Motivating Contributions to Public Information Goods

by

Fangzhou Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in The University of Michigan
2018

Doctoral Committee:

Professor Yan Chen, Chair
Associate Professor David Miller
Assistant Professor Daniel M. Romero
Associate Professor Tanya Rosenblat

arkzhang@umich.edu
ORCID ID: 0000-0002-3282-5811

To all those who contribute to the social good of making the sum of all human knowledge freely available to everyone

ACKNOWLEDGEMENT

First and foremost, I owe my deepest gratitude to my advisor, Dr. Yan Chen, who has been the greatest mentor in my academic life. I still remember the time when I met Yan for her advice on my very first research project in the summer of 2012. It is her continuous support, guidance and encouragement along the journey in that past four years that make me where I am today. It is an honor and privilege to have her as my advisor, mentor and friend.

I am also indebted to all my committee members. I want to extend my deep gratitude to Dr. Tanya Rosenblat for her guidance and support. She has always been providing me with succinct but insightful comments for my research. I also appreciate the encouragement and guidance from Dr. Daniel Romero. The two-year-long collaboration with him inspired me tremendously on this dissertation. I want to thank Dr. David Miller, who has provided many helpful comments and suggestions on this dissertation.

I am grateful to all my coauthors, who have provided me with great collaboration experience. I want to thank Dr. Daniel Romero, Dr. Ceren Budak, Dr. Lionel Robert, Dr. Robert Kraut, Dr. Rosta Farzan, Iman Yeckezhaare, Danielle Livneh, Ruihan Wang and Eric Blohm for their invaluable input.

I would to like to thank all faculty and student members in the Behavioral and Experimental Economics Lab. I appreciate the support and advice from Dr. Erin L. Krupka, Dr. Yusufcan Masatlioglu, Dr. Neslihan Uler, Dr. Steve Leider, Ming Jiang, Carrie Xu, Daphne Chang, Evgeny Kagan, Linfeng Li and Yingzhi Liang. I also benefit substantially from my interactions with other members in the School of Information, including Dr. Paul Resnick, Dr. Qiaozhu Mei, Dr. Eytan Adar, Wei Ai, Teng Ye, Yue Wang, Samuel Carton and Ed Platt. I am also grateful for the financial support from Rackham Graduate School at the University of Michigan and the National Science Foundation.

I also want to thank Dr. Dominic Coey, my internship manager at Facebook core data science. Dominic provided great mentorship during my internship in the summer of 2017 and much support for my job search process.

Finally, I would like to thank my family. It is their love and support along the way that drives me to move forward.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

LIST OF APPENDICES

APPENDIX

ABSTRACT

This dissertation investigates how to motivate contributions to public information goods, characterized by non-rivalry by nature and non-excludability by choice. With the running example of Wikipedia - the online encyclopedia that grants free access to everyone and relies entirely on the inputs from volunteers, I present three designed-based empirical studies that explore how individual contributions are affected by 1) reduction in the number of collaborators, 2) private benefit, number of readers, and 3) group membership.

Chapter II studies how the reduction in group size affects the collaborative behavior in contributing to Wikipedia articles. Exploiting a natural experiment at the Chinese Wikipedia, we find that the level of contribution and conflict within the group drop on articles that face a shock, whereas centralization increases. Interestingly, the impact of a shock on activity increases with shock level, whereas the impact on centralization and conflict is higher for moderate shock levels than for very small or very high shock levels.

Chapter III examines how the private benefit of citation and acknowledgement and the social impact from the public goods can motivate contributions from domain experts. Using a randomized field experiment, we invite 3,974 academic economists to contribute to Wikipedia articles relevant to their researches. The results show that experts are significantly more interested in contributing when citation benefit is mentioned. Furthermore, cosine similarity between a Wikipedia article and the expert's paper abstract is the most significant factor leading to more and higher-quality contributions, indicating that better matching is a crucial factor in motivating contributions to public information goods.

Chapter IV examines the potential of team-based approach on motivating individual contributions to Wikipedia. Employing a panel data set across over 9,000 Wikipedia editors and exploiting the variations in the exposure to WikiProject due to quasi-randomness in the association of Wikipedia articles to WikiProject, we find that joining a WikiProject has a sizable impact on the level of contribution, measured by both the number and size of revisions. Further analysis on the patterns in the behavior of WikiProject members indicates that the recommendation of articles can be the factor driving the impact of WikiProject.

CHAPTER I
Introduction

The advent and rise of user-generated content platforms mark a transition in how information goods are produced on the Internet. Instead of having one single publisher produce and distribute the content, platforms of various kinds – such as software development (e.g., Github), photo-sharing community (e.g., Flicker), Q&A sites (e.g., Stack Overflow), online health support groups (e.g.,BreastCancer.org) – allow and rely on individual users to contribute.

These peer-produced public goods enabled by information technology, which we call *public information goods*, have distinct characteristics. They are information goods with free and open access to the general public. Unlike textbook examples of pure public goods, such as national defense, which makes exclusion technically costly or infeasible, public information goods are technically easy to exclude by requiring authentication. However, they are provided to the general public for free. Therefore, *public information goods* are non-rivalrous by nature and non-excludable by choice. The differentiation of public information goods from the traditional public goods is beyond this. Unlike charitable giving where donations are perfect substitutes, public information goods are usually non-substitutes (e.g., a piece of code on Github or an answer on StackOverflow). The production cost and quality of the public information goods depend on whether a contributor is the right person for the task regarding her interest and expertise.

Motivating voluntary contributions to public (information) goods is of interest across many social sciences. A large body of economic research has been dedicated to understanding strategies to encourage efficient provision of public goods since the seminal work by Bergstrom et al. (1986); Samuelson (1954). Recognizing the inherent challenge to maintain efficient provision of public goods due to the classic free-rider problem, many theoretical and applied work focus on identifying factors reducing the incentives to free ride and therefore understanding strategies to encourage individual provision. Ledyard (1995) and Vesterlund (2015) provide surveys summarizing results from laboratory experiments investigating effective strategies to encourage contribution to public goods (charity). In the social psychology literature, the tendency to lower one's own effort when working collectively is termed as social loafing. To understand approaches mitigating social loafing, Karau and Williams (1993) present the collective effort model, which integrates a variety of factors that have been shown to play an important role. The three chapters in the dissertation start with factors that have been shown to affect motives for traditional public goods contributions in these two strands of literature (e.g., group size in chapter 2, private benefits and number of recipients in chapter 3, and group identity in chapter 4) and explore how they can be applied to public information goods with the example of Wikipedia. Unlike many traditional repositories of scientific knowledge, such as most academic journals that require individual or institutional subscription fees for access, Wikipedia was created to provide web-based, free-content encyclopedia to the public. Ever since its establishment in 2001, Wikipedia has developed the most comprehensive encyclopedia in history. By May 2018, it has accumulated more than 5.6 million articles, with over 838 million revisions from 33 million registered users.

This dissertation consists of three designed-based empirical studies that examine factors influencing the incentives to contribute to Wikipedia articles. In the second chapter, "Shocking the Crowd: The Effect of Censorship Shocks on Chinese Wikipedia", along with my co-authors, Danielle Livneh, Daniel Romero, Ceren Budak, Lionel Robert, we investigate how the reduction in group size affects collaborative behavior in contributing to Wikipedia articles. This work is motivated by the block of the Chinese Wikipedia at the mainland of China in 2005. The qualification of the causal empiricism hinges on the exogeneity of the block, which induces rich variations in the fraction of collaborators lost for each article. We find that the level of contribution and conflict within the group drop for articles that face a shock, whereas centralization increases. Interestingly, the impact of a shock on activity increases with shock level, whereas the impact on centralization and conflict is higher for moderate shock levels than for very small or very high shock levels. These findings provide support for threat rigidity theory – originally introduced in the organizational theory literature.

The third chapter, “Motivating Contributions to Public Information Goods: A Field Experiment at Wikipedia”, is a joint work with Yan Chen, Rosta Farzan, Robert Kraut and Iman YeckehZaare. Using a randomized field experiment, we invite 3,974 academic economists to contribute to Wikipedia article relevant to their research and examine the incentives which might affect their motivations to contribute their expertise. The experiment achieves the manipulation of social impact and private benefit by vary the mentioning of likely citation, public acknowledgement and the number of views an article receives. We find that the matching precision between the recommended Wikipedia article and the expert significantly and substantially increase both the quantity and quality of contribution. Furthermore, the provision of private benefit leads to better but not necessarily longer comments, which suggests the promise of non-monetary incentive in inducing high-quality content.

In the final chapter, “Group Membership and Contributions to Wikipedia - The Case of WikiProject”, coauthored with Yan Chen and Iman YeckehZaare, we evaluate the potential of team-based approach on motivating individual contributions to Wikipedia. Employing a panel data set across over 9,000 Wikipedia editors, we find that joining a WikiProject - a group of editors who are interested in a specific topic - is significantly associated with more active contribution behavior. To causally identify and measure the influence of WikiProject membership, we use an instrumental variable approach that exploits the variation in the exposure to WikiProject due to quasi-randomness in the association of Wikipedia articles to WikiProject. The estimates confirm that joining a WikiProject has a sizable impact on the level of contribution. Further analysis on the patterns in the behavior of WikiProject members indicates that the recommendation of articles can be the factor driving the impact of WikiProject.

Taken together, the three chapters provide analyses on motivating contributions to public information goods through various aspects, including the coordination among contributors, the efficacy of private and social impact, the importance of matching between contributors and goods and the promise of group identity. Our findings provide important design implications for not only Wikipedia but many other online platforms of public information goods that rely on the contributions from individual contributors. For example, the analyses on coordination among contributes can be generalized to Github, where inputs from contributors are highly dependent; chapter 3 suggests that private acknowledgement can be of promise in inducing high-quality for content platforms such as Stack Overflow. In additional, our study also sheds light on strategies encouraging contributions to the general class of public information goods. For example, the findings on accurate matching between contributors and tasks in chapter 3 suggest that directing a potential donor to the charity that she cares most might be helpful in motivating her contribution.

CHAPTER II

Shocking the Crowd: The Effect of Censorship Shocks on Chinese Wikipedia

Abstract

Collaborative crowdsourcing has become a popular approach to organizing work across the globe. Being global also means being vulnerable to shocks – unforeseen events that disrupt crowds – that originate from any country. In this study, we examine changes in collaborative behavior of editors of Chinese Wikipedia that arise due to the 2005 government censorship in mainland China. Using the exogenous variation in the fraction of editors blocked across different articles due to the censorship, we examine the impact of reduction in group size, which we denote as the shock level, on three collaborative behavior measures: volume of activity, centralization, and conflict. We find that activity and conflict drop on articles that face a shock, whereas centralization increases. The impact of a shock on activity increases with shock level, whereas the impact on centralization and conflict is higher for moderate shock levels than for very small or very high shock levels. These findings provide support for threat rigidity theory – originally introduced in the organizational theory literature – in the context of large-scale collaborative crowds.

