrap 95% confidence intervals. This replicates Figure 2 in RAD.

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.

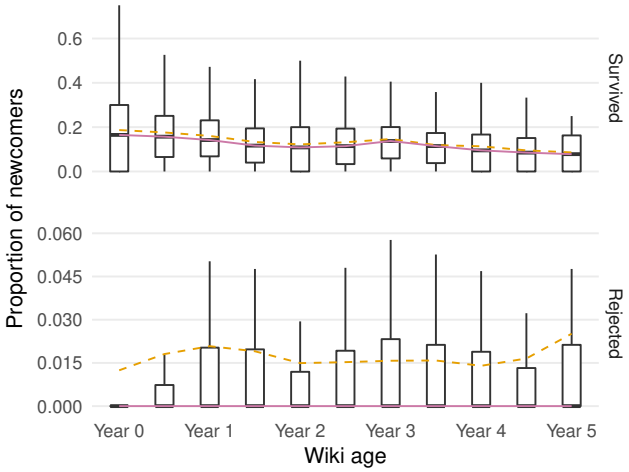


Figure 2. Newcomer survival and rejection over time. The orange dashed line shows the mean and the pink solid line shows the median. Years with data from at least 20 wikis are shown.

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 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 4$^{\text{th}}$ 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*.

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

	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 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.

[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.

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.

ACKNOWLEDGMENTS

The authors would like to thank our anonymous reviewers and associate chairs at CHI for their thoughtful and detailed feedback. We would also like to thank Wikia for providing public access to data from their wikis and other members of the Community Data Science Collective for sharing the software, data, and research infrastructure necessary to complete this work. We thank Jonathan Morgan for providing help in planning this study and Amanda TeBlunthuis, Kaylea Champion, Wm Salt Hale, and Sayamindu Dasgupta for feedback on drafts of our manuscript. This project was completed using the Hyak high performance computing cluster at the University of Washington. Support for this work came from the National Science Foundation (grants IIS-1617129, IIS-1617468, and GRFP-2016220885), Northwestern University, the Center for Advanced Study in the Behavioral Sciences at Stanford University, and the University of Washington.

ACCESS TO DATA

A replication dataset has been placed in the Harvard Dataverse archive and is available at the following URL: https://doi.org/10.7910/DVN/SG3LP1.

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

Part 4: Teaching Materials

This section contains sample syllabuses and Course and Teacher Evaluation Council (CTEC) reports for every course I have taught since arriving at Northwestern. I have also included the Final Report I created as part of my Searle Center for Teaching Junior Faculty Fellowship.

Learning Informal and Ad-hoc Organization: Tactics of Teamwork and Problem-solving in Small Groups

Aaron Shaw[1]

Northwestern University
Searle Center for Teaching
Junior Faculty Fellowship
Final Report

Communication Studies 378, Online Communities and Crowds

Abstract

How can college instruction support students' cultivation of the teamwork and collaboration skills necessary to understand, critically analyze, and participate in informal and ad-hoc organizations? In this project, I describe the design, implementation, and results of a pilot learning module on informal teamwork and problem-solving implemented during Fall, 2013, as part of a Searle Center Faculty Fellowship. The module, which takes place over several weeks of my Communication Studies 378 course, engages students in the synthesis of central research themes, the production of theoretical knowledge and practical skills, task-oriented small group work, and multiple forms of assessment and critical evaluation.

Keywords: teams, collaboration, organization, group work, pedagogy

1 School of Communication, Department of Communication Studies, aaronshaw@northwestern.edu

Introduction

Teamwork and collaboration constitute a core focal area of research and pedagogy within Communication, Management, Innovation, Design, Organizational research, and Human-Computer Interaction (HCI). Across these different research traditions, the empirical findings are unequivocal: great collaborations are carefully and intelligently made, not born by accident or luck. Effective teams and collectives emerge as the product of institutions and practices that support motivation, participation, collective action, leadership, and accountability (Hackman 1987, 2011; Olson 1965; Ostrom 1990). With the rise of the Internet and digitally-networked communication technologies, the organization of collaboration has become increasingly virtual, distributed, and informal (Benkler 2006; Benkler, Shaw, and Hill 2014; Contractor, et al., 2012; Contractor 2013, Shirky 2008). At the same time, these radical technological shifts have not altered the "fundamentals" of effective teams and collaborative systems.

How can college instruction support students' knowledge and cultivation of the teamwork and collaboration skills necessary to participate in such networked and virtual organizations? Previous research has pointed to the importance of small group work and problem-oriented projects as pedagogical strategies to increase student engagement, build commitment and identification within the classroom, facilitate active learning, and promote critical problem-solving skills (Johnson, Johnson, and Smith 1991; Nilson 2010; Smith 2000). These findings extend a core tenet of constructivist pedagogy (Vygotsky 1980), that teaching which engages students in the active creation and enactment of knowledge promotes higher order creative thinking and learning (Anderson, Krathwohl, and Bloom 2001). As many programs and instructors move their teaching online under the assumption that students, as "digital natives," will intuitively adapt to virtual pedagogy, prior literature on digital skills and learning has shown that inequalities in Internet use skills and experiences will likely prevent many from participating fully or equally in these online modes of education (Hargittai 2010). As a result, classroom learning experiences still present an ideal opportunity to enable students of different abilities to blend abstract theories of teamwork with collaborative practice (Bogaard et al. 2005; Cooper and Robinson 2000).

In the rest of this document, I describe the design, implementation, and results of a pilot learning module on informal teamwork and problem solving that I implemented over several weeks in my Communication Studies 378 course, Online Communities and Crowds, at Northwestern University during Fall, 2013. The module, created as part of a Searle Center Faculty Fellowship, synthesizes several central themes from the course, engages students in the production of theoretical knowledge and practical skills, and involves multiple forms of assessment and critical evaluation. This module also builds directly on my research studying virtual and informal organizations, online participation, peer production, and crowdsourcing (Benkler et al. 2014; Kittur et al. 2013; Shaw 2012; Shaw and Hill 2014; Shaw et al. 2014). In collaboration with my faculty mentor, Noshir Contractor, the Searle Fellowship has also led to the development of a novel line of research to further investigate the factors that determine the effectiveness and success of informal and ad-hoc teams.

Course and Learning Module Description

The learning module developed for this project occupies a position of central importance within the design of Communication Studies 378, "Online Communities and Crowds." Collaboration and teamwork increasingly happen in informal, distributed organizations connected by means of digitally networked communication technologies. The most innovative, informal organizations are massive "peer production" communities and other online "crowds" that pursue the production of shared information resources (Benkler 2006; Benkler et al. 2014; Contractor et al. 2012; Contractor 2013; Shirky 2008). Distributed groups collaborate over the Internet to write free encyclopedias (Wikipedia), launch social movements (Avaaz, MoveOn), create software (GNU/Linux), share music and films (the Pirate Bay), raise vast sums of money (Kickstarter), and conduct advanced scientific research (Galaxy Zoo). Why do some of these large-scale collaborative groups succeed? How can the sources of online communities' and crowds' successes be harnessed and reproduced? What can be learned from their shortcomings?

Communication Studies 378, Online Communities and Crowds[2], engages students in an interdisciplinary introduction to the study of distributed collaboration, collective action, and organization. The course places a particular emphasis on understanding how and why some systems are so wildly effective at mobilizing and organizing people in ways that seemed impossible a few decades ago (no monetary incentives, loose organizations with no boundaries, radically distributed participation). The syllabus draws on current research in the social sciences (Communication, Sociology, Economics, Political Science, Organizational research), Human Computer Interaction, Law, and social theory. The class utilizes a mixed format, combining short lectures, seminar-style discussion, small group work, team-based classroom activities, as well as individual and collaborative assignments.

In Fall, 2013, Online Communities and Crowds enrolled 20 advanced undergraduates (mainly Juniors and Seniors) from Communication Studies, Radio TV and Film, Sociology, Political Science, and other majors. In future years, the course will expand to accommodate as many as 50 students, including some from Computer Science and related disciplines. The course met for 80 minutes twice a week in a small seminar room where the seating arrangement could be easily organized depending on the format of the day's session (rows for lecture, round or horse-shoe for discussion, small clusters for group work or team activities). The students tended to have substantial first-hand knowledge of the subject matter through their familiarity with social media and related online systems for information sharing such as Facebook, Reddit, Wikipedia, and Twitter. At the same time, few (if any) had encountered any of the formal theoretical, empirical, or applied perspectives on informal collaboration and teamwork.

The Learning Module

The learning module on informal teamwork and small group collaboration takes place over three class sessions roughly in the middle of the course. In Session one, the instructor introduces the main themes and concerns of the module; students are randomly assigned to four-person teams;

2 Details of the course, including assignments, readings, and the syllabus, are available at http://aaronshaw.org/occ.

teams complete the first team-based challenge in class; and a short lecture with interactive discussion prompts synthesizes core conclusions. Session two then builds on this experience, presenting the same teams with a novel in-class challenge activity and randomly assigning half of the teams to use more formal organizational structure during the completion of the puzzle. Teams are also asked to debrief and conduct a collective formative self-assessment immediately after the completion of the challenge. Then, in preparation for the third Session, students are asked to individually write up more elaborate responses to a series of self-assessment questions along with analytical critiques of their team's performance in light of the theories and findings they have encountered in specific prior readings. During the third class session, the instructor facilitates a full-class debrief and synthesis, combining small group conversations, seminar discussion, a brief lecture presenting aggregated data and commentary on the different groups' performance in the challenges. Finally, the course TA comments on students' written evaluations, assessing the work along several dimensions. Below, I elaborate on the module learning objectives, teaching, methodology and the substance of the three sessions in greater detail.

Learning Objectives

The problem-based module advances several learning objectives. These objectives span Bloom's widely-used taxonomy (Anderson, Krathwohl, and Bloom 2001), with a particular focus on engaging students in “higher-order” cognitive and learning processes.

By the conclusion of the module, students will recognize problems, contexts, and approaches in which to apply theories and techniques of team-based collaboration and informal organization. They will also describe how teams and informal collaborations work with reference to appropriate theories and empirical examples from the course readings. Individually, as well as in groups, the students will explain how particular teamwork strategies and informal organizational techniques may support effective collaboration or not. In written, spoken, and enacted work, they describe and compare he performance of teams with reference to relevant theories and examples.

Together with peers in their assigned teams, students will also apply their theoretical knowledge of online collaboration and collective action to their own activities. Students will engage in iterative rounds of application and reflection across the different parts of the module. At multiple points, students engage in critical analysis of the groups' performance in light of the different theoretical and empirical examples engaged in the class. Students also compare their team's performance on multiple challenges, allowing them to organize their thinking in more formal, theoretically-informed ways.

Students engage in multiple rounds of self- and group-assessment, evaluating their team's performance along several dimensions. Following the completion of the module, the students evaluate the module itself, and are asked to assess how effectively (or not) the different classroom exercises and small group challenges engage them in deeper, higher-order forms of learning that go beyond merely consuming the reading materials and lectures. In the final session of the module, students will participate in a collective debriefing session in which the instructor leads a collaborative group assessment of the module as an experimental learning experience.

In the context of the teamwork challenges themselves, students engage in original, difficult problem-solving that combines design, construction, collaborative planning and execution. The work is highly creative, presenting students with completely novel situation in which they are asked to apply materials from the course as a small group.

In these ways, the module thus advances intellectual, practical, personal, and social dimensions of student learning and engagement. Throughout the three sessions, the module requires students to connect theoretical knowledge with practical, collaborative behavior in the context of concrete problem solving and creative work. The small groups develop camaraderie and collective identity through their experiences together, ideally leading to increases in classroom participation as well as extracurricular interaction. The module also demonstrates the broader utility and salience of the course material to a wide range of contexts and concerns. This facilitates students' ability to recognize and create additional opportunities in which to use skills and knowledge drawn from the course material.

Learning Module Teaching Methods & Assessment

The learning module employs multiple teaching methods and techniques within an iterative framework. The methods include: small group work, problem-based learning, active learning, lecture, open-ended discussion, and analytical writing. Various kinds of assessment also occur throughout the learning module, such as formative and summative assessment, all coming from students themselves, their peers (group members) as well as the instructor and course Teaching Assistant. Below, I discuss the three sessions in greater detail as well as the accompanying assignments, follow-up assessment and evaluation procedures.

Session 1- Introduction and Team Problem-Solving Challenge #1

In the first session, the instructor uses a micro-lecture (5 minutes) to briefly introduce some of the key concepts related to teamwork, informal organization, and collaboration. The micro-lecture incorporates a single slide, listing these concepts alongside a single, symbolic image. In preparation for the session, students were assigned readings on effective teamwork by Richard Hackman (e.g., 1987, 2011). These readings extend and elaborate other materials from prior weeks in the course, in which we have analyzed how informal, virtual organizations such as Free Software communities engage in the creation of complex information goods (Fogel 2005, von Krogh et al. 2012). Students have also been told in advance that they will be performing a problem-solving challenge, but the teaching team has not provided them with any further details about the content of the challenge.

Following the micro-lecture, the instructor announces that students have been randomly assigned to four-person teams, in which they will now be asked to complete a problem-solving challenge. The task, in this case, is a team-building and collaboration exercise called “the marshmallow challenge” originally developed and popularized by Tom Wujec (2010)[3]. As the student teams cluster together around tables, the Teaching Assistant distributes the following materials to each team: one marshmallow, 20 strands of uncooked spaghetti, one yard of masking tape, and one

3 The marshmallow challenge is described in detail online at: http://marshmallowchallenge.com.

yard of twine. The instructor then presents the instructions and rules for the challenge: With these materials (and only these materials), each team's objective is to build a free-standing structure that supports the marshmallow as high off the table as possible within 18 minutes. Teams are not explicitly in competition, but they are not allowed to coordinate with each other (i.e. each team must work independently). No Internet access is allowed during the challenge. Upon receiving the instructions (which are projected on a slide and left up throughout the challenge), students are given the opportunity to ask clarifying questions and then a timer is started to signal the beginning of the challenge.

While the challenge takes place, the instructor and teaching assistant take notes on the teams' different approaches to the challenge as they relate to questions of teamwork and informal collaboration. One of the key lessons of the marshmallow challenge is the value of iterative experimentation for creative and innovative work, and the students generally tend to spend too large a proportion of the time (more than half) planning and theorizing about their marshmallow structures before they ever attempt to "get the marshmallow off the ground." After the time is up, the teaching team records the heights of the different structures and the instructor facilitates a formative assessment of team performance. First, students are asked to individually respond to a series of open-ended questions about their group's work. Then, volunteers share their thoughts with the class in a larger discussion. The instructor follows this with another micro-lecture, explaining some of the background and relevant findings from previous research with the marshmallow challenge. The core concepts included are the importance of team leadership, effective work processes, and the utilization of domain-specific skills. At the conclusion of the lecture, the instructor eliciting students to explain how and why those findings do or do not apply to their own experience with the challenge.

Session 2 – Elaboration and Team Problem-Solving Challenge #2

The second session of the module extends the first by asking students to solve a creative, collaborative puzzle in their teams[4]. Prior to the session, the instructor assigns half the teams to have formal leaders. Upon arriving in the class, each team is provided with a unique set of instructions and an identical complex puzzle. For those teams assigned to have leaders, the instructions state that a particular team member (who is identified by name) will be the leader of the group. Teams are once again instructed that they cannot confer on their work and that while their completion of the puzzle is not, in itself, being graded, they will be timed. They are also provided with written assessment prompts that they are individually asked to writeup and hand in before the following class session. The prompts incorporate self and group assessment as well as several analytical questions, asking students to compare their group's performance across the two tasks and to apply their knowledge of teamwork and organizations to explain their group's work.

After receiving these instructions, the instructor then asks the groups to spread out in study areas near the classroom and to use the remainder of the class period to complete the puzzle. As groups

4 All puzzles were adapted from the archives of the MIT Mystery Hunt (see: http://web.mit.edu/puzzle/www/, for examples and details). These puzzles are designed to be extremely difficult and solved by large teams collaborating intensively. Each puzzle has one correct answer and, in almost all cases, Internet-based searches do little to help find solutions.

complete the puzzle, the instructor checks their answer (there is only one correct solution) records their time and, in the cases where time remained before the end of class, asks the teams to begin working on their written assignments (which will be graded)[5].

Session 3 – Self and Peer-Assessment, Group Debrief and Synthesis

Session 3 consists of a collective debrief and discussion at the beginning of the subsequent class following Session 2. Students also hand in their written self- and group-assessments, as well as their responses to the analysis questions asking them to apply the theoretical readings to evaluate their teams' performance. In the debrief and discussion session, the instructor reviews the design of the different teamwork exercises and presents representations of the aggregated group outcomes. The entire class then discusses the results as well as what conclusions can (and cannot) be drawn based on their experience. Finally, the instructor elicits formative assessments and suggestions about the exercise itself.

Follow-up Assessment and Evaluation

The teaching team grades the written analyses submitted by the students, focusing on the clarity of the arguments and the degree to which students demonstrate relevant, theoretically-informed knowledge of teamwork and organization in their discussion of their team's performance. In conjunction with the two in-class exercises and the class discussion, these written assignments provide an opportunity for the teaching team to evaluate the students' understanding of the material and their progress towards the learning objectives.

Learning Module Evaluation

The module itself was evaluated in a formative manner during the written assignment and during the debrief discussion in the third class session. Students were asked to provide open-ended feedback, to describe the purpose of the two challenges and to comment on its effectiveness in synthesizing theoretical concerns from the course. These open-ended questions were subsequently complemented by closed-ended survey questions in the mid-quarter course evaluation (conducted in the class meeting immediately following the third session), in which they were asked to rate the effectiveness of the in-class teamwork challenges alongside the other class assignments, lectures, and readings.

Based on these multiple stages of assessment and evaluation, the module produced several positive learning outcomes. Overall, the module generated progress towards many of the learning objectives. The quality of discussion and engagement in the classroom increased noticeably across the three sessions. The written assignments students submitted reflected a high level of overall comprehension, effective comparisons across theoretical texts and group experiences, as well as creative interpretive analysis. In addition, nearly half of the students gave the in-class exercises a rating of 5 out of 6 in terms of their overall effectiveness (the rest of the ratings were distributed equally around this modal response). In open-ended course evaluation questions, several students singled out the in-class exercises as a particularly effective way of facilitating

5 The questions are available at: http://aaronshaw.org/occ/assignments/team.html.

participation and collaboration.

At the same time, much room for improvement remains in nearly every facet of the module. First, in future iterations, I aim to refine the design of the second collaboration challenge as well as the method by which I assign team structure to some of the groups. The puzzles used in the second challenge did not prove difficult enough to elicit much variation in the groups' experiences. The puzzles also did not require a deeply collaborative process – making it easy for teams to simply divide the work very quickly and work without much further coordination. In addition, the assignment of leaders needed to be more prominent in order to trigger more substantial behavior changes among the groups involved. Harder, more deeply team-oriented puzzles will also increase the salience of teamwork and group organization during the process of completing the task.

In addition, I plan to incorporate more peer assessment into the module. team and self-assessment components (Lejk and Wyvill 2001). In the current version of the module, students focused primarily on providing formative assessments their own performance as well as the performance of their team as a whole. Implemented effectively, I believe that peer assessment could help students engage each other around the particular roles and responsibilities they performed within the group. It would also complement the summative forms of evaluation such as the written assignment and formal performance measures from each team.

Conclusions

The process of developing this learning module has allowed me to reflect and improve upon my teaching in several ways. First, and foremost, the extended opportunity to design and refine the module in light of empirical research on pedagogy and learning has enhanced the quality of the module. Second, the feedback from Searle Center staff has helped me to focus and clarify the module content around concrete learning objectives. Finally, the process of developing the module in conversation with my faculty mentor, Noshir Contractor, has also led the two of us to develop a novel line of collaborative research analyzing the performance of informal and ad-hoc teams. While this work is still in its early stages, we have begun preliminary field studies and anticipate that the work will generate external grant funding and peer reviewed publications, create research opportunities for graduate students and undergraduates, and make its way into our teaching as part of courses such as Communication Studies 378 that involve study of teams, collaboration, and leadership.

Based on these experiences, I expect that I will continue to refine many additional aspects of my teaching in order to enhance student learning outcomes. Using the pilot version of the team-based learning module described here as a template, I feel that I can approach my course design in a much more systematic way. The early indications I have received suggest that this process will enhance my students' learning as well as their overall enjoyment of class. Subsequent iterations of the team-based learning module, Communication Studies 378, and my other undergraduate and graduate courses will all benefit from this experience.

References

Anderson, Lorin W., David R. Krathwohl, and Benjamin Samuel Bloom. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Allyn & Bacon.

Benkler, Yochai. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.

Benkler, Yochai, Aaron Shaw, and Benjamin Mako Hill. 2014. "Peer Production: A Modality of Collective Intelligence." in *The Handbook of Collective Intelligence*, edited by Michael Bernstein and Thomas Malone. Cambridge, MA: MIT Press.

Bogaard, Amy, Sabine C. Carey, Gwilym Dodd, Ian D. Repath, and Richard Whitaker. 2005. "Small Group Teaching: Perceptions and Problems." *Politics* 25(2):116–25.

Contractor, Noshir. 2013. "Some Assembly Required: Leveraging Web Science to Understand and Enable Team Assembly." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371(1987):20120385.

Contractor, Noshir S., Leslie A. DeChurch, Jay Carson, Dorothy R. Carter, and Brian Keegan. 2012. "The Topology of Collective Leadership." *The Leadership Quarterly* 23(6):994–1011.

Cooper, James L., and Pamela Robinson. 2000. "Getting Started: Informal Small-Group Strategies in Large Classes." *New Directions for Teaching and Learning* 2000(81):17–24.

Fogel, Karl. 2005. *Producing Open Source Software: How to Run a Successful Free Software Project*. Sebastopol, CA: O'Reilly Media.

Hackman, J. Richard. 1987. "The Design of Work Teams." *Handbook of organizational behavior* 315:342.

Hackman, J. Richard. 2011. *Collaborative Intelligence: Using Teams to Solve Hard Problems*. San Francisco, CA: Berrett-Koehler Publishers.

Hargittai, Eszter. 2010. "Digital Na(t)ives: Variation in Internet Skills and Uses among Members of the 'Net Generation.'" *Sociological Inquiry* 80(1):92–113.

Johnson, David W., Roger T. Johnson, and Karl A. Smith. 1991. *Active Learning: Cooperation in the College Classroom*. Edina, MN: Interaction Book Co.

Kittur, Aniket et al. 2013. "The Future of Crowd Work." Pp. 1301–18 in *Proceedings of the 2013 conference on Computer supported cooperative work*, *CSCW '13*. New York, NY, USA: ACM. Retrieved April 4, 2013 (http://doi.acm.org/10.1145/2441776.2441923).

von Krogh, Georg, Stefan Haefliger, Sebastian Spaeth, and Martin Wallin. 2012. "Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development."

Management Information Systems Quarterly 36(2):649–76.

Lejk, Mark, and Michael Wyvill. 2001. "Peer Assessment of Contributions to a Group Project: A Comparison of Holistic and Category-Based Approaches." *Assessment & Evaluation in Higher Education* 26(1):61–72.

Nilson, Linda B. 2010. *Teaching at Its Best: A Research-Based Resource for College Instructors*. John Wiley & Sons.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York, NY: Cambridge University Press.

Shaw, Aaron. 2012. "Centralized and Decentralized Gatekeeping in an Open Online Collective." *Politics & Society* 40(3):349–88.

Shaw, Aaron et al. 2014. "Computer Supported Collective Action." *Interactions* 21(2):74–77.

Shaw, Aaron, and Benjamin Mako Hill. 2014. "Laboratories of Oligarchy? How the Iron Law Extends to Peer Production." *Journal of Communication* n/a–n/a.

Shirky, Clay. 2008. *Here Comes Everybody: The Power of Organizing Without Organizations*. New York: Penguin Press.

Smith, Karl A. 2000. "Going Deeper: Formal Small-Group Learning in Large Classes." *New Directions for Teaching and Learning* 2000(81):25–46.

Vygotsky, L. S. 1980. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wujec, Tom. 2010. "The Marshmallow Challenge." *The Marshmallow Challenge*. Retrieved May 22, 2014 (http://marshmallowchallenge.com/Welcome.html).

Online Communities & Crowds

Creating Effective Online Organization and Collaboration

Northwestern University
Comm Studies 378
Mon/Wed 2:00-3:20pm; Fri 2:00-3:00pm
Frances Searle 2-107
Fall, 2016

Aaron Shaw (Professor)
Office location: Frances Searle Building, room 2-142
Office hours: Mon. 10am-12pm, Wed. 3:30-5:30pm, and by appointment

Jeremy Foote (Teaching Assistant)
Office location: Frances Searle Building, room 2-419
Office hours: Mon. 1-2pm, Fri. 3-4pm, and by appointment

Course materials:

> **Description**
>
> **Objectives, requirements, assignments & grading**
>
> **Policies**
>
> **Schedule**

This Web site contains all the most up-to-date information about this class. The schedule page contains the day-to-day details you'll probably need most often.

Please email Aaron or Jeremy with any questions and make sure to include `[occ]` in the subject line. This will help us keep track of class-related messages and respond more promptly.

Course Description

Many of the most innovative and ground-breaking organizations today are online "peer production" communities and crowds. Distributed groups collaborate over the Internet to write free encyclopedias (Wikipedia), launch social movements (Avaaz, Change.org), create software (GNU/Linux), share music and films (the Pirate Bay), develop innovative products (CHDK), and conduct advanced scientific research (Zooniverse).

When and why do these efforts succeed? What motivates participants to join, contribute, and sustain these communities? How can online communities' and crowds' successes be harnessed and reproduced? What can be learned from their shortcomings?

This course presents an intensive and interdisciplinary introduction to the study of online communities and crowds, with a particular emphasis on how and why some of these systems are so wildly effective at mobilizing and organizing people in ways that seem to have been impossible a few decades ago. Throughout the quarter, we will analyze these and other conceptual

puzzles, studying many different communities in the process. Readings and assignments will draw on current research in the social sciences (Communication, Sociology, Economics, Psychology), Human Computer Interaction, law, and social theory.

Please send questions, corrections, or concerns to Aaron Shaw and make sure to include `[occ]` in the subject line.

This site licensed under a cc-by-sa 4.0 license and created using Markdown and Kevin Burke's markdown.css stylesheet.

Learning Objectives, requirements, & assignments

Online Communities & Crowds
Northwestern University
Comm Studies 378
Fall, 2016

Learning Objectives

My primary objectives for all students taking this course are that you:

> - Analyze, discuss, understand, and critically engage with central concepts, examples, and issues relevant to online communities and crowds.
> - Experience and apply practical approaches to online participation, collaboration, innovation, teamwork, and collective action.
> - Reflect on (assess) and iteratively improve upon your own work and that of your peers in light of the concerns analyzed in class.
> - Identify and elaborate original problems and approaches to online collaboration, extending and expanding on the material presented in class.

In order to accomplish these objectives, the course has a syllabus with different kinds of assignments and exercises requiring you to approach online communities and crowds from multiple perspectives.

Organization

I will use the course material to advance an overarching argument: *the participants, architecture, culture, and environment of online communities and crowds determine their character and impact.*

To support this argument, the course covers four overlapping dimensions of online communities and crowds:

> - Foundations: key underlying concepts, histories, and narratives.
> - Dynamics: issues communities face and solve every day.
> - Cases: key examples of online communities and crowds.
> - Challenges: deep, hard problems nobody has solved (yet!).

We start with a heavier emphasis on foundations and dynamics then move towards the discussion of more cases and challenges. I try to weave my central arugment throughout and expect that you will develop the knowledge and skills necessary to understand, challenge, extend, and improve upon my claims.

Format and materials

The course sessions will follow a mixed format, including some lectures, discussions, group projects, and visits from guest speakers.

Up-to-date information about readings, scheduling, assignments and more can be found on the schedule page of the course website.

Most of the readings are or will be available online. In general, all assignments will be finalized and posted to the course website at least one

week before they are due. In other words, make sure to check the course website consistently as things may change. I will plan to communicate changes to you in class and (when appropriate) via email. If you want to complete assignments more than a week in advance, please discuss your plans with me so that I can confirm the assignments with you.

Please note: We have multiple readings from one book. *Building Successful Online Communities: Evidence-Based Social Design* by Robert Kraut and Paul Resnick. You may want to purchase a copy. It is available in a variety of formats (hardcover, paperback, epub, etc.) from a variety of fine booksellers. Pre-publication versions of all the chapters are also available from one author's website.

Requirements

There are no prerequisites for the class. You are expected to perform all the usual activities: prepare for, attend, and participate actively in all class sessions; complete assignments in a timely manner; and abide by all applicable norms and policies. If you have questions about any of this, please get in touch.

Assignments and evaluation

You will be asked to submit the following, all of which will be evaluated as part of your grade for the course (percentage in parentheses):

- Wikipedia project and other labs (25%)
- Critical community analysis #1 (15%)
- Peer assessments (5%)
- Final project (Critical community analysis #2) proposal (5%)
- Final project (critical community analysis #2) (20%)

Brief descriptions of these assignments are provided in the following section. Additional details including evaluation rubrics and deadlines will be provided via the course website and in class.

An additional 30% of your final grade will evaluate your participation and overall engagement with the course material. As part of this, you are required to meet with Jeremy Foote, the Teaching Assistant for the course, *at least once before the final project proposal is due*. My strong recommendation is that you meet with Jeremy sooner rather than later -- both because he'll have helpful feedback and because his schedule will undoubtedly fill up closer to the deadline (when it will be too late for you to take full advantage of his advice anyway). **Jeremy's office hours will be Wednesdays 12-2pm (or by appoitment) in Frances Searle 2-419.** You can contact him at jdfoote@u.northwestern.edu.

In general, you should plan to submit all your assignments via Canvas. You can always also email your assignments to the teaching team if you encounter technical difficulties or other surprises.

Several of these assignments (and especially the Wikipedia project) either obligate or allow you to work in teams. Whenever you collaborate with one or more of your peers on an assignment, you are required to include the names of your collaborators on the assignment at the time you hand it in. In some cases, all individuals will be asked to submit their own copy of the assignment even when they have collaborators.

Teaching team email protocol

When emailing either member of the teaching team about anything related to the course, please put "[occ]" in the subject line. We both get a lot of email and this will help us stay on top of it.

Friday Labs

Additional details will be provided about the Friday lab sessions in class. Please note that several of the activities in the lab sessions will require (or at least be much easier with) a networked computing device with a decent size keyboard (e.g., laptop or tablet). If you would like our help securing access to such a device during the lab session, please contact the teaching team.

Brief descriptions of assignments

Wikipedia project and labs includes the output of all the lab assignments (several to-be completed in groups) as well as (individually) written responses to reflection questions about your experiences.