2.1 Introduction

Crowdsourcing is now poised to fundamentally transform the way we coordinate work (Anya, 2015). Online collaborative crowdsourcing platforms such as Wikipedia present a unique opportunity to tackle complex problems (Baldwin and von Hippel, 2011; Yu and Nickerson, 2011). Shocks are unforeseen events that can disrupt and even threaten crowds (Cohendet and Simon, 2016; Jackson and Dutton, 1988; Ocasio, 2011). Examples of such shocks include massive influxes or outflows of members, platform or government policies, or exogenous events (like the death of a celebrity) that increase the importance and visibility of the crowd's work.

A number of studies have explored shocks and threats in organizations (Cohendet and Simon, 2016; Dutton et al., 2006) as well as in small groups through experimental approaches (Argote et al., 1989; Gladstein and Reilly, 1985; Harrington et al., 2002). However, less is known about how shocks affect online crowds, which often face distinct challenges to effectively respond to shocks (Kittur and Kraut, 2010; Robert and Romero, 2017). For example, online crowds have a much more fluid membership than offline groups and organizations, which makes a potential response to an unexpected shock much more difficult to organize. Therefore, a shock is likely to have a different impact on online crowds than their organizational work group counterparts. It is not clear how, or whether, online crowds respond to a shock. Part of the reason for this gap in the literature is the lack of adequate instances where the same type of shock affects a large number of online crowds – a phenomenon that would allow for a systematic analysis of how crowds typically respond to shocks.

In this paper, we take a step to fill this knowledge gap by examining the impact of the 2005 Chinese government censorship block of Chinese Wikipedia. This event presents an exogenous shock to all Wikipedia articles that have contributors from mainland China, because these articles lose some and in some cases most of their contributors as a result of the censorship block. Using the exogenous variation in the fraction of editors blocked across different articles, we investigate the impact of shocks of different magnitudes on articles with varying numbers of editors.

How do we expect the Chinese Wikipedia community to respond to this censorship shock? The literature on threat rigidity suggests that groups respond to an external threat or shock by centralizing their decision-making and decreasing internal conflict (Staw et al., 1981). Should we expect Chinese Wikipedia crowds to behave like traditional offline groups? Our study aims to answer this question. We examine the crowds' response to shocks with respect to three collaborative behavior measures: activity, centralization, and conflict.

Our main contributions are the following: *(i)* We find that the overall activity level drops after a shock, but the exact drop in activity depends on the crowd's size;

(ii) as predicted by thread rigidity theory, centralization increases and conflict decreases when crowds are faced with moderate shocks. But surprisingly, the effects are less profound for more severe shocks; *(iii)* our findings contribute to the organization theory literature by providing a large-scale validation of threat rigidity in a new emerging context; *(iv)* our findings contribute to the crowdsourcing literature by providing analysis that could have important implications to the design and management of crowdsourcing platforms.

2.2 Related Work

Threat Rigidity and Centralization: Threat rigidity is often used to explain how groups behave when faced with an external shock (Kamphuis et al., 2011; Staw et al., 1981). This theory suggests that groups will seek to overcome external threats by increasing both the centralization of decision-making and group cohesion (Staw et al., 1981). Centralization helps the group better coordinate its response to the threat during a time when coordination is difficult (Cohendet and Simon, 2016). Centralization also makes the group more efficient by leveraging its existing work practices while resources are low (Argote et al., 1989). Increases in cohesion reduce conflict (Windeler et al., 2015), which further facilitates coordination (Cummings et al., 2009; Hinds and Bailey, 2003). Both increases in centralization and decreases in conflict allow the group to focus more on responding to the threat.

Threat rigidity has been found to be consistent with behaviors observed in organizations (Cohendet and Simon, 2016; Dutton et al., 2006) but less so in experimental studies of groups (Gladstein and Reilly, 1985; Harrington et al., 2002; Kamphuis et al., 2011). Other experimental studies found no evidence of centralization under threat (Argote et al., 1989; Driskell and Salas, 1991).

There are several gaps in the current literature, which our study aims to fill. One, in the previously mentioned studies, due to the experimental methods employed, all threats were artificial. Our crowds are faced with a real external threat that could undermine their long term viability. Two, previous studies employed ad-hoc groups of people who had never worked together. Threats may not have much of an impact when members have no real history or future with their group. We overcome this limitation by examining crowds with both a history and a potential long term future. Three, group sizes have had little or no variance. This is particularly problematic given that previous research on crowds has shown that size is often related to both centralization and conflict (Arazy et al., 2011; Kittur and Kraut, 2008). We study crowds of different sizes which allows us to examine the interaction of crowd size and the effect of the shock on collaborative dynamics. Four, past studies examined the short-term impact of shocks on groups. Thus, even when these findings show a link between shocks and centralization it is difficult to know if such effects are lasting. In our study, we examined the impacts of shocks over a much longer period – 1 year. Finally, other studies do not vary the intensity of the shock. Groups were either exposed to a shock or not exposed. In natural settings, shocks are likely to vary in intensity. In this study, shocks greatly vary in intensity, which allows us to examine their impacts over a range of levels.

Conflict and Crowds: Coordinating work in large online crowds can be particularly difficult for several reasons. Unlike organizational work groups, online crowds often lack hierarchical structures, formal boundaries, stable memberships, and formal training (Keegan et al., 2012; Kittur and Kraut, 2010; Robert and Romero, 2015). Additionally, these crowds are typically composed of members who work at a distance and rely on electronic communication, which further increases the prevalence of conflict (Filippova and Cho, 2016; Hinds and Bailey, 2003; Windeler et al., 2015). For example, Filippova and Cho (2016) find that as task interdependence and geographic dispersion increase, so does conflict in Github crowds. Kittur and Kraut (2010) examine Wikipedia crowds and find that as crowd size increases coordination becomes more difficult and conflict increases.

Other studies have focused on identifying ways to reduce conflict in online crowds. For example, Kittur and Kraut (2010) discover that the positive relationship between crowd size and conflict diminishes when either communication between editors increases or crowds centralize their work. Filippova and Cho (2016) find that leadership style and member participation in the decision-making reduce conflict in Github crowds. Arazy et al. (2011) and Arazy et al. (2013) both find that crowds with more administrators have less conflict.

This strand of literature offers an important and rich understanding of conflict and centralization in crowds. However, it has focused exclusively on conflict constructed under stable and static conditions. In online crowdsourcing platforms, crowds operate in environments that are far more chaotic and susceptible to disruptions than traditional organizational work groups. In this work, we aim to fill this gap by studying how centralization and conflict levels in collaborative online crowds change as a result of disruptive shocks.

2.3 Background

Chinese Wikipedia, the Chinese-language version of Wikipedia, was established on October 24, 2002. As of October 2016, Chinese Wikipedia has accumulated over 4.8 million articles, with 43 million revisions contributed by over 2 million registered users. Aiming to provide a free online encyclopedia for Chinese-speaking users, Chinese Wikipedia has benefited from contribution by editors from mainland China (20.9%), Hong Kong (26.3%), Taiwan (36.9%), the Unites States (5.6%), and Canada (1.9%).[1]

Due to the censorship of online content by the Great Firewall System of the Chinese government, Chinese Wikipedia had been blocked massively in mainland China three times by 2008. These blocks denied users from mainland China the access to Chinese Wikipedia. The first block took place on July 2, 2004, and was lifted on July 21, 2004. On Sep. 23, 2004, the Chinese government issued the second block, which lasted for 5 days. The third block of Chinese Wikipedia started on Oct. 19, 2005. Unlike the first two blocks, both of which lasted for only a short period of time, this block spanned for almost 1 year and was not lifted until Oct. 10, 2006 (Zhang and Zhu, 2011a).

In this study, we focus on the impact of the shocks due to the third block on the collaborative behavior of editors of Chinese Wikipedia. There are two reasons for focusing on the third block. First, it was deployed without any prior announcement or warning and the Chinese government offered no official explanation afterward. Hence, it serves as an exogenous shock largely unexpected by editors of Chinese Wikipedia. Moreover, unlike the prior two blocks, this block spanned a relatively

[1] http://en.wikipedia.org/wiki/Chinese_Wikipedia.

long period of time, which allows us to overcome the difficulty resulting from the overall sparsity of contribution to Wikipedia.

2.4 Identifying Blocked Users

In order to study the effect of the block on collaborative groups of editors who maintain a specific article, we need to identify the blocked users. To provide a reliable identification of the blocked users, we make use of three criteria to decide whether a specific editor of Chinese Wikipedia is from mainland China and is therefore blocked during the censorship period of the third block: editing behavior, linguistic patterns, and temporal patterns.

Editing Behavior: We first restrict the set of Wikipedia editors to those who made edits before the block. We then inspect the edits from these editors to filter out those who made edits during any of the three blocks – those editors are either from outside mainland China, and thus unblocked, or have found methods to circumvent the censorship.

社交媒体
社交媒體

Figure 2.1: Chinese characters for the word "social media": The first line is the simplified Chinese and the second line is the traditional Chinese.

Linguistic Patterns: Our second check exploits the unique feature of the linguistic pattern of the Chinese language. There exist two encoding systems for Chinese language: the simplified Chinese and the traditional Chinese. Among all Chinese characters, there are approximately 2,000 for which the simplified version differs from the traditional version. For example, Figure 2.1 shows the Chinese characters for the word "social media" in the simplified version (the first line) and those in the traditional version (the second line). Note that the two versions share the first three characters, but differ from each other in the last one. The simplified Chinese

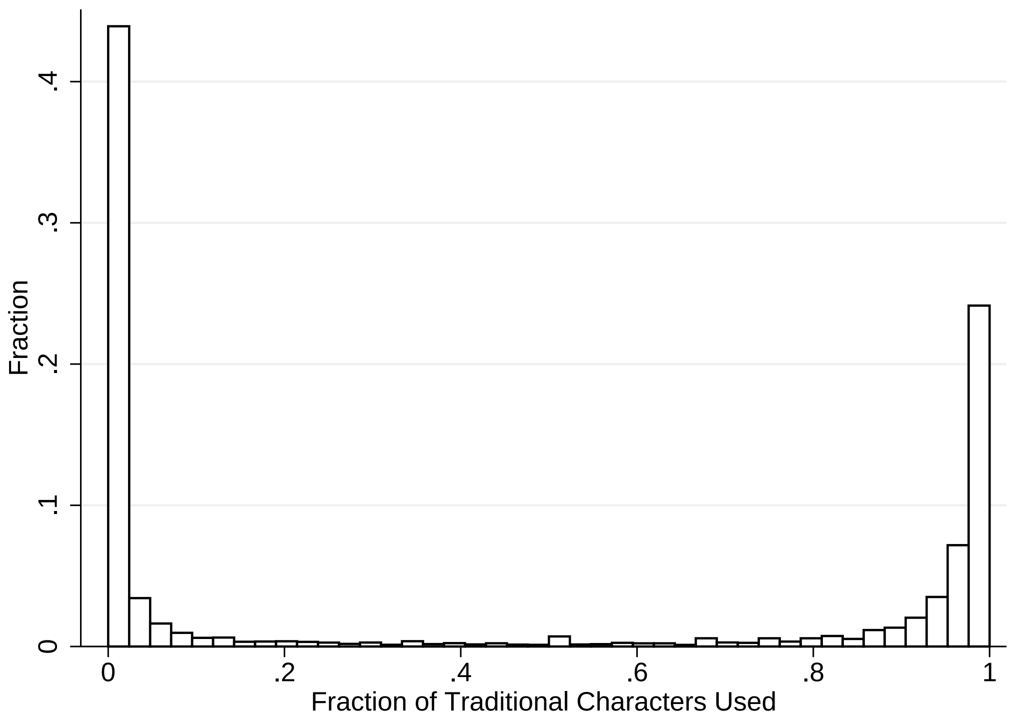


Figure 2.2: Traditional character usage: The y-axis denotes the fraction of the editors of Chinese Wikipedia that have x% of their total characters written in traditional characters