Critical community analysis #1 takes the form of a short (about 1000 words) written analysis of a particular dynamic or problem in a particular community (to-be-assigned by the teaching team) in which you describe the community, how the dynamic/problem works, and propose improvements or solutions based on material discussed in class. You will have about 72 hours (3 days) to complete this assignment.

Peer assessments contains several different peer assessment assignments that you will complete throughout the quarter. In general, these involve providing written assessments of and feedback on the work of your peers.

Final Project, a.k.a. Critical Community Analysis #2 consists of an original research paper of **no more than 2000 words** that also includes relevant citations and sources (which do not need to be counted in the word total). The objective is for you to analyze a particular puzzle or dynamic in the context of a specific community that you observe, participate in, and/or care about. You should aim to (a) describe the community; (b) explain how and why the community illustrates or engages with the puzzle or dynamic; and (c) explain how you would apply insights from the class to improve the community's approach to the puzzle or dynamic. You will have several opportunities throughout the quarter to develop your project ideas, discuss them, and receive feedback on them. This includes the **Final Project Proposal**, which will consist of a short (about 500 words) written description of the community you plan to analyze, some discussion of why you care about it, as well as the reasons you believe it would present an important site for critical analysis and engagement with the themes of the course. **Final projects are due via Canvas by 5pm Central Time on Wednesday, December 7, 2016**.

Course Policies

Online Communities & Crowds
Northwestern University
Comm Studies 378
Fall, 2016

The policies below outline fundamental groundrules for the class. If you have questions or concerns about any of them, please contact me.

Attendance

Attendance in class is expected of all participants. If you need to miss class for any reason, please contact me ahead of time (email is best). Multiple unexplained absences will likely result in a lower grade or (in extreme circumstances) a failing grade. In the event of an absence, you are responsible for obtaining class notes, handouts, assignments, etc.

In-class Use of Networked Digital Devices with Screens

I ask that you refrain from using your laptops, tablets, phones, and pretty much any networked or digital device with a screen during the class meetings that are not labs. If you have a compelling reason or need to use such a device during class time, please contact me.

The goal of this policy is to help you stay focused and avoid distractions for yourself and your peers in the classroom. This turns out to be difficult in the presence of powerful computing devices with brightly glowing screens and fast connections to the Internet. For more on the rationale behind this policy, please read Clay Shirky's thoughtful discussion of his approach to this issue.

Deadlines & Extensions

Deadlines for all assignments are non-negotiable. Any assignment you submit after the deadline may be penalized or returned (un-graded) at my discretion. That said, I understand that emergencies happen. Unanticipated obstacles arise. If you absolutely cannot meet a deadline, please contact me. Keep in mind that I will be more sympathetic to your situation if you do this before the deadline instead of afterwards.

Email protocol

I receive too much email. Anytime you contact me or Jeremy about any questions or concerns you may have related to the course, please include `[occ]` in the subject line of your message. This will help both of us filter and locate class-related messages so that we can stay on top of them more effectively. If, for some reason, I do not respond to your message within 48 hours, please do not take it personally and feel free to re-send the message.

Confidentiality of Peers' Work

Throughout the course, you will receive, read, and comment on classmates' work. These assignments are for class use only. You may not share them

with anybody outside of class without explicit written permission from the work's author pertaining to the specific piece.

Confidentiality of In-Class Discussions

It is essential to the success of this class that all participants feel comfortable sharing their questions, thoughts, ideas, confusions, and experiences during class discussions. Therefore, you may not create any audio or video recordings during class time nor share verbatim comments with those not in class nor are you allowed to share using other methods — e.g., social media — comments linked to people's identities unless you get appropriate permissions ahead of time. If you want to share general impressions or specifics of in-class discussions with those not in class, ask for permission first.

Academic Integrity

You are responsible for reading and abiding by the Northwestern University Principles Regarding Academic Integrity.

Make sure to document all of your work and acknowledge the ideas and work of others. Possible sanctions, as per the university guidelines, include reduced or failing grade, a defined period of probation or suspension, exclusion from the university and notation on the official record. You must not, in any way, misrepresent your work or be party to another student's failure to maintain academic integrity. Do not ever copy other people's words without quotation marks (do not do this even if you are "just" taking notes) and always use proper citation. Do not ever refer to other people's work without attribution. DO NOT cheat, plagiarize or disregard the University Principles Regarding Academic Integrity in any way, it is NOT worth it! When in doubt, err on the side of giving more credit to the original source rather than less. Please feel free to ask me questions about this and any related matters.

Accommodations

Any student requesting accommodations related to a disability or other condition is required to register with AccessibleNU (847-467-5530) and provide professors with an accommodation notification from AccessibleNU, preferably within the first two weeks of class. All information will remain confidential. For more information, visit the AccessibleNU site.

Sexual Harassment

All participants in this class are bound by the Northwestern University sexual harassment policy. Please note that the core of the policy states, "no member of the Northwestern community may sexually harass any other member of the community." I take this policy very seriously. Please review the policy and speak to me if you have any questions or concerns.

Incompletes

I strongly discourage incompletes. If you wish to request an incomplete for the class, you should explain your reasons to me **in writing** at least two weeks before the end of the quarter and propose a timeline for completing the course. Please note that I may decline your request.

Grades

If you have questions or concerns about any grade you receive, please contact me. If you wish to dispute a grade, please provide a written explanation of your rationale along with any supporting documentation you think is necessary.

Course Schedule

Online Communities & Crowds
Northwestern University
Comm Studies 378
Fall, 2016

Please note: This document will evolve a great deal over the quarter. You should check it frequently to ensure that you have the most current information about assignments, readings, etc.! All reading assignments will be updated no less than 1 week prior to their due dates. If you wish to read further ahead than that, please get in touch and I'll help you out.

Send any questions or course-related emails to Aaron or Jeremy and please include `[occ]` in the subject line. This will help us keep track of them and respond more promptly.

Sept 21 — Introductions, the Internet, and Sharing (Slides)

Sept 23 — Lab: Internet Protocols and/as Public Goods (Slides)

- (Optional) Watch and/or read Steve Jobs' 2005 Stanford Commencement Address.
- (Optional) Stephenson,N. (1996). Mother earth mother board. Wired Magazine, 4(12).

Sept 26 — Cyberculture! Early Online Communities (Slides and Soundtrack)

- Hafner, K. (1997). The epic saga of the WELL.
- Turner, F. (2005). Where the counterculture met the new economy: The WELL and the origins of virtual community.
- (Optional) Stanford 2011 symposium: From Counterculture to Cyberculture: the Legacy of the Whole Earth Catalog.
- (Optional) The BBS Documentary.
- (*very* Optional) Kevin Driscoll's USC dissertation, Hobbyist Internetworking.

Sept 28 — Motivation (Slides)

- Kraut & Resnick. *Building Successful Online Communities,* "Encouraging contributions to online communities" (Chapter 2) (pdf from Kraut's Web site).

Sept 30 — Lab: Wikipedia Task #1 (Slides)

In class: begin Wikipedia Task #1
(here's a direct link to our course dashboard on Wikipedia).

- Wales, J. (2007). TED Talk: How a ragtag band created Wikipedia.

Reminder: please bring a computing device with an Internet connection (e.g., a laptop or a tablet) to class!

Oct 3 — Commons, Public Goods, and Collective Action (Slides)

- Ostrom, E. (2004). Understanding Collective Action (pdf via Canvas). International Food Policy Research Institute (IFPRI) Focus Brief 11(2).
- Olson, M. (1965). The logic of collective action: public goods and the theory of groups. pp. 5-22, 33-43, 46-52 (pdf via Canvas).

Oct 5 — Organizational culture & institutions (Slides)

- Reagle, J. (2010), *Good Faith Collaboration*, Chapter 1 and Chapter 3, Section 5 (Wikipedia Policy, Guidelines, and the Five Pillars) and Section 6 (Neutral Point of View and Good Faith: An Example).
- Schein, E. H. (2010), *Organizational Culture and Leadership*, "How Culture Emerges in New Groups" (Chapter 12). (Please note: I have included Chapter 13 in the pdf. I will talk a little bit about this chapter in class, but consider it optional reading).
- Examine Wikipedia Project Pages (links below). For each of these, spend about 10 minutes poking around to get a feel for what the project is about, who is there, and how you could characterize the culture of each "organization."

    - WikiProject Military History (flyer).
    - WikiWomen's Collaborative (flyer).
    - Wikipedia Guild of Copyeditors (flyer).
    - Wikipedia Department of Fun.
    - WikiProject Chicago.

Oct 7 — Lab: Wikipedia Task #2: Big Shoulders Edit-a-thon

occ Wikipedia course dashboard.

We will meet in Fiske Hall, Room 111 (the Knight Lab)

Oct 10 — Code and Other Laws: Community Rules & Governance (Slides)

- Kraut & Resnick. *Building Successful Online Communities*, "Regulating behavior in online communities" (Chapter 4) (pdf from Kraut's Web site).
- Example codes of conduct:

    - Ubuntu code of conduct
    - GNOME code of conduct
    - Geek Feminism code of conduct
    - Valerie Aurora's essay, HOWTO design a code of conduct for your community, Ada Initiative (For context, you should know that Aurora is one of the authors of the the Geek Feminism code.)

Oct 12 — Teamwork and Leadership (Slides)

**Today's class will meet in Frances Searle Room 1-180 **

- Hackman, R. (2004). What makes for a great team?.
- Hackman, R. (1998). Why teams don't work.

Oct 14 — Lab: Wikipedia Task #3

occ Wikipedia course dashboard.

Oct 17 — n00bs! Encouraging Newcomers in Online Communities (Slides)

- Kraut & Resnick. *Building Successful Online Communities*, "Dealing with newcomers" (Chapter 5) (pdf from Kraut's Web site).
- Morgan et al. (2013). Tea and sympathy: Crafting positive new user experiences on Wikipedia.

Critical Community Analysis #1 posted today

Oct 19 — Free/Libre and Open Source Software (FLOSS)(Slides)

- Fogel, K. (2009). Producing Open Source Software. Read
  - The Table of Contents (Yes, I'm serious).
  - Why Write This Book?.
  - Introduction.
- Fogel, Karl. (2005). What is Free Software
- Stallman, R. (1984). The GNU manifesto.
- Stallman, R. (1989). The GNU general public license, version 1.
- Debian (1997). Debian social contract, version 1.0.
- (Optional) Preston-Werner. (2011). Open Source Everything.

Oct 21 — Lab: Wikipedia Task #4

occ Wikipedia course dashboard.

Oct 24 — Peer Production (Slides)

Critical Community Analysis #1 Due Today via Canvas

- Benkler, Y. (2006). Wealth of Networks. pp. 29-34 & Ch. 3 (all).
- Shirky, C. (2008). Here Comes Everybody. Chapter 10.

Oct 26 — Innovation Communities(Slides)

- von Hippel, E. (2012). Video: The Paradigm Shift from Producer to User Innovation.
- (Optional) von Hippel, E. (2012). Video: Users Working Together in Communities are Powerful Innovators.
- (Optional) von Hippel, E. (2005). Democratizing innovation. Chapters 1 & 2.

Oct 28 — Lab: Wikipedia Task #5

occ Wikipedia course dashboard.

Oct 31 — Hackers and Pirates(Slides)

Peer assessment of Critical Community Analysis #1 Due Today before class

- Rosenbaum. (1971). Secrets of the Little Blue Box (reprinted in Slate in 2011 — there's also a very large PDF scan of the original

Esquire Magazine article which includes the arguably NSFW and offensive magazine cover image).
- Larkin. (2004). Degraded images, distorted sounds: Nigerian video and the infrastructure of piracy.

Nov 2 — Movements & Civic Media (Slides)

- Monroy-Hernández et al. (2012). Narcotweets.
- (Optional) Watch Monroy-Hernández and Metaxas Berkman Center luncheon series talk on Narcotweets on Narcotweets.
- (Optional) Bennett & Segerberg. (2012). The Logic of Connective Action.

Nov 4 — Lab: Microtasks

- Microtasks Lab Assignment

Nov 7 — Crowd Work & Citizen Science (Slides)

Microtasks Lab reflection due before class via Canvas

- Fortson et al. (2012). Galaxy Zoo: Morphological Classification & Citizen Science. (Focus on Sections 1-2, 6-7, skim the rest).
- von Ahn, Luis. TEDx Talk: Massive-scale online collaboration.

Nov 9 — Human Computation (with a human face)? (Slides)

- Future of Crowd Work paper (Focus on Introduction, Section on "Future of Crowd Workers", and Conclusion).
- Zittrain, J. (2009) Talk: Minds for Sale.

Nov 11 — Lab: Wikipedia Task #6

occ Wikipedia course dashboard.

Nov 14 — Remix Culture

Guest lecture: Sneha Narayan

- Lessig, L. Ted talk on remix and creativity.
- Taylor, A. (2016). The People's Platform. Chapter 5. (via Canvas).

Nov 16 — Tough Problems #1: Starting New Communities

Final project proposals due today

- Kraut & Resnick, *Building Successful Online Communities*, Chapter 6.

Nov 18 — Lab: Final Project Workshop Session

Due today:

- Wikipedia Reflection Essay (See Wikipedia Task #6)
- Final project proposal peer assessment

Nov 21 — No class

Nov 23 — No class

Nov 28 — Tough Problems #2: Building Inclusive Communities (Slides)

> - Buechley, L. and Eisenberg, M. (2008). The LilyPad arduino: Toward wearable engineering for everyone.
> - Buechley, L. and Hill, B. M. (2010). LilyPad in the wild: How hardware's long tail is supporting new engineering and design communities.

Nov 30 — Closing Argument: The Future of the Internet (Slides)

December 7 — Final Projects due by 5pm Central Time via Canvas

> - Final projects due via Canvas.

Introduction to Graduate Research in TSB & MTS

Northwestern University
MTS 501 Wednesday 9:00-11:50am
Frances Searle Building, 2-378
Fall, 2016

Professor Aaron Shaw
Frances Searle Building, room 2-142
M 10am-12pm, W 3:30-5:30pm or by appointment

Contents

Description & Overview

Learning Objectives

Assignments & Evaluation

Course Schedule

Resources

Policies

Description & Overview

The goal of this seminar is to introduce first-year students in the MTS and TSB Ph.D. programs to the fields they cover and how to be a productive and responsible graduate student. (↑)

Learning Objectives

Throughout the quarter, participants in the seminar will achieve the following goals:

- Understand, interpret, discuss, and assess academic career strategies and development.
- Develop, articulate, and apply effective academic career development strategies in the context of your own research program.
- Reflect on (assess) and iteratively improve upon your own work and that of your peers in light of the concerns analyzed in class. (↑)

We will do this through a combination of readings, writings, activities, in-class discussions, and guest visits from MTS and TSB faculty.

Assignments & Evaluation

In addition to the reading and active participation in seminar discussions, there will be frequent written assignments for this course. Details of the assignments are provided below (or will be soon).

You should complete assigned readings before beginning any written assignments for each week. **All written assignments are due 24 hours**

before class (a.k.a. 9am on Tuesdays each week) via Canvas. Meeting these deadlines each week is important because you will also frequently be assigned a colleague's work to comment on and assess. Submitting your own work on time ensures that you can give and receive the full benefits of peer feedback.

In general, your written assignments for this class can take a somewhat informal, but nevertheless professional tone. When submitting assignments via Canvas, I recommend either uploading files formatted as pdfs or using the WYSIWYG text editor provided in the Canvas discussion threads (I'll show you what I mean). It is your responsibility to ensure that you complete assignments fully and on time. Failure to do so reflects poorly on you and undermines the objectives of the class.

You will be evaluated on your engagement with the material and participation in our class sessions (40%); your timely and effective completion of assignments throughout the quarter, including peer assessment assignments (40%); as well as the final project (20%). (↑)

Final Project

Link to final project page

The final project for the course will be a ~3000 word written document consisting of two parts: (1) a strategic plan for your graduate school career and (2) a research planning document for a project you plan to conduct. I will provide additional details of the Final project requirements and expectations well before the end of the quarter. Final projects are due via Canvas at 12pm (Central time) on December 7, 2015. (↑)

A Note About the Course Schedule

You should expect this syllabus to be a dynamic document and you will notice that there are a few places left blank or marked "To Be Determined." Although the core expectations for this class are fixed, the details of readings and assignments will shift. As a result, there are three important things to keep in mind:

> - Details on this syllabus will change, but I will not change readings or assignments less than one week before they are due. If I don't fill in a "To Be Determined" one week before it's due, it is dropped. If you plan to read more than one week ahead, contact me first.
> - Keep an eye out for emails and announcements I send through Canvas re: updates to the syllabus.
> - If you anticipate scheduling conflicts or require accommodations of any kind, please do your best to notify me well in advance.

You can always give me feedback and suggestions related to what works and what doesn't about the course. I will explicitly solicit your input a few times during the quarter, but **be bold** and feel free to submit your feedback to me at any time in any format. In the past, I have made substantive changes to courses on-the-fly in response to student feedback. I expect to do the same here. (↑)

Course Schedule

September 21 — Introductions

- (In class reading) McPhee, Sarah. 2015. Habits of a highly effective graduate student. Ellison Center, University of Washington.

September 28 — Build professional community & networks

In-class guest: Darren Gergle

Readings

- Agre, Phil. 2002. Networking on the network: A guide to professional skills for PhD students (Sections 1-6).
- Rackham Graduate School. 2014. How to get the mentoring you want: A guide for graduate students (pdf). University of Michigan.

Writings & Activities

- Submit (via Canvas) a brief reflection (250-350 words or so) about the readings. Consider what you find surprising, insightful, anxiety provoking, puzzling, obnoxious, etc.
- Include 3 discussion provocations or questions with your reflection. These provocations should be designed to elicit an interesting and meaningful conversation in class.
- Talk to a 3rd or 4th year student in either the MTS or TSB program. Ask them about the professional communities they belong to, the conferences they attend, how they learned about these, and what they like/dislike about these communities. Come prepared to discuss your findings in class.

October 5 — Get funding, fellowships, grants

Readings

- The National Science Foundation Graduate Student Research Fellowship Program (NSFGRFP) Web site (Read fellowship description, application requirements, eligibility restrictions, and especially the Merit Review Principles and Criteria).
- Review the Northwestern TGS Fellowships site to identify at least three internal or external fellowships to which you realistically plan/want to apply.
- Example fellowship application (pdf on Canvas).
- Hargittai, Eszter. 2012. Learning from Others' CVs. Inside Higher Ed: Ph.Do column, December 14.

Writings & Activities

- Write a draft the "Research Statement" component of a fellowship application based on NSF GRFP requirements (length, formatting, content, etc.) or the requirements of another fellowship application of your choosing. Submit this via Canvas as a pdf.
- Review the CVs of one person from each of the following categories whose work resembles the kind you do or plan to do: (a) a current graduate student; (b) someone who received their PhD in 2012-2015; (c) a tenured faculty member at a peer institution. Identify the fellowships, grants, and funding sources each has received. Create a written version of this list for each of these three persons and bring it to class.
- Use the Nortwhestern Fellowships site and your investigation of CVs to build a list of fellowships you might apply for within the

next year or so.

October 12 — Get things done

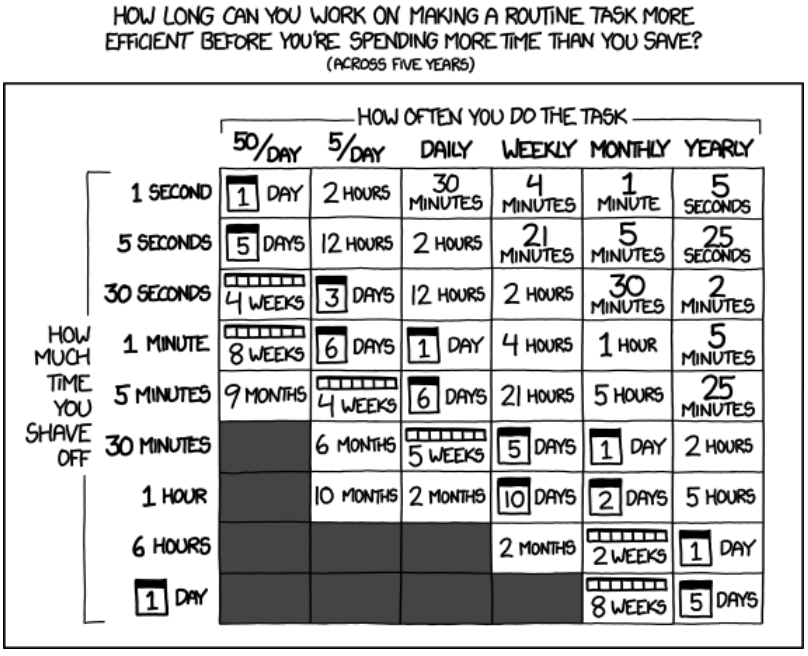
In-class guest: Anne Marie Piper

Readings

- Review your TSB or MTS program handbook (especially the milestones/timeline parts) and bring it to class.
- Henry, Alan. 2014. Productivity 101: A Primer to the Getting-Things-Done Philosophy. Lifehacker, March 26.
- Healy, Kieran. 2016. The Plain Person's Guide to Plain Text Social Science. Duke University.
- Interviews with at least three people from the "Professor" category of Usesthis.com.

Writings & Activities

- Keep a time diary for two work days prior to this week's class. Your time diary should record information about what you do when (including non-work things). Format/record the information however you like.
- Write a 250-400 word (or so) reflection on what you learn from your time diary. How did you spend your time? How much of it was work-related? How would you categorize how you spent your work time (e.g., classes, research, service)? How typical were these days for you? What changes do you anticipate/plan in the future? Submit your reflection via Canvas.



Is it worth the time? (xkcd, by Randall Munroe)

October 19 — Craft your professional identity

In-class guest: Larissa Buchholz

Readings

- Abbott, Andrew. 2004. Ideas and Puzzles, Chapter 7 in *Methods of Discovery*, W.W. Norton, NY. pp 211-248.
- Merton, Robert K. 1968. The Matthew Effect in Science. *Science* 159(3810): 56-63.
- Brewer, Marilynn B. 1991. The Social Self: On Being the Same and Different at the Same Time. *Personality and Social Psychology Bulletin* 17:475-482.
- (Optional) Cross, Nohria, and Parker. 2002. Six Myths About Informal Networks — and How to Overcome Them. *MIT Sloan Management Review* 43(3):67-75.
- (Optional) Hargittai, Eszter and Brayden King. 2013. You need a website. Insider Higher Ed: Ph.Do column, November 11.

Writings & Activities

- You will be assigned a partner for one of this week's activities. You and your partner should do the following (independently):
    - Search your partner's full name. See what results come up.
    - Using a browser with no search/browsing history and without being logged in to any services (Gmail, Facebook, etc.), conduct another search and see what comes up.
    - Conduct yet another search for your classmate's name using a search engine that you do not use frequently/ever. A good option for many of you is probably DuckDuckGo.
    - Repeat your searches with different configurations of your classmate's name (e.g., with/without middle names, with/without quotations, with/without institutional affiliation(s), etc.). Be creative.
    - Write a brief analysis documenting what you learned. What did you learn through the quick/initial search? What did you learn through more involved/elaborate searches? Include links/screenshots as you deem appropriate. Also, be sure to conclude your analysis with at least three concrete recommendations for how this person can improve their online image? Share this analysis with your partner.
- Find 3 examples of current graduate students (at NU or elsewhere in any field) who you believe have a good online presence. Bring your examples with you to class and be prepared to explain/defend why you believe they make good examples as well as ways in which they might improve.
- Identify a topic you (might) study and how you would go about it. Identify the key findings and phenomena that motivate this work. What is the puzzle? Write the puzzle up in about a paragraph and submit it via Canvas.

October 26 — Collaborate and play well with others

In-class guest: Haoqi Zhang

Readings

- Hargittai, Eszter. 2010. The Case for Collaboration. Inside Higher Ed: Ph.Do column, August 27.
- APA Science Student Council. 2006. A Graduate Student's Guide to Determining Authorship Credit and Authorship Order. American Psychological Association.
- Shaw, Aaron. 2015. Hired Hands and Dubious Guesses: Adventures in Crowdsourced Data Collection, from *Digital Confidential: The Secrets of Studying Behavior Online*, edited by Eszter Hargittai and Christian Sandvig. Cambridge, MA: MIT Press.

Writings & Activities

- Find a more advanced graduate student who has published at least one paper involving at least one coauthor. Conduct a short (~20 minutes or so?) open-ended interview using the questions below as a guide. Make sure to take notes (or record the conversation with permission). Write up the key findings from your interview in about 500 words and post this to Canvas. Be sure to address any key takeaways or insights that you found especially new, surprising, or confusing.
    - How were responsibilities divided in this collaboration (who was responsible for what)? How were decisions made about who would be responsible for what?
    - How did you and your collaborator(s) manage your work across different stages of the project (from research design through data collection, analysis, writing up, reference management, peer review, and publication)? What social/technical systems did you use?
    - What are things that went well in this collaboration? What are things that you would change next time you collaborate (with the same coauthor(s) or others)?

November 2 — Write and publish

- King, Brayden. 2011. "The editors speak: what makes a good review?" (read the entire post and all the statements from the journal editors). OrgTheory Blog.
- Read the submission guidelines for the most important (according to you / your mentor(s)) venue in your field.
- Submissions and reviews from Hargittai and Shaw 2015:
    - Initial submission — (HCI conference in 2013).
    - Final publication — *Information, Communication & Society* 2015.
    - Full text of all reviews (including plenty of rejection letters!).

Writings & Activities

- Solicit a colleague or mentor for a high quality peer review (a good review, not necessarily one that was positive!) they have received recently from an important venue in your/their field.

- Write a brief (about 300-500 words) reflection on the reviewer comments distributed as part of the readings, those you solicited from a colleague or mentor, and any reviewer comments you have received. Be sure to identify and comment on elements of the reviews you find (1) most surprising; (2) most useful; (3) least useful and (4) any aspects of peer review that you want to discuss in class.

November 8 — Take the show on the road

In-class guest (9am!): Michelle Shumate

Readings

- Hargittai, Eszter. 2009. The Conference Scene. Inside Higher Ed: Ph.Do column, September 14.
- Hargittai, Eszter. 2009. Conference Do's and Don't's. Inside Higher Ed: Ph.Do column, October 19.
- Howard, Phil. 2015. A Dozen Slides. Inside Higher Ed, September 16.

Writings & Activities

- Spend at least *an hour or two* browsing the websites of at least three conferences of prominence in your field (some suggestions below). Also discuss one of these conferences with a colleague who has attended it in the past. Be able to answer the following questions about each:
    - How is the organization running the conference structured? Who is in charge? Are there divisions? Administrative roles? Some other system?
    - How is the conference itself organized? When are submissions due? What format are they due in? How do decisions about paper acceptance/rejection get made? How do reviews happen? Who determines what winds up on which panels and who goes where? How long are presentations?
    - Are the proceedings published? Are they archival (count as publications)? Are there awards?
- Write up a brief (300-500 words) reflection on your observations. Emphasize points for further discussion in our class meeting.

November 16 — Class cancelled.

November 23 — Thanksgiving holiday vacation (no class)

November 30 — Plan for the long haul

In-class guest: Dan O'Keefe

Reading

- Varian, Hal. 1997. How to Build an Economic Model in Your Spare Time (pdf). In *Passion and Craft: Economists at Work*. Edited by Michael Szenberg, Univeristy of Michigan Press, Ann Arbor, MI.

December 7 — Final Papers Due at 12pm

- Submit your final project via Canvas.
- Closing ceremonies & festivities (6pm, off-site location TBA).

Resources

I will add links to useful/fun resources here as they come up in conversation or become relevant during the quarter. Please send me any suggestions! (↑)

- Hargittai, Eszter. Archives of Ph.Do column. (*Insider Higher Ed*).
- Piled Higher and Deeper (PhD) Comics.
- Dance your PhD contest (check out past winners).
- "Power Pose" co-author Dana Carney statement (pdf) that she no longer believes the effects are real. Current story and background on problems with the original study.
- Overview of the Michael LaCour academic fraud scandal.

Policies

In-class device usage policy

Please refrain from any uses of digitally networked devices or other distraction machines that do not directly contribute to your engagement with the course material. If you struggle to comply with this policy, I may recommend you temporarily put away your device(s) or leave the classroom. (↑)

Confidentiality of Peers' Work

Throughout the course, you will be receiving, reading and commenting on classmates' writing. These writing assignments are for class use only. You may not share them with anybody outside of class without explicit written permission from the document's author and pertaining to the specific piece. (↑)

Confidentiality of In-Class Discussions

It is essential to the success of this class that all participants feel comfortable discussing questions, thoughts, ideas, fears, reservations, apprehensions and confusion about works-in-progress, writing, the research process, and scholarly experiences more generally. Therefore, you may not create any audio or video recordings during class time nor share verbatim comments with those not in class nor are you allowed to share using other methods -- e.g., social media -- any comments linked to people's identities unless you get clear and explicit permission. If you want to share general impressions or specifics of in-class discussions with those not in class, please do so without disclosing personal identities or details. (↑)

Academic Integrity

You are responsible for reading and abiding by the Northwestern University Principles Regarding Academic Integrity.

Make sure to document all of your work and acknowledge the ideas and work of others. Possible sanctions, as per the university guidelines, include reduced or failing grade, a defined period of probation or suspension, exclusion from the university and notation on the official record. You must not, in any way, misrepresent your work or be party to another student's failure to maintain academic integrity. Do not ever copy other people's words without quotation marks (do not do this even if you are "just" taking notes) and always use proper citations. Do not ever refer to other people's work without attribution. DO NOT cheat, plagiarize or disregard the University Principles Regarding Academic Integrity in any way, it is NOT worth it! When in doubt, err on the side of giving more credit to the original source rather than less. Feel free to ask me (the instructor) for clarification about this or related matters. (↑)

Deadlines, Absences, etc.

Emergencies happen. Unanticipated obstacles arise. If you cannot make a deadline, please contact me to figure out a schedule that will work. If you must miss a class, contact me. You are responsible for obtaining class notes, handouts, assignments, etc. from fellow students in case of an absence. You are also responsible for delivering any assignments in a timely way even if you are not present in class.

An additional word about extensions and incompletes: I strongly discourage them. In principle, I have no problem with extensions or incompletes. In practice, they tend to be a pain for everybody involved. If you absolutely must submit an assignment late, assume that I will require at least 1 month (4 weeks) to grade it. Please take this into account if you will need me to to submit a grade in order to receive your fellowship/diploma/visa/etc. by a particular date. (↑)

Accommodations

I am totally happy to provide accommodations as needed. Any student requesting accommodations related to a disability or other condition is required to register with AccessibleNU (847-467-5530) and provide professors with an accommodation notification from AccessibleNU, preferably within the first two weeks of class. All information will remain confidential. For more information, visit AccessibleNU. (↑)

Sexual Harassment

All participants in this class are bound by the Northwestern University sexual harassment policy. Please note, that the core of the policy states, "no member of the Northwestern community may sexually harass any other member of the community." I take this very seriously. Please review the policy and speak to me if you have any questions or concerns. (↑)

Email protocol

I receive too much email and I sometimes fail to keep up. If, for some reason, I do not respond to your message within 48 hours, please do not take it personally and feel free to re-send the message with a polite reminder. This will help me and I will not resent you for it. (↑)

This site licensed under a cc-by-sa 4.0 license and created using Markdown and Kevin Burke's markdown.css stylesheet. Special thanks to Eszter Hargittai whose previous MTS 501 syllabuses and generous mentoring heavily influenced this course design. (↑)

Practice of scholarship (Spring 2016)

Media, Technology, and Society (MTS) 503

Northwestern University

Tuesdays 9am-12pm

Frances Searle Building Room 2-107

Instructor: Aaron Shaw (http://aaronshaw.org) (aaronshaw@northwestern.edu (mailto:aaronshaw@northwestern.edu))

- Frances Searle 2-142
- Office Hours: M 1-3pm, T 3-5pm, and by appointment.

Course resources:

- We will use Canvas (https://canvas.northwestern.edu/courses/36533) for announcements, submitting assignments, and maybe discussions.
- Everything else will be linked to from this page.

Contents

Overview & objectives

A note about this syllabus

Assignments
- Readings
- Writing assignments
- Feedback assignments
- Research journal
- Final Project: Manuscript submission

Evaluation and grades

Course schedule
- Week 1: April 5 — Introductions
- Week 2: April 12 — Planning your work & work your plan
- Week 3: April 19 — Research question: Where's the puzzle?
- Week 4: April 26 — Prior Work: Interrupting a conversation
- Week 5: May 3 — Method: Research design & justification
- Week 6: May 10 — Results & Discussion
- Week 7: May 17 — Introduction & Conclusion: End up at the beginning
- Week 8: May 24 — Revise, revise, revise
- Week 9: May 31 — Submission, reviews, and revision in publication
- Week 10: June 7 — Final projects due

Resources

Policies & protocols

Overview & objectives

The goal for this course is simple: submit a piece of academic research for publication by the end of the quarter. The piece should (obviously) be original. You should be the primary person responsible for the research and should be the lead author of the submission.

The course and assignments are structured to help you cultivate (more of) the skills, wisdom, and experience necessary to publish independent, original, and high-quality scholarship in relevant venues for your work. There are several milestones to help you measure your progress towards manuscript submission at the end of the quarter. The seminar will be run as a workshop in which you will produce written work and provide feedback on each other's work every week. Most weeks, we will also read and discuss materials related to the crafts of designing, conducting, writing, submitting, reviewing, revising, and publishing scholarly research. The experience will probably feel like a combination of a writing bootcamp and an extended group therapy session.

A note about this syllabus

You should expect this syllabus to be a dynamic document and you will notice that there are a few places left blank or marked "To Be Determined." Although the core expectations for this class are fixed, the details of readings and assignments will shift. As a result, there are three important things to keep in mind:

1. Details on this syllabus will change, but I will not change readings or assignments less than one week before they are due. If I don't fill in a "To Be Determined" one week before it's due, it is dropped. If you plan to read more than one week ahead, contact me first.
2. Keep an eye out for emails and announcements I send through Canvas (https://canvas.northwestern.edu/courses/36533) re: updates to the syllabus. You can also review the edit history of this page (http://wiki.communitydata.cc/index.php?title=Practice_of_scholarship_(Spring_2016)&action=history) to track what has changed recently and compare it against earlier versions.
3. You can *always* give me feedback and suggestions related to what works and what doesn't about the course. I will explicitly solicit your input a few times during the quarter, but **be bold** and feel free to submit your feedback to me at any time in any format. In the past, I have made substantive changes to courses on-the-fly in response to student feedback.

Assignments

The assignments are designed to facilitate your progress toward the course objectives. Due dates for assignments are included in the course schedule below.

Readings

Almost every week will have some required readings. In general, I will provide links to readings or distribute them via Canvas (https://canvas.northwestern.edu/courses/36533). You are expected to have read these before you come to class and to have prepared for discussion. There are also some suggested readings and other resources you might find useful.

There are two books that we will read multiple selections from. I recommend you buy both (in theory, you already own one of them from MTS 501). There are multiple editions of each and I don't care which edition you use (I think the chapter numbers and titles are consistent across them all):

- Booth, Wayne C., Gregory G. Colomb, Joseph M. Williams. *The Craft of Research.* Chicago: University of Chicago Press.
- Becker, Howard S. *Writing for Social Scientists.* Chicago: University of Chicago Press.

Writing assignments

Every week, I ask you to produce and submit some written work. Unless otherwise noted, you should plan to upload this work to Canvas by 9am on Monday morning each week (24 hours before the class meets). This will make it possible for me and your colleagues to read and comment on your work before class.

Feedback assignments

Almost every week, I ask you to review and comment on colleagues' work prior to class. In general, you should provide your comments as a response to their post on Canvas and come to class prepared to discuss the work and your feedback.

Research journal

Throughout the quarter, you should keep a research journal documenting your effort, progress, and reflections on your project in this course. I encourage you to write brief daily entries (or as close to daily as you can) and, at minimum, two entries per week. Entries can be brief and might simply record what you worked on that day, how long you worked on it, and a sentence or two reflection on your work experience. You may also find yourself inspired to write more. I have asked you to submit journal entries to me twice during the quarter for review.

Final Project: Manuscript submission

Due: Tuesday, June 7

Your final project for the course is a submission-ready manuscript for a peer reviewed conference or journal of your choosing. It should follow the style, length, and formatting guidelines of the venue in which you seek to publish it.

Evaluation and grades

In addition to the assignments and frequent feedback you will provide and receive on your work, you will also perform self, peer, and course evaluations at several points throughout the quarter. Your final grades for the course will be constructed based on an aggregation of all these materials with the following weights:

- Participation 20%
- Written assignments 20%
- Feedback assignments 20%
- Peer and self evaluations 15%
- Final manuscript 25%

Course schedule

Week 1: April 5 — Introductions

Assignment: Please bring two printed hard copies of the following materials with you:

- A brief synopsis (800 words MAX) of the research project you plan to pursue in this course.
- A target venue (peer reviewed journal or archival conference) to which you plan to submit your work.
- Bibliographic information (the citation) for an exemplary paper (probably authored by someone else) after which you plan to model critical aspects of your own work (e.g., the topic of study, the theoretical framing/contribution, the research design, the empirical analysis, the writing style or structure).
- A couple of sentences summarizing the aspects of your exemplary paper that make it a useful model for the work you are pursuing.
- A paragraph or two summarizing the status of the work you have pursued on this project to-date, including any prior papers you may have written, data you may have collected, and resources you may have or need to complete the work.
- A timeline (with whatever milestones you deem relevant) for completing the project and submitting it for review by the end of the Spring Quarter (or a tiny bit after). If you like, you might find it useful to create a Gantt chart using spreadsheet software.

Week 2: April 12 — Planning your work & work your plan

Reading Part I:

- Becker, Howard. *Writing for Social Scientists.* Chapters 1 ("Freshman English for Graduate Students") & 7 ("Getting It out the Door").
- Booth et al. Prologue to Section IV ("Planning Again") and Quick Tip on Outlining (pp. 185-188).

Reading Part II (pick any two):

- Ko, Andrew. How do I write a good research paper? (https://faculty.washington.edu/ajko/advice#goodpaper) (HCI-oriented).
- Landers, Richard N. 2014. How to Write a Publishable Social Scientific Research Article: Exploring Your "Process." (http://neoacademic.com/2014/07/16/how-to-write-a-publishable-social-scientific-research-article-exploring-your-process/) *NeoAcademic Blog.*
- Pasek, Josh. 2012. "Writing the Empirical Social Science Research Paper: A Guide for the Perplexed" (https://www.apa.org/education/undergrad/empirical-social-science.pdf)(pdf). *Psychology Teacher Network*, *21*(4).
- Wobbrock, Jacob O. Catchy Titles Are Good: But Avoid Being Cute (http://faculty.washington.edu/wobbrock/pubs/Wobbrock-2015.pdf)(pdf). An HCI research paper writing guide formatted as an HCI paper...

Assignment:

- Create an outline of your exemplary paper. Be sure to include section word counts.
- Create an outline or a research planning document (http://aaronshaw.org/teaching/2015/peer/planning_document.html) for your project. Submit this to the corresponding "Discussion" in Canvas (https://canvas.northwestern.edu/courses/36533/discussion_topics/230615).
- Review a peer's outline or planning document. Write your comments as a response to the peer's "Discussion" posting.
- Complete weekly Research journal entries.

Week 3: April 19 — Research question: Where's the puzzle?

Reading:

- Booth et al., Chapter 3 ("From Topics to Questions") & Chapter 4 ("From Questions to Problems").
- Durkheim, Émile. 1897. *Suicide*. Excerpt — final section of the Introduction (available via Canvas (https://canvas.northwestern.edu/courses/36533/files)).
- Kahn, C. Ronald. 1994. "Sounding Board: Picking a Research Problem — The Critical Decision (http://www.nejm.org/doi/full/10.1056/NEJM199405263302113)." *The New England Journal of Medicine 330*(21):1530-1533.

Assignment:

- A synopsis of your research project that includes the following elements (submitted, once again, via the corresponding "Discussion" in Canvas (https://canvas.northwestern.edu/courses/36533/discussion_topics/230622)):
  - A description of the topic and clear statement of the claim.
  - A list of questions derived from the topic and claim. Underscore the most interesting one(s) that you will address.
  - A brief statement posing your research around a problem (or puzzle).
  - A brief statement of the significance or application of your project.
- Review a peer's project synopsis. Evaluate whether it effectively articulates a research topic, question, significance, and problem using the criteria described by Booth et al. Post your review as a response to your peer's post in the appropriate "Discussion" on Canvas.
- Complete weekly research journal entries.

Week 4: April 26 — Prior Work: Interrupting a conversation

Reading:

- Becker, Chapter 8 ("Terrorized by the Literature").
- Booth et al., Chapter 6 ("Engaging Sources").
- Becker, Howard. 1953. "Becoming a Marihuana User." (http://www.jstor.org/stable/pdf/2771989.pdf)(pdf) *American Journal of Sociology*, *(59)*3: 235-242.
- **Optional:** Healy, Kieran. 2016. Fuck Nuance (http://kieranhealy.org/files/papers/fuck-nuance.pdf)(pdf). (forthcoming in *Sociological Theory*).

Assignment:

- Identify two or three most important existing theories/findings/systems that your work will test/synthesize/extend/enhance. Briefly (in no more than 200 words per theory/finding/system!) explain the relevant claims of the prior work, how it connects to your project, and what differentiates your project from it. As usual, post this to the appropriate "Discussion" page on Canvas (https://canvas.northwestern.edu/courses/36533/discussion_topics/230623).
- Review a peer's posting. For each existing theory/finding/system they discuss, do they provide an effective, compelling rationale that justifies their project in relation to prior work? Are you convinced that they are addressing an important question in their domain of study?
- Complete weekly research journal entries. Submit 2 or 3 journal of your favorite entries so far to Aaron via email (mailto:aaronshaw@northwestern.edu).
- Complete mid-quarter course evaluation (https://docs.google.com/forms/d/1w2gGNzBxH2ZW0TJdd6Obq4t7dR85gSuttzol9MtsCYE/viewform) (by Sunday, April 24).

Week 5: May 3 — Method: Research design & justification

Reading:

- Small, Mario Luis., 2009. How many cases do I need? On science and the logic of case selection in field-based research (http://eth.sagepub.com/content/10/1/5.short)." *Ethnography (10)*:1, 5-38.
- **Optional:** Booth et al., Chapter 9 ("Reasons and Evidence").

Assignment:

- Complete mid-course self-assessment and reflection (*tbd*).
- Write up the methodological approach you (plan to) pursue in your project and your justification for the approach. Make sure to restate your research question and explain why the data/evidence you (will) collect and the method(s) of analysis you (will) use provide insight into the problem you are addressing. Make sure that your argument will convince a skeptical reader that your approach is sensible, well-thought through, and compelling (500-800 words) Post to the discussion page (https://canvas.northwestern.edu/courses/36533/discussion_topics/236976).
- Review a peer's write-up of their methodological approach & justification. Does it make sense? Has the author provided a clear and compelling rationale for the analytical approach they take to their research problem and the data they use? Is there a mismatch between the research questions and the data? Between the methods of analysis and the focus of the inquiry? Be a skeptical (but nonetheless generous) reviewer.
- Complete weekly research journal entries.

Week 6: May 10 — Results & Discussion

Reading assignment goals: This week you will use *one of the instructional readings* and *your model paper* to extract general guidelines for presenting results and analysis. If you would like suggestions for additional model papers, please ask Aaron.

Reading: choose your own adventure. Because the presentations and discussions of results vary so widely across methods and research communities, you should chose *one* of the instructional readings below. Each one is aimed at writing up and discussing results gathered through a specific method (participant observation, interviews, and field experiments respectively. Copies of the text(s) can be made available if we need them.

- Emerson, Fretz & Shaw. 1995. *Writing Ethnographic Fieldnotes,* Chapter 7.
- Weiss. 1994. *Learning from Strangers,* Chapter 7 (Available on Canvas (https://canvas.northwestern.edu/courses/36533/files)).
- Gerber & Green. 2012. *Field Experiments,* Chapter 13.

Please note: Aaron will add other potential instructional readings to this list as he becomes aware of them. If you know of another instructional reading that you would like to use because it fits your purposes better, please ask Aaron so he can review it and confirm that it's suitable for the assignment.

Optional reading:

- Booth et al., Chapter 10 ("Acknowledgments and Responses") and Chapter 15 ("Communicating Evidence Visually").

Assignment:

- Based on your instructional reading *and* your model paper, prepare a check-list (or some similarly concise, usable representation) of attributes of excellently presented research evidence/findings. Your list (or whatever) should be the kind of thing you will use to guide your own work. Upload this to the corresponding Canvas "Discussion."

(https://canvas.northwestern.edu/courses/36533/discussion_topics/239310) We will use these to compile lists and common themes in class (using this google drive file (https://docs.google.com/document/d/19Xlpip5JC9Q-GREOgjJHAA1doLaByPiNiPI0Uso2Ndw).

- Write up about 1000 words synthesizing the (anticipated) findings and discussing the significance of your research and upload that to the corresponding Canvas "Discussion." (https://canvas.northwestern.edu/courses/36533/discussion_topics/239311) I recommend doing this in two parts:
  - Write ~500 words explaining the (anticipated) findings from your study. Quite literally, explain what you (expect to) find. What patterns of evidence (would) support these findings? If appropriate, include any data visualizations or tables you (plan to) present.
  - Write ~500 words discussing the findings in the context of the research questions and prior literature that frames your project. What is the (expected) contribution of your research? What do you (expect to) know at the conclusion of your study that was unknown or misunderstood before your study?
- Provide feedback to your peer on their findings and discussion write up (and *only* their findings and discussion write up).

Week 7: May 17 — Introduction & Conclusion: End up at the beginning

Reading:

- Little, Andrew T. 2016. "Three Templates for Introductions to Political Science Articles (http://www.andrewtlittle.com/papers/little_intros.pdf)." Manuscript, Cornell University.

Assignment:

- Pick two articles from the February, 2016 issue (http://onlinelibrary.wiley.com/doi/10.1111/jcom.2016.66.issue-1/issuetoc) of *Journal of Communication* (Volume 66, Issue 1) or two papers from CHI 2016 (http://dl.acm.org/citation.cfm?id=2858036) (or one from each).
  - If you choose JoC pieces, do not pick the Vorderer article. Do not pick a book review.
  - If you choose CHI pieces, do not choose a Note or a Panel or something else that is not a full, peer reviewed paper.
- Read the Introduction and Conclusion for both articles (ideally, don't read anything else — not even the abstract!) and respond to the following questions (in writing, submitted via Canvas (https://canvas.northwestern.edu/courses/36533/discussion_topics/241563)):

1. Provide a link/citation to the paper.
2. Briefly summarize the papers' respective central claims, evidence, and contributions in your own words.
3. According to Little's templates (See above), what type of introduction does each paper have?
4. For your favorite of the two, identify something you think it does well in the introduction and something you think it does well in the conclusion. Justify these choices/preferences.
5. For the same article (your favorite), what suggestions would you make to the author(s) for improving the introduction? the conclusion?
6. What can you take away from this favorite article for introducing/concluding your own work?

- Write an introduction for your project and submit it to the corresponding "Discussion" on Canvas (https://canvas.northwestern.edu/courses/36533/discussion_topics/241564). Keep the Introduction under 600 words. Have it reflect your anticipated findings and contribution (from last week's assignment).
- Provide feedback on your partner's Introduction.

Week 8: May 24 — Revise, revise, revise

Reading:

- Becker, Chapter 3 ("One Right Way") and Chapter 4 ("Editing by Ear").
- Strunk & White. Chapter 2 ("Elementary Principles of Composition") and Chapter 5 ("An Approach to Style").

Assignment:

- Revision assignment: Using Becker and Strunk & White as inspirations, please prepare to line-edit the rough draft texts that Aaron circulates via email/canvas (one by Silvia (https://canvas.northwestern.edu/courses/36533/discussion_topics/246614) and one by Aaron (https://canvas.northwestern.edu/courses/36533/discussion_topics/246615)). Read them, maybe bring a hard copy with you if you like to edit that way. In class, we will focus on improving the tone, style, and organization of the texts.
- Work on accomplishing your goals for your final project for this week (*no written assignment to submit or provide feedback on*). Note that you will be asked to provide an update on your progress to your discussant from the May 17 class.

Week 9: May 31 — Submission, reviews, and revision in publication

Reading:

- King, Brayden. 2011. "The editors speak: what makes a good review? (https://orgtheory.wordpress.com/2011/05/31/the-editors-speak-what-makes-a-good-review/) (read the entire post and all the statements from the journal editors). *OrgTheory*.
- Robin et al's CHI reviews and rebuttal (link tbd).
- Jeremy et al's *Social Science Research* reviews and response letter (link tbd).

Assignment:

- Make progress on your final projects!

Week 10: June 7 — Final projects due

No class meeting today. Submit your final projects via Canvas.

Resources

This section can accumulate helpful resources on topics generally related to the content of the course.

- A Dozen Slides (https://www.insidehighered.com/advice/2015/09/16/essay-preparing-effective-presentation-academic-job-talk) Philip N. Howard gives advice on preparing a social science job talk that might also help you organize your thinking and writing for any project.

Policies & protocols

Confidentiality of Peers' Work

Throughout the course, you will be receiving, reading and commenting on classmates' writing. These writing assignments are for class use only. You may not share them with anybody outside of class without explicit written permission from the document's author and pertaining to the specific piece.

Confidentiality of In-Class Discussions

It is essential to the success of this class that participants feel comfortable sharing questions, thoughts, ideas, fears, reservations, apprehensions and confusion about works-in-progress, writing, the research process and scholarly experiences more generally speaking during discussions. Therefore, you may not create any audio or video recordings during class time nor share verbatim comments with those not in class nor are you allowed to share using other methods -- e.g., social media -- comments linked to people's identities unless you get a person's permission. If you want to share general impressions or specifics of in-class discussions with those not in class, ask for permission first.

Academic Integrity

You are responsible for reading and abiding by the Northwestern University Principles Regarding Academic Integrity (http://www.northwestern.edu/provost/students/integrity/).

Make sure to document all of your work and acknowledge the ideas and work of others. Possible sanctions, as per the university guidelines, include reduced or failing grade, a defined period of probation or suspension, exclusion from the university and notation on the official record. You must not, in any way, misrepresent your work or be party to another student's failure to maintain academic integrity. Do not ever copy other people's words without quotation marks (do not do this even if you are "just" taking notes) and always use proper citation. Do not ever refer to other people's work without attribution. DO NOT cheat, plagiarize or disregard the University Principles Regarding Academic Integrity in any way, it is NOT worth it! When in doubt, err on the side of giving more credit to the original source rather than less. Feel free to ask me (the instructor) for clarification about related matters.

Deadlines, Absences, etc.

Emergencies happen. Unanticipated obstacles arise. If you cannot make a deadline, please contact me to figure out a schedule that will work. If you must miss a class, contact me. You are responsible for obtaining class notes, handouts, assignments, etc. from fellow students in case of an absence.

An additional word about extensions and incompletes: In principle, I have no problem with extensions or incompletes. In practice, they tend to be a pain for everybody involved and I strongly discourage them. If you absolutely must submit an assignment late, assume that I will require at least 1 month (4 weeks) to grade it. Please take this into account if you will need me to to submit a grade in order to receive your fellowship/diploma/visa/etc. by a particular date.

Accommodations

I am totally happy to provide accommodations. Any student requesting accommodations related to a disability or other condition is required to register with AccessibleNU (847-467-5530) and provide professors with an accommodation notification from AccessibleNU, preferably within the first two weeks of class. All information will remain confidential. For more information, visit AccessibleNU (http://http//www.northwestern.edu/accessiblenu/%7D%7BAccessibleNU).

Sexual Harasssment

All participants in this class are bound by the Northwestern University sexual harassment policy (http://www.northwestern.edu/sexual-harassment/policy/). Please note, that the core of the policy states, "no member of the Northwestern community may sexually harass any other member of the community." I take this very seriously. Please review the policy and speak to me if you have any questions or concerns.

Email protocol

I receive too much email. If, for some reason, I do not respond to your message within 48 hours, please do not take it personally and feel free to re-send the message. This will help me and I will not resent you for it.

Retrieved from "https://wiki.communitydata.cc/index.php?title=Practice_of_scholarship_(Spring_2016)&oldid=86944"

This page was last edited on 24 May 2016, at 12:38.

Content is available under Attribution-Share Alike 3.0 Unported unless otherwise noted.

Peer Production: Collective Action & Organization Online

MTS 525

Fall, 2014
Thurs. 3-5:50pm
Frances Searle Building (FSB) 2-370

Aaron Shaw
aaronshaw@northwestern.edu
FSB 2-142
Office hours by appointment

Course Description & Organization

Peer production and related modes of online collective action are the most significant organizational innovations that have emerged from Internet-mediated social practice. The best known examples of peer production include free/libre and open source software (FLOSS) such as the GNU/Linux operating system, and free culture projects such as Wikipedia. This graduate seminar will survey some of the most influential threads of empirical and theoretical research on peer production and related practices of online cooperation and collective action.

The syllabus is structured around four central puzzles or concerns that encompass the major areas of scholarship related to peer production to-date:

1. Why do (some) individuals participate in peer production systems? (Motivation)
2. How are peer production systems organized and governed in the apparent absence of formal bureaucracy, relational contracts, or organizational boundaries? (Organization)
3. Why are some peer produced resources of such high/low quality? (Quality)
4. What are the the social (cultural, economic, technological, and political) effects of peer production? (Impact)

Insofar as existing peer production research involves a variety of approaches to these overlapping sets of questions and topics, one of the goals for the course is to learn to think with the tools, concepts, and methods of a number of different disciplinary and interdisciplinary traditions. The course includes research from Communication, Sociology, Design, Human Computer Interaction, Economics, Social Psychology, Political Science, and Law. As a result, we will consider how standards of evidence, knowledge production, and what constitutes a meaningful object of research vary

across these fields. These sorts of skills will translate well even if the primary focus of your own research does not turn out to be peer production *per se*.

In addition, throughout the quarter, you will iteratively develop a final project (see below for further details). The structure of these final projects is very flexible and I hope you will take advantage of this flexibility to try something adventurous. As they say on Wikipedia, **Be bold!**

Learning Objectives

By participating actively in and completing this seminar, I believe you can go far beyond merely consuming the course material. I encourage you to work towards becoming active producers of knowledge and ideas in this domain. Throughout the quarter, you should aim to cultivate the following skills in the context of research on peer production and related topics:

- Analyze, interpret, synthesize, and critique empirical social science research.
- Describe, understand, and engage with practical challenges relevant to online collaboration and collective action.
- Develop and articulate novel theoretical and empirical perspectives.
- Identify and elaborate compelling research and design problems, as well as pathways towards solving them.
- Produce creative, original research and design projects.
- Reflect on (assess) and iteratively improve upon your own work and that of your peers in light of the concerns analyzed in class.

Prerequisites

There are no prerequisites for this course. You may find that the readings and ideas we discuss in class are easier if you have some background in social scientific theories and/or research methods. That said, I do not assume that you will have prior experience with the material. If you have questions or concerns, please email me.

Requirements

The day-to-day course requirements consist in performing standard graduate seminar activities in a standard timely fashion: doing readings, attending and participating in seminar meetings, completing assignments, and completing the final project. I believe it is crucial for *all* participants of the seminar to do *all* of these things. If you believe you require some sort of exemption or exception from any of these activities, you must negotiate that with me before the end of the first week of the quarter.

Course Materials

The course materials, including an updated schedule, the latest version of this syllabus, some readings, handouts, etc. will be maintained online to facilitate distribution. The url for these materials is:

http://aaronshaw.org/teaching/2014/peer

Note that you will need a username and password (to be distributed in class) to access some resources on the site.

Assignments

In addition to the reading and active participation in seminar discussions, there will be four kinds of assignments for this course: (1) weekly discussion provocations; (2) group "position paper" write-ups; (3) a peer-review of an unpublished article-length manuscript; and (4) a final project. Details about the final project are included in the following section and will be discussed at greater length in class.

Weekly Discussion Provocations: One day before each class meeting, you are required to circulate a very brief discussion provocation to the other members of the class. Your provocation should be no more than a paragraph or two in length and should respond to central concerns raised by the week's reading assignments. You can address one of the readings or some thematic issue that connects more than one of the readings. It's up to you. The key here is that you should address something interesting and of substantive importance in a way that will engage others in discussion or debate.

Group Position Papers: Roughly every two weeks, I'll assemble, remix, and distill the provocations you've written into a set of about 3-4 prompts that synthesize key themes and issues raised by the readings and discussions. You'll then form groups of 2-3 people to craft a collaborative response — a brief position paper — to one of the prompts (there will be as many groups as there are prompts). You'll circulate these position papers among each other and are responsible for reading the positions of your peers prior to the beginning of the corresponding class.

Note that both the provocations and the collaborative position papers should be submitted one full day (24 hours) before the class meets (Wednesday at 3pm!). By "submitted" I mean that you have circulated or posted whatever you're submitting. We'll sort out the logistical details for this during our first meeting of the quarter.

In general, whenever you are submitting something to me, please send PDF files. I don't really want hard copies and I strongly prefer open/free standards for all file formats. If you have questions about what I mean by this, please ask.

Final Projects: A minimal template for a final project is an empirically-oriented research paper approximately 4,000 words (15 pages) in length. You may also consider writing a research design, system design, prototype assessment, or research/design proposal of roughly the same length. In any scenario, I discourage you from writing more than 20 pages unless you have a very compelling

reason to do so. Writing concisely (and then iterating on revisions) will serve you well in the long run.

Final projects should be emailed to me as a PDF file **no later than December 9, 2014 at 5pm, Central Time**. Other than that, I will not impose speficic, formal, or formatting requirements on the final project, preferring instead that it provide an opportunity for you to generate a research product that best suits your particular needs and ambitions. *You should plan to meet with me to negotiate the shape and scope of your project early in the quarter or even before the quarter*.

Evaluation

You will be evaluated on your engagement with the material in class (25%); your timely completion of assignments throughout the quarter (35%); as well as your success in producing a final project that applies knowledge of relevant subject matter in an original way (40%). You may submit a collaborative final project with up to two other members of the class.

My evaluation of your final project will take into account not only the quality of your execution, but also the scope and difficulty of your ambition (think of it like competitive diving). You are *very strongly* encouraged to work with me as well as your advisors and peers to ensure that your project meets standards and requirements that best correspond to your goals.

Confidentiality of Peers' Work

Throughout the course, you will be receiving, reading and commenting on classmates' writing. These writing assignments are for class use only. You may not share them with anybody outside of class without explicit written permission from the document's author and pertaining to the specific piece.

Confidentiality of In-Class Discussions

It is essential to the success of this class that participants feel comfortable sharing questions, thoughts, ideas, fears, reservations, apprehensions and confusion about works-in-progress, writing, the research process and scholarly experiences more generally speaking during discussions. Therefore, you may not create any audio or video recordings during class time nor share verbatim comments with those not in class nor are you allowed to share using other methods – e.g., social media – comments linked to people's identities unless you get a person's permission. If you want to share general impressions or specifics of in-class discussions with those not in class, ask for permission first.

Academic Integrity

You are responsible for reading and abiding by the Northwestern University (Principles Regarding Academic Integrity). Make sure to document all of your work and acknowledge the ideas and work of others. Possible sanctions, as per the university guidelines, include reduced or failing grade, a defined period of probation or suspension, exclusion from the university and notation on the official record. You must not, in any way, misrepresent your work or be party to another student's failure

to maintain academic integrity. Do not ever copy other people's words without quotation marks (do not do this even if you are "just" taking notes) and always use proper citation. Do not ever refer to other people's work without attribution. DO NOT cheat, plagiarize or disregard the University Principles Regarding Academic Integrity in any way, it is NOT worth it! When in doubt, err on the side of giving more credit to the original source rather than less. Feel free to ask me (the instructor) for clarification about related matters.

Deadlines, Absences, etc.

Emergencies happen. Unanticipated obstacles arise. If you cannot make a deadline, please contact me to figure out a schedule that will work. If you must miss a class, contact me. You are responsible for obtaining class notes, handouts, assignments, etc. from fellow students in case of an absence.

An additional word about extensions and incompletes: In principle, I have no problem with extensions or incompletes. In practice, they tend to be a pain for everybody involved and I strongly discourage them. If you absolutely must submit an assignment late, assume that I will require at least 1 month (4 weeks) to grade it. Please take this into account if you will need me to to submit a grade in order to receive your fellowship/diploma/visa/etc. by a particular date.

Course Outline

Please note: All readings are due the date they are listed. Weekly provocations and position papers must be submitted no less than one day (24 hours) prior to the beginning of the class for which they are due.

September 25 – Introductions

Introductions. Preliminaries for the course. WikiHow mini-tutorial, edit-a-thon, and discussion. Research brainstorming and design activity. Research speed dating.

Assignments

Create a WikiHow account.

Take the WikiHow Tour.

Read the WikiHow Writer's Guide.

October 2 – Collective Action and Public Goods

Reading

Fulk, J., Flanagin, A. J., Kalman, M. E., Monge, P. R., and Ryan, T. (1996). Connective and communal public goods in interactive communication systems. *Communication Theory*, 6(1):60–87.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press. Capters 1-2.