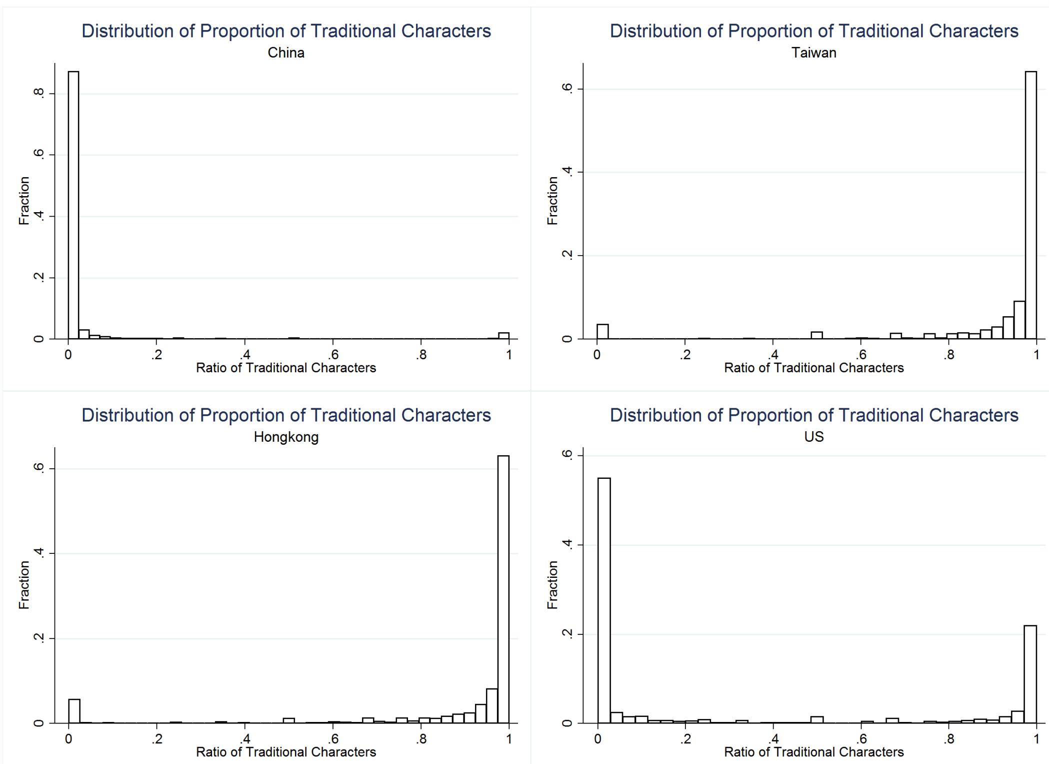


Figure 2.3: Geographical Pattern of Traditional Character Usage

is mainly used in mainland China, whereas the traditional Chinese is mainly used in Taiwan, Hong Kong and Macau. This feature provides a reliable measure for identifying mainland China editors, and has been used in related work (Zhang and Zhu, 2011a). However, Zhang and Zhu (2011a) define an editor as non-blocked if more than 50 percent of the editor's additions are in traditional Chinese, a threshold that we ultimately find to be arbitrary. In comparison, we identify the optimal cut-off from the data. To motivate this point further, we present in Figure 2.2: the distribution of editors in terms of their traditional character usage.[2] The plot reveals a prominent bimodal pattern – editors consistently use either traditional or simplified Chinese encoding. In addition, we observe that the optimal cutoff lies closer to 20%. Next, we demonstrate that the use of encodings does indeed vary across different countries in Figure 2.3. We produce this plot by considering edits from anonymous users, whose contributions are recorded by their IP addresses instead of usernames. This set of IP addresses allows us to map these editors and their use of encoding to their respective countries.[3] The editors from mainland China consistently use the simplified version while those from Taiwan and Hong Kong use the traditional version. Given our results, we classify an editor as blocked only if 20% or less of the characters used in their contributions are traditional characters.

Temporal Patterns: While Figure 2.3 justifies the use of linguistic patterns to identify blocked users, Figure 2.3 also presents a challenge when considering the U.S. (or other countries with large Asian populations that are not included due to space limitations). We observe that while the U.S. population consists of both editors who use the simplified Chinese and editors that use the traditional Chinese, most of them use the simplified Chinese. Therefore the encoding technique might falsely classify a large number of editors from the U.S. as being from mainland China, where simplified encoding is also predominantly used. In an effort to remove these conflating editors, we also consider the daily editing patterns of editors from different countries (Figure 2.4). We find that the editors in the U.S. contribute in different time frames from those in mainland China. We observe the sharpest difference for

[2]In this analysis we ignore characters that have the same representation under the two versions.

[3]Anonymous users account for 78.89% of all editors in our dataset, whose contribution represents no more than 13% of all revisions prior to the third block.

18:00-24:00. Based on this finding, we classify an editor as being from mainland China, and therefore blocked, only if $y\%$ or less of their edits are contributed during this time frame.

In summary, we classify editors who pass the first test (described in editing behavior), use a traditional character at most $x\%$ of the time, and have at most $y\%$ of their edits contributed during the idle mainland China hours as blocked and the rest as unblocked.[4] Given that we have ground truth for 49,051 editors with IP addresses, we choose the values for these parameters that maximize the $F1$ measure when classifying this population. We find that the optimal $x = 0.2$ and $y = 1.0$. This setting results in $recall = 1$, $precision = 0.74$ and $F1 = 0.85$.

2.5 Collaboration and Shock Measures

Our goal is to characterize the effect that the block of Chinese Wikipedia in mainland China has on the dynamics of collaboration within an article. We consider each article as a unit of analysis and the set of Wikipedia users who edit the article as a collaborative crowd or team. We compare activity during the pre-block period (Oct. 19, 2004, to Oct. 19, 2005) and post-block period (Oct. 19, 2005, to Oct. 19, 2006). Because the number of editors and type of editors affected by the block varies across articles, we analyze the relationship between the fraction of edits made by blocked editors in an article and impact on three collaboration measures: level of activity, centralization of workload, and conflict. We now define these measures precisely.

Shock Level: Given an article a, we define the *weighted blocked ratio* B_a of article a as the fraction of edits contributed by the editors blocked among all the edits during the pre-block period. This measure quantifies the intensity of the shock caused by the block on an article.

Level of Activity: We first consider the effect that the block has on editing volume. For each article a, we let EV_a^{pre} and EV_a^{post} be the number of edits of a made

[4] We also attempted to differentiate between the truly blocked users and those who simply dropped out by fitting time lapses between two edits from a given editor to a Poisson distribution. Given the fitted distribution, we determine the likelihood of an editor to make a contribution during the block and identify an editor as blocked only if the likelihood of edit was above a threshold. This method does not improve the accuracy and therefore was not included in our final analysis.

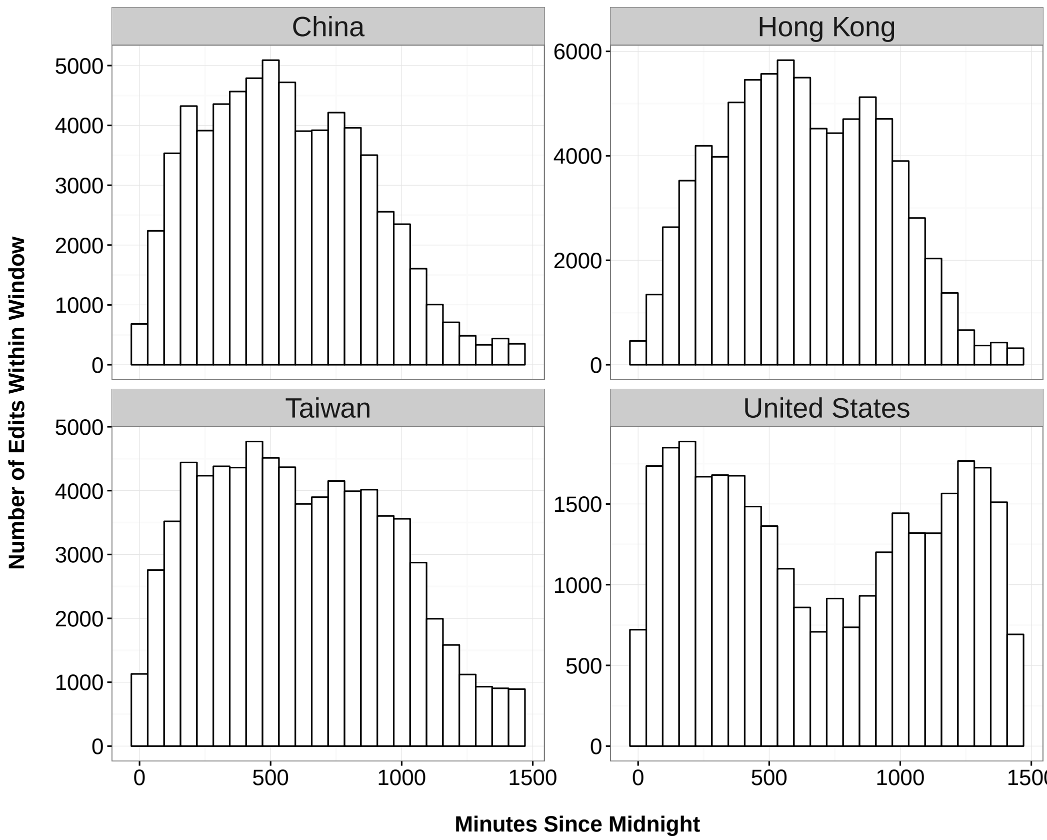


Figure 2.4: Daily temporal patterns in editing across countries and areas

during the pre-block period and post-block period, respectively. We then measure the relative change in number of edits as $EV_a^{\Delta} = (EV_a^{\text{post}} - EV_a^{\text{pre}})/EV_a^{\text{pre}}$. Because collaborative crowds lose members due to the block, we intuitively expect a decrease in the total number of edits after the block. Previous literature suggests, however, that efforts to compensate for shock can perhaps prove effective (Kamphuis et al., 2011; Sutcliffe and Vogus, 2003). Thus, while it is unclear how the block may impact levels of activity, it is even less clear how activity levels interact with weighted blocked ratio.

Centralization: It is common for Wikipedia articles to have a skewed distribution of editors' contributions (Romero et al., n.d.). Centralization is a form of coordination, where a few editors take charge of the majority of the work and rely on a large number of peripheral users to take on minor tasks, and make the crowd more effective by reducing the cost of explicit coordination (Kittur and Kraut, 2008). Meanwhile, a centralized article is less likely to be exposed to diverse expertise and points of view, which could limit the quality of the crowd's output (Arazy et al., 2006). Overall, centralization can have an important impact on the coordination dynamics of the crowd and on the quality of its output.