Poteete, A. R. and Ostrom, E. (2004). Heterogeneity, group size and collective action: The role of institutions in forest management. *Development and Change*, 35(3):435–461.

October 9 – Peer Production & Social Production

Reading

Benkler, Y. (2002). Coase's penguin, or, linux and the nature of the firm. *Yale Law Journal*, 112(3).

Benkler, Y. (2004). Sharing nicely: On shareable goods and the emergence of sharing as a modality of economic production. *The Yale Law Journal*, 114(2):273–358.

Benkler, Y., Shaw, A., and Hill, B. M. (2014). Peer production: A modality of collective intelligence. In Bernstein, M. and Malone, T., editors, *The Handbook of Collective Intelligence*. MIT Press, Cambridge, MA.

October 16 – Motivation I

Group Position Paper #1 due Wednesday, October 15 at 3pm.

Reading

Re-read: Benkler et al. section on Motivation.

Antin, J., Cheshire, C., and Nov, O. (2012). Technology-mediated contributions: editing behaviors among new wikipedians. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 373–382, New York, NY, USA. ACM.

Restivo, M. and van de Rijt, A. (2014). No praise without effort: experimental evidence on how rewards affect wikipedia's contributor community. *Information, Communication & Society*, pages 1–12.

Shah, S. K. (2006). Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science*, 52(7):1000–1014.

Zhang, X. M. and Zhu, F. (2010). Group size and incentives to contribute: A natural experiment at chinese wikipedia. *American Economic Review*.

October 23 – Motivation II

Final project proposal due Thursday, October 23 before class. No more than 800 words, please.

Reading

Bimber, B. A., Flanagin, A. J., and Stohl, C. (2012). *Collective Action in Organizations: Interaction and engagement in an Era of Technological Change*. Communication, society and politics. Cambridge University Press, New York.

October 30 – Organization I

Group Position Paper #2 due Wednesday, October 29 at 3pm.

Peer review of "Almost Wikipedia" due Thursday, October 30 before class (additional instructions TBA).

Reading

Re-read: Benkler et al. section on Organization.

Forte, A., Larco, V., and Bruckman, A. (2009). Decentralization in wikipedia governance. *J. Manage. Inf. Syst.*, 26(1):49–72.

Hill, B. M. (2013). Almost wikipedia: What eight early online collaborative encyclopedia projects reveal about the mechanisms of collective action.

Ransbotham, S. and Kane, K. (2011). Membership turnover and collaboration success in online communities: Explaining rises. *Management Information Systems Quarterly*, 35(3):613–627.

Wang, X., Butler, B. S., and Ren, Y. (2013). The impact of membership overlap on growth: An ecological competition view of online groups.

November 6 – Organization II

Reading

Schweik, C. M. and English, R. C. (2012). *Internet success: a study of open-source software commons*. MIT Press, Cambridge, Mass.

November 13 – Quality I

Group Position Paper #3 due Wednesday, November 12 at 3pm.

Final project précis due Thursday, November 13 before class. No more than 1500 words please.

Reading

Re-read: Benkler et al. section on Quality.

Hill, B. M. and Monroy-Hernández, A. (2013). The cost of collaboration for code and art: evidence from a remixing community. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 1035–1046, New York, NY, USA. ACM.

Gorbatai, A. (2012). Aligning collective production with demand : Evidence from wikipedia.

Reagle, J. and Rhue, L. (2011). Gender bias in wikipedia and britannica. *International Journal of Communication*, 5(0):21.

Reich, J., Murnane, R., and Willett, J. (2012). The state of wiki usage in u.s. k–12 schools leveraging web 2.0 data warehouses to assess quality and equity in online learning environments. *Educational Researcher*, 41(1):7–15.

November 20 – Quality II

Reading

Hindman, M. (2008). *The Myth of Digital Democracy*. Princeton University Press, Princeton, NJ.

November 27 – Thanksgiving (no class)

December 4 – Impact and Implications

Group Position Paper #4 due Wednesday at 3pm.

In class: lightning talks (5 minutes each!) on your final projects.

Reading

Taylor, A. (2014). *The People's Platform: Taking Back Power and Culture in the Digital Age*. Metropolitan Books, New York.

Final Papers Due: December 9, 5pm

Additional Sources

For each topic we address in this course, we are barely scratching the surface. There are many papers, articles, and books that could have been included, but which I left out for various reasons. Below, I've collected some of my favorite readings that did not make the final cut for the syllabus. They provide great starting points for follow-up reading, research project ideas, and qualifying exams. I have organized them into the same thematic groupings used in the course as a whole.

Collective Action and Public Goods

Bennett, W. L. and Segerberg, A. (2013). *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. Cambridge University Press.

Hardin, G. (1968). The tragedy of the commons. *Science*, 162(December 13):1243–1248.

Marwell, G., Oliver, P. E., and Prahl, R. (1988). Social networks and collective action: A theory of the critical mass. III. *American Journal of Sociology*, 94(3):502–534.

Oliver, P., Marwell, G., and Teixeira, R. (1985). A theory of the critical mass. i. interdependence, group heterogeneity, and the production of collective action. *American Journal of Sociology*, 91(3):522–556.

Oliver, P. E. and Marwell, G. (1988). The paradox of group size in collective action: A theory of the critical mass. II. *American Sociological Review*, 53(1):1–8.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, New York, NY.

Ostrom, E. (2009). Beyond markets and states: Polycentric governance of complex economic systems. *Nobel Prize in Economics Lecture*.

Peer Production & related phenomena

Crowston, K., Wei, K., Howison, J., and Wiggins, A. (2010). Free/libre open source software: What we know and what we do not know. *ACM Computing Surveys*, 44(2):2012. Working Paper.

Evans, D. and Schmalensee, R. (2008). Markets with two-sided platforms. In *Issues in Competition Law and Policy*, volume Volume 1, page Chapter 28. ABA Section of Antitrust Law.

Jenkins, H. (2006). *Convergence culture: where old and new media collide*. New York University Press, New York.

Jullien, N. (2012). What we know about wikipedia: A review of the literature analyzing the project(s). SSRN Scholarly Paper ID 2053597, Social Science Research Network, Rochester, NY.

von Hippel, E. (2005). *Democratizing innovation*. The MIT Press, Cambridge, Massachusetts.

Weber, S. (2004). *The Success of Open Source*. Harvard University Press.

Motivation

Algan, Y., Benkler, Y., Fuster Morell, M., and Hergueux, J. (2013). Cooperation in a peer production economy: Experimental evidence from wikipedia.

Cheshire, C. (2007). Selective incentives and generalized information exchange. *Social Psychology Quarterly*, 70:82–100.

Coleman, E. G. (2012). *Coding Freedom: The Ethics and Aesthetics of Hacking*. Princeton University Press.

Halfaker, A., Keyes, O., and Taraborelli, D. (2013). Making peripheral participation legitimate: reader engagement experiments in wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 849–860, New York, NY, USA. ACM.

Kollock, P. (1999). The economies of online cooperation: Gifts and public goods in cyberspace. In Kollock, P. and Smith, M., editors, *Communities in Cyberspace*, pages 220–239. Routledge, London.

Kraut, R. E. and Resnick, P. (2012). *Building Successful Online Communities: Evidence-Based Social Design*. The MIT Press.

Lakhani, K. and Wolf, B. (2005). Why hackers do what they do: Understanding motivation and effort in free/open source software projects. In Feller, J., Fitzgerald, B., Hissam, S. A., and Lakhani, K. R., editors, *Perspectives on Free and Open Source Software*, pages 3–22. MIT Press.

von Krogh, G., Haefliger, S., Spaeth, S., and Wallin, M. (2012). Carrots and rainbows: Motivation and social practice in open source software development. *Management Information Systems Quarterly*, 36(2):649–676.

Organization

Bryant, S. L., Forte, A., and Bruckman, A. (2005). Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, GROUP '05, pages 1–10, New York, NY, USA. ACM.

Burke, M. and Kraut, R. (2008). Mopping up: modeling wikipedia promotion decisions. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, pages 27–36, San Diego, CA, USA. ACM.

Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, 12(4):346–362.

Fogel, K. (2009). *Producing Open Source Software: How to Run a Successful Free Software Project*. O'Reilly Media, Inc.

Kittur, A., Lee, B., and Kraut, R. E. (2009). Coordination in collective intelligence: the role of team structure and task interdependence. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 1495–1504.

Luther, K., Caine, K., Zeigler, K., and Bruckman, A. (2010). Why it works (when it works): Success factors in online creative collaboration. In *Proceedings of the ACM Conference on Supporting Group Work.*, Sanibel Island, Florida, USA. ACM.

Luther, K., Fiesler, C., and Bruckman, A. (2013). Redistributing leadership in online creative collaboration. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 1007–1022, New York, NY, USA. ACM.

O'Mahony, S. and Ferraro, F. (2007). The emergence of governance in an open source community. *Academy of Management Journal*, 50(5):1079–1106.

Roth, C., Taraborelli, D., and Gilbert, N. (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*, pages 1–5, Porto, Portugal. ACM.

Shaw, A. (2012). Centralized and decentralized gatekeeping in an open online collective. *Politics & Society*, 40(3):349–388.

Shaw, A. and Hill, B. M. (2014). Laboratories of oligarchy? how the iron law applies to peer production. *Journal of Communication*, 64(2):215–238.

Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations*. The Penguin Press.

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., and Smith, M. (2011). Finding social roles in wikipedia. In *Proceedings of the 2011 iConference*, iConference '11, pages 122–129, New York, NY, USA. ACM.

Wilkinson, D. M. (2008). Strong regularities in online peer production. In *Proceedings of the 9th ACM conference on Electronic commerce*, EC '08, pages 302–309, New York, NY, USA. ACM Press.

Wu, F., Wilkinson, D. M., and Huberman, B. A. (2009). Feedback loops of attention in peer production. In *Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 04*, CSE '09, pages 409–415, Washington, DC, USA. IEEE Computer Society.

Zhu, H., Kraut, R., and Kittur, A. (2012). Organizing without formal organization: group identification, goal setting and social modeling in directing online production. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 935–944, New York, NY, USA. ACM.

Quality

Arazy, O. and Nov, O. (2010). Determinants of wikipedia quality: the roles of global and local contribution inequality. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, CSCW '10, pages 233–236, New York, NY, USA. ACM.

Bao, P., Hecht, B., Carton, S., Quaderi, M., Horn, M., and Gergle, D. (2012). Omnipedia: bridging the wikipedia language gap. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 1075–1084, New York, NY, USA. ACM.

Halfaker, A., Kittur, A., Kraut, R., and Riedl, J. (2009). A jury of your peers: quality, experience and ownership in wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, WikiSym '09, pages 15:1–15:10, New York, NY, USA. ACM.

Halavais, A. and Lackaff, D. (2008). An analysis of topical coverage of wikipedia. *Journal of Computer-Mediated Communication*, 13(2):429–440.

Hecht, B. and Gergle, D. (2009). Measuring self-focus bias in community-maintained knowledge repositories. In *Proceedings of the fourth international conference on Communities and technologies*, pages 11–20, New York, NY, USA. ACM.

Hill, B. M. and Monroy-Hernández, A. (2013). The cost of collaboration for code and art: evidence from a remixing community. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 1035–1046, New York, NY, USA. ACM.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901.

Greenstein, S. and Zhu, F. (2012). Is wikipedia biased? *American Economic Review*, 102(3):343–348.

Kittur, A. and Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. pages 37–46, San Diego, CA, USA. ACM.

Kittur, A. and Kraut, R. E. (2010). Beyond wikipedia: coordination and conflict in online production groups. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 215–224, Savannah, Georgia, USA. ACM.

Priedhorsky, R., Chen, J., Lam, S. T. K., Panciera, K., Terveen, L., and Riedl, J. (2007). Creating, destroying, and restoring value in wikipedia. In *Proceedings of the 2007 international ACM conference on Supporting group work*, GROUP '07, pages 259–268, New York, NY, USA. ACM.

Raymond, E. S. (1999). *The cathedral and the bazaar: Musings on Linux and open source by an accidental revolutionary*. O'Reilly and Associates, Sebastopol, CA.

Impact

N.b., the following list is slanted towards politics and cultural production, which are two areas where this topic has provoked quite a lot of debate (and which I happen to know a bit more about). Please talk to me if you are interested in starting points for thinking about the impact of peer production in other domains.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.

Bennett, W. L. and Segerberg, A. (2013). *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. Cambridge University Press.

Castells, M. (2012). *Networks of Outrage and Hope: Social Movements in the Internet Age*. John Wiley & Sons.

Jenkins, H. (2006). *Convergence culture: where old and new media collide*. New York University Press, New York.

Karpf, D. (2012). *The MoveOn Effect: The Unexpected Transformation of American Political Advocacy*. Oxford University Press, USA.

Shaw, A. and Benkler, Y. (2012). A tale of two blogospheres: Discursive practices on the left and right. *American Behavioral Scientist*, 56(4):459–487.

Shaw, A. and Hill, B. M. (2014). Laboratories of oligarchy? how the iron law applies to peer production. *Journal of Communication*, 64(2):215–238.

Research Design for Causal Inference

MTS 525-0-20

Spring, 2015
Wed. 3-5:50pm
Location TBA

Aaron Shaw
aaronshaw@northwestern.edu
FSB 2-142
Office hours TBA

Course Description

This course is a Ph.D. seminar focused on analyzing, designing, and conducting empirical causal inference using quantitative methods. Inquiries into cause and effect lie at the heart of almost all social and behavioral research. For example: What are the effects of media exposure on children's educational attainment? Which kinds of outreach and incentives promote participation in voluntary organizations? Can social pressure increase political engagement? Answering these questions requires methods of analysis that identify the effects of particular causes.

Learning Objectives and Course Design

By the end of the quarter, students should be able to: (1) discuss and compare multiple research methods of causal inference; (2) assess causal identification strategies and claims in empirical research; (3) design research to address causal theories and questions; (4) manipulate and analyze empirical data sets to identify, estimate, and interpret causal effects; (5) identify and address threats to the validity of findings; and (6) apply methods of causal inference to empirical questions and data beyond those covered in course readings and assignments.

The course consists of two "units." The first unit focuses on the fundamentals of causal inference. Then, in the second unit, we will survey a variety of specific methodological approaches, such as field experiments, natural experiments, regression discontinuity designs, difference in differences, matching, and instrumental variable estimation. For each of these approaches, I recommend you read *at least* one text providing an instructional overview of the method as well as *at least* one empirical application of the method.

Every week, I will assign problem sets and in-class activities aimed at helping you cultivate practical experience applying the methods in question. In general, the problem sets will lag a week behind the readings so that you can get a feel for each approach before you apply it.

During the course, you will also develop two larger-scale projects. The first of these is a class exercise in which you will work (perhaps in a small team) over several weeks to plan, design, execute,

analyze, and report the results of a field experiment. The second is a planning document for an empirical research project that you plan to execute (or perhaps have already begun executing). The first assignment is intended to help you cultivate experience with the craft of designing, conducting, and writing up an empirical study to address a causal question. We will workshop these group projects throughout the quarter. The second assignment is intended to provide an opportunity to apply your skills and experiences from the class to the design of an independent project that directly advances your research interests. I especially encourage you to use this second assignment to pursue something bold and creative.

Prerequisites

There are no formal prerequisites for this course. However, you should be familiar with (i.e. have taken at least one graduate-level class on) social scientific research methods and should be comfortable with (i.e. have taken at least one graduate-level class on) quantitative analysis techniques including multivariate regression. Some experience conducting empirical research will be useful, but is not strictly necessary. Knowledge of some statistical software and/or programming will also prove very helpful, but is not required. If you have questions or concerns, please email me.

Requirements

The day-to-day course requirements consist in performing standard graduate seminar activities in a standard timely fashion: attending and participating in seminar meetings, as well as completing all assignments. I believe it is crucial for *all* participants of the seminar to do *all* of these things. If you believe you require some sort of exemption or exception from any of these activities, please negotiate with me before the end of the first week of the quarter.

Course Website and Materials

Most materials for the class (including a "live" version of the course schedule, readings, problem sets, datasets, and solution code) will be made available through the course website.

http://aaronshaw.org/teaching/2015/causal

Once the quarter begins, I will not issue revised PDF versions of the syllabus, so check the website! I may distribute some readings and materials via Canvas.

Assignments

There are four kinds of assignments for this class: readings, problem sets, your field experiment, and your research planning document. Along the way, I will also ask you to submit incremental portions of the two larger assignments (the field experiment and the planning document).

Any assignments you turn in must be submitted to me via Canvas no less than one day (24 hours) before the class meeting for that week. I strongly prefer PDFs and plain-text formats because they don't require special software. For problem sets, please include your (well-commented) code as well as the final output or answers. I will evaluate problem sets at random throughout the quarter.

Mostly, you are on your own recognizance as far as checking your solutions and following up when you do not understand any content.

Problem sets will encourage conceptual understanding, provide experience applying techniques, elicit analytical interpretation, and (to much a lesser extent) facilitate computational skill-building. I strongly encourage you to complete the problem sets in a small group (ideally no more than 2-3 people). If you work in a group, I also recommend that you (1) try to complete the work on your own first; (2) submit your work individually and acknowledge your collaborators by name.

Readings

During the first few weeks of the course we will use the following as a textbook:

Gerber, A. S. and Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. WW Norton, New York.

This book provides a clear, practically-oriented introduction to the potential outcomes framework for causal inference focused on field experiments. I recommend that you get a copy.

For all subsequent topics I will assign at least one instructional reading along with at least one empirical reading. I assign several instructional readings about natural and quasi-expermental methods from:

Murnane, R. J. and Willett, J. B. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. Oxford University Press, Oxford ; New York.

These authors emphasize conceptual understanding and examples over formal math. If their approach works for you (and it may not!), get a copy of this book too. If Murnane & Willett does not work, I still recommend that you acquire a "textbook" treatment that covers all of the methods we talk about in a consistent way. All of the following are widely used:

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Morgan, S. L. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, New York.

Imbens, G. and Rubin, D. B. (2014). *Causal inference for statistics, social, and biomedical sciences: an introduction*. Cambridge University Press, New York.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth, Belmont, CA.

Depending on your training and aspirations, you may prefer these other approaches and you should feel free to substitute corresponding selections for instructional readings throughout the latter half of the quarter.

For each topic, I have also listed additional, "optional" readings that you may find useful. If you have trouble locating any of the readings, please let me know.

Last, if you want to revisit some of the statistical foundations of the material covered in this course (e.g. Regression, Probability, Linear Algebra), I recommend the following textbooks:

Freedman, D. (2005). *Statistical Models: Theory And Practice*. Cambridge University Press.

Wooldridge, J. (2012). *Introductory Econometrics: A Modern Approach*. Cengage Learning, 5th edition edition (earlier editions are also good).

Verbeek, M. (2012). *A Guide to Modern Econometrics*. John Wiley & Sons, 4th edition edition.

Statistical Computing

You will need to use statistical computing software in this class. You are welcome to use any statistical software you like (Stata, SPSS, Python, etc.); however, I will use (heavily commented) R for all in-class examples and solutions to the problem sets.

I do all of my statistical computing and analysis in R (http://r-project.org), which is free software available for use on any operating system and extremely powerful for just about any kind of statistical application. If you have not yet invested heavily in another statistical computing language/environment (and sometimes even if you have) I highly recommend learning R and using it to complete assignments for this class. I'll say more during the first class session.

Major Projects

You are required to complete two major projects in this course: (1) a field experiment and (2) a final project consisting of a detailed research planning document for another project. The field experiment provides an opportunity to design, plan, execute, and write-up a research project estimating a causal effect. The final project provides an opportunity to apply the course material to a research design that will lead to a major research product (e.g. journal/conference submission, qualifying paper, dissertation chapter, replication study, research proposal). I will provide additional details about my requirements and expectations for each project early in the quarter.

Field experiments are due via Canvas by May 20, 2015 at 5pm, CDT.
Final projects are due via Canvas by June 9, 2015 at 5pm, CDT.

Evaluation

Your final grade for the course will be based on my evaluation on your engagement with the material in class (25%) as well as the successful and timely completion of the field experiment project (25%), final project (25%), and all other assignments throughout the quarter (25%).

Course Policies

Confidentiality of Peers' Work

You will be receiving, reading and commenting on classmates' writing. These writing assignments are for class use only. You may not share them with anybody outside of class without explicit written permission from the document's author pertaining to the specific piece.

Confidentiality of In-Class Discussions

It is essential to the success of this class that all participants feel comfortable sharing questions, ideas, fears, reservations, apprehensions, and confusions about works-in-progress, writing, the research process, and related experiences during discussions. Therefore, you may not create any audio or video recordings during class time nor share verbatim comments with those not in class nor are you allowed to share using other methods – e.g., social media – comments linked to people's identities unless you get a person's permission.

Academic Integrity

You are responsible for reading and abiding by the Northwestern University Principles Regarding Academic Integrity as well as the applicable school-specific academic integrity policies. The bottom line: make sure to document all of your work and acknowledge the ideas and work of others. When in doubt, err on the side of giving more credit to the original source rather than less. The sanctions for violations of these principles are severe. Feel free to ask me (the instructor) for clarification about related matters.

Deadlines, Absences, etc.

If something causes you to miss a deadline or a class, please contact me. If you request, and obtain, an incomplete for the course and/or an extension on the final project (note: I strongly discourage this!), please allow at least 1 month (4 weeks) after you submit your completed work for me to submit a grade. Keep this in mind if you will need the grade in order to receive your fellowship/diploma/visa/etc. by a particular date.

Students with Disabilities

Any student requesting accommodations related to a disability or other condition is required to register with AccessibleNU (accessiblenu@northwestern.edu; 847-467-5530) and provide professors with an accommodation notification from AccessibleNU, preferably within the first two weeks of class. All information will remain confidential.

Sexual Harassment

All participants in this class are bound by the Northwestern University sexual harassment policies. Please note that the core of the policy states, "no member of the Northwestern community may sexually harass any other member of the community." Review the policy and speak to me (or some other appropriate person) if you have any questions or concerns.

Course Outline

As described above, all readings and assigned tasks are due *prior* to the class alongside which they are listed. Items listed in the "Additional Readings" subsection for each week are not required. I may adjust the list of readings or the schedule as needed throughout the quarter, so please consult the online version of the schedule for the most up-to-date information.

April 1 – Introductions & Causality

Readings

Gerber and Green – Chapter 1.

Tasks

Describe three research questions addressing some causal relationship relevant to your research interests. Write the research questions down for the first class meeting and be prepared to discuss at least one of them.

Make sure you have a satisfactory setup for statistical computing. If you are working with R, Download & install it. You will also want to pick and install a development environment for R: I recommend Rstudio unless you use Emacs already, in which case Emacs Speaks Statistics (ESS) may be more appealing.

Learn (or review) statistical computing basics: R users: if you're completely new, take the Try R course. Also, this Getting Started with R page from York University contains a variety of tutorials, tips, and instructional resources. You may also want to check out the startR resources page that I created for this course.

Optional: Attend an "Intro to R" session that I will lead on **Friday, April 3, 2:00-4:00pm in Frances Searle 2-378**. We'll cover basic data input and manipulation; some operators and mathematical functions; as well as writing loops and simple functions of your own.

Additional Readings

Hidalgo, F. D. and Sekhon, J. (2011). Causality. In Badie, B., Berg-Schlosser, D., and Morlino, L., editors, *International Encyclopedia of Political Science*, pages 204–211. Sage Publications, Inc, Thousand Oaks, CA. [Available from sekhon.berkeley.edu]

Murnane and Willett – Chapters 1-2, 13.

Morgan and Winship – Chapters 1-2.

Watch Ned Hall and L.A. Paul discuss causation on Philosophy TV.

April 8 – Causal Inference & Experimentation

Readings

Gerber and Green – Chapter 2; Appendix B.

Tasks

Complete Problem Set 1.

Complete (if you have not done so already) the required CITI training for human subjects research through the Northwestern Institutional Review Board (IRB).

Additional Readings

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396):945–960. [Available via JSTOR]

Little, R. J. and Rubin, D. B. (2000). Causal Effects in Clinical and Epidemiological Studies Via Potential Outcomes: Concepts and Analytical Approaches. *Annual Review of Public Health*, 21:121–145. [Available via Annualreviews.org]

April 15 – Sampling Distributions, Inference, & Hypothesis Testing

Readings

Gerber and Green – Chapter 3.

Bowers, J. and Panagopoulos, C. (2012). Do Newspaper Ads Raise Voter Turnout? *Unpublished working paper*. [Available via jakebowers.org]

Tasks

Complete Problem Set 2.

Complete Field Experiment Assignment 1.

Additional Readings

Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer series in statistics. Springer, New York – Chapter 2, through section 2.3.2. [Available via Springer]

Splawa-Neyman, J. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472. [Available via JSTOR]

Fisher, S. R. A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh and London. Chapters 1-2. [Available online]

Bowers, J. and Panagopoulos, C. (2011). Fisher's Randomization Mode of Statistical Inference, Then and Now. *Unpublished working paper*. [Available via jakebowers.org]

Rubin, D. B. (1990). Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science*, 5(4):472–480. [Available via JSTOR]

April 22 – Covariates in Experimental Design & Analysis

Readings

Gerber and Green – Chapter 4 (& skim Chapter 9).

Wantchekon, L. (2003). Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin. *World Politics*, 55(3):399–422. [Available via JSTOR]

Tasks

Complete Problem Set 3.

Additional Readings

Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer series in statistics. Springer, New York – Chapter 2, section 2.3.3-end. [Available via Springer]

Bloom, H. S. (2006). The Core Analytics of Randomized Experiments for Social Research. *MDRC Working Papers on Research Methodology*. [Available via MDRC]

Howell, W. G., Wolf, P. J., Campbell, D. E., and Peterson, P. E. (2002). School vouchers and academic performance: results from three randomized field trials. *Journal of Policy Analysis and Management*, 21(2):191–217. [Available via Wiley]

Krueger, A. B. and Zhu, P. (2004). Another Look at the New York City School Voucher Experiment. *American Behavioral Scientist*, 47(5):658–698. [Available via Sage]

Myers, D. E. and Mayer, D. P. (2003). Comments On "Another Look at the New York City Voucher Experiment". Mathematica Policy Research, Inc. [Available online]

Freedman, D. A. (2008). Randomization Does Not Justify Logistic Regression. *Statistical Science*, 23(2):237–249. [Available via Project Euclid]

Rosenbaum, P. R. (2002). Covariance Adjustment in Randomized Experiments and Observational Studies. *Statistical Science*, 17(3):286–327. [Available via Project Euclid]

April 29 – Validity: Planning, Executing, & Archiving Research Effectively

Readings

Gerber and Green. Chapter 13.

Gelman, A. and Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking"and the research hypothesis was posited ahead of time. [Available via Gelman's website]

Tasks

Complete Problem Set 4.

Complete Mid-quarter course assessment (distributed in class).

Additional Readings

Inter-university Consortium for Political and Social Research (ICPSR) (2009). Guide to Social Science Data Preparation and Archiving. [Available from ICPSR website]

Boutron, I., John, P., and Torgerson, D. J. (2010). Reporting Methodological Items in Randomized Experiments in Political Science. *The ANNALS of the American Academy of Political and Social Science*, 628(1):112–131. [Available from Sage]

Stodden, V. and Miguez, S. (2013). Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. SSRN Scholarly Paper ID 2322276, Social Science Research Network, Rochester, NY. [Available from SSRN]

2008. MDRC. Program Design and Evaluation Strategy for Opportunity NYC-Family Rewards. [Available from FEDAI website]

Miller, C., Riccio, J., and Smith, J. (2009). A Preliminary Look at Early Educational Results of the Opportunity NYC—Family Rewards Program: A Research Note for Funders. *New York: MDRC*. [Available from MDRC]

Casey, K., Glennerster, R., and Miguel, E. (2012). Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan*. *The Quarterly Journal of Economics*, 127(4):1755–1812. [Both the paper and online supplement available via Oxford Journals]

Gerber and Green. Chapter 11.

May 6 – Natural Experiments

Readings

Murnane and Willett – Chapter 8.

Blattman, C. and Annan, J. (2010). The Consequences of Child Soldiering. *Review of Economics and Statistics*, 92(4):882–898. [Available via MIT Press]

Tasks

Complete Problem Set 5.

Complete Field Experiment Assignment 2 (Submit Planning Document for your field experiment).

Additional Readings

Dunning, T. (2008). Improving Causal Inference Strengths and Limitations of Natural Experiments. *Political Research Quarterly*, 61(2):282–293. [Available via Sage]

Zhang, X. M. and Zhu, F. (2010). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review*, 101:1601–1615. [Available via mikezhang.com]

DellaVigna, S. and Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics*, 122(3):1187–1234. [Available via Oxford Journals]

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, 114(2):533–575. [Available via Oxford Journals]

Dynarski, S. M. (2003). Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *American Economic Review*, 93(1):279–288. [Available via AEA]

Tyler, J. H., Murnane, R. J., and Willett, J. B. (2000). Estimating the Labor Market Signaling Value of the GED. *The Quarterly Journal of Economics*, 115(2):431–468. [Available via Oxford Journals]

May 13 – Regression Discontinuity & Difference-in-Differences

Readings

Murnane and Willett – Chapter 9.

Card, D. and Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, 84(4):772–793. [Available via JSTOR]

Tasks

Complete Problem Set 5.

Additional Readings

Morgan and Winship – Chapter 9.

Jacob, R. T., Zhu, P., Somers, M.-A., and Bloom, H. (2012). A Practical Guide to Regression Discontinuity. *MDRC Working Papers on Research Methodology*. [Available via MDRC]

Lee, D. S. and Lemieux, T. (2009). Regression Discontinuity Designs in Economics. Working Paper 14723, National Bureau of Economic Research. [Available via SSRN]

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, 114(2):533–575. [Available via Oxford Journals]

Neumark, D. and Wascher, W. (2000). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment. *The American Economic Review*, 90(5):1362–1396. [Available via JSTOR]

Card, D. and Krueger, A. B. (2000). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply. *The American Economic Review*, 90(5):1397–1420. [Available via JSTOR]

Malamud, O. and Pop-Eleches, C. (2011). Home Computer Use and the Development of Human Capital. *The Quarterly Journal of Economics*, 126(2):987 –1027. [Available via Oxford Journals]

May 20 – No class meeting

No class meeting today.

Submit all materials for your field experiment (planning document; dataset and reproducible analysis code; research report).

May 27 – Instrumental Variables

Readings

Murnane and Willett – Chapters 10-11.

Dittmar, J. E. (2011). Information Technology and Economic Change: The Impact of The Printing Press. *The Quarterly Journal of Economics*, 126(3):1133–1172. [Available via Oxford Journals]

Additional Readings

Morgan and Winship – Chapter 7.

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, 114(2):533–575. [Available via Oxford Journals]

Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *The Quarterly Journal of Economics*, 115(4):1239–1285. [Available via Oxford Journals]

Tasks

Complete peer evaluation of field experiments.

Work on final projects.

Download & read through example IVE analysis.

June 3 – Lightning talks & Assessment

In-class $\leq$ 5 minute presentations of your final projects and discussion. Collective reflection on the design of the course and other related questions of profound significance. Much rejoicing.

Final Projects Due June 9

Final projects due no later than 5pm Central Time, June 9, 2015. Please submit your project via Canvas.

References

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, 114(2):533–575.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Blattman, C. and Annan, J. (2010). The Consequences of Child Soldiering. *Review of Economics and Statistics*, 92(4):882–898.

Bloom, H. S. (2006). The Core Analytics of Randomized Experiments for Social Research. *MDRC Working Papers on Research Methodology*.

Boutron, I., John, P., and Torgerson, D. J. (2010). Reporting Methodological Items in Randomized Experiments in Political Science. *The ANNALS of the American Academy of Political and Social Science*, 628(1):112–131.

Bowers, J. and Panagopoulos, C. (2011). Fisher's Randomization Mode of Statistical Inference, Then and Now. *Unpublished working paper*.

Bowers, J. and Panagopoulos, C. (2012). Do Newspaper Ads Raise Voter Turnout? *Unpublished working paper*.

Card, D. and Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, 84(4):772–793.

Card, D. and Krueger, A. B. (2000). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply. *The American Economic Review*, 90(5):1397–1420.

Casey, K., Glennerster, R., and Miguel, E. (2012). Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan*. *The Quarterly Journal of Economics*, 127(4):1755–1812.

DellaVigna, S. and Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Dittmar, J. E. (2011). Information Technology and Economic Change: The Impact of The Printing Press. *The Quarterly Journal of Economics*, 126(3):1133–1172.

Dunning, T. (2008). Improving Causal Inference Strengths and Limitations of Natural Experiments. *Political Research Quarterly*, 61(2):282–293.

Dynarski, S. M. (2003). Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *American Economic Review*, 93(1):279–288.

Fisher, S. R. A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh and London.

Freedman, D. (2005). *Statistical Models: Theory And Practice*. Cambridge University Press.

Freedman, D. A. (2008). Randomization Does Not Justify Logistic Regression. *Statistical Science*, 23(2):237–249.

Gelman, A. and Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking"and the research hypothesis was posited ahead of time.

Gerber, A. S. and Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. WW Norton, New York.

Hidalgo, F. D. and Sekhon, J. (2011). Causality. In Badie, B., Berg-Schlosser, D., and Morlino, L., editors, *International Encyclopedia of Political Science*, pages 204–211. Sage Publications, Inc, Thousand Oaks, CA.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396):945–960.

Howell, W. G., Wolf, P. J., Campbell, D. E., and Peterson, P. E. (2002). School vouchers and academic performance: results from three randomized field trials. *Journal of Policy Analysis and Management*, 21(2):191–217.

Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *The Quarterly Journal of Economics*, 115(4):1239–1285.

Imbens, G. and Rubin, D. B. (2014). *Causal inference for statistics, social, and biomedical sciences: an introduction*. Cambridge University Press, New York.

Inter-university Consortium for Political and Social Research (ICPSR) (2009). Guide to Social Science Data Preparation and Archiving.

Jacob, R. T., Zhu, P., Somers, M.-A., and Bloom, H. (2012). A Practical Guide to Regression Discontinuity. *MDRC Working Papers on Research Methodology.*

Krueger, A. B. and Zhu, P. (2004). Another Look at the New York City School Voucher Experiment. *American Behavioral Scientist*, 47(5):658–698.

Lee, D. S. and Lemieux, T. (2009). Regression Discontinuity Designs in Economics. Working Paper 14723, National Bureau of Economic Research.

Little, R. J. and Rubin, D. B. (2000). Causal Effects in Clinical and Epidemiological Studies Via Potential Outcomes: Concepts and Analytical Approaches. *Annual Review of Public Health*, 21:121–145.

Malamud, O. and Pop-Eleches, C. (2011). Home Computer Use and the Development of Human Capital. *The Quarterly Journal of Economics*, 126(2):987 –1027.

Miller, C., Riccio, J., and Smith, J. (2009). A Preliminary Look at Early Educational Results of the Opportunity NYC—Family Rewards Program: A Research Note for Funders. *New York: MDRC*.

Morgan, S. L. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, New York.

Murnane, R. J. and Willett, J. B. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. Oxford University Press, Oxford ; New York.

Myers, D. E. and Mayer, D. P. (2003). Comments On "Another Look at the New York City Voucher Experiment". Mathematica Policy Research, Inc.

Neumark, D. and Wascher, W. (2000). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment. *The American Economic Review*, 90(5):1362–1396.

Rosenbaum, P. R. (2002). Covariance Adjustment in Randomized Experiments and Observational Studies. *Statistical Science*, 17(3):286–327.

Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer series in statistics. Springer, New York.

Rubin, D. B. (1990). Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science*, 5(4):472–480.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth, Belmont, CA.

Splawa-Neyman, J. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472.

Stodden, V. and Miguez, S. (2013). Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. SSRN Scholarly Paper ID 2322276, Social Science Research Network, Rochester, NY.

Tyler, J. H., Murnane, R. J., and Willett, J. B. (2000). Estimating the Labor Market Signaling Value of the GED. *The Quarterly Journal of Economics*, 115(2):431–468.

Verbeek, M. (2012). *A Guide to Modern Econometrics*. John Wiley & Sons, 4th edition edition.

Wantchekon, L. (2003). Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin. *World Politics*, 55(3):399–422.

Wooldridge, J. (2012). *Introductory Econometrics: A Modern Approach*. Cengage Learning, 5th edition edition.

Zhang, X. M. and Zhu, F. (2010). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review*, 101:1601–1615.

Course and Teacher Evaluation Summary for Instructors

Subject: COMM_ST Communication Studies
Academic Term: Fall 2013
Class Number: 18491

Enrollment: 19
Responses: 13
% Responses 68.42

Class: 378-0-20 Online Communities and Crowds
Instructor: Aaron Shaw

Combined Course: COMM_ST 378-0-20 LEC (18491)
Combined Course: SOCIOL 376-0-22 LEC (18814)

Demographic Questions

	Education & SP	Communication	Graduate School	KGSM	McCormick	Medill	Music	Summer	SCS	WCAS
Your School	0	9	0	0	0	0	0	0	0	4

	a. Freshman	b. Sophomore	c. Junior	d. Senior	e. Graduate	f. Other
Your Class	0	4	5	4	0	0

	a. Distribution requirement	b. Major requirement	c. Minor requirement	d. Elective requirement	e. Other requirement	f. No requirement
Your reason for taking course (mark all that apply).	2	8	0	0	2	2

	1	2	3	4	5	6	Total Response	Average Response
Interest in subject before taking course.	0	0	0	7	2	4	13	4.77

Core Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
1. Provide an overall rating of the instruction	12	0	0	0	1	7	4	5.25
2. Provide an overall rating of the course.	13	0	0	0	2	8	3	5.08
3. Estimate how much you learned in the course.	13	0	0	0	3	5	5	5.15
4. Rate the effectiveness of the course in challenging you intellectually.	13	0	0	0	7	3	3	4.69
5. Rate the effectiveness of the instructor(s) in stimulating your interest in the subject.	12	0	0	0	3	4	5	5.17

Time-Survey Question

	Total Response	a. 3 or fewer	b. 4 - 7	c. 8 - 11	d. 12 - 15	e. 16 - 19	f. 20 or more
6. Estimate the average number of hours per week you spent on this course outside of class and lab time.	12	3	7	2	0	0	0

School and Department Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
7. Rate how well prepared the instructor was for the class.	11	0	0	1	0	3	7	5.45
8. Rate the effectiveness with which the instructor communicated course content and ideas.	12	0	0	0	1	8	3	5.17
9. Rate the instructional materials (texts, audiovisual materials, etc.) used in this course.	12	0	0	0	3	6	3	5.00
10. Rate the instructor's enthusiasm in teaching this class.	12	0	0	0	0	2	10	5.83
11. Rate the clarity of examples.	8	0	0	0	0	3	5	5.63
12. Rate the helpfulness of the instructor outside of class.	10	0	0	1	1	5	3	5.00
13. Rate how well the instructor encouraged you to think for yourself.	12	0	0	0	1	6	5	5.33
14. Rate how well the exams stress important points of the lectures/readings/discussions.	12	0	0	1	5	4	2	4.58
15. Rate how much effort you put into this course in comparison to other courses.	11	0	1	2	6	0	2	4.00

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
16. Rate the instructor's efforts to highlight the interesting nature of the subject.	12	0	0	0	0	4	8	5.67
17. Rate how well the instructor contrasted the implications of various theories.	11	0	0	0	4	4	3	4.91
18. Rate how well this course helped you improve your ability to read critically.	12	0	1	2	5	3	1	4.08

Essay Questions

1. What are the primary teaching strengths of the instructor(s)?

Aaron always seemed really excited about the class and showed a genuine interest in the material and students./Having first hand experience and understanding of the topics that were being discussed and ability to involve students in discussion./- Inspiring Energy
- Current research knowledge in the subject area
- Great personal connections with great speakers/Enthusiasm and accessibility./Prof. Shaw is very enthusiastic about the subject of the class, and frequently encourages students to share their own thoughts during in-class discussions.

2. What are the primary weaknesses, if any, of the instruction?

The course sometimes felt a little disorganized, although this improved dramatically as the quarter progressed./Sometimes it was a little bit difficult to see the overall theme of the course./- Occasional focus on whole class discussion that dragged on/I felt like there was a wide variety in how productive each class session was; sometimes class was extremely interesting, and sometimes it was forgettable. (However, this partly has to do with the fact that the class is new, and Prof. Shaw is still trying to figure out how to best structure class time.)

3. Can you offer suggestions for improvement?

Having more evenly distributed assignments for preparing for class. Some of them took much longer than others./- I really enjoyed when Prof Shaw switched from full class-period discussions to a main-points lecture with more pointed discussion questions./Sometimes class time wasn't used in the most productive way, especially early in the quarter; for instance, some of the in-class discussions were a little too open-ended and student-directed (although I think Prof. Shaw realized this partway through the quarter). I also wasn't a fan of the occasional group activities; they each seemed to take up far more class time than they were worth.

4. Did the course help you learn? Why or why not?

Yes. I learned a great deal about online communities, relevant historical details, and a range of interesting theories that are absolutely applicable to any subject./Yes. The course definitely made me think about Internet communities in a different way. I frequently consider ideas from the class during my daily life./The course introduced me to many interesting topics that I had not heard of before and it made want to study some of the topics more in depth./I learned a lot in this

course from the variety of material, but also from the many perspectives we viewed OCC from./Yes and, more importantly, it helped me adopt new ways of thinking about critical problems./Yes, absolutely! The course covered a wide variety of very interesting topics, and the accompanying readings were typically very good). Furthermore, the structure of the class allowed plenty of room for us to think about and interact with the material on our own; rather than just being lectured at, we were able (and encouraged) to contribute to class discussions and to explore parts of the material that interested us.

5. Please summarize your reaction to this course focusing on the aspects that were most important to you.
This was an interesting class that piqued my interest in different types of online communities, the role they play in society, implications, and more. Aaron did a good job of keeping the class interesting with intriguing material and encouraging discussion./I would recommend this class to those who want to understand the history and infrastructure that governs online communities. This is not a class for people who are interested in creating or participating in online communities in an intensive fashion./Overall, this was a really great class. The workload is not bad at all, although some of the readings are a bit lengthy. Aaron definitely makes the material seem interesting and made me think about the Internet in a completely different way./I would highly recommend taking this course because it was very interesting and even allowed me the flexibility to study the things that were the most interesting to me in more depth./If you are at all interested in Online Communities, TAKE THIS CLASS! Professor Shaw is the most energetic professor that I've ever had, and you can see his passion for the subject. There is quite a bit of outside reading for the class, so be prepared to spend an hour or two to read/skim all of it./I took this course because I needed a last minute comm class and it seemed kind of interesting, but I didn't really know what to expect. I found the class pretty interesting, but I didn't understand every single concept (some just seemed too technical for me) and since he wrote the syllabus as the class went on throughout the quarter, I never really knew what to expect from future class topics. The final paper was way too long./This is the best class I have ever taken at Northwestern. Aaron clearly loves this class and it shows in the way he teaches. If you don't do the readings it's not the end of the world but you have to pay attention in class and the discussion questions will save your life. Definitely recommend this class and this professor./Aaron is a great professor with a ton of enthusiasm. Even if you don't feel the same passion for the course material he does, his energy will inspire you to learn. Take this class!/This was a really interesting class in a subject that is so important to living in our increasingly internet-centered world. Before taking the class I didn't have a strong interest in online communities and I didn't consider myself very much a part of the online community culture, but now I've definitely gotten more involved and Ive come to understand it better. This class gave me a good overall survey of internet history and online communities today, and made me realize the importance of getting involved in the internet./This was a fascinating and thought-provoking class that did a good job summarizing a lot of the contemporary issues surrounding the Internet and online communities. Professor Shaw was extremely enthusiastic about the course material, and frequently encouraged his students to contribute their own opinions during class discussions. Furthermore, in the course of the quarter, he brought in several fascinating, high-profile guest speakers who had firsthand experience with the subjects we were covering. Although the structure of the class often felt unbalanced (for instance, the occasional in-class group activities were a waste of time, in my opinion), this was forgivable, since it's a brand-new course and Prof. Shaw was experimenting a bit. Overall, while there were problems with this class, it was ultimately a very worthwhile experience, and has introduced me to topics of study that I intend to explore more in the future.

Course and Teacher Evaluation Summary for Instructors

Subject: COMM_ST Communication Studies
Academic Term: Fall 2014
Class Number: 15707

Enrollment: 41
Responses: 29
% Responses 70.73

Class: 378-0-20 Online Communities and Crowds
Instructor: Aaron Shaw

Demographic Questions

	Education & SP	Communication	Graduate School	KGSM	McCormick	Medill	Music	Summer	SCS	WCAS
Your School	3	23	0	0	0	0	0	0	0	3

	a. Freshman	b. Sophomore	c. Junior	d. Senior	e. Graduate	f. Other
Your Class	0	9	13	6	0	0

	a. Distribution requirement	b. Major requirement	c. Minor requirement	d. Elective requirement	e. Other requirement	f. No requirement
Your reason for taking course (mark all that apply).	4	19	0	1	0	3

	1	2	3	4	5	6	Total Response	Average Response
Interest in subject before taking course.	0	2	4	10	5	7	28	4.39

Core Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
1. Provide an overall rating of the instruction	26	0	0	2	3	10	11	5.15
2. Provide an overall rating of the course.	26	0	0	2	8	10	6	4.77
3. Estimate how much you learned in the course.	26	0	0	3	7	12	4	4.65
4. Rate the effectiveness of the course in challenging you intellectually.	28	0	0	4	15	8	1	4.21
5. Rate the effectiveness of the instructor(s) in stimulating your interest in the subject.	28	0	1	3	2	11	11	5.00

Time-Survey Question

	Total Response	a. 3 or fewer	b. 4 - 7	c. 8 - 11	d. 12 - 15	e. 16 - 19	f. 20 or more
6. Estimate the average number of hours per week you spent on this course outside of class and lab time.	28	7	16	4	0	1	0

School and Department Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
7. Rate how well prepared the instructor was for the class.	28	0	0	1	2	8	17	5.46
8. Rate the effectiveness with which the instructor communicated course content and ideas.	27	0	0	1	6	12	8	5.00
9. Rate the instructional materials (texts, audiovisual materials, etc.) used in this course.	28	0	1	2	8	9	8	4.75
10. Rate the instructor's enthusiasm in teaching this class.	28	0	0	0	1	6	21	5.71
11. Rate the clarity of examples.	27	0	1	1	2	15	8	5.04
12. Rate the helpfulness of the instructor outside of class.	25	0	0	2	2	10	11	5.20
13. Rate how well the instructor encouraged you to think for yourself.	28	0	0	0	5	11	12	5.25
14. Rate how well the exams stress important points of the lectures/readings/discussions.	19	0	0	1	6	8	4	4.79
15. Rate how much effort you put into this course in comparison to other courses.	28	0	1	2	15	8	2	4.29

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
16. Rate the instructor's efforts to highlight the interesting nature of the subject.	27	0	0	0	6	9	12	5.22
17. Rate how well the instructor contrasted the implications of various theories.	28	0	0	1	6	13	8	5.00
18. Rate how well this course helped you improve your ability to read critically.	27	1	1	12	8	1	4	3.70

Essay Questions

1. What are the primary teaching strengths of the instructor(s)?

prof shaw is super enthusiastic and very knowledgeable --- he is an active participant in some of the fields that we discuss and the array of guest speakers was fascinating. We had a lot of class exercises that engaged us in participatory ways that other classes don't utilize. Also the groupwork section, although at times frustrating, helped demonstrate some of the topics the course is attempting to deal with as well./Enthusiasm, knowledge of topic, connecting ideas, challenging the class to think for themselves, building assignments that challenged the class/Aaron's great. He really cares about his students and tries his best to cater to each student's needs./Really captivating lecturer and incredibly knowledgable about current founders/innovators of online communities./Aaron will always try and engage the class with the materials and is an amazing lecturer. I had little interest in the course subject prior to taking the class, but he made it more than bearable./I liked his enthusiasim./Aaron is a dynamic speaker who brought in multiple facets of online communities and crowds to make for a very interesting and applicable course not just for online C&Cs but offline communities as well./Aaron is very passionate about this class. He is also really great at engaging students to discuss in class, and at taking key points out of jumbled students' reponses/He was well-prepared for class. He had a good mix of lectures and group discussion/activities./Aaron is SUPER nice and SUPER helpful! He's really enthusiastic about the material which keeps you interested. He seems to really understand students and breakdown the material nicely during class./Enabling discussion/I enjoyed Shaw's teaching style he's very enthusiastic and loves audience participation. I loved the guest speakers./Aaron is an exceptionally nice professor, and this allowed me to learn in more relaxed manner, which actually helped me to digest his lectures more. He also focused much on open discussion, which really brought out the strength of this course: the study in ever-changing field of online communities. His occasional activities that he showcases to demonstrate how some communities behave were a lot of fun as well while equally being instructive./Aaron diverges from the typical lecture style class and tries to incorporate games and activities with lessons associated with the material. He also brings in fabulous guests. He is always enthusiastic and very engaging./I could see that Prof. Shaw clearly cared about what he was teaching. The teamwork and experiential learning portion was not very pleasant at first, but I realized in the end that they were necessary to really understand the course material.

2. What are the primary weaknesses, if any, of the instruction?

Maybe trying to cover too much (or assuming we know more than we already do)? I think that a lot of the technical stuff at the beginning went over my head, and probably the heads of other students too, so maybe more time discussing exactly what the internet is...or...well, that part of the course was the most hazy./None./None./I felt like some of the assignments could have been constructed better and I felt like sometimes the class discussions didn't really focus on the topics at hand/Class discussions could be more engaging./Inserting the big picture at the beginning of lecture and coming back to it at the end./Some concepts could have been explained better because I'm not very sure about certain code and Internet concepts./Few weaknesses./Some of the key concepts needed more clarification/explanation or link to the rest of the course/class. For example, I was not sure what Phreaks/pirates portion had to do with the rest of the course. While

the variety of opinions from students was insightful, the discussions would have benefited from a bit more structure.

3. Can you offer suggestions for improvement?
More TED talks!! Enjoyed watching those a lot more than some of the readings./I would suggest taking the group work out of the course/The group work was great, but my group finished the wikipedia assignment early, and as such the friday classes were sort of a waste of time/Less readings./Explain everything and don't assume everyone knows what certain acronyms mean./The grading system seemed spontaneous at times, and we were not really sure where we stood in terms of grading scale./Having students get involved with other online communities, not just Wikipedia./The course got better toward the end with the midquarter evaluation.

4. Did the course help you learn? Why or why not?
so much material that isnt covered comprehensively or in context together in other parts of the university. very grateful./Yes, unfamiliar topics (e.g. open source) were covered and critical thinking was required./Yes. I did not know much about online communities and this class was an incredible introduction/overview to the topic./Yes, the course did help me learn especially because of the experiential learning portion of the course./Definitely. I was made more aware of the dynamics and challenges faced by crowds and communities not just in the online context but offline as well. The guest speakers came from a variety of backgrounds which lent itself strongly the applicability of the course across disciplines./Absolutely, aaron was extremely engaging/Yes. I didn't know much about this subject before. I liked that it didn't focus on social media sites, because many other communication studies class do focus on them./Yes but more so if you read the articles and watch the assigns videos./Yes; Aaron's instructions opened up a new line of thinking in terms of online communities in ways that I haven't really taken a time to think about before./By doing assignments such as Wikipedia writing, we must actively engage with online communities, instead of thinking about them hypothetically./The course provided with me the critical thinking tools with which to tackle issues on my own. I think the greatest strength of this course is that it focused more on empowering students to think on their own than stuffing them with knowledge.

5. Please summarize your reaction to this course focusing on the aspects that were most important to you.
this is a FANTASTIC course - it will teach you many things that no other course at NU will teach you about---if you are at all interested in the way that society is changing due to the internet's all-pervasive influence (socially, economically, politically, legally, it's all covered!) - than take/TAKE THIS CLASS if you have any interest in the subject. If you hate group projects, don't take it./Great class! Aaron is super awesome and very understanding. Great guy. 10/10 would take again~/Great course. Positive experience. Awesome professor (the most jolly person I've ever met) and TA/Great class- Shaw is awesome. One of the best courses I've taken in Comm Studies so far!/Really intriguing class to take. Enjoyed being able to actually create wikiHow and wikipedia articles. Learned so much about online communities and methods to improve socialization of online communities./I wasn't very interested in the course material but Aaron was a great professor and was able to get people excited about the material. I would highly recommend taking this class to learn about a new world and be taught by someone who really cares about it!/I liked the course!/I am pleased with how much the course covered and how relevant it is to our technology-saturated society today. However, I felt it difficult to contribute to the discussions at the beginning of the course because it seemed to require prior knowledge of economics or the Commons (something I have no prior knowledge in). It made the beginning topics in class a bit inchoate. However, things really pulled together and the end and the key ideas at the beginning began to relate to the present ideas and formed a clearer big picture./This class is way better than I thought it would be. Both the professor and the TA or enthusiastic about the subject. Professor Shaw doesn't just lecture the whole time, but involves the class. There is a lot of group work, which sometimes gets annoying. The majority of your grade is based of individual writing, though./This class was pretty interesting and not that difficult. Both Aaron and Sneha were great. The readings weren't too long and were usually pretty interesting, and if you can't finish some of the readings occasionally, they are pretty good about covering the material in class./Some readings are interesting while others get a bit technical. No exams. Only exercises, papers and projects./I learned to look at the social media platforms I participate in and online communities I utilize for my own benefits in a different light. I mostly saw the complexity of some

seemingly mundane communities, and learned about the challenges they encountered and the tools to overcome them./I would highly recommend this course. The subject matter is interesting and Aaron and Sneha were always excited to be in class with us. They explained concepts well and introduced me to many new areas of study./Mostly interesting, especially in actually getting involved on sites like Wikipedia but the course can get very technical at times, which I struggled with/Shaw is the best. He knows a lot about the subject and is able to portray that in a very clear, manageable fashion. Some of the readings can be a bit long, however you can get away with skimming them. Make sure to go to class and stay actively engaged. The assignments were a nice change from the typical paper/exam format of a lot of classes./I enjoyed Shaw's teaching style he's very enthusiastic and loves audience participation. I loved the guest speakers. The entire class you have to do work in assigned groups of 4 people and you have multiple collaborative assignments but the wikipedia assignment is the worst and most annoying. The final paper you get to choose your own online community to research. Sneha was very accessible and approachable and open to questions at anytime./Aaron is possibly one of the most laid-back and nicest professors at NU. His method of teaching relies on not just pure lectures and distribution of knowledge, but open discussions that promotes the education in the ever-changing field of online communities. His lectures can be monotonous at times if students are not participating fully, however./A fabulous and engaging course that anyone interested in community building online and offline should take. Project, essay, and reflection based class style instead of exam based./This course taught me to think understand, critically about, and analyze online communities and analyze community dynamics.

Northwestern

Individual Report for Aaron Shaw (COMM_ST_378-0_20: Online Communities and Crowds)

Course and Teacher Evaluations CTEC Fall 2016

Project Audience 36
Responses Received 24
Response Ratio 66.7%

Report Comments

Course and Teacher Evaluations are intended solely for the use of faculty, staff and students of Northwestern University. Any reproduction , republication or redistribution of this site's content is prohibited without the express permission of Northwestern University.

For more information about the CTEC Guidelines please visit our website: http://www.northwestern.edu/ctec/guidelines/index.html

Creation Date Tue, Jan 03, 2017

Powered by blue

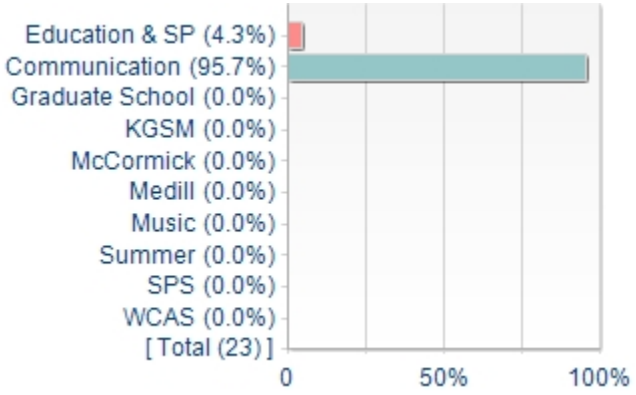
Northwestern University

Course Evaluations

Instructor	Course
Aaron Shaw	**COMM_ST_378-0_20: Online Communities and Crowds**

DEMOGRAPHICS

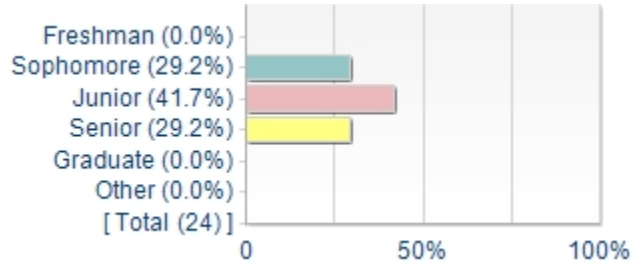
Your School



Options	Count	Percentage
Education & SP	1	4.3%
Communication	22	95.7%
Graduate School	0	0.0%
KGSM	0	0.0%
McCormick	0	0.0%
Medill	0	0.0%
Music	0	0.0%
Summer	0	0.0%
SPS	0	0.0%
WCAS	0	0.0%

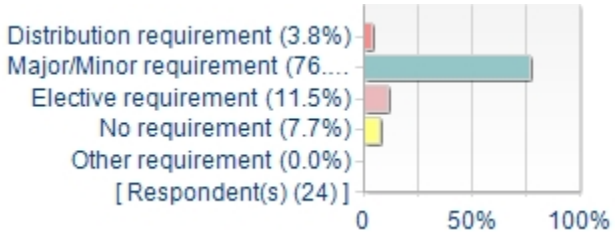
Copyright Northwestern University

Your Class



Options	Count	Percentage
Freshman	0	0.0%
Sophomore	7	29.2%
Junior	10	41.7%
Senior	7	29.2%
Graduate	0	0.0%
Other	0	0.0%

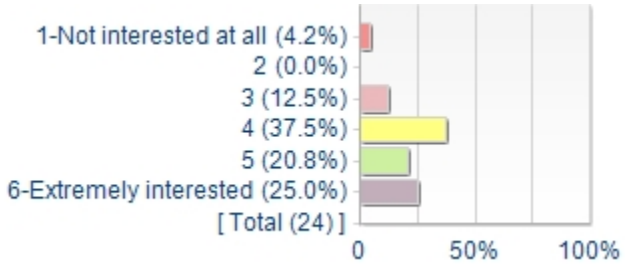
What is your reason for taking the course? (mark all that apply)



Options	Count	Percentage
Distribution requirement	1	3.8%
Major/Minor requirement	20	76.9%
Elective requirement	3	11.5%
No requirement	2	7.7%
Other requirement	0	0.0%
Respondent(s)	**24**	

Copyright Northwestern University

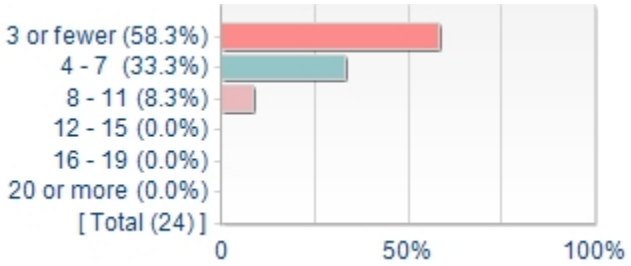
What was your Interest in this subject before taking the course?



Options	Count	Percentage
1-Not interested at all	1	4.2%
2	0	0.0%
3	3	12.5%
4	9	37.5%
5	5	20.8%
6-Extremely interested	6	25.0%

TIME-SURVEY QUESTION

Estimate the average number of hours per week you spent on this course outside of class and lab time.

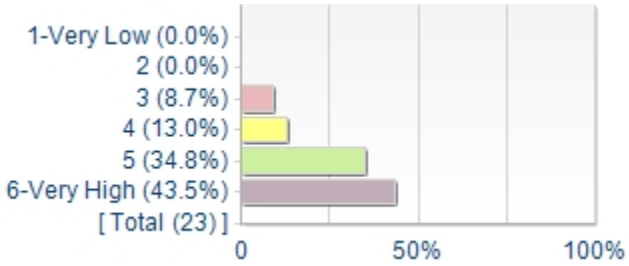


Options	Count	Percentage
3 or fewer	14	58.3%
4 - 7	8	33.3%
8 - 11	2	8.3%
12 - 15	0	0.0%
16 - 19	0	0.0%
20 or more	0	0.0%

Copyright Northwestern University

COURSE QUESTIONS

Provide an overall rating of the course.