To measure centralization, we use the Gini coefficient, a statistical measure of dispersion to quantify the level of inequality in a distribution (Dorfman, 1979). We let E_a^{pre} and E_a^{post} be the set of editors of article a in the pre-block and post-block period, respectively, and N_a^{pre} and N_a^{post} be the number of editors in E_a^{pre} and E_a^{post}. We let $W_a^{\text{pre}}(e)$ and $W_a^{\text{post}}(e)$ be the number of times editor e contributed to article a in the respective time periods. We begin by computing the G_a^{pre}, the Gini coefficient of the set $\{\cup_{e \in E_a^{\text{pre}}} W_a^{\text{pre}}(e)\}$:

$$G_a^{\text{pre}} = \frac{\sum_{i \in E_a^{\text{pre}}} \sum_{j \in E_a^{\text{pre}}} |e_i - e_j|}{2 \sum_{i \in E_a^{\text{pre}}} \sum_{j \in E_a^{\text{pre}}} e_j}$$

For example, an article where every editor contributes the same number of edits has a Gini coefficient of 0, whereas an article with five editors who contribute 1 edit and one editor who contributes 20 edits has a Gini coefficient of 0.63. We calculate the corresponding Gini coefficient for the post-block period similarly.

Because the value of Gini coefficient depends on the number of editors and edits in the article, we normalize G_a^{pre} and G_a^{post} by their maximum possible values given the number of editors and edits in article a during the period for which we are calculating. We define the centralization of article a during the pre-block period, C_a^{pre}, as the fraction of G_a^{pre} and the maximum value of G_a given E_a^{pre} and EV_a^{pre}. We also define the corresponding measures of centralization of an article during the post-block period in the same manner. Finally, we define the change in centralization as $C_a^{\Delta} = C_a^{\text{post}} - C_a^{\text{pre}}$.

Conflict: Wikipedia editors have access to a feature known as reverting that allows them to undo any other edit. When editors have disagreements with one another, they often engage in "edit wars", where they repeatedly revert one another's edits (Tsvetkova et al., 2016; Viegas et al., 2007). We use the fraction of edits that are reverts as a measure of conflict in an article during a given time period. We let R_a^{pre} and R_a^{post} be the number of reverts in article a in the pre-block and post-block periods, respectively. We define the change in conflict in an article as $R_a^{\Delta} = R_a^{\text{post}} - R_a^{\text{pre}}$.

2.6 Results

Here we present the results of our analysis on the change in activity, centralization, and conflict due to the block. For all subsequent analysis, we only consider articles with at least two editors before the block, as our goal is to understand how crowds respond to unexpected shocks. In addition, for all three measures we distinguish between the articles that have no editors from mainland China before the block from the articles that have at least one. These two populations are qualitatively different in important ways – the former group does not appeal to a specific culture (mainland China). Indeed, articles from these two groups exhibit pre-block difference in characteristics that are relevant to our study. The articles that have at least one editor blocked have 9.06 editors contributing 19.50 revisions on average before the block, while those with no editors blocked have 4.45 editors contributing 9.29 revisions on average. In addition, the articles with at least one blocked editor tend to be contentious, on average exhibiting roughly a 50% increase in rate of reverting compared to the group of articles with no blocked editors.

	mean	min	max	std.dev	skewness
Activity	-0.2927	-1.0000	46.2857	1.0440	11.9280
Centralization	-0.0522	-1.0000	1.0000	0.2169	0.2739
Conflict	-0.0235	-0.6667	0.5000	0.1004	0.1827

Table 2.1: *Descriptive statistics of the collaboration measures.*

Given these pre-block differences between the articles with and without any of their editors blocked, we provide two steps of analysis for each of our three measures. First, we compare the change in behavior between the articles with no editors blocked and the articles with at least one editor blocked. This allows us to understand whether being exposed to the shock, regardless of its level, has an effect on the articles. In total, we have 49,945 articles in our dataset and 27,856 among them have no editors blocked. Then, we examine how variations in the shock level affect articles that have at least one editor blocked. In short, we find that the shock negatively affects activity within groups, and that our findings for centralization and conflict are in agreement with literature on threat rigidity theory.

2.6.1 Activity

Table 2.1 provides descriptive statistics of the relative level of activity. An average article has a 29% decrease in the level of activity, with the standard deviation of 1.044. We first compare the level of activity between articles with no editors blocked and those with at least one editor blocked. To this end, we regress the relative change in number of revisions of an article (EV_a^{Δ}) over a dummy variable indicating whether or not the article has at least one editor blocked, denoted by I_a^{Block}, controlling for the number of editors, denoted by N_a^{pre}.

$$EV_a^{\Delta} = \beta_0 + \beta_1 I_a^{Block} + \beta_2 N_a^{\text{pre}} + \epsilon_a \tag{2.1}$$

Table 2.2 presents the result for regression 2.1. We see that articles on average become less active after the block, with a nearly 30% decrease in the number of revisions. This might be explained by the reduction in the number of readers, which

Figure 2.5: Change in editing volume as a function of shock level: The blue curve denotes the average EV_a^{Δ} within articles that were exposed to the shock level provided in the x axis. 1.96 standard errors are plotted for each point. The green hyphens indicate the regression fit without controlling for number of editors and the red dashes indicate the regression fit with control.

reduces individuals' incentives to contribute to the public good (Zhang and Zhu, 2011a). Specifically, the volume of revisions for articles with no editor from mainland China shrinks by nearly 37%. Compared with that, articles with at least one editor blocked experience an additional 3% decrease (statistically significant at the 5% level), which leads to a total 40% decline in activity.[5]

Next, we examine the impact of various shock levels on the relative changes in activity across articles. To this end, we regress the relative change in number of revisions over the shock level (weighted blocked ratio) denoted by B_a, controlling for the number of editors before the block N_a^{pre}.

$$EV_a^{\Delta} = f(B_a) + \beta N_a^{\text{pre}} + \epsilon_a \tag{2.2}$$

In Table 2.3, we report the regression results using both the linear specification and the quadratic specification for $f(B_a)$.[6] To compare the fitness of the two models, for each specification we perform a likelihood ratio (LR) test of the quadratic model against the linear model. This shows that the quadratic model does not provide a significantly better fit to the change in activity than the linear model does.

[5]The variance inflation factor of the regression is 1.1, which indicates that the potential collinearity between the number of editors before the block and whether an article has any editor blocked does not severely affect the variance of the estimated coefficients.

[6]Controlling for lifetime of articles in the regressions does not qualitatively affect the results.

Figure 2.6.1 illustrates the relative change in activity due to various levels of shocks. The blue line plots the relative change in number of revisions as well as the error bars denoting 1.96 standard error above and below the mean. The red line and the green line represent the regression results with and without the number of editors before the block as a control. Hence, the discrepancy between the two regression lines illustrates how much of the change in articles' behavior is a result of the size of the crowd. We see from figure 2.6.1 that articles subject to a higher level of shock experience a larger decrease in activity. Specifically, the regression results show that a 10% loss in number of editors leads to a nearly 7.5% decline in volume of revisions for an average article. We further separate the analysis between small crowds with at most 5 editors and large crowds with over 5 editors. Although the shock level still has a significant impact on both types of crowds, it appears to affect large crowds more substantially. Articles with more than 5 editors before the block exhibit non-linear decreases in number of revisions, whereas those with at most 5 editors respond to the shock in a linear way. Specifically, when the shock level is below 0.45, articles with more than 5 editors experience smaller changes in activity than those with no more than 5 editors do. When the shock level exceeds 0.45, the change in activity decreases faster in articles with more than 5 editors than those with less than 5 editors. This suggests that the vulnerability of a crowd to shocks depends on the size of the crowd. When the shock level is low, a large crowd is resilient toward it. However, beyond the threshold of 0.45, as the shock level increases, the change in the effect of a shock on activity level is more severe for large crowds.

2.6.2 Centralization

We regress the change in centralization C_a^{Δ} over the indicator variable I_a^{Block} controlling for the number of editors:

$$C_a^{\Delta} = \beta_0 + \beta_1 I_a^{Block} + \beta_2 N_a^{\text{pre}} + \epsilon_a \tag{2.3}$$

We find that the coefficient β_1 is not significantly different from zero, suggesting that there is no difference in the change in centralization between the two types of

	Activity	Centralization	Conflict
$N^{p}re_a$	0.0119**	-0.0013***	-0.0001
	(0.0008)	(0.0002)	(0.0001)
$I^{B}lock$	-0.0319**	0.0007	-0.0474***
	(0.0127)	(0.0029)	(0.0030)
constant	-0.3691***	0.0637***	0.0253***
	(0.0079)	(0.0018)	0.0024

The standard errors of the parameter estimates are provided in parentheses. ***and **denote significance at 1% and 5%, respectively.

Table 2.2: *Regressions of (relative) change in activity, centralization and conflict over the dummy variable indicating whether an article has at least one editor blocked.*

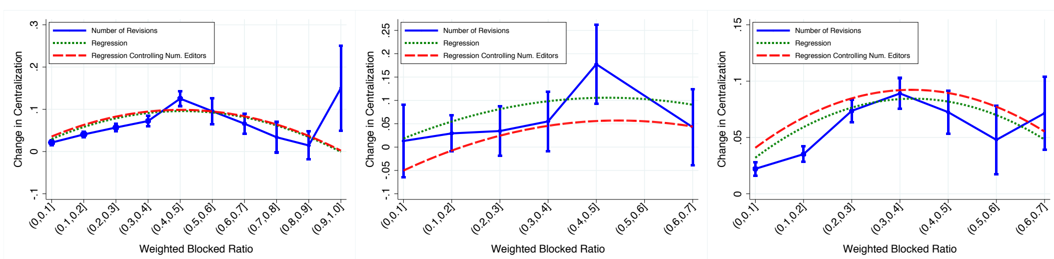


Figure 2.6: Change in gini as a function of shock level.

articles. To further investigate how the level of shock, B_a, relates to the changes in centralization of articles with at least one editor blocked, we regress C_a^{Δ} over the shock level, B_a, controlling for the number of editors in the pre-block period:

$$C_a^{\Delta} = f(B_a) + \beta N_a^{\text{pre}} + \epsilon_a \tag{2.4}$$

We fit both the linear and the quadratic models in the regression, and find that the quadratic one provides a significantly better fit than the linear model according to the likelihood ratio test (p-value $< 0.01\%$).

Figure 2.6 illustrates the relationship between C_a^{Δ} and B_a for all articles (a), articles with a small number of editors (b), and articles with a large number of editors (c). We find a consistent inverse U-shaped pattern – that the change in

	Activity		Centralization		Conflict	
E^{pre}	0.0055***	0.0055***	-0.0006***	-0.0005***	0.0002	0.0002
	(0.0008)	(0.0008)	(0.0002)	(0.0002)	(0.0001)	(0.0001)
B	-0.7471***	-0.7463***	0.1116***	0.3916***	0.0013	-0.2699***
	(0.0478)	(0.1472)	(0.0108)	(0.0327)	(0.0105)	(0.0293)
B^2	-	-0.0011	-	-0.3898***	-	0.4430***
		(0.1900)		(0.0430)		(0.0448)
$\chi^2(1)^\dagger$		0.00		81.95		96.69
LR test††		0.9954		0.0000		0.0000

† Reports the test statistics for the likelihood ratio test.
†† Reports the p-value for the likelihood ratio test.
For each measure, the left column represents the result from the linear regression and the right column represents that from the quadratic regression. The standard errors of the parameter estimates are provided in parentheses. ***denotes significance at 1%.