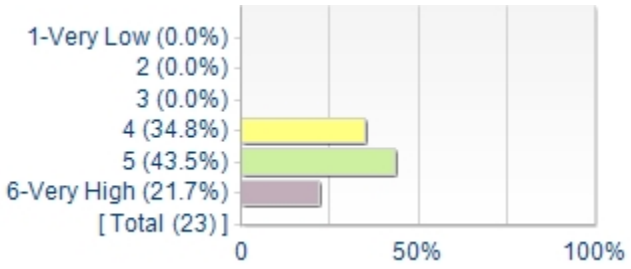


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	2	8.7%
4	4	3	13.0%
5	5	8	34.8%
6-Very High	6	10	43.5%

Statistics	Value
Response Count	23
Mean	5.13
Median	5.00
Standard Deviation	0.97

Copyright Northwestern University

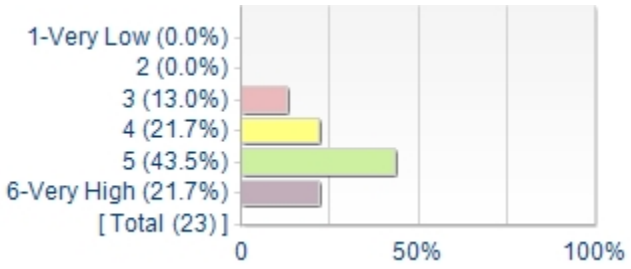
Estimate how much you learned in the course.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	8	34.8%
5	5	10	43.5%
6-Very High	6	5	21.7%

Statistics	Value
Response Count	23
Mean	4.87
Median	5.00
Standard Deviation	0.76

Rate the effectiveness of the course in challenging you intellectually.

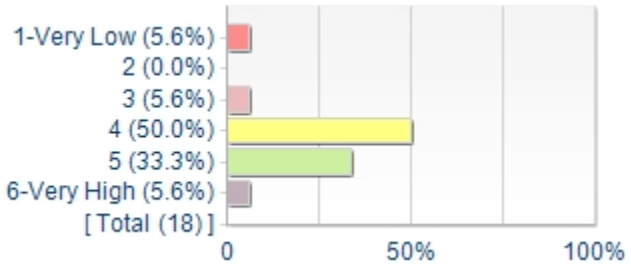


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	3	13.0%
4	4	5	21.7%
5	5	10	43.5%
6-Very High	6	5	21.7%

Statistics	Value
Response Count	23
Mean	4.74
Median	5.00
Standard Deviation	0.96

Copyright Northwestern University

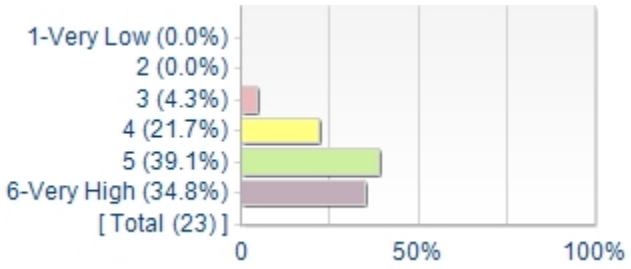
Rate how well the prerequisite course prepared you for this course.



Options	Score	Count	Percentage
1-Very Low	1	1	5.6%
2	2	0	0.0%
3	3	1	5.6%
4	4	9	50.0%
5	5	6	33.3%
6-Very High	6	1	5.6%

Statistics	Value
Response Count	18
Mean	4.22
Median	4.00
Standard Deviation	1.06

Rate how well this course helped you to think clearly about this subject.



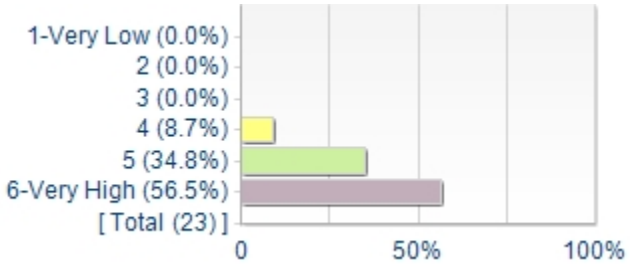
Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	1	4.3%
4	4	5	21.7%
5	5	9	39.1%
6-Very High	6	8	34.8%

Statistics	Value
Response Count	23
Mean	5.04
Median	5.00
Standard Deviation	0.88

Copyright Northwestern University

INSTRUCTOR QUESTIONS

Provide an overall rating of the instruction.

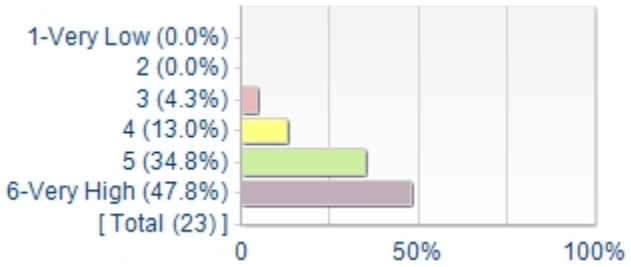


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	2	8.7%
5	5	8	34.8%
6-Very High	6	13	56.5%

Statistics	Value
Response Count	23
Mean	5.48
Median	6.00
Standard Deviation	0.67

Copyright Northwestern University

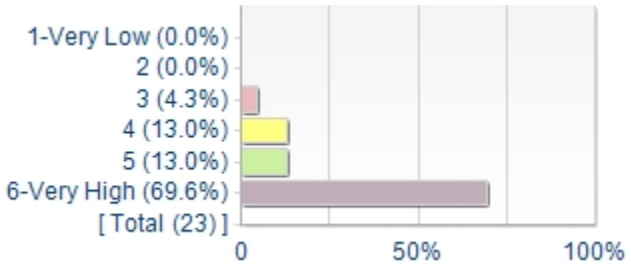
Rate the effectiveness of the instructor in stimulating your interest in the subject.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	1	4.3%
4	4	3	13.0%
5	5	8	34.8%
6-Very High	6	11	47.8%

Statistics	Value
Response Count	23
Mean	5.26
Median	5.00
Standard Deviation	0.86

Rate how well the instructor displayed interest in whether students learned.

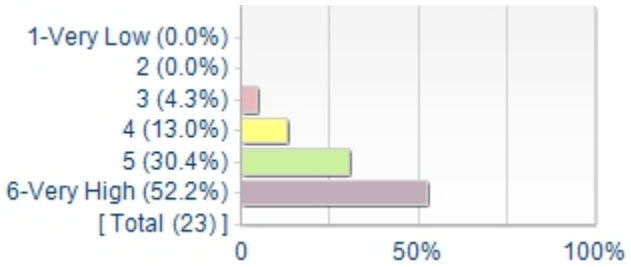


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	1	4.3%
4	4	3	13.0%
5	5	3	13.0%
6-Very High	6	16	69.6%

Statistics	Value
Response Count	23
Mean	5.48
Median	6.00
Standard Deviation	0.90

Copyright Northwestern University

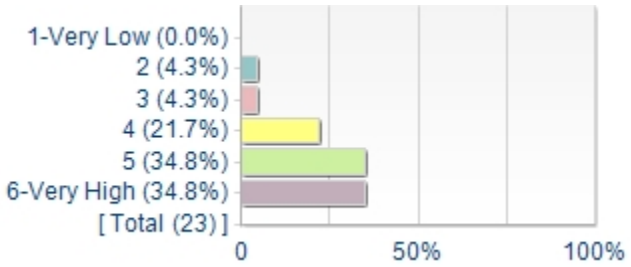
Rate how well the instructor provided constructive evaluations/feedback.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	1	4.3%
4	4	3	13.0%
5	5	7	30.4%
6-Very High	6	12	52.2%

Statistics	Value
Response Count	23
Mean	5.30
Median	6.00
Standard Deviation	0.88

Rate how well the instructor used class time.

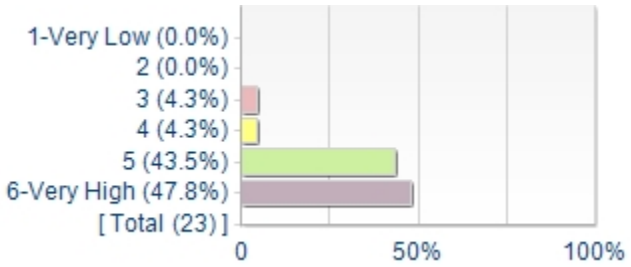


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	1	4.3%
3	3	1	4.3%
4	4	5	21.7%
5	5	8	34.8%
6-Very High	6	8	34.8%

Statistics	Value
Response Count	23
Mean	4.91
Median	5.00
Standard Deviation	1.08

Copyright Northwestern University

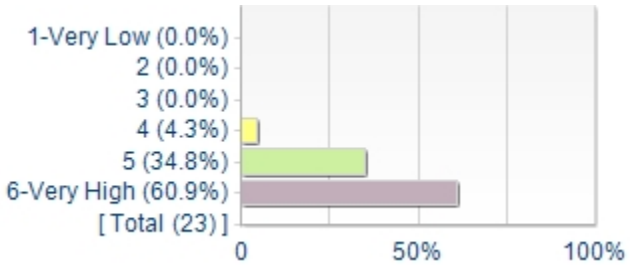
Rate the clarity with which the instructor communicated his/her ideas.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	1	4.3%
4	4	1	4.3%
5	5	10	43.5%
6-Very High	6	11	47.8%

Statistics	Value
Response Count	23
Mean	5.35
Median	5.00
Standard Deviation	0.78

Rate the effectiveness of the instructor in communicating with the class.

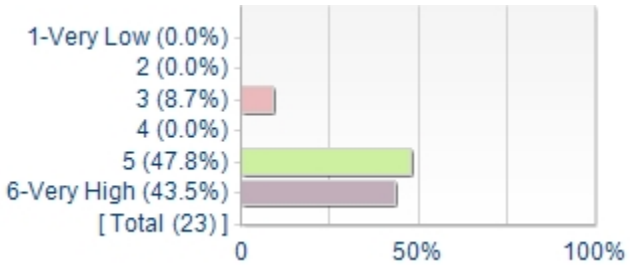


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	1	4.3%
5	5	8	34.8%
6-Very High	6	14	60.9%

Statistics	Value
Response Count	23
Mean	5.57
Median	6.00
Standard Deviation	0.59

Copyright Northwestern University

Rate how well the instructor presented class activities in an organized fashion.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	2	8.7%
4	4	0	0.0%
5	5	11	47.8%
6-Very High	6	10	43.5%

Statistics	Value
Response Count	23
Mean	5.26
Median	5.00
Standard Deviation	0.86

OPEN-ENDED QUESTIONS

Copyright Northwestern University

Did the course help you learn? Why or why not?

Comments
Yes. Great discussions and new perspectives.
Introduced core concepts that really allow you to examine OCCs with a fresh eye.
Yes a lot
The amount that I ended up learning within this class was amazing. The mix of history that was taught and working with real-time examples (like Wikipedia) allowed for a really interesting experience of observation.
somewhat yes - all of our discussions were around online communities which is so applicable and relevant to today's world. My awareness about these communities increased but sometimes ideas felt jumbled
OCC was definitely one of my favorite courses I've taken at Northwestern. The content was stimulating, engaging, and eye-opening.
Yes, but not as much as I could have learned had there not been as many longnse readings that took time away from actually thinking about the material.
Since this class relies more on exams and student exploration than exams and specific information students are meant to retain and regurgitate, this course feels designed for students to get out of it however much they put into it, which seems like a good approach for the subject material.
Definitely. It provided theoretical frameworks for thinking about online communities, and the examples helped provide points of comparison.
Yes - highly involved students and proper classroom management really made this class vibrant and participatory. Professor Shaw's allocation of 30% of the grade to participation helps with this... I thought it was way too much at first but can really see how it made the class as enjoyable as it was.
Yes, I learned a lot. Aaron is such a great teacher, the material was clear and explained simply and step-by-step and in various contexts.
Yes! It was super interesting and different.

Copyright Northwestern University

Please summarize your reaction to this course focusing on the aspects that were most important to you.

Comments
Aaron is a passionate, engaging professor who really cares about his students. I definitely recommend taking this class. The work is manageable and interesting. You learn so much. The TA, Jeremy, is awesome too.
Great class! Aaron and Jeremy are both amazing. In class discussions were the best and the class was overall fun yet insightful.
Do the readings or at least skim cause otherwise class is confusing. SO cool that you get to join wikipedia and actually write an article! Great class great prof
If you're interested in Digital Media, the internet, or just interesting digital history and interaction in general- this is the class for you. Aaron is an amazing professor, and you can tell that he is incredibly knowledgeable and excited about this topic. His lectures are always filled with lots of interesting facts, and the minimal assignments are all helpful to understanding the content presented.
Great course overall. Wasn't what I was necessarily expecting. He goes into a lot of depth about the dynamics of online communities and what makes them work rather than the people behind the communities themselves. IMPORTANT STUFF, If you go to class and occasionally contribute, you'll get the grade. 2 papers, 4 and 8 pages. A couple of assignments on the side and a group project that isn't too much work either. Also the course is basically all about Wikipedia so be prepared for that
If you're interested in online communities, take this class. If not, it's kind of boring. Aaron is great though and super enthusiastic and personable
Take OCC if you can. Aaron is a wonderful professor, and the course will open your eyes to so many aspects of online communities and crowds that most internet users take for granted. 10 would recommend.
Aaron is really enthusiastic about the material, which helps make it interesting. The readings are sometimes impossibly long and/or dense and there isn't as much focus on current online communities as I expected. Cool to learn about the history though and the underlying ideas of communication at play.
A pretty cool course if you're interested in this sort of thing. If the class description appeals to you, you'll probably like it, and if it doesn't, you probably won't. Not a huge workload, although a few of the projects can be a hassle if you procrastinate. Aaron's a great teacher who really cares about the content.
Great class! It definitely helps you think clearly and critically about modern social networks and websites and the class discussion is very engaging.
Brilliant stuff if you're interested in the subject matter. Professor Shaw encourages his students to take their learning and apply it to big-picture problems, even in other domains. Don't worry about the group project... Professor Shaw is one of the few at Northwestern to run proper group projects where each individual controls their own destiny: you will only be graded on your own actions, not those of others.
Aaron is great and Jeremy was an excellent TA. I do have to say though that this class is kind of like Aaron's scatterbrained self; the lack of a classic structure might be hard to work with for some people, but this was also what sparked really interesting conversations and tangents. You'll learn a lot from Aaron since this is his field of expertise, but he does try to pack in a lot of things into 10 weeks. For instance, part of the assignments is to edit Wikipedia which is kind of rushed. Overall, very interesting balance of hands-on and discussion based work. Aaron also gives you a lot of freedom to write about a topic that really interests you for your final assignment.
This is such a fun class, Aaron is an incredible teacher, I looked forward to every class.
Fantastic class that covers a whole lot, taught by a fantastic Professor (Aaron Shaw). Definitely worth taking.
This class is very little time outside of class, the lectures are not really relevant to the papers but really interesting, the teacher is great, and the papers are not graded too harshly if they are coherent and draw on necessary concepts!

Copyright Northwestern University

What are the primary teaching strengths of the instructor?

Comments
Receptive to feedback! So knowledgable. passionate. Love the no stakes quizzes, they get you really thinking / analyzing but in a non-stressful environment.
Class flight plans. Conductive feedback. Facilitating discussion. Writing on the board. Class website. Smiling and laughter. So many things. You'll be amazed how much the small things really made a difference.
High energy, great at helping us understand the material
Very knowledgeable and willing to work with students in a very constructive way. His discussions are always very thought provoking and consist of lots of interesting content presented in an engaging way.
Enthusiastic, knows his stuff and interacts very well with the class and audience
very enthusiastic and passionate about teaching the subject and loves working with students and hearing their ideas
Aaron is extremely organized, passionate, helpful, and genuinely cares about his students' learning experiences. There's not much I can say here that I haven't already mentioned in the mid-term course evaluation.
Genuine enthusiasm about the topic, very approachable and helpful, was excellent at getting and maintaining attention/interest of students
Aaron is very passionate about the subject material, and does a good job of helping instill that passion in students who weren't all that interested at the beginning of the course, like myself.
Enthusiastic and knowledgeable! He was very good at facilitating class discussion as well.
Great lectures, content, and readings. Everything seemed to create vivid and active class participation, with brilliant discussions going around the room. The two main papers were an effective way to allow us to apply our learning. The professor took our feedback seriously and used it to meaningfully improve his course throughout the quarter - how refreshing! :) A small thing that makes a big difference: Professor Shaw is one of the few at Northwestern to run proper group projects where each individual controls their own destiny: you will only be graded on your own actions, not those of others. Thank you for doing this, it seriously reduces stress and prevents the horrible free-loading which seems to permeate through the SoC at times whenever group work comes up.
He's the best.
He rocks!

What are the primary weaknesses, if any, of the instruction?

Comments
n`
I truly have no negative comments about the class or teaching, I thought it was structured perfectly for the topic at hand.
BUT, there were way too many class activities, the occasional one is fine but constantly having them gets overwhelming.
seems a bit disorganized
N`.
At times, he can rely a bit too much on student discussion, whether its amongst ourselves in small groups or as a part of lecture.
None really
Friday Labs were inconsistent - see suggestions for improvement below.
None.
None

Copyright Northwestern University

Can you offer suggestions for improvement?

Comments
n`
None really. Unless you have the answer to how to test your students to do all the readings to help fuel discussion, keep focusing on making small incremental improvements (check out http://jamesclear.com/marginal-gains) here and there and your class would be even more stellar.
You changed this at the midpoint of the quarter, but outlining and highlighting what is important and what to look for in the articles was very helpful!
narrow down specific focal points for the readings and intertwine Wikipedia article creation into the rest of class
Like a lot of us have said before, perhaps bring in elements from current and relevant online communities as application points throughout the quarter to keep students (who may not be as interested in the WELL or Wikipedia as they would be in Tumblr or reddit) engaged.
Don't divide the class into small groups to discuss the readings on our own as often (although that might just be a personal preference)
Nope
Literally only one thing. The Friday Labs: pain in the butt to get to, and half the weeks we didn't even do anything that required being physically present. Attendance to these was also poor. I would seriously suggest changing the class to 2 1 hour 50 minute sessions per week and incorporating the lab material into there. I get that there are benefits to separating the classes, but think proper time management would allow this to happen and make things better for students.
Nope.
Would have liked more feedback from the Professor himself
No he rocks

Copyright Northwestern University

Northwestern

Individual Report for Aaron Shaw (MTS_501-0_20: Intro to Grad Research in MTS)

Course and Teacher Evaluations CTEC Fall 2016

Project Audience 16
Responses Received 12
Response Ratio 75.0%

Report Comments

Course and Teacher Evaluations are intended solely for the use of faculty, staff and students of Northwestern University. Any reproduction , republication or redistribution of this site's content is prohibited without the express permission of Northwestern University.

For more information about the CTEC Guidelines please visit our website: http://www.northwestern.edu/ctec/guidelines/index.html

Creation Date Tue, Jan 03, 2017

Powered by blue

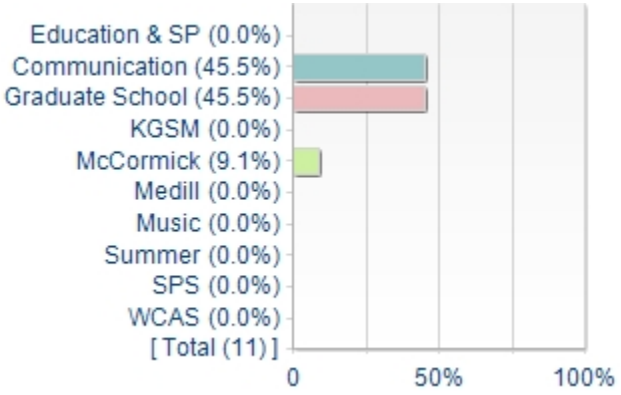
Northwestern University

Course Evaluations

Instructor	Course
Aaron Shaw	**MTS_501-0_20: Intro to Grad Research in MTS**

DEMOGRAPHICS

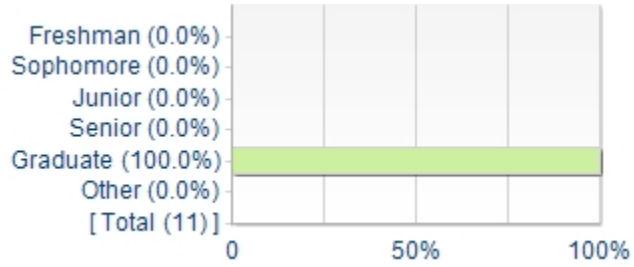
Your School



Options	Count	Percentage
Education & SP	0	0.0%
Communication	5	45.5%
Graduate School	5	45.5%
KGSM	0	0.0%
McCormick	1	9.1%
Medill	0	0.0%
Music	0	0.0%
Summer	0	0.0%
SPS	0	0.0%
WCAS	0	0.0%

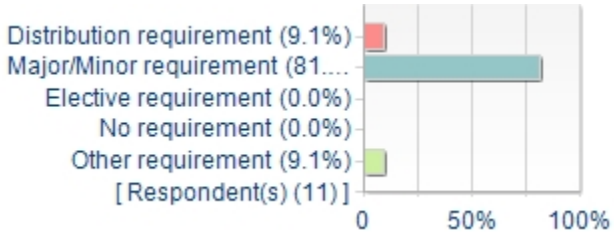
Copyright Northwestern University

Your Class



Options	Count	Percentage
Freshman	0	0.0%
Sophomore	0	0.0%
Junior	0	0.0%
Senior	0	0.0%
Graduate	11	100.0%
Other	0	0.0%

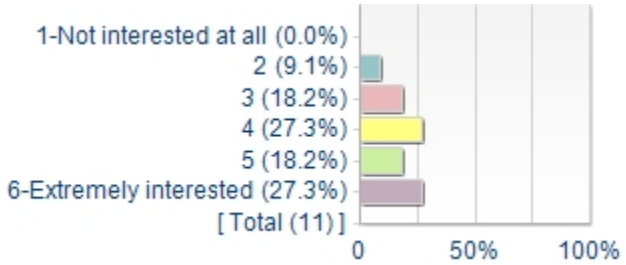
What is your reason for taking the course? (mark all that apply)



Options	Count	Percentage
Distribution requirement	1	9.1%
Major/Minor requirement	9	81.8%
Elective requirement	0	0.0%
No requirement	0	0.0%
Other requirement	1	9.1%
Respondent(s)	**11**	

Copyright Northwestern University

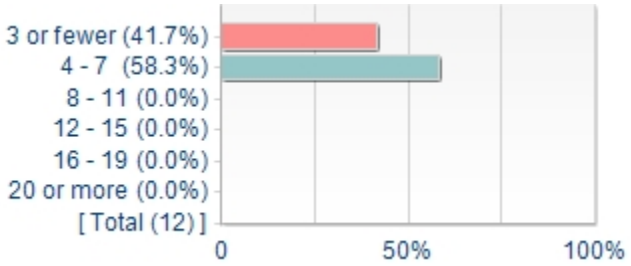
What was your Interest in this subject before taking the course?



Options	Count	Percentage
1-Not interested at all	0	0.0%
2	1	9.1%
3	2	18.2%
4	3	27.3%
5	2	18.2%
6-Extremely interested	3	27.3%

TIME-SURVEY QUESTION

Estimate the average number of hours per week you spent on this course outside of class and lab time.

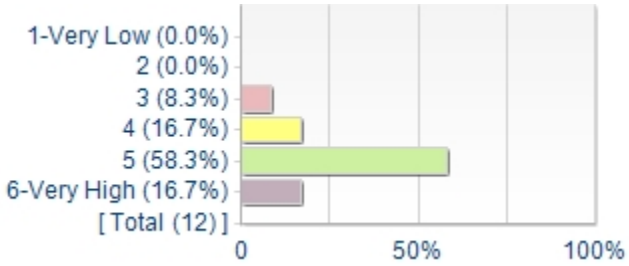


Options	Count	Percentage
3 or fewer	5	41.7%
4 - 7	7	58.3%
8 - 11	0	0.0%
12 - 15	0	0.0%
16 - 19	0	0.0%
20 or more	0	0.0%

Copyright Northwestern University

COURSE QUESTIONS

Provide an overall rating of the course.

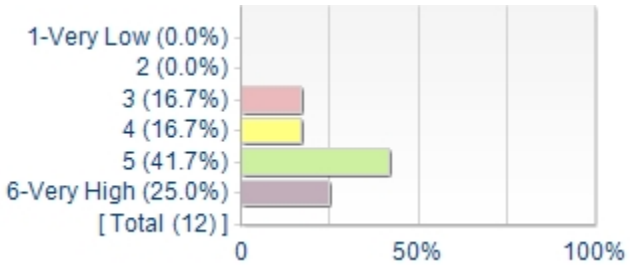


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	1	8.3%
4	4	2	16.7%
5	5	7	58.3%
6-Very High	6	2	16.7%

Statistics	Value
Response Count	12
Mean	4.83
Median	5.00
Standard Deviation	0.83

Copyright Northwestern University

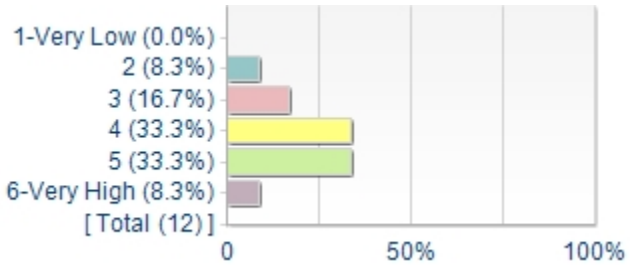
Estimate how much you learned in the course.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	2	16.7%
4	4	2	16.7%
5	5	5	41.7%
6-Very High	6	3	25.0%

Statistics	Value
Response Count	12
Mean	4.75
Median	5.00
Standard Deviation	1.06

Rate the effectiveness of the course in challenging you intellectually.

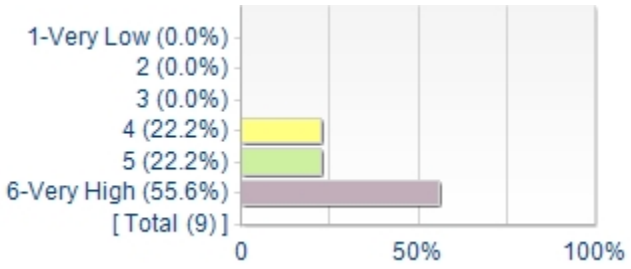


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	1	8.3%
3	3	2	16.7%
4	4	4	33.3%
5	5	4	33.3%
6-Very High	6	1	8.3%

Statistics	Value
Response Count	12
Mean	4.17
Median	4.00
Standard Deviation	1.11

Copyright Northwestern University

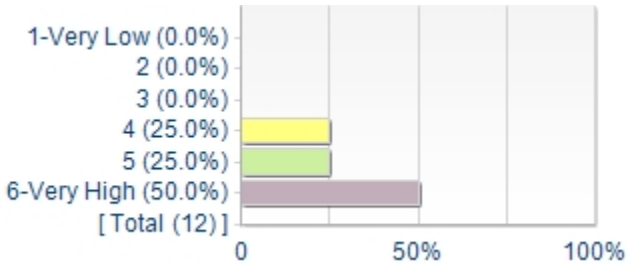
Rate how well the prerequisite course prepared you for this course.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	2	22.2%
5	5	2	22.2%
6-Very High	6	5	55.6%

Statistics	Value
Response Count	9
Mean	5.33
Median	6.00
Standard Deviation	0.87

Rate how well this course helped you to think clearly about this subject.



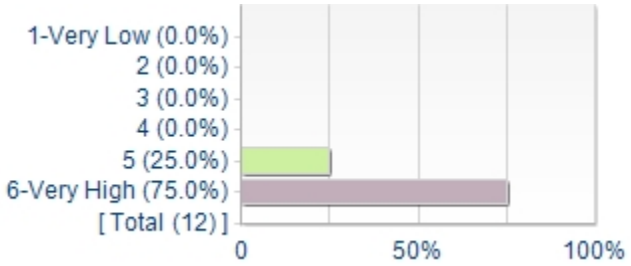
Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	3	25.0%
5	5	3	25.0%
6-Very High	6	6	50.0%

Statistics	Value
Response Count	12
Mean	5.25
Median	5.50
Standard Deviation	0.87

Copyright Northwestern University

INSTRUCTOR QUESTIONS

Provide an overall rating of the instruction.

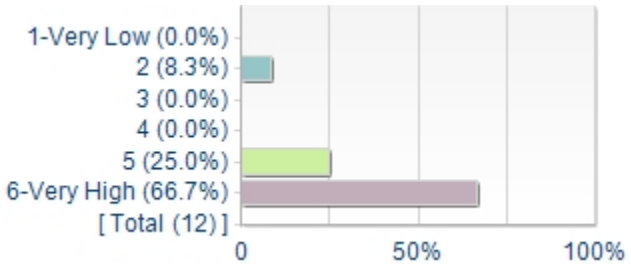


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	0	0.0%
5	5	3	25.0%
6-Very High	6	9	75.0%

Statistics	Value
Response Count	12
Mean	5.75
Median	6.00
Standard Deviation	0.45

Copyright Northwestern University

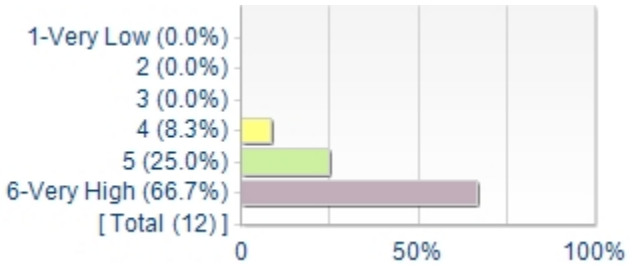
Rate the effectiveness of the instructor in stimulating your interest in the subject.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	1	8.3%
3	3	0	0.0%
4	4	0	0.0%
5	5	3	25.0%
6-Very High	6	8	66.7%

Statistics	Value
Response Count	12
Mean	5.42
Median	6.00
Standard Deviation	1.16

Rate how well the instructor displayed interest in whether students learned.

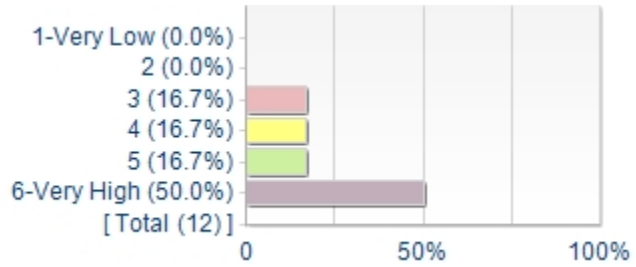


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	1	8.3%
5	5	3	25.0%
6-Very High	6	8	66.7%

Statistics	Value
Response Count	12
Mean	5.58
Median	6.00
Standard Deviation	0.67

Copyright Northwestern University

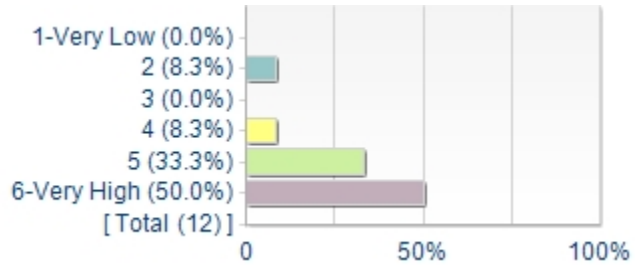
Rate how well the instructor provided constructive evaluations/feedback.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	2	16.7%
4	4	2	16.7%
5	5	2	16.7%
6-Very High	6	6	50.0%

Statistics	Value
Response Count	12
Mean	5.00
Median	5.50
Standard Deviation	1.21

Rate how well the instructor used class time.

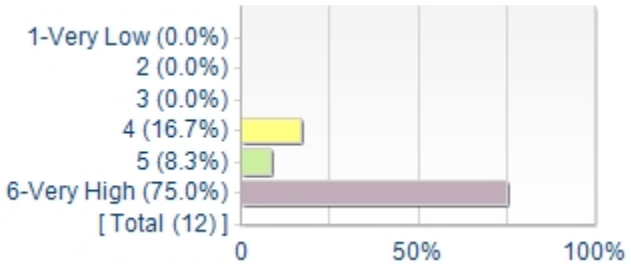


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	1	8.3%
3	3	0	0.0%
4	4	1	8.3%
5	5	4	33.3%
6-Very High	6	6	50.0%

Statistics	Value
Response Count	12
Mean	5.17
Median	5.50
Standard Deviation	1.19

Copyright Northwestern University

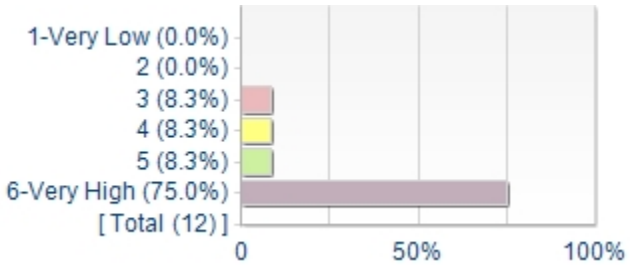
Rate the clarity with which the instructor communicated his/her ideas.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	2	16.7%
5	5	1	8.3%
6-Very High	6	9	75.0%

Statistics	Value
Response Count	12
Mean	5.58
Median	6.00
Standard Deviation	0.79

Rate the effectiveness of the instructor in communicating with the class.

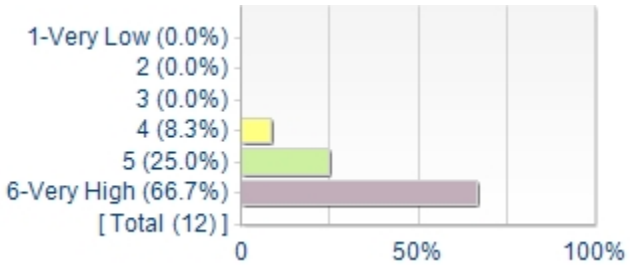


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	1	8.3%
4	4	1	8.3%
5	5	1	8.3%
6-Very High	6	9	75.0%

Statistics	Value
Response Count	12
Mean	5.50
Median	6.00
Standard Deviation	1.00

Copyright Northwestern University

Rate how well the instructor presented class activities in an organized fashion.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	1	8.3%
5	5	3	25.0%
6-Very High	6	8	66.7%

Statistics	Value
Response Count	12
Mean	5.58
Median	6.00
Standard Deviation	0.67

OPEN-ENDED QUESTIONS

Copyright Northwestern University

Did the course help you learn? Why or why not?

Comments
The course undoubtedly taught me many of the nuances of graduate school that would otherwise have to be learned through experience. I greatly appreciated this since it saves a lot of the hardship and frustration that I might have experienced had I not taken this course. However, there were times when the class was slightly disengaging and I struggled to focus, but this may stem from all my mentee experiences in undergrad.
This course helped me to understand the specifics of academic process in the graduate school and beyond.
Yes, I thought the course did a fine job of covering a breadth of topics relevant to students.
Yes - readings and topics were pertinent and relatable. In a supportive seminar environment, I felt free to acknowledge my "known-unknowns" and discuss questions and concerns with colleagues in a similar position.
501 provided a great, structured space for reflection and time for planning for our academic career.
Yes, I really liked the guest faculty, and the discussions on conferences, grant writing and review processes.
I enjoyed the breadth of topics that this course tried to cover. I also appreciated how having comments due the day before class made everyone accountable -- this ensured that everyone had done enough of the readings and thought through the topic in advance enough to have a worthwhile discussion. I think that all of these topics would have eventually come up at some point during my experience as a grad student (probably in discussions with my labmates and adviser), and am glad that we, as a first-year cohort, had a chance to take a first stab at reflecting on these topics in my first quarter of the PhD. I will continue to reflect on these topics throughout my academic career.
Not really. I really enjoyed the assignments that had us talking to upperclassmen because I learned a lot from them and got great advice. The guest speakers were really nice because I got to meet many professors in TSB that I don't normally get the opportunity to interact with. However, the in class time I found quite ineffective. I didn't benefit from many of the discussions and felt like I came to this program with the skills that we were trying to teach in this course. There were a few very helpful things, like when we discussed the peer review process in depth, but those were pretty few.
Yes. This course definitely helps orient PhD students to the career of academia and the various aspects of how to remain as productive and creative as possible. I Liked …. - That this class allowed for open and honest dialogue that discussed the necessary tools to being a scholar without pretention. - That this class pointed to different literature that provided some guidelines for how to handle life outside of the classroom -That this class allowed us to meet faculty and hear what their specific interests are and learn about their academic journeys without awkward introduction.

Copyright Northwestern University

Please summarize your reaction to this course focusing on the aspects that were most important to you.

Comments
I really enjoyed the class and appreciate all the insight it offered into the intricacies of grad school. While some of the lectures were a little more challenging to focus in, they all provided a lot of valuable knowledge about my next five years.
I very much enjoyed class conversations and discussion sessions with invited speakers. I also found the readings extremely relevant and intellectually stimulating. Overall, this was a great course.
Good course that lays a solid foundation for starting graduate school
It gave a great overview of the career of an academic (e.g. the importance of interpersonal relations all the way to conference submission basics).
Aaron did a great job providing structured space for us to think about our goals for our graduate career and how we can achieve them. I appreciate having time to reflect on what I've accomplished thus far, and what skills I need to acquire moving forward in order to be successful. I also really liked having an opportunity to interact with MTSB faculty.
I think it's a valuable course which provides a good introduction to graduate school. I do, however, think that it can be shrunk to 2.5 hours versus its current 3-hour structure, as some of the discussions on the readings take place prior to class via Canvas.
* Reflect on your academic identity early and often. What excites you and what sparks your curiosities as a researcher? How do these sentiments then become research questions (which may be crystallized as puzzles, design challenges, or other forms)? * Be thoughtful about how you communicate your research for the sake of (in no particular order) publications, grants, job talks, networking, teaching, etc. * Collaboration has the strong potential for good but is not easy -- be prepared to discuss team communication, authorship expectations, tools, conflict resolution strategies, time commitments, the stakes of the situations (open-ended side project vs. heavy-duty conference publication), etc. whenever you consider collaborating with other academics. * Find mentorship in different places (the lab, student cohort, adviser, other faculty on campus, people one meets off-campus at conferences, etc) * The pipeline of peer-reviewed publications at conference and journal venues
I don't think this course answered the questions that I had about graduate research. Since everyone has different requirements and is in different places in their research it is hard to teach a full class that discusses this. However, I think we spent most of the time in 'therapy' mode and that didn't help me because I felt like I came in to school with organization and time-management experience. I would really wish that we had less open ended conversation because I feel like only a few people end up speaking, and it only benefits a few people and we end up going down rabbit-holes. I wish we had more direction with our guests. Maybe giving us something they wrote so we can engage with them so I don’t feel like it is a random person walking in the room and giving us their time that I haven’t prepared for. I also wish that we got any feedback. I’m not sure if I’m meeting any expectations for the course or for you or on how my reflections are.
I think this is a great introduction for people who have been removed from research for a bit, but is likely redundant for those who are less removed. My favorite aspect of the class were the visiting lecturers from the program coming to class and talking about things beyond their research interests. It was helpful to hear an array of tips for successes and work strategies. I wish each visiting lecturer had recommended their favorite publication from their own work. A step up, step back approach had made the discussions less tedious. A lot of the class is discussion based and the same voices typically dominated the conversation. This can happen in any class, but because it happened so frequently for us, I wish step up (if you have an idea that hasn’t been expressed or disagree with what’s already been said, please share) and step back (just nodding along if you have a similar experience/sentiment that's already been expressed) had been implemented.
This course is a great introduction to the expectations of Northwestern University and allows students the chance to create concrete goals for their tenure as a PhD student.

Copyright Northwestern University

What are the primary teaching strengths of the instructor?

Comments
Aaron brings a lot of energy to lecture and shows a genuine concern for student learning. I think this helps people stay engaged inside and outside of class regardless of their interest in the material itself.
Ability to engage everyone in a productive discussion, sense of humour.
Participatory style, very supportive, useful materials and assignments, great guest speakers
Conveys interest in students and their learning.
Highly engaged and prepared for discussion. Very clear and timely instructions on his website.
Aaron is awesome. Very engaging and asks great questions that challenge us to think more critically and realistically about a topic. He also organized the class in a way that made sense and was a great introductory course for the program. I really liked the course website, and that the readings were relevant to each class session and actually discussed in class. Aaron provided space that made us feel comfortable opening up while also sharing his experiences and expertise.
I appreciated how Aaron framed his assignments as personalized and open-ended -- everyone was writing about their own research interests, academic identity, prior experiences, etc. Overall I think most of the readings were relevant and provided insights that I would not have encountered in readings on my own (i.e. if we did not have a graduate student seminar). I also appreciated how there was a mix of large group discussion, pair discussion, and faculty guests. Aaron allowed open-endedness and flexibility in letting students think through their ideas by talking out loud and pursue different trains of thoughts throughout the class session, and chiming in as facilitator to recenter the discussion on relevant topics as needed.
Aaron is a great instructor and it was great getting the opportunity to get to know him and learn about how he approaches research.
Down to earth and patient.

What are the primary weaknesses, if any, of the instruction?

Comments
I can't think of any.
I haven't received any feedback or evaluations throughout the quarter and so do not know what to expect as a grade. That being said, I could have deliberately sought him out during office hours to discuss this mid-quarter.
Understanding Aaron's background, it makes sense but some aspects of the HCI-world were not as relevant to the work some of us do so having a balance that includes more examples of what happens in the social sciences would be helpful. I really liked having the opportunity to share personal stories and experiences in class, but I think some people dominated the conversation and there could have been more managing of the discussion that would have encouraged some people who stayed quiet to speak up. Particularly in a class that size, it would have been nice to hear from more than just a handful of people in the room.
Aaron's in-class comments and questions tend to be verbose -- this shows that he is engaged and is sincerely thinking through the discussion contents, but can occasionally be difficult to follow his train of thought. However he's definitely approachable, friendly enough to ask for clarification, so this isn't a problem in the long run.
I don't think Aaron conducted class time in the best way. We had many conversations that were long and drawn out and I don't feel they were particularly beneficial to anyone in the room. I didn't appreciate the 'therapy' emphasis that he put on the class.
Class often didn't stay on topic stated in syllabus.

Copyright Northwestern University

Can you offer suggestions for improvement?

Comments
I would maybe like an extra lecture on developing your research questions/interests and identifying publication venues.
Maybe invite more guest speakers with diverse background.
Even though publishing is very dependent on the field and venue, I think it would have been great to spend time on general rules for publishing and how to navigate that world in terms of picking a journal and the actual submission process (i.e., what information should you include in the cover letter, how to list affiliations). During the session on professional identity and CVs, it would be nice for the assignment to be working on creating your CV and getting feedback on what should be included and where. It would be nice in this course to cover very basic things like email signatures and LinkedIn profile information so that we know what is appropriate to include or how present ourselves to the world in a professional manner.
Firstly, here's a comment that's not about Aaron specifically but more about the class size␣ynamic: I wonder if this class would be better with a smaller class size -- our class size of eighteen-ish students had pros and cons. On the one hand, this meant that there was never an awkward silence, and you never felt pressured to speak just because no one else was speaking. On the other hand, it meant that we didn't get to know each of our classmates very well (unless they tended to be vocal with personal examples in the large group discussions, you already connected outside of class settings, or you happened to have pair discussions with them very often), and the interpersonal/social bonds weren't very deep. I don't have a simple solution to the class size topic, but I think that experimenting with different conventions for seminar discussions (e.g. nonverbal cues to express agreement or disagreement, establishing more opportunities to make it easier for less talkative␣xtraverted folks to chime in) could help tackle some of the cons of having a seminar with such a large student group.
If we went into more detail about writing grants, fellowships, IRB or other tangible things that would be great. Instead, I felt like we glanced over the material and just talked about our reactions. I wish we had more direction in some of these conversations. I feel like often they go into a direction with a few people talking a lot and not many people benefiting. I wish we talked more about things going on in the community, like MTS speaker series.
I think Aaron did great with this course, but the course itself has some limitations and constraints. This course is so broad that there's no one size fits all approach that would work for every student. The style that would work best for one would likely be disastrous for another. I think this is a difficult course to teach, but that Aaron did as great of job with it as could be done.
I Wish …. -That this class was not so early in the morning. -That this class allowed students the opportunity to present their potential research ideas. -We were able to hear from selected MTS/TSB Postdocs or Senior Students during the course as guest speakers.

Copyright Northwestern University

Northwestern

Individual Report for Aaron Shaw (MTS_503-0_1: The Practice of Scholarship)

Course and Teacher Evaluations CTEC Spring 2016

Project Audience 8
Responses Received 5
Response Ratio 62.5%

Report Comments

Course and Teacher Evaluations are intended solely for the use of faculty, staff and students of Northwestern University. Any reproduction , republication or redistribution of this site's content is prohibited without the express permission of Northwestern University.

For more information about the CTEC Guidelines please visit our website: http://www.northwestern.edu/ctec/guidelines/index.html

Creation Date Wed, Dec 07, 2016

Powered by blue

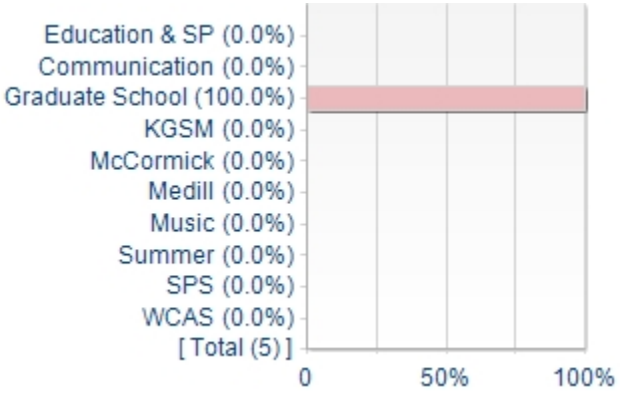
Northwestern University

Course Evaluations

Instructor	Course
Aaron Shaw	**MTS_503-0_1: The Practice of Scholarship**

DEMOGRAPHICS

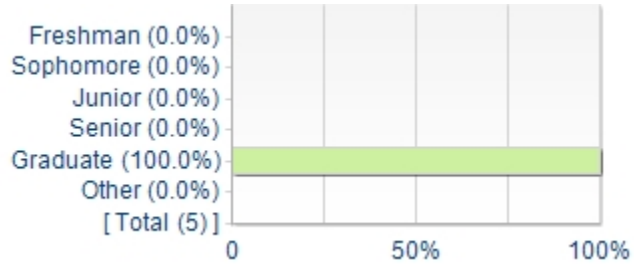
Your School



Options	Count	Percentage
Education & SP	0	0.0%
Communication	0	0.0%
Graduate School	5	100.0%
KGSM	0	0.0%
McCormick	0	0.0%
Medill	0	0.0%
Music	0	0.0%
Summer	0	0.0%
SPS	0	0.0%
WCAS	0	0.0%

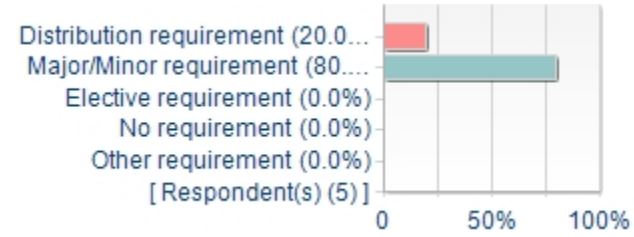
Copyright Northwestern University

Your Class



Options	Count	Percentage
Freshman	0	0.0%
Sophomore	0	0.0%
Junior	0	0.0%
Senior	0	0.0%
Graduate	5	100.0%
Other	0	0.0%

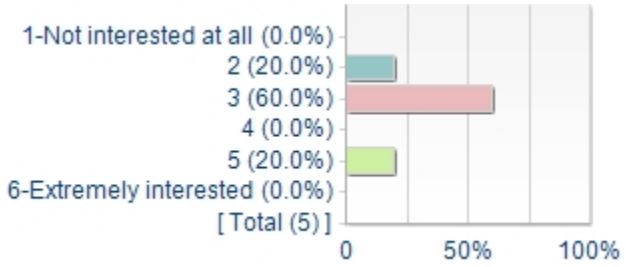
What is your reason for taking the course? (mark all that apply)



Options	Count	Percentage
Distribution requirement	1	20.0%
Major/Minor requirement	4	80.0%
Elective requirement	0	0.0%
No requirement	0	0.0%
Other requirement	0	0.0%
Respondent(s)	**5**	

Copyright Northwestern University

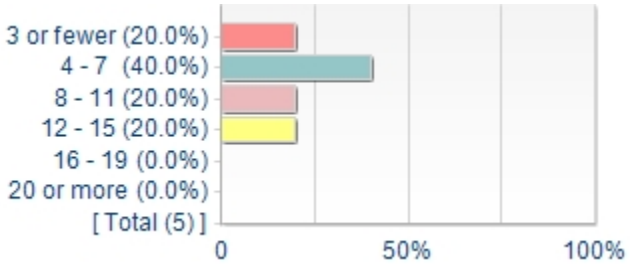
What was your Interest in this subject before taking the course?



Options	Count	Percentage
1-Not interested at all	0	0.0%
2	1	20.0%
3	3	60.0%
4	0	0.0%
5	1	20.0%
6-Extremely interested	0	0.0%

TIME-SURVEY QUESTION

Estimate the average number of hours per week you spent on this course outside of class and lab time.

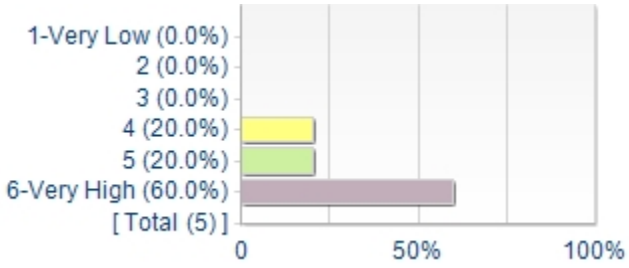


Options	Count	Percentage
3 or fewer	1	20.0%
4 - 7	2	40.0%
8 - 11	1	20.0%
12 - 15	1	20.0%
16 - 19	0	0.0%
20 or more	0	0.0%

Copyright Northwestern University

COURSE QUESTIONS

Provide an overall rating of the course.

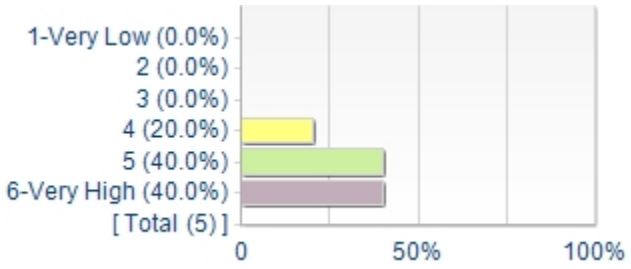


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	1	20.0%
5	5	1	20.0%
6-Very High	6	3	60.0%

Statistics	Value
Response Count	5
Mean	5.40
Median	6.00
Standard Deviation	0.89

Copyright Northwestern University

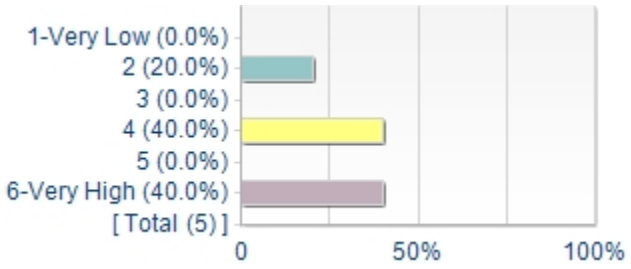
Estimate how much you learned in the course.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	1	20.0%
5	5	2	40.0%
6-Very High	6	2	40.0%

Statistics	Value
Response Count	5
Mean	5.20
Median	5.00
Standard Deviation	0.84

Rate the effectiveness of the course in challenging you intellectually.

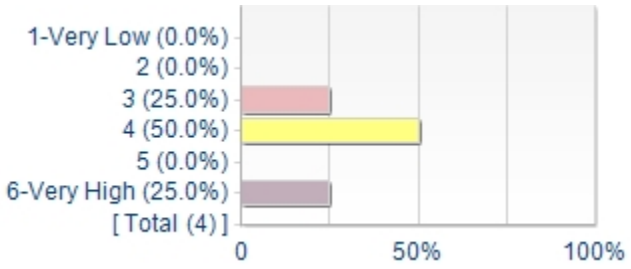


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	1	20.0%
3	3	0	0.0%
4	4	2	40.0%
5	5	0	0.0%
6-Very High	6	2	40.0%

Statistics	Value
Response Count	5
Mean	4.40
Median	4.00
Standard Deviation	1.67

Copyright Northwestern University

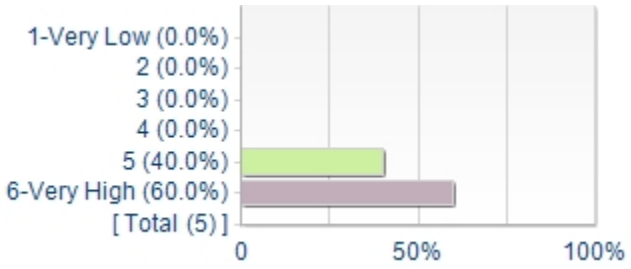
Rate how well the prerequisite course prepared you for this course.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	1	25.0%
4	4	2	50.0%
5	5	0	0.0%
6-Very High	6	1	25.0%

Statistics	Value
Response Count	4
Mean	4.25
Median	4.00
Standard Deviation	1.26

Rate how well this course helped you to think clearly about this subject.



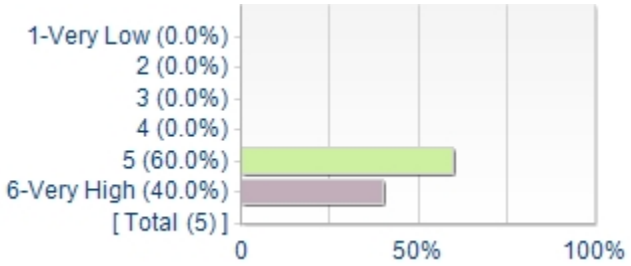
Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	0	0.0%
5	5	2	40.0%
6-Very High	6	3	60.0%

Statistics	Value
Response Count	5
Mean	5.60
Median	6.00
Standard Deviation	0.55

Copyright Northwestern University

INSTRUCTOR QUESTIONS

Provide an overall rating of the instruction.

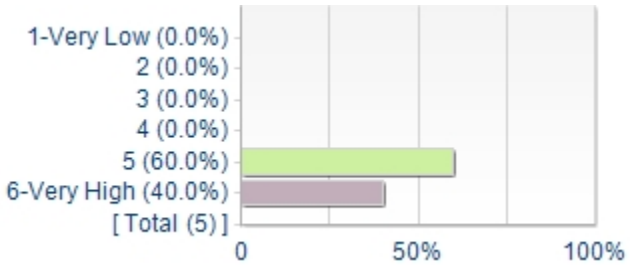


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	0	0.0%
5	5	3	60.0%
6-Very High	6	2	40.0%

Statistics	Value
Response Count	5
Mean	5.40
Median	5.00
Standard Deviation	0.55

Copyright Northwestern University

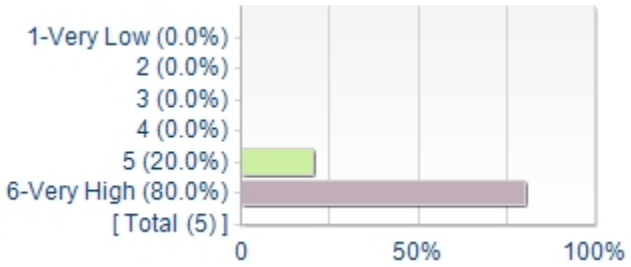
Rate the effectiveness of the instructor in stimulating your interest in the subject.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	0	0.0%
5	5	3	60.0%
6-Very High	6	2	40.0%

Statistics	Value
Response Count	5
Mean	5.40
Median	5.00
Standard Deviation	0.55

Rate how well the instructor displayed interest in whether students learned.

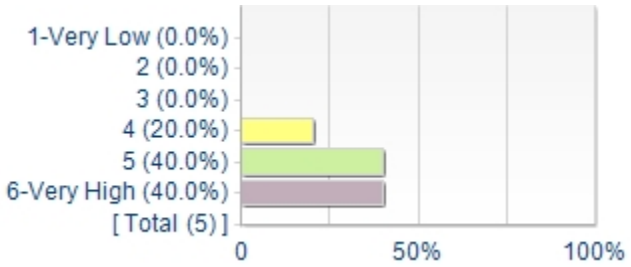


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	0	0.0%
5	5	1	20.0%
6-Very High	6	4	80.0%

Statistics	Value
Response Count	5
Mean	5.80
Median	6.00
Standard Deviation	0.45

Copyright Northwestern University

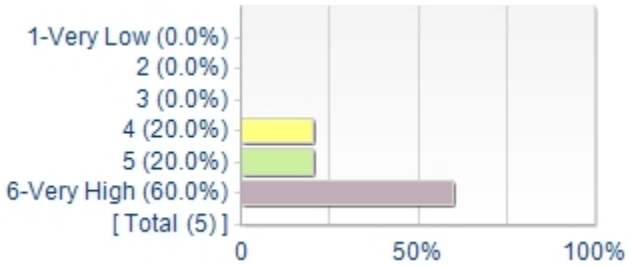
Rate how well the instructor provided constructive evaluations/feedback.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	1	20.0%
5	5	2	40.0%
6-Very High	6	2	40.0%

Statistics	Value
Response Count	5
Mean	5.20
Median	5.00
Standard Deviation	0.84

Rate how well the instructor used class time.

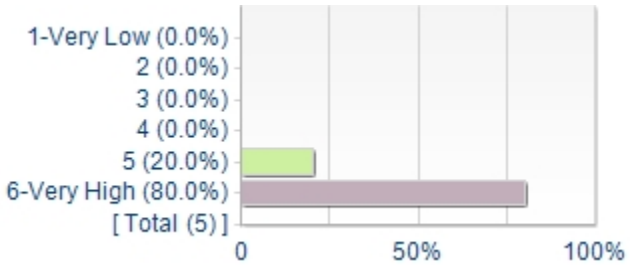


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	1	20.0%
5	5	1	20.0%
6-Very High	6	3	60.0%

Statistics	Value
Response Count	5
Mean	5.40
Median	6.00
Standard Deviation	0.89

Copyright Northwestern University

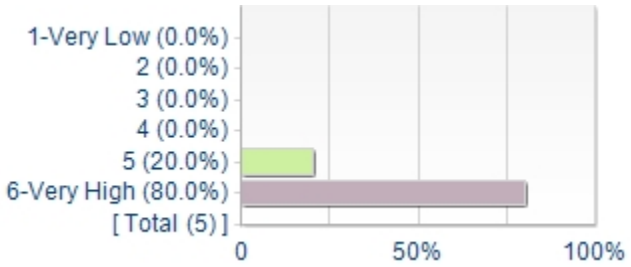
Rate the clarity with which the instructor communicated his/her ideas.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	0	0.0%
5	5	1	20.0%
6-Very High	6	4	80.0%

Statistics	Value
Response Count	5
Mean	5.80
Median	6.00
Standard Deviation	0.45

Rate the effectiveness of the instructor in communicating with the class.

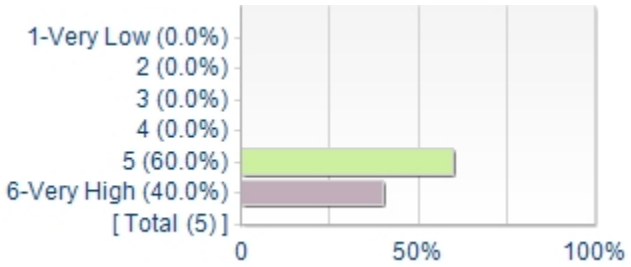


Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	0	0.0%
5	5	1	20.0%
6-Very High	6	4	80.0%

Statistics	Value
Response Count	5
Mean	5.80
Median	6.00
Standard Deviation	0.45

Copyright Northwestern University

Rate how well the instructor presented class activities in an organized fashion.



Options	Score	Count	Percentage
1-Very Low	1	0	0.0%
2	2	0	0.0%
3	3	0	0.0%
4	4	0	0.0%
5	5	3	60.0%
6-Very High	6	2	40.0%

Statistics	Value
Response Count	5
Mean	5.40
Median	5.00
Standard Deviation	0.55

OPEN-ENDED QUESTIONS

Did the course help you learn? Why or why not?

Comments
It was good to focus on the big picture of research and writing, and to think about the craft and the purposes. The readings were good, and the weekly exercises really helped.
This course has done a great job at helping us workshop through the process of writing an article for publication.
The course did help me learn. I feel that I'm leaving with a better understanding of the details involved in each step of pulling together an academic paper for publication.

Please summarize your reaction to this course focusing on the aspects that were most important to you.

Comments
I think that having the goal of preparing a paper for publication made this a much more interesting and useful class. It was good to think about research and writing in the context of an actual paper, and to be able to have an improved paper at the end is great.
This is a great class to help think strategically about preparing, framing and writing an article. If you do not have a project underway, it will definitely require a lot of time. At the same time, it is extremely useful in helping push out an article for publication. There are also abundant opportunities to offer and receive feedback on the writing among students.
I was positively surprised by this course. I had heard poor reviews from students in previous years (different instructor) and wasn't sure how much was the instructor and how much was that the class was just not a good idea. I think Aaron did a great job at structuring our time together and keeping us accountable and productive.

Copyright Northwestern University

What are the primary teaching strengths of the instructor?