Table 2.3: *Regressions of the collaboration measures.*

centralization tends to increase with the shock level initially but decrease afterward. Using a back-of-the-envelope calculation based on the parameter estimates for the regression results in Table 2.3, we find that the break-point at which the change in centralization begins to decrease is consistently around the point where $B_a = 0.5$ for the three sets of articles shown in Figure 2.6.[7]

The initial increase in centralization is consistent with the threat rigidity theory, which suggests that when groups have a perceived threat they become more centralized. So what explains the change in behavior beyond $B_a = 0.5$? We present the reasoning below.

We now investigate the change in composition of a group – in terms of new versus old editors – as a function of the shock level. We let $Comp_a^{\text{pre}}$ and $Comp_a^{\text{post}}$ be the number of editors who were active during the post-block period and who edited article a for the first time during the pre-block period and post-block period, respectively. We then measure the fraction of new editors during the post-block period for an article a as $Comp_a = \frac{Comp_a^{\text{post}}}{Comp_a^{\text{post}} + Comp_a^{\text{pre}}}$. In Figure 2.7, we show how this measure varies across articles with different shock levels. The x-axis in this figure denotes the shock level and the y-axis denotes the fraction of new users ($Comp_a$). We provide boxplots and means (diamond shape) for articles of varying shock levels. The

[7]The break point is given by $-\beta_B/2\beta_{B^2}$.

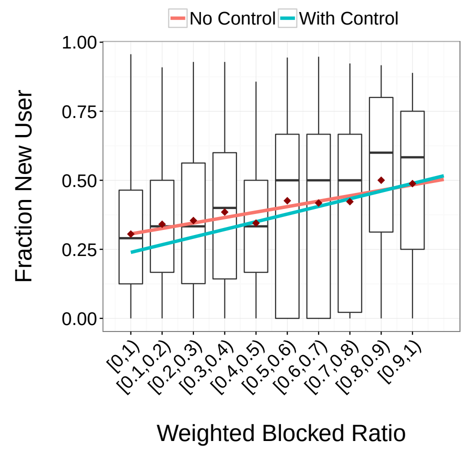


Figure 2.7: Change in composition as a function of shock levels.

red line and the green line are defined in the same manner as in the analysis of activity. The results show that as the shock level increases, the composition of a group post-block tends to include more new editors. This suggests that for the high shock levels, the composition of a group tends to be dominated by editors who joined the group later and therefore did not experience the shock. In fact, for $B_a >= 0.5$, the majority of editors for more than half of the articles are new to the group.

Given the finding on compositional effects, the break point in Figure 2.6 is now easier to interpret. As we move beyond $B_a \geq 0.5$, for instance, the majority of a group consists of users who joined the article after the shock and thus did not experience the shock. It is natural that for such articles, the changes in concentration are not as strong as in the cases where most group members experienced the shock and hence behave according to threat rigidity theory.

2.6.3 Conflict

We analyze conflict by comparing articles with and without editors blocked in the following regression:

$$R_a^{\Delta} = \beta_0 + \beta_1 I_a^{Block} + \beta_2 N_a^{\text{pre}} + \epsilon_a \tag{2.5}$$

We find that articles with at least one editor blocked experience a 2.2% drop in conflict, while those with no editors blocked experience a 2.5% increase in conflict. Given that articles with no editors blocked are not directly affected by the shock, this poses a conundrum. However, this can be explained once the trend of conflict is estimated during the pre-block period. Indeed, we find that articles with no editors blocked already experience a 2.5% increase when comparing time periods October 2004-May 2005 to May 2005-October 2005, while those with at least one editor blocked experience $< 0.001\%$ change in the same time period. This shows that the trend in conflict is unchanged for articles with no editors blocked, while those with at least one editor blocked shifts from constant conflict to a decreasing one.

Next, we examine the effect of the shock level on conflict in articles. To that end, we regress the amount of conflict of an article to the weighted ratio of blocked editors, controlling for the number of editors as follows:

$$R_a^{\Delta} = f(B_a) + \beta N_a^{\text{pre}} + \epsilon_a \tag{2.6}$$

Here, we consider articles that had at least one revert either one year before or after the block as this limits the analysis to articles with editors who are aware of the reverting feature. The results are consistent with the overall qualitative findings if all articles are included. Note that we also evaluated the fit of a linear model and find that the quadratic model provides a significantly better fit for the data given the likelihood ratio test (p-value $<$ 0.01%). The findings presented in Figure 2.8 are intriguing. We observe that for both small and large articles, small shocks result in a decrease in conflict. This finding is in agreement with threat rigidity theory, which suggests that when groups face an external threat they become more cohesive and hence exhibit less conflict Staw et al. (1981).

Moving from small to big shocks, we find an inflection point – after around $B_a > 0.3$, a larger blocked ratio results in a smaller reduction in conflict. As shown in the analysis of centralization, when large shocks occur, most of the current group members disappear and the new composition of the team consists of new group members. Thus, the increase in the change in conflict when B_a is large is likely due to the fact that most group members in these cases did not experience the shock.

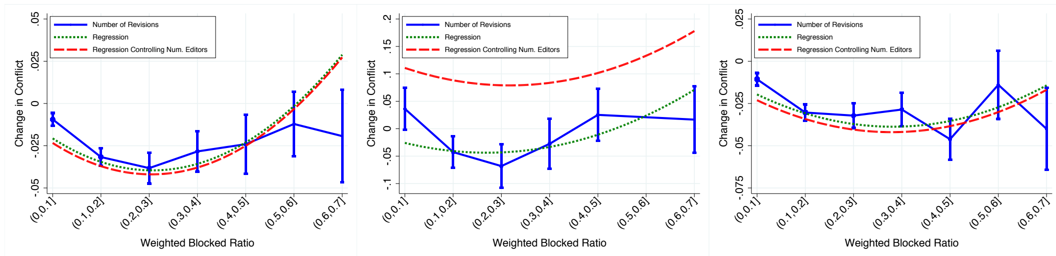


Figure 2.8: Change in conflict as a function of shock level.

2.6.4 Mediation Analysis

Centralization and conflict are likely to relate to each other. When a group is highly centralized, explicit coordination is less costly and it is easier to complete tasks without engaging in conflict (Kittur and Kraut, 2010). We observe that shock level had the opposite relationship with centralization than it does with conflict. Indeed, controlling for the shock level and the number of editors, we find a negative and significant relationship between centralization and conflict. It is possible that the shock affects conflict indirectly through its impact on centralization. To separate out the direct effect of weighted blocked ratio on conflict and any indirect effect through centralization, we conduct a mediation analysis (MacKinnon et al., 2007) among weighted blocked ratio, centralization, and conflict.

Figure 2.9 shows the model and Table 2.4 summarizes the decomposition of the direct effect and the indirect effect from the mediation analysis. The only significant effect that weighted blocked ratio has on conflict is the direct effect and there is no significant indirect effect through concentration. Indeed, the direct effect accounts for over 99% of the total effect that the weighted blocked ratio has on conflict. This suggests that while centralization directly impacts conflict, the observed non-linear effect that weighted blocked ratio has on conflict is independent of the effect of centralization.

2.7 Discussion

Through this research we seek to understand the impact of external shocks on crowds. To do so, we examine the 2005 Chinese government censorship of

	B^2	B
Direct Effect	0.4222***	-0.2471***
	(0.0448)	(0.0295)
Indirect Effect	0.0012	-0.0011
	(0.0035)	(0.0035)
Total Effect	0.4234***	-0.2482***
	(0.0446)	(0.0293)

Table 2.4: Results from the mediation analysis. The standard errors of the parameter estimates are provided in parentheses. ***denotes p-value < 0.01 respectively.

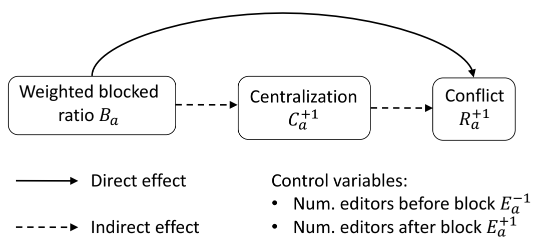


Figure 2.9: Mediation analysis diagram

Wikipedia. Results from our analysis provide four overarching findings, which have implications for research and design.

First, group size matters. Although size is not a key element in the threat rigidity literature on groups, it had an important role in our study. Larger crowds were able to maintain similar levels of activity when they experienced moderate shocks. Smaller crowds experienced more dramatic drop-offs in their level of activity. This supports the idea of resiliency through size. However, the opposite was true when shocks were more severe. For severe shocks, smaller crowds experienced smaller decreases in their level of activity, while larger crowds had dramatic drop-offs in their activity. The importance of size in understanding how groups respond to shocks may have been de-emphasized in prior literature, which did not significantly vary group size. However, our results suggest that size is vital to understanding how groups respond to threats.

Second, in the context of crowds the impact of shocks on centralization and conflict is not as straightforward as the literature suggests. Surprisingly, moderate shocks had a much more profound and lasting impact than severe shocks. In cases of severe shocks large portions of the crowd were lost and later replaced with newcomers. The greater the influx of newcomers into the crowd, the less the crowd displayed evidence of the shock. More specifically, these crowds are more decentralized and have more conflict compared to crowds that experience more moderate shocks and retain more of their previous members. Newcomers did not experience the shock and are likely to be less willing to support increases in centralization and decreases in conflict. Although this finding is novel, it is unclear whether it only applies to crowds or it could generalize to other settings.

Third, this study extends research on threat rigidity to include a large-scale validation in the context of online groups. As predicted by threat rigidity, crowds become more centralized and conflict decreased after those crowds experience a moderate shock. The fact that these are real groups and face a genuine threat may explain why our findings support threat rigidity while some prior studies do not (Argote et al., 1989; Gladstein and Reilly, 1985; Harrington et al., 2002). We also find that threat rigidity in the context of crowds appears to be much more complex than what we would expect to find in traditional groups. Nonetheless, this study presents a distinct opportunity to extend the research on threat rigidity in a more natural setting.

Finally, the results of this study have implications for design. The literature on threat rigidity suggests that there is not one correct way for groups to respond to a shock. Therefore, systems should be designed to support sudden changes because these are likely to fluctuate with exogenous shocks. Results of our study demonstrate that crowd size, the severity of the shock, and the availability of newcomers are key factors that designers have to consider when designing systems to support crowds.

Compared to other groups, crowds operate in uniquely volatile environments where coordination is difficult and conflict is probable. We examine the impact of an external shock on crowds by analyzing the effects of the 2005 Chinese government block of Chinese Wikipedia. This event provides a natural experiment that allows us to systematically analyze the effects of a real external shock on real crowds. We find compelling evidence that both generally supports threat rigidity and contextualizes it to crowds. Our findings can help to inform both theory and design of crowdsourcing systems.