Comments
Dr. Shaw is an enthusiastic and committed profesor with lots of great feedback. He is also more than willing to help accommodate the class requirements to the need of each student.
Aaron was organized, approachable and able to create a comfortable environment where students were able to share doubts and concerns about the writing process freely.

What are the primary weaknesses, if any, of the instruction?

Comments
Can't think of any.

Copyright Northwestern University

Course and Teacher Evaluation Summary for Instructors

Subject: MTS Media, Technology & Society
Academic Term: Fall 2013
Class Number: 15451

Enrollment: 8
Responses: 8
% Responses 100.00

Class: 525-0-22 Special Topics Research Seminar Collective Action & Org Online
Instructor: Aaron Shaw

Combined Course: MTS 525-0-22 SEM (15451)
Combined Course: SOCIOL 476-0-24 LEC (18815)

Demographic Questions

	Education & SP	Communication	Graduate School	KGSM	McCormick	Medill	Music	Summer	SCS	WCAS
Your School	1	3	3	0	1	0	0	0	0	0

	a. Freshman	b. Sophomore	c. Junior	d. Senior	e. Graduate	f. Other
Your Class	0	0	0	0	8	0

	a. Distribution requirement	b. Major requirement	c. Minor requirement	d. Elective requirement	e. Other requirement	f. No requirement
Your reason for taking course (mark all that apply).	0	1	1	2	1	3

	1	2	3	4	5	6	Total Response	Average Response
Interest in subject before taking course.	0	0	0	1	2	5	8	5.50

Core Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
1. Provide an overall rating of the instruction	6	0	0	0	0	1	5	5.83
2. Provide an overall rating of the course.	5	0	0	0	0	1	4	5.80
3. Estimate how much you learned in the course.	7	0	0	0	0	2	5	5.71
4. Rate the effectiveness of the course in challenging you intellectually.	8	0	0	0	1	1	6	5.63
5. Rate the effectiveness of the instructor(s) in stimulating your interest in the subject.	8	0	0	0	0	2	6	5.75

Time-Survey Question

	Total Response	a. 3 or fewer	b. 4 - 7	c. 8 - 11	d. 12 - 15	e. 16 - 19	f. 20 or more
6. Estimate the average number of hours per week you spent on this course outside of class and lab time.	8	1	4	2	1	0	0

School and Department Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
7. Rate how well the instructor used class time.	8	0	0	0	2	2	4	5.25
8. Rate the effectiveness of the instructor in communicating with the class.	8	0	0	0	0	4	4	5.50
9. Rate the clarity with which the instructor communicated his/her ideas.	8	0	0	0	0	3	5	5.63
10. Rate how well the instructor provided constructive evaluations/feedback.	8	0	0	0	1	2	5	5.50
11. Rate how well the instructor displayed interest in whether students learned.	8	0	0	0	0	0	8	6.00
12. Rate how well the prerequisite course prepared you for this course.	2	0	0	0	1	1	0	4.50
13. Rate how well the instructor presented class activities in an organized fashion.	8	0	0	0	1	2	5	5.50
14. Rate how well this course helped you to think clearly about this subject.	8	0	0	0	0	2	6	5.75

Essay Questions

1. What are the primary teaching strengths of the instructor(s)?

Aaron is an expert on peer production, so learning from him felt like getting a good grounding in the overall field. He chose great readings, organized them to build on each other, offered plenty of supplementary materials. He guided the students through discussion, by connecting to our interests and questions, and avoided simply demonstrating his own point of view and expertise, instead letting us construct our own shared knowledge and wrestle with interesting problems ¿ which is what we should do in a graduate seminar!/Great facilitator of discussion with a very solid knowledge of a very new area of research. It's awesome to have someone who has been working in the forefront of this new research area, peer production. I loved the way the course was laid out, especially the position papers./Aaron got us excited about the content. He chose great readings for us and lead interesting discussions about the material./Critical thinking, listening, stimulating thoughts, ideas, conversation. Allowing conversation to flow between students, and bringing us back on track when we derailed./Very engaging and engaged. Takes a lot of interests in the students and any topics we want to discuss, even if it's a bit of a tangent from the main discussion point (though he is also good at keeping the conversation from getting too tangential). Makes the material very relatable and understandable and has a lot of experience and expertise he brings to the material.

2. What are the primary weaknesses, if any, of the instruction?

Sometimes there were too many readings. One week had 6. Most weeks had around 3-4 articles, which was a manageable amount. Sometimes it felt like the class was guided too much by the students' questions and personal interests. During the first few classes, Aaron had some questions for us to help us think about the readings, which was god./Sometimes the class discussions around the readings could have been more focused.

3. Can you offer suggestions for improvement?

1 - It would be helpful to have the instructor share some questions every week that we should be thinking about as we are reading. We did have that through the position papers.

2 - I liked it when the position papers required us to write in a specific professional genre - editorial response, research design; I think it could be a great to do that more.

3 - I would have liked it if we had more benchmarks to work on our final seminar project - such as submitting different pieces for it thoughout the course of the class. This was very loosely planned and required the most work at the end, which was challenging with other final requirements./Perhaps having more guided points for the discussions would be helpful. Or having one person prepare to present on 1 reading to make sure that they thought through how we should think about the discussion./Maybe a little more structure on administrative matters (like creating google docs for discussion questions) though this didn't detract from my learning, really./The specification of responding with questions felt artificial, and by the end of the course we had clearly abandoned it in favor of a mixture of comment, response, critique, and questions. I'd make that the explicit model.

4. Did the course help you learn? Why or why not?

Yes. The readings were well organized, the position papers helped me reflect on and synthesize different trends in the literature. My other classmates asked good questions and were great collaborators./Yes, definitely. I was introduced to a diverse and comprehensive literature about something I'd been meaning to get into./Yes, it definitely helped me learn. I found myself very interested in the material and it helped me realize that I want to focus my dissertation to look more closely at peer production to some extent./Without question, this course helped me learn. It provided me with a structured approach to a body of literature that ensured that I'd get a well-rounded perspective on the field, even the parts of the field I am less enthusiastic about exploring, but really ought to know about./This was a world I knew was out there, but one that I never really explored. It was great to see the connection between this course and my passions for designing for user interaction.

5. Please summarize your reaction to this course focusing on the aspects that were most important to you.
This was a great seminar about an emerging field/This class is an excellent course on peer production: something that is becoming more and more present in the online world. The instructor is very knowledgeable and a very engaging discussant./Overall, I really enjoyed this course. Aaron was fantastic and really wanted us to find interest and learn about the subject. I'd recommend this to anyone interested in peer production or online communities!/Very interesting material. Aaron really knows his stuff. Learned a lot about a really fascinating topic./This class was a fantastic introduction to the field of collective action. Prof. Shaw helped us stay on course and structured the class such that we were able to get a broad overview of the field while still seeing some of the good and the bad of nitty gritty research. I not only learned about collective action, but also about scholarship. Highly recommend.

Course and Teacher Evaluation Summary for Instructors

Subject:	MTS Media, Technology & Society	**Enrollment:**	8
Academic Term:	Fall 2014	**Responses:**	6
Class Number:	15607	**% Responses**	75.00

Class: 525-0-22 Special Topics Research Seminar Collective Action & Org Online
Instructor: Aaron Shaw

Demographic Questions

	Education & SP	Communication	Graduate School	KGSM	McCormick	Medill	Music	Summer	SCS	WCAS
Your School	1	3	2	0	0	0	0	0	0	0

	a. Freshman	b. Sophomore	c. Junior	d. Senior	e. Graduate	f. Other
Your Class	0	0	0	0	6	0

	a. Distribution requirement	b. Major requirement	c. Minor requirement	d. Elective requirement	e. Other requirement	f. No requirement
Your reason for taking course (mark all that apply).	0	0	0	0	1	5

	1	2	3	4	5	6	Total Response	Average Response
Interest in subject before taking course.	0	0	0	2	0	4	6	5.33

Core Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
1. Provide an overall rating of the instruction	5	0	0	0	0	2	3	5.60
2. Provide an overall rating of the course.	5	0	0	0	0	1	4	5.80
3. Estimate how much you learned in the course.	5	0	0	0	0	1	4	5.80
4. Rate the effectiveness of the course in challenging you intellectually.	6	0	0	0	0	2	4	5.67
5. Rate the effectiveness of the instructor(s) in stimulating your interest in the subject.	6	0	0	0	0	2	4	5.67

Time-Survey Question

	Total Response	a. 3 or fewer	b. 4 - 7	c. 8 - 11	d. 12 - 15	e. 16 - 19	f. 20 or more
6. Estimate the average number of hours per week you spent on this course outside of class and lab time.	6	0	1	4	1	0	0

School and Department Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
7. Rate how well the instructor used class time.	6	0	0	0	0	2	4	5.67
8. Rate the effectiveness of the instructor in communicating with the class.	6	0	0	0	0	2	4	5.67
9. Rate the clarity with which the instructor communicated his/her ideas.	6	0	0	0	0	3	3	5.50
10. Rate how well the instructor provided constructive evaluations/feedback.	6	0	0	0	2	3	1	4.83
11. Rate how well the instructor displayed interest in whether students learned.	6	0	0	0	0	2	4	5.67
12. Rate how well the prerequisite course prepared you for this course.	1	0	0	0	0	0	1	6.00
13. Rate how well the instructor presented class activities in an organized fashion.	6	0	0	0	0	1	5	5.83
14. Rate how well this course helped you to think clearly about this subject.	6	0	0	0	0	1	5	5.83

Essay Questions

1. What are the primary teaching strengths of the instructor(s)?

Clear outline of the course. Good mixture of giving the class context of the readings and encouraging us to develop our ideas of the concepts. I really liked how Aaron used our comments construct future activities./- Well organized class discussions

- Innovative way to organize and structure homework assignments
- The topic of the course, peer production, was implemented into the very structure of the course (the students actually peer produced essays on given topics)
- Efficient usage of canvas and site web page
- Willingness of the instructor to meet with students during his office hours and give thoughtful recommendations on their research/Knows the topic very well; treats all students with tons of respect, and gives opportunities for them to develop their own ideas about the concepts presented in class. He also uses the course as a way for students to develop their own projects, and provides helpful feedback in how to do that.

2. What are the primary weaknesses, if any, of the instruction?

At times I think we could have used more help in terms of discussing the readings -- see suggestions./- No feedback on precis of the final paper/Some of the feedback, particularly the précis, was slow to be returned.

3. Can you offer suggestions for improvement?

I think that there are two main areas that the course can be improved

1) Having us do more 'hands on' peer production work or analysis. Specifically doing this using actual peer production systems (e.g. Wikipedia, Mozilla). (this was done in week 1 - expanding this would be awesome)

2) A bit more guidance in the readings would be great. I think in future years it might be nice to have some extra guidance around what to look for in the texts -- I understand that this probably wasn't feasible given this is the first year the course has been taught in this way. However, in future this is definitely something that could be developed over the course of a number of years to get some really pin-pointed questions.

I know that this is sometimes viewed as 'beneath' graduate students, but in my experience the best graduate school classes I've taken have a number of reading aides that help to focus a discussion around a number of key concepts that the professor has identified (while also allowing some time and flexibility to move away from these discussions). Making these as explicit as possible going in helped me know what part of the reading to really focus on and led to a better class discussion (not that our class discussions were bad).

There is a great bit in the book 'Zen and the Art of Motorcycle Maintenance' that gets at focusing learners attentions can be profitable. The main character is having a hard time encouraging students to write more in classes. He found that if you give a really broad topic like Money then students really had a hard time writing. When you narrow the scope of they find it much easier to write a lot in a more precise and focused manner. Eventually he has them writing about the side of a quarter, and they write pages and pages of interesting and detailed stuff. I think discussions work the same way -- really focusing in on some key aspects allows learners to explore in detailed and meaningful ways./I wish that there was more lecture - Aaron is an expert on the subject, and I wish he'd done a little

more to tell us about his perspective on the topics.

4. Did the course help you learn? Why or why not?
Yes - I really got the chance to develop my knowledge and thinking in a key area because the readings were well chosen, organized, and Aaron really knows his stuff and can facilitate discussions./- I learned a lot. Although upon taking the course I was not sure about whether the topic quite matched my research interests, in the end I benefited from the course, its reading materials, and discussions we had in class. I think that the class also benefited from the interdisciplinary backgrounds of its students, and the diversity of their views. In-class discussions were very well moderated by the instructor./Yes - the readings were chosen carefully, the assignments were helpful, and the instructor was engaged and helpful.

5. Please summarize your reaction to this course focusing on the aspects that were most important to you.
Excellent class./Wonderful topic range. Good discussions. Great facilitation. In future years Aaron has a great opportunity to make this into one of the legendary courses by 1) having us experience or analyze more actual peer production systems (e.g. Mozilla, Wikipedia) and, 2) providing a bit more guidance going in on what to focus on in the readings (number 2. will take time to develop the materials/key questions)./Overall, I liked the course and I am glad I took it. The in-class discussions were interesting and skillfully moderated by the instructor, the homework assignments were well-designed and thought-provoking. It was interesting to study the subject of peer production from various perspectives, in different contexts, and with a variety of different disciplines in view. Aaron is a great professor, who is able to create a friendly and relaxed atmosphere in class, but at the same time encourage students to work hard. I appreciate his genuine interest in students' input, and his willingness to always work towards making every possible improvement of the course. Excellent class, excellent professor, highly recommended./Great class, Aaron is very enthusiastic about the topic and it reflects on his teaching/A very good, interesting course, with a great instructor. Interesting readings, very good class discussions, and helpful assignments.

Course and Teacher Evaluation Summary for Instructors

Subject: MTS Media, Technology & Society
Academic Term: Fall 2015
Class Number: 12512

Enrollment: 9
Responses: 9
% Responses 100.00

Class: 525-0-20 Special Topics Research Seminar Collective Action Online
Instructor: Aaron Shaw

Demographic Questions

	Education & SP	Communication	Graduate School	KGSM	McCormick	Medill	Music	Summer	SCS	WCAS
Your School	0	6	2	0	1	0	0	0	0	0

	a. Freshman	b. Sophomore	c. Junior	d. Senior	e. Graduate	f. Other
Your Class	0	0	0	0	9	0

	a. Distribution requirement	b. Major requirement	c. Minor requirement	d. Elective requirement	e. Other requirement	f. No requirement
Your reason for taking course (mark all that apply).	1	1	0	1	1	3

	1	2	3	4	5	6	Total Response	Average Response
Interest in subject before taking course.	0	0	0	0	5	4	9	5.44

Core Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
1. Provide an overall rating of the instruction	6	0	0	0	0	3	3	5.50
2. Provide an overall rating of the course.	6	0	0	1	1	1	3	5.00
3. Estimate how much you learned in the course.	7	0	0	1	1	3	2	4.86
4. Rate the effectiveness of the course in challenging you intellectually.	9	0	0	0	2	5	2	5.00
5. Rate the effectiveness of the instructor(s) in stimulating your interest in the subject.	9	0	0	0	1	3	5	5.44

Time-Survey Question

	Total Response	a. 3 or fewer	b. 4 - 7	c. 8 - 11	d. 12 - 15	e. 16 - 19	f. 20 or more
6. Estimate the average number of hours per week you spent on this course outside of class and lab time.	7	0	3	1	3	0	0

School and Department Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
7. Rate how well the instructor used class time.	9	0	0	0	0	5	4	5.44
8. Rate the effectiveness of the instructor in communicating with the class.	9	0	0	0	2	2	5	5.33
9. Rate the clarity with which the instructor communicated his/her ideas.	9	0	0	0	1	4	4	5.33
10. Rate how well the instructor provided constructive evaluations/feedback.	9	0	0	0	1	2	6	5.56
11. Rate how well the instructor displayed interest in whether students learned.	9	0	0	0	0	3	6	5.67
12. Rate how well the prerequisite course prepared you for this course.	1	0	0	0	0	0	1	6.00
13. Rate how well the instructor presented class activities in an organized fashion.	9	0	0	0	3	3	3	5.00
14. Rate how well this course helped you to think clearly about this subject.	8	0	0	0	1	2	5	5.50

Essay Questions

1. What are the primary teaching strengths of the instructor(s)?

Open to working with students, communicates clearly, accurate in instruction, knows a broad body of research, extremely proficient in the subject material./guiding discussion, making students feel an active part of the creation of knowledge/He is enthusiastic about the topic and very knowledgeable about research design. /Inclusivity, course design/engaging, obviously cares about the material, provides thoughtful and constructive feedback/very good at being able to synthesize and respond to comments from students in nearly any context/Enthusiasm, clear instructions, helpful feedback, giving us freedom to choose subjects, organizing groups to help each individual with their work, very helpful individual meetings, very helpful suggested readings, well-managed discussion facilitation/The instructor leads class discussions well. Also, the class is designed as very flexible that the instructor allows students to choose readings. He is very cool and friendly.

2. What are the primary weaknesses, if any, of the instruction?

None/timeliness of getting prompts for weekly responses out./He is very casual, which I appreciate to some degree, but I also would appreciate more formal authority in the classroom because although there are opportunities to learn from one another we also want to learn from the expert.I felt that too much of the learning (for example choosing articles to read or sharing research) was offloaded / outsourced to the students in the class instead of being driven by the instructor./Hesitant to make authoritative statements (which is a strength to a certain extent)/you definitely get out what you put into the course--it's possible to slide by without much effort. somewhat unorganized and hard to contact at times./timing of communications about required coursework could be more consistent/Time management with second half of the class (presenting readings late)

3. Can you offer suggestions for improvement?

None/Just opening all of the weekly discussion fora at the beginning of the semester./An underview at the end of class (if possible) to illustrate what the discussion might have missed or obscured, i.e., Here are the main points that scholars in the field will recognize from this article/book that we didn't talk about.../more consistent timing of coursework communications would be good, for example. Every tuesday a communication goes out about what the required readings for the week are so that preparations for the week/weekend can be made

4. Did the course help you learn? Why or why not?

Immensely, the course was structured open enough to encourage projects and research collaboration./Yep. In particular, helped with the transition to PhD scholarship as different from undergrad and master's scholarship/Yes/yes, the material was excellent and the assignments were general enough that you could tailor the work to specific projects/Absolutely, there is a lot to think about in this space and I can appreciate that because of this professor's work/Yes, the readings gave me a great understanding of the topics, the discussions helped me broaden my view and get to know challenges in the existing subjects better, the weekly writing helped my practice my articulation of thoughts, the final project helped me with narrowing my interest and learning more about them./I was interested in the subject before taking this course, and I am more interested in it by learning this area of research.

5. Please summarize your reaction to this course focusing on the aspects that were most important to you.

This course help me develop research iteration and incubation skills./This class was great and was particularly good at providing an entry point to many of the different problems in collective action literature, although I think the name should be changed as I think that some of the people in the class were hoping that the focus would be more one social movements in specific./Very inclusive and open-ended, tailoring itself to student needs. The workshop was of vital importance to my learning experience and professional growth through the class. During the first half of the course, I would have liked more of a traditional lecture incorporated./It helped me to think of my work in the context of collective action and social behavior in general/learning about the basics of collective action, subjects related to it and methods used.

Question	Response 1	Response 2	Response 3	AVERAGES:
Provide an overall rating of the instruction	5	5	5	5
Provide an overall rating of the course	5	5	5	5
Estimate how much you learned in the course	5	6	6	5.666666667
Rate the effectiveness of the course in challenging you intellectually	5	5	6	5.333333333
Rate the effectiveness of the instructor(s) in stimulating your interest in the subject	6	5	5	5.333333333
Please rate the overall usefulness of the problem sets	5	6	5	5.333333333
Please rate how useful you found the assigned readings	5	5	4	4.666666667
Please rate the overall usefulness of the research design exercises	4	4	4	4
Please rate the overall usefulness of the in-class discussions	6	5	5	5.333333333
Estimate the average number of hours you spent per week on this course outside of class and lab time	5	10?	8	6.5
What are the primary teaching strengths of the instructor(s)?	Overall, I think you did a great job deciphering the material and helping us absorb it over the course of the quarter. I found you to be both knowledge about the subjects covered in the course and humble/helpful/open to suggestions, new interpretations, etc. Your greatest teaching strengths include your down-to-earth attitude and your clear interest and passion for the material covered in the course.	Aaron is engaging, interesting and knowledgable professor (instructor?). I really enjoyed how he worked through the examples from the problem sets with R files as well as how we worked through these in class as well.	No slides! Seriously, when it comes to understanding concepts, nothing beats a marker on a board. The format also meant that instead of lecturing at us, you were able to engage with the material and let us engage with it with you.	
What are the primary teaching weaknesses, if any, of the instructor? Can you offer suggestions for improvement?	There aren't many. One potential weakness might be the ability to spend 100% of each class period covering relevant material in a clear, directed, and engaging way. I would say this was the case 95% of the time, but every once and while I feel like we hit a bit of a lull for 5-10 minutes (which I think is almost inevitable in a graduate-level seminar with three students). Not sure how to remedy this without a bigger class.	Weaknesses? I'm probably not one to judge. For some reason or another I wanted some slides to look at at a few points in this class. Or rather just some quick reference sheets on the material. Looking across the web for some good slides on this material, there seems to be a great opportunity for some quality, not rehashed material on these topics.	This isn't about you, per se, but three hours is too long, even with the generous break. 1.5 hours twice a week would be much more manageable.	
Did the course help you learn? Why or why not?	The course absolutely helped me learn. I think the instructional readings, in-class discussion, outside problem sets, and application to our research projects made for the perfect (if time consuming) combination of course work.	Yes, for sure. I found the Psets to be the most useful when we were working through real world problems, although I found some of the R code suggestions to be a bit unhelpful (not Aaron's fault that R can be a real pain, and someone else's R code can really be indecipherable at times).	Yes. The small size of the class meant that when there were statistical concepts I didn't understand we could talk about them until I did. Directly applying the concepts in the problem sets was also really helpful.	
Any additional comments or summary statements you'd care to provide?	Nope--great class, hope you continue to offer it in the future.	This class was originally presented as a "workshop" class, although I didn't feel that this was the case throughout the course. Not that I was disappointed, but that it might mislead others in the future.		
Any general suggestions for improving the assignments?	None other than the ones I mentioned in class (e.g. streamline some of the readings, doing the first problem set in-class, being agnostic about the software used to complete problem sets, and focusing on 2-3 research design exercises during the first month of the quarter [e.g. one focused on idea generation, one focused on answering some fundamental design questions, and the precis/proposal]).	As we said in the discussion, the research design excercices should probably be rethought a bit. I found them useful for applying the material in the readings, but they could probably be spaced out more to support the ramp up to the final projects.	Sometimes the problem sets felt more like fighting with R than grappling with statistics. The hints helped, but then it felt like following a series of steps rather than building an understanding of the material. I'm not sure how to balance these, but if I think of anything I'll let you know.	

Course and Teacher Evaluation Summary for Instructors

Subject: MTS Media, Technology & Society
Academic Term: Spring 2014
Class Number: 38193

Enrollment: 9
Responses: 8
% Responses 88.89

Class: 525-0-25 Special Topics Research Seminar Research Design for Casual Inf
Instructor: Aaron Shaw

Demographic Questions

	Education & SP	Communication	Graduate School	KGSM	McCormick	Medill	Music	Summer	SCS	WCAS
Your School	0	2	4	1	0	0	0	0	0	0

	a. Freshman	b. Sophomore	c. Junior	d. Senior	e. Graduate	f. Other
Your Class	0	0	0	0	7	0

	a. Distribution requirement	b. Major requirement	c. Minor requirement	d. Elective requirement	e. Other requirement	f. No requirement
Your reason for taking course (mark all that apply).	0	0	0	3	1	3

	1	2	3	4	5	6	Total Response	Average Response
Interest in subject before taking course.	0	0	0	3	2	2	7	4.86

Core Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
1. Provide an overall rating of the instruction	5	0	0	0	1	3	1	5.00
2. Provide an overall rating of the course.	5	0	0	0	2	2	1	4.80
3. Estimate how much you learned in the course.	6	0	0	1	0	4	1	4.83
4. Rate the effectiveness of the course in challenging you intellectually.	7	0	0	0	0	2	5	5.71
5. Rate the effectiveness of the instructor(s) in stimulating your interest in the subject.	7	0	0	0	0	4	3	5.43

Time-Survey Question

	Total Response	a. 3 or fewer	b. 4 - 7	c. 8 - 11	d. 12 - 15	e. 16 - 19	f. 20 or more
6. Estimate the average number of hours per week you spent on this course outside of class and lab time.	7	0	5	1	1	0	0

School and Department Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
7. Rate how well the instructor used class time.	7	0	0	1	3	1	2	4.57
8. Rate the effectiveness of the instructor in communicating with the class.	7	0	0	0	2	1	4	5.29
9. Rate the clarity with which the instructor communicated his/her ideas.	7	0	0	0	1	3	3	5.29
10. Rate how well the instructor provided constructive evaluations/feedback.	7	0	0	0	1	0	6	5.71
11. Rate how well the instructor displayed interest in whether students learned.	7	0	0	0	0	1	6	5.86
12. Rate how well the prerequisite course prepared you for this course.	3	0	0	0	0	2	1	5.33
13. Rate how well the instructor presented class activities in an organized fashion.	7	0	0	1	1	1	4	5.14
14. Rate how well this course helped you to think clearly about this subject.	7	0	0	0	2	2	3	5.14

Essay Questions

1. What are the primary teaching strengths of the instructor(s)?

Aaron is a very patient and concerned teacher. No complaints!/Dr. Shaw is great at effectively using discussion as opposed to lecture. He is great about making time outside of class to help explain concepts./You can really tell that Aaron is very highly invested in helping his students learn and conduct quality research./The instructor is passionate, gives very constructive feedback and stimulates critical thinking

2. What are the primary weaknesses, if any, of the instruction?

Could be a little more directed in his approach/The class materials are dry and sometimes difficult to understand, I hope the instructor could give more time to help understand the various methods and techniques

3. Can you offer suggestions for improvement?

Maybe spend a little more time on R right at the beginning? The lab sessions were awesome and very helpful, but I think a better base understanding might make the class go a lot more smoothly./Maybe an one-hour lecture session for each class on the class readings; more tutorials on R

4. Did the course help you learn? Why or why not?

Yes!/Yes, it helps me reflect the experiment design and how to do better research

5. Please summarize your reaction to this course focusing on the aspects that were most important to you.

This class covers A LOT of information. There is little chance that you can master any of the covered topics, but it's very helpful to have a basic introduction to these challenging topics. The instructor is still kind of figuring out his teaching style, but you can definitely tell that he's very committed to his students and very knowledgeable about the topics!/great course!i think everyone in grad school needs to learn causal inference/Prof. Shaw's enthusiasm and knowledge of causal inference stood out throughout this course. He also did a great job at making sure that we focused on learning and understanding, rather than rote learning and grade chasing. I would encourage students to brush up on multivariate regression and model building before taking this class./interesting, mind-blowing, and reflective.

Course and Teacher Evaluation Summary for Instructors

Subject: MTS Media, Technology & Society
Academic Term: Spring 2015
Class Number: 33135

Enrollment: 5
Responses: 4
% Responses 80.00

Class: 525-0-25 Special Topics Research Seminar Research Design for Casual Inf
Instructor: Aaron Shaw

Demographic Questions

	Education & SP	Communication	Graduate School	KGSM	McCormick	Medill	Music	Summer	SCS	WCAS
Your School	0	0	4	0	0	0	0	0	0	0

	a. Freshman	b. Sophomore	c. Junior	d. Senior	e. Graduate	f. Other
Your Class	0	0	0	0	3	0

	a. Distribution requirement	b. Major requirement	c. Minor requirement	d. Elective requirement	e. Other requirement	f. No requirement
Your reason for taking course (mark all that apply).	0	1	0	2	1	0

	1	2	3	4	5	6	Total Response	Average Response
Interest in subject before taking course.	0	0	1	1	2	0	4	4.25

Core Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
1. Provide an overall rating of the instruction	3	0	0	0	0	2	1	5.33
2. Provide an overall rating of the course.	3	0	0	0	0	1	2	5.67
3. Estimate how much you learned in the course.	4	0	0	0	0	2	2	5.50
4. Rate the effectiveness of the course in challenging you intellectually.	4	0	0	0	0	2	2	5.50
5. Rate the effectiveness of the instructor(s) in stimulating your interest in the subject.	4	0	0	0	0	2	2	5.50

Time-Survey Question

	Total Response	a. 3 or fewer	b. 4 - 7	c. 8 - 11	d. 12 - 15	e. 16 - 19	f. 20 or more
6. Estimate the average number of hours per week you spent on this course outside of class and lab time.	3	0	1	1	1	0	0

School and Department Questions

	Total Response	1 Very Low	2	3	4	5	6 Very High	Average Response
7. Rate how well the instructor used class time.	4	0	0	0	1	1	2	5.25
8. Rate the effectiveness of the instructor in communicating with the class.	4	0	0	0	2	1	1	4.75
9. Rate the clarity with which the instructor communicated his/her ideas.	4	0	0	0	0	2	2	5.50
10. Rate how well the instructor provided constructive evaluations/feedback.	4	0	0	0	0	1	3	5.75
11. Rate how well the instructor displayed interest in whether students learned.	4	0	0	0	0	0	4	6.00
12. Rate how well the prerequisite course prepared you for this course.	1	0	0	0	0	0	1	6.00
13. Rate how well the instructor presented class activities in an organized fashion.	4	0	0	0	0	2	2	5.50
14. Rate how well this course helped you to think clearly about this subject.	4	0	0	0	0	2	2	5.50

Essay Questions

1. What are the primary teaching strengths of the instructor(s)?
Aaron is very willing and able to explain concepts in ways that are accessible, and he takes the time to make sure that everyone understands the material to their satisfaction.

The readings, and the exercises, were carefully chosen and were helpful./Explaining complicated subjects easily with examples, and choosing good books that do the same

2. What are the primary weaknesses, if any, of the instruction?
I wish there was more class time to learn R if we are expected to use it for the assignments

3. Can you offer suggestions for improvement?
I think it would be helpful to go over the homeworks in a more organized manner, since that can remind students of questions/misunderstandings./If the goal is not to master R, have more conceptual questions than R assignments

4. Did the course help you learn? Why or why not?
Yes. Although I had come across many of the concepts, they are now much more clear and make more intuitive sense./Yes. I didn't know ANY of what the course taught before minus a tiny bit of regression and (incorrect) knowledge of statistical testing

5. Please summarize your reaction to this course focusing on the aspects that were most important to you.
I really enjoyed this course - Aaron provides great resources for thinking through and understanding causality, and how different techniques attempt to provide evidence of causality. He is a patient and clear instructor./Great course for learning about causal inference. You think you may not need it in your work, but the course has made me realize that I will and that it's possible to do. The professor explains information in more than one way and in simpler terms than I have heard in any stats class.