CHAPTER III

Motivating Contributions to Public Information Goods: A Field Experiment at Wikipedia

Abstract

Motivating experts to contribute to public information goods can improve its quality. In a large-scale field experiment on Wikipedia, we find that experts are more interested in contributing when the likelihood of citation is mentioned. Conditional on a positive response, we find that citation benefit together with public acknowledgement increases contribution quality, whereas matching accuracy between a Wikipedia article and an expert's paper abstract, measured by cosine similarity, significantly increases both contribution quantity and quality, suggesting the potential of combining the predictive accuracy of machine learning with the causal inference of field experiments in promoting prosocial behavior.

3.1 Introduction

Online communities, social networking sites and other online social environments are increasingly being used to bring together labor and resource contributions to create public goods. The Wikipedia community has developed history's most comprehensive encyclopedia (Lih, 2009). Members of open source software development projects have created the software that runs the Internet and many other valuable software artifacts (Weber, 2004). Technical question and answer sites like the StackOverflow provide users with often highly specific advice about technical problems. Online health support groups, like BreastCancer.org and the American Cancer Society's Cancer Support Network, provide members both informational and emotional support to deal with serious illnesses (Wang et al., 2012).

These peer-produced public goods enabled by information technology, which we call *public information goods*, have distinct characteristics. They are information goods with free and open access to the general public. Unlike textbook examples of pure public goods, such as national defense, which makes exclusion technically costly or infeasible, public information goods are technically easy to exclude by requiring authentication. However, they are provided to the general public for free. Therefore, *public information goods* are non-rivalrous by nature and non-excludable by choice. Unlike charitable giving where everyone's contributions are perfect substitutes, accurately matching potential contributor's expertise with the right task can simultaneously improve the quality and lower the cost of contributions. Furthermore, accurate matching can even invoke a contributor's personal or professional identity which can also motivate contributions. For example, a game theorist working on equilibrium selection might find it less costly to comment on the Wikipedia article, "Coordination games," than on "Business cycle." Because of the expertise he has developed over the years, his contribution quality on coordination games will be higher than it would be in a poorly matched area. Additionally, he might be motivated to contribute to this article as he cares about this subject being introduced correctly to the general public.[1]

In this paper, we investigate individual motivations to contribute to public information goods from several perspectives. First, individuals may care about the social impact of the public good (Andreoni, 2007). For example, she might be more motivated to contribute if many recipients benefit from her contributions. Second, she might care about her private benefit from her contribution, such as being cited or publicly acknowledged. In addition, we investigate the effect of matching accuracy between the recommended public information good and potential contributors' expertise, as well as the social distance between the contributor and the askers.

Related research on charitable giving has identified two reasons why people donate to charity: first, they have been asked; and second, they have been asked by someone they care about (Castillo et al., 2014). In this paper, we add a third reason which is critically important in the context of public information goods - what they have been asked to do is important. We use machine learning techniques to match

[1]We thank David Cooper for helpful discussions.

expertise with tasks, which can scale to arbitrarily large communities and to any field with open-access content.

We conduct our study in the context of the English language Wikipedia. The English Wikipedia was founded in January 2001 and is operated by the Wikimedia Foundation. It is among the most important information sources for the general public.[2] As of the February of 2018, the English Wikipedia has accumulated over 5.5 million articles, with open and free access to all Internet users. The non-excludability property of Wikipedia distinguishes it from traditional repositories of scientific knowledge, such as most academic journals that require individual or institutional subscription fees for access. Recent field experiments demonstrate that Wikipedia not only reflects the state of scientific knowledge it shapes science (Thompson and Hanley, 2017).

We design our field experiment to investigate what motivates domain experts to contribute their expertise to public information goods. We exogenously vary the social impact of the public information good and potential private benefit of contributing using a 2-by-3 factorial design. Along the social impact dimension, we provide the experts with either the average number of views of a Wikipedia article or, in addition, a greater number of views which we use as a cutoff for all Wikipedia articles in our sample. Along the private benefit dimension, we vary the mentioning of the likelihood of citation of an expert's research with or without a public acknowledgement of their contributions.

We invited 3,974 academic economists who had at least five papers posted in a public research paper repository which we used for expertise matching. We find that the baseline positive response rate is 45%, much higher than the 2% positive response rate from a comparable field experiment inviting academic psychologists to review Wikipedia articles.[3] Compared to the baseline, mentioning citation benefit at high view further increase the positive response rates by 6 percentage points

[2]According to Alexa Internet, Wikipedia ranks among the top five most popular websites globally, with over 262 million daily visits. See `https://www.alexa.com/siteinfo/wikipedia.org`.

[3]In an unpublished field experiment, authors Farzan and Kraut emailed 9,532 members of the American Psychological Society inviting them to review Wikipedia articles, with a 2% positive response rate. They manipulated two main factors: who has done the work and who will benefit from the reviews provided by APS members.

(pp), whereas citation at high view, with or without public acknowledgement, reduces negative response rate by about 6 pp. Conditional on a positive response, the matching precision between the recommended Wikipedia article and the expert's research paper abstract, measured by cosine similarity, has a substantial and significant impact on both contribution quantity and quality. Furthermore, citation benefit and public acknowledgement together significantly increases contribution quality. Our findings suggest that precise matching of volunteers to tasks is critically important in encouraging contributions to public information goods, and likely to public goods provision and volunteering in general.

Our paper makes novel and important contributions to the vast experimental public goods literature (Ledyard, 1995; Vesterlund, 2015). First, it identifies public information goods as an increasingly important class of public goods and explores factors which encourage domain experts' contributions. Second, using machine learning and natural language processing techniques to match experts to Wikipedia articles, we identify matching accuracy between volunteers and tasks as a robust and significant predictor of both contribution quantity and quality. We expect this finding to generalize to other scholarly communities as well as other types of volunteer activities where expertise matter. On the methodology front, our approach synthesizes the predictive accuracy of machine learning with the causal inference of theory-guided field experiments (Kleinberg et al., 2015), representing a new wave of personalized intervention, analogous to the recent development of precision medicine (Collins and Varmus, 2015). Lastly, our field experiment has generated valuable public information goods, i.e., 1,188 expert comments on Wikipedia articles in economics, all of which have been posted on the Talk Pages of the corresponding Wikipedia articles, where Wikipedians coordinate with each other in the production process. These comments help improve the quality of Wikipedia articles during the Wikipedia Year of Science, an unprecedented initiative to improve articles related to STEM and the social sciences.[4]

[4] Our field experiment is neither funded nor influenced by the Wikimedia Foundation.

3.2 Literature Review

Neoclassical theories of public goods provision predict that rational individuals have an incentive to under-contribute to public goods as they do not internalize the positive externalities of their contributions on others (Bergstrom et al., 1986; Samuelson, 1954). Numerous experiments have been conducted to test and expand the theories. We refer the readers to Ledyard (1995) for a survey of laboratory experiments using the voluntary contribution mechanism in a wide range of environments, and to Vesterlund (2015) for a more recent survey of laboratory and field experiments on charitable giving.

Economists have developed several perspectives to address the problem of under-contribution. The mechanism design perspective relies on incentive-compatible tax-subsidy schemes enforced by a central authority.[5] Therefore, they cannot be directly applied to communities which rely on voluntary participation and contribution. In comparison, the social norms and identity perspective applies insights from theories of social identity to the study of economic problems (Akerlof and Kranton, 2000a, 2010). This body of research shows that when people feel a stronger sense of common identity with a group, they exert more effort and make more contributions to reach an efficient outcome (Chen and Chen, 2011; Eckel and Grossman, 2005).

Since the seminal work of Bénabou and Tirole (2011), a growing literature investigates how pro-social behavior can be motivated by the private benefits from acknowledgement and awards (Andreoni and Bernheim, 2009; Ariely et al., 2009; Rege and Telle, 2004). In the context of Wikipedia, Algan et al. (2013) study a diverse sample of 850 contributors and find that reciprocity and social image are both important motivations to foster individual contributions. Using naturally occurring data, Kriplean et al. (2008) categorize the types of barnstars, symbolic awards given to and received by Wikipedia editors as an appreciation of their work.[6] Following this line of research, Gallus (2016) uses a natural field experiment at the German

[5]See Groves and Ledyard (1987) for a survey of the theoretical literature and Chen (2008) for a survey of the experimental literature.

[6]A barnstar is an image accompanied by a short and often personalized statement of appreciation for the work of another Wikipedia editor. See `http://en.wikipedia.org/wiki/Wikipedia:Barnstar`.

Wikipedia to test the impact of symbolic awards on contribution levels. She finds that a purely symbolic award has a sizable and persistent impact on the retention of new editors. Building on these prior findings, our experimental design will explore the effects of private benefits, such as citation of one's own work and public acknowledgement, on contributions.

Another potentially important factor that influences contributions to public goods is the social impact, or the number of beneficiaries of the public goods. In the linear public goods environment with voluntary contribution mechanisms, laboratory experiments find a positive effect of group size on total contribution levels with certain parameter configurations (Goeree et al., 2002; Isaac and Walker, 1988; Isaac et al., 1994). By comparison, in the non-linear public goods environment, where the production function is concave in the sum of players' contributions, Guttman (1986) finds evidence that increasing the group size leads to an increase in aggregate contributions to the group, but a decrease in average contribution. More recently, Chen and Liang (2018) prove theoretically and find evidence in the lab that the effects of group size on public goods contributions depend on the complementarity of the production function. In the context of a congestable public good, Andreoni (2007) finds that although an increase in the number of recipients encourages a higher contribution, it does not lead to an equivalent increase in total contributions. The most closely related prior work on the effect of social impact on contributions to public information goods utilizes government blocking of the Chinese Wikipedia as exogenous shocks to the size of the readership, and find that it leads to a 42.8% decrease in the level of contribution by overseas Wikipedia editors who were not blocked during that time (Zhang and Zhu, 2011b). This paper indicates that a reduction in the social impact of the public information good discourages contributions.

In addition to the public goods literature in economics, we also benefit from the insights, techniques and measurements in the machine learning and natural language processing literature (Jordan and Mitchell, 2015; Manning et al., 1999). Cosley et al. (2007) deploy an intelligent task-routing agent, SuggestBot, to study how Wikipedia workload distribution interfaces affect the amount of work editors undertake and complete. They use SuggestBot to pre-process a dump of Wikipedia data to build a learning model of what articles a user might be interested in editing based on their past editing behavior. SuggestBot then recommends editing tasks to users through their talk pages. Their findings show that personalized recommendations lead to nearly four times as many actual edits as random suggestions. While Cosley et al. (2007) utilize Wikipedia editors' existing editing history to recommend articles, we motivate domain experts who have never edited Wikipedia articles to contribute, using their publications and working papers to infer their expertise and make personalized recommendations. Like Cosley et al. (2007), we also deploy a bot, ExpertIdeas Bot to post the experts' comments to the corresponding Wikipedia talk pages to standardize the format of the posts. Our approach demonstrates the potential of combining the predictive accuracy of machine learning with the causal inference of theory-guided field experiments in promoting prosocial behavior (Ai et al., 2016), representing a new wave of personalized intervention in economics.

3.3 A Theoretical Framework

In this section, we outline a simple theoretical model for contributions to public information goods. While our theoretical framework is closely related to the literature on voluntary contributions to public goods, we incorporate features of public information goods production into our model to better represent the context of our field experiment.

We study the behavior of potential contributors who choose whether and how much to contribute to a public information good, $y \geq 0$. To simplify notation, we use a single public information good and it is straightforward to generalize the results to multiple public goods. Let the set of potential contributors be I, and the number of consumers of this public good be $n \geq 0$. Each agent, $i \in I$, selects a contribution level, $y_i \in [0, T]$, where $T > 0$ is the total resources or time available to agent i. For simplicity, we assume that the quantity and quality of the public information good is the sum of individual contributions, $y = \sum_{j \in I} y_j$. A contributor's utility function is comprised of several components. Let the social impact of the public good be the product of y and the value derived from the number of consumers, $v(n)$, where $v'(\cdot) > 0$ and $v''(\cdot) \leq 0$. Thus, the first component of a contributor's utility function is $v(n)y$, which we call the social impact of the public information good. Incorporating the social impact of contributions is supported by the effects of the exogenous blocking of the Chinese Wikipedia on the contribution behavior of editors who were not blocked (Zhang and Zhu, 2011b).

The second component is the private benefit from the act of contribution, which could be the warm glow from contributing (Andreoni, 1989, 1990), or increased visibility of the contributor's own work, which should be an increasing function of the number of consumers. The private benefit from contributing can also be a function of the social distance between the potential contributor and the asker (Castillo et al., 2014). If a potential contributor is asked by someone she knows, she might be more likely to contribute due to a number of reasons, such as social image concern (Bénabou and Tirole, 2011), or social pressure (DellaVigna et al., 2012). We capture this type of motives by a social distance parameter, $s \in [0, 1]$, where zero denotes that the contributor does not care about the askers, and one denotes the maximum extent a contributor cares about the askers. We use a specification that is general enough to encompass various types of private benefit, $w(n, s)$, where $\partial w/\partial n \geq 0$, and $\partial w/\partial s \geq 0$. Thus, the private benefit of contribution is captured by $w(n, s)y_i$.

In comparison, a contributor's cost of contribution has two components. Contributing $y_i \geq 0$ entails a cost, $c(y_i)$, which is assumed to be convex in y_i. Let $r_i \geq 0$ be her marginal opportunity cost. We assume that her work time not spent on contribution to the public information good is devoted to improving her scholarship or paid work, yielding private benefit of $r_i(T - y_i)$.

Crucially for public information goods, we can use a recommender system to infer the expertise of a potential contributor based on her prior work, and recommend tasks which match her expertise. Let $m_i \in [0, 1]$ be the matching quality between an expert's expertise and the public information good. A good match can potentially have two effects. First, it reduces the cost of contributions as the expert is asked to contribute to content in her areas of expertise. Second, matching an expert to tasks in her domain of expertise might also invoke her professional identity, which could also increase the value she places on the public good. For simplicity, we focus on the former and omit the latter. Matching quality is primarily determined by the state of art of the recommender system. Let $f(m)$ be the probability distribution function of match quality. We assume that experts share the same common prior with regard to the distribution of matching quality.

We consider a two-stage process, participation and contribution, in a similar spirit as DellaVigna et al. (2012).

The first stage: participation. In the first stage, we elicit the expert's interests in contributing to a public information good in her area of expertise. In this stage, matching accuracy is not realized and the expert forms an expectation of the matching quality, $\bar{m}$. Therefore, an expert will decide to participate if the expected utility from participation dominates that of nonparticipation. Those who express interests in participation move to the second stage.

The second stage: contribution. In the second stage, upon observing recommended task and hence, the realized matching accuracy, m_i, expert i decides how much to contribute to the public information good. The accuracy with which the recommended work matches her expertise, m_i, reduces the contribution cost, $c(y_i)/m_i$. Therefore, the more accurate the match is, the lower the contribution cost will be. Expert i solves the following optimization problem:

$$\max_{y_i \in [0,T_i]} v(n)y + w(n, s_i)y_i + r_i(T - y_i) - \frac{c(y_i)}{m_i}. \tag{3.1}$$

Using backward induction, we first solve the optimal contribution level in the second stage, and then solve the participation decision in the first stage. Proofs are relegated to Appendix A. We obtain the following comparative statics for each stage.

Proposition 1 (Participation)**.** *Ceteris paribus, a potential contributor is more likely to participate if*

(a) more people consumer the public information good; or

(b) the private benefit of contribution is more salient; or

(c) her expertise overlaps with the askers; or

(d) her opportunity cost of time is lower.

In the second stage, after the realization of individual matching quality, we obtain the following comparative statics.

Proposition 2 (Contribution)**.** *After an expert agrees to participate, she will contribute more if*

(a) more people consume the public information good; or

(b) the private benefit of contribution is more salient; or

(c) her expertise overlaps with the askers; or

(d) her opportunity cost of time is lower; or

(e) matching quality between the public information good and her expertise is higher, when the marginal public and private benefit is at least as great as the marginal reputation cost.

These results provide guidance to our experimental design and form the basis for our hypotheses.

3.4 Experimental Design

We design our field experiment to explore factors that motivate domain experts to contribute to public information goods. We choose the English Wikipedia as the site for our experiment as it is one of the best known information resources used by the general public. We choose academic economists as participants, as we know the subject area well, and we have access to a public repository of economic research. In what follows, we present our sample selection strategies, design of treatments and experimental procedures.

3.4.1 Sample Selection: Experts and Articles

The experts whom we invite to contribute to Wikipedia are academic economists registered on *Research Papers in Economics* (*RePEc*).[7] *RePEc* is a public repository of working papers and journal articles in the field of economics. It maintains

[7] See `https://ideas.repec.org`.

a profile for each registered economist, including information about her research, such as fields of expertise and a list of publications and working papers. The recommendation of Wikipedia article is based primarily on experts' fields of expertise. For each expert, we identify her most recent field of expertise based on her most recent publications and working papers. Appendix B provides more details on the algorithm we use.

The initial sample consists of 31,670 economists who maintain a research profile at *RePEc*, of which 13,261 do not provide an email address or research specialization. We exclude these experts and reach a sample of 18,409 economists. To obtain our final sample size, we consider a minimum sample size based on (1) our power calculation, and (2) what is required by the recommendation algorithm to guarantee matching accuracy. In our power calculation, we specify $\alpha = 0.05$, $\beta = 0.10$, and ability to detect a 10 p.p change between two treatments holding one factor constant. Using the positive response rates from the data in our pilot conducted in the summer of 2015 ($N = 142$), we would need at least 636 participants per experimental condition, or 3,816 participants in all six experimental conditions, in the first stage. To further guarantee the accuracy of recommendation, we further restrict the experiment to the 3,974 experts with at least five research papers in English archived in *RePEc*.

The Wikipedia articles recommended to an expert are selected according to their relevance to her research. For each expert, we first use the Google custom search API to narrow down a list of Wikipedia articles that appear to be the most relevant to the keywords in the expert's research papers. Among these articles, we filter out those with fewer than 1,500 characters. We further eliminate articles viewed less than 1,000 times in the past 30 days. Therefore, all articles in our sample have a minimum amount of content for experts to comment on, with more than twice as many views as the average Wikipedia article at the time of our experiment, which was 426 views. The average number of views is computed using a Wikipedia data dump the month before the launch of our experiment.

In sum, our dataset contains 3,974 experts and 3,304 unique Wikipedia articles. For each expert, the dataset includes the number of times the abstracts for her research papers on *RePEc* have been viewed in 2016, whether she is ranked within the top 10 percentile at *RePEc*, and the affiliated institution.[8] For each Wikipedia article, our dataset includes the quality and importance class assessed by Wikipedia, the number of characters, the number of revisions, and the number of times it has been viewed over the past 30 days.[9]

3.4.2 Treatments

We implement a 2×3 between-subject factorial design in which we vary two factors in the emails inviting experts to contribute to Wikipedia (see Table 3.1). Along the social impact dimension, we vary the number of views of Wikipedia articles. In the Average View (AvgView) condition, we provide the experts with only the average number of views a typical Wikipedia article received in the past 30 days, which is 426. This information serves to set the experts' expectation on the social impact of a typical Wikipedia article. In the High View (HighView) condition, we provide an expert with the additional information that we will only recommend articles which have been viewed at least 1,000 times in the past 30 days. Recall that every Wikipedia article in our sample has been viewed at least 1,000 times. Along the private benefit dimension, we vary the expert's expectation on the amount of private benefit they might receive from their contribution. We include three conditions: a No Citation (NoCite) condition as the baseline, a Citation (Cite) condition, and a Citation & Acknowledgement (CiteAckn) condition.

The NoCite condition serves as a control and no private benefit is mentioned in the email. Only the average number of views is mentioned. For each condition, we send one of six personalized email messages. Each email consists of three sections. The first section is common to all treatments (with words in square brackets

[8]*RePEc* assigns a percentile ranking for each expert based on her number of publications and citations, and list the top 10 percentile in its public data base.

[9]The quality scale at Wikipedia contains the following six classes in increasing order: *Stub*, *Start*, *C*, *B*, *Good Article* and *Featured Article*. The criteria for various quality classes range from "little more than a dictionary definition" for the *Stub* class to "a definitive source for encyclopedic information" for the *Featured Article* class. The *Good Article* class is sometimes labeled *A class*. The importance scale at Wikipedia contains four classes: *Low*, *Mid*, *High* and *Top*. The criteria for various importance classes range from "not particularly notable or significant even within its field of study" for the *Low* class to "extremely important, even crucial, to its specific field" for the *Top* class. See detailed information at `https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Wikipedia/Assessment`.

Table 3.1: Features of Experimental Conditions

	No Citation	Citation	Citation & Acknowledge
Average View (426 times)	AvgView-NoCite ($N = 678$)	AvgView-Cite ($N = 669$)	AvgView-CiteAckn ($N = 671$)
High View ($\geq 1,000$ times)	HighView-NoCite ($N = 637$)	HighView-Cite ($N = 661$)	HighView-CiteAckn ($N = 658$)

personalized for each expert), starting with a brief introduction of Wikipedia and including the average number of views of a typical Wikipedia article:

> Dear Dr. [Chen],
>
> Would you be willing to spend 10-20 minutes providing feedback on a few Wikipedia articles related to [behavioral and experimental economics]? Wikipedia is among the most important information sources the general public uses to find out about a wide range of topics. A Wikipedia article is viewed on average 426 times each month. While many Wikipedia articles are useful, articles written by enthusiasts instead of experts can be inaccurate, incomplete, or out of date.

Depending on the experimental condition, the second section provides information regarding the readership of the articles to be recommended to the expert and the private benefits she expects to receive. In the HighView condition, we mention that we will select articles with over 1,000 views. In the Cite condition, we mention that the articles recommended to the experts are likely to cite their research, randomly choosing one of the following three messages: "*may include some of your publications in their references*", "*might refer to some of your research*", or "*are likely to cite your research*". Results from χ^2 tests show that the null hypothesis of independence between the actual realization of the email messages and the experts' first-stage responses cannot be reject for the Cite condition (p-value $= 0.564$) or the CiteAckn condition (p-value $= 0.435$).

> If you are willing to help, we will send you links to a few Wikipedia articles in your area of expertise. We will select only articles, with over

> 1,000 views in the past month, so that your feedback will benefit many Wikipedia readers.
>
> These articles might include some of your publications in their references.

The CiteAckn condition strengthens the private benefit by mentioning public acknowledgement as an additional benefit. In this condition, the experts are told in the email message that their contributions will be acknowledged on the WikiProject Economics page at Wikipedia (see Figure C.2).[10] WikiProject Economics is a group of Wikipedia editors who work together as a team to improve articles related to economics. Being acknowledged for one's contribution in the WikiProject Economics page thus serves as a private benefit in additional to the citation benefit. To avoid potential confound due to the (likely) asynchronous timing of experts' contributions, we freeze the acknowledgement page to include only contributions from our pilot phase throughout the main experiment. This way, every expert in this condition sees the same page.

The last section of the email inquiries whether the expert is willing to contribute by commenting on Wikipedia articles. The experts are provided with two options: "*Yes, please send some Wikipedia articles to comment on.*" and "*No, I am not interested.*" Authors Chen and Kraut signed the email with their respective titles and institutional affiliations. A screen shot of an example email in the HighView-Cite condition is included in Appendix C as Figure C.3.

Table 3.2 reports the summary statistics for the pre-treatment characteristics, broken down into the six experimental conditions. Panels A and B present the characteristics of the experts and recommended Wikipedia articles, respectively. Columns 1 through 6 report average values as well as standard deviations. We perform χ^2 tests on joint orthogonality across the treatments and report the associated p-values in column 7. Our results show that the randomization yields balanced experimental groups along most characteristics. However, recommended Wikipedia articles in the HighView-NoCite condition are longer and of higher quality class, compared to those in the other conditions.

[10] See detailed information at `https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Economics/ExpertIdeas`.

Table 3.2: Characteristics of Experts and Recommended Wikipedia Articles, by Experimental Conditions

	Average View			High View			
	NoCite	Cite	CiteAckn	NoCite	Cite	CiteAckn	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Characteristics of Experts							
Abstract Views	1,610	1,633	1,764	1,697	1,810	1,644	0.493
	(1,763)	(1,875)	(2,637)	(2,106)	(2,652)	(1,764)	
Top 10%	0.360	0.378	0.358	0.347	0.371	0.386	0.712
	(0.480)	(0.485)	(0.480)	(0.476)	(0.483)	(0.487)	
English Affiliation	0.417	0.457	0.434	0.452	0.477	0.407	0.103
	(0.493)	(0.499)	(0.496)	(0.498)	(0.500)	(0.492)	
Observations	678	669	671	637	661	658	
Panel B: Characteristics of Article Recommendations							
Article Length	34,266	33,973	34,579	36,269	35,000	34,150	0.044
	(33,552)	(33,194)	(34,269)	(36,399)	(34,875)	(33,582)	
Number of Edits	725	725	708	754	750	712	0.273
	(997)	(1,081)	(1,000)	(1,066)	(1,102)	(1,036)	
Views in Past Month	14,409	14,023	14,013	14,348	14,471	13,934	0.732
	(17,086)	(19,842)	(19,956)	(18.108)	(19,955)	(21,391)	
Quality:							
Featured Article	0.054	0.050	0.046	0.058	0.047	0.048	0.095
	(0.227)	(0.217)	(0.210)	(0.235)	(0.211)	(0.213)	
Good Article	0.216	0.211	0.215	0.226	0.205	0.201	0.120
	(0.412)	(0.408)	(0.411)	(0.418)	(0.404)	(0.401)	
B	0.594	0.604	0.601	0.581	0.613	0.613	0.037
	(0.491)	(0.489)	(0.490)	(0.493)	(0.487)	(0.487)	
C	0.127	0.125	0.126	0.123	0.122	0.127	0.978
	(0.333)	(0.331)	(0.332)	(0.328)	(0.328)	(0.333)	
Start & Stub	0.009	0.010	0.011	0.012	0.013	0.011	0.582
	(0.094)	(0.099)	(0.106)	(0.109)	(0.113)	(0.103)	
Importance:							
Top	0.168	0.160	0.158	0.173	0.152	0.153	0.077
	(0.374)	(0.367)	(0.365)	(0.378)	(0.359)	(0.360)	
High	0.350	0.339	0.353	0.347	0.358	0.348	0.630
	(0.477)	(0.474)	(0.478)	(0.476)	(0.480)	(0.476)	
Mid	0.255	0.270	0.256	0.245	0.264	0.263	0.192
	(0.436)	(0.444)	(0.437)	(0.430)	(0.441)	(0.440)	
Low	0.064	0.073	0.070	0.067	0.067	0.071	0.664
	(0.245)	(0.260)	(0.256)	(0.251)	(0.251)	(0.257)	
Observations	3,924	3,872	3,845	3,693	3,779	3,794	

Note. Columns 1 through 6 report average values in each experimental condition, whereas column 7 reports the p-value testing the joint orthogonality across treatments. Standard errors are provided in parentheses. There are four articles for which the quality class is unassigned.

3.4.3 Experimental Procedure

Our experiment consists of two stages. In the first stage, we send an initial email inquiring whether an expert is willing to provide comments on Wikipedia articles related to her domain of expertise. The subject line of the email contains the expert's area of expertise as identified by Algorithm 1 in Appendix B.

The experts who respond positively (i.e., clicking "*Yes*") to the first-stage email are then sent a second email immediately. This email starts by thanking the expert. It then presents a table listing the articles recommended to the expert for her participation. If the expert is in the HighView condition, the table also shows the actual number of views each recommended article has received in the past month. For each article, there is a hyperlink directing the expert to a webpage in which to put in her comments (see Figure C.1 in Appendix C).

To minimize entry cost so that an expert can comment on an article without having to learn how to edit a wiki, the webpage consists of a mirror image of the Wikipedia article on the right side of the screen and a dashboard on the left. In the mirror image of the article, we disable all the hyperlinks which can direct the expert to another page. The dashboard contains a textbox in which an expert can leave her comments on the article, while reading the article by scrolling it up or down. We display the mirror image of the article and the text box side by side so that the experts can input their comments without switching between browser pages. After the expert submits her comment, a thank-you email is sent to her immediately and her comments are posted on the talk page associated with the corresponding Wikipedia article by our bot, the ExpertIdeas Bot.[11]

The experiment started on May 6, 2016 and ended on December 22, 2016, during the Wikipedia Year of Science. The emails are sent between 6:00 AM and 7:00 PM on weekdays based on the local time of an expert's primary institutional affiliation. To avoid the emails being filtered as spam, we send up to 10 emails every four hours. Throughout the experiment, we use a tracking tool to monitor whether emails sent to an expert are opened. If the expert does not respond after two weeks, we send a reminder for at most four times. If the expert declines in any stage, no

[11] See `https://en.wikipedia.org/wiki/User:ExpertIdeasBot`.

more email will be sent to her. All emails are sent from the first author's Michigan email address.

3.5 Results

We first investigate the treatment effects on experts' participation decisions in the first stage. Conditional on a positive response, we next explore the impact of treatments, matching quality, opportunity cost of time and social distance on the experts' contribution behavior in the second stage.

3.5.1 First Stage: Participation

Among the 3,974 experts to whom we sent the first-stage email (our intent-to-treat sample), a total of 3,346 (84%) opened it, constituting our treated sub-sample. We find no significant difference in the likelihood to open the first-stage email between any pair of the six experimental conditions ($p > 0.10$ using proportion tests). Using the χ^2 tests, we confirm that the treated experts in the six treatments are balanced on every observable characteristics ($p = 0.561$ for Abstract Views, 0.490 for Top 10% and 0.383 for English Affiliation).

Figure 3.1 presents the proportion of positive responses from the treated experts, with the error bars denoting one standard error above and below the mean. The white bars plot the fraction of positive responses in the first stage, whereas the grey bars plot the fraction of experts who actually contribute during the second stage which we will explore in greater detail in the next subsection.

We first notice that the baseline willingness to participate is surprisingly high. In the baseline condition with average view and no mentioning of any citation benefit (NoCite-AvgView), 44.8% experts respond positively to our invitation, much higher than the 2% positive response rate from a comparable field experiment.[12] While there are several differences in the experimental design, we conjecture that our more accurate inference of the expert's domain of expertise in the email subject

[12]In an unpublished field experiment, authors Farzan and Kraut emailed 9,532 members of the American Psychological Society inviting them to review Wikipedia articles, and obtained a 2% positive response rate.

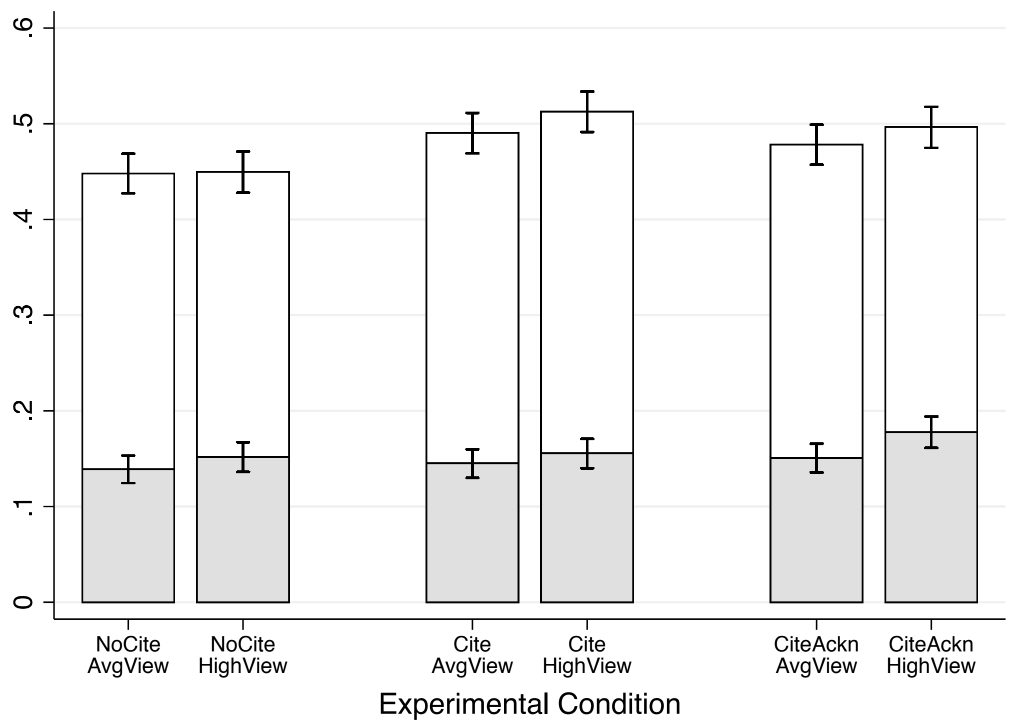


Figure 3.1: Proportion of positive responses in the first stage (white bar) and actual contributions in the second stage (grey bar).

line and the first paragraph of the first email might have led to the high response rate. Comparing the baseline with the other conditions, we find that including citation benefit in the email leads to at least a 4 p.p. increase in positive response rate ($p < 0.05$ using a proportion test), whereas adding public acknowledgment to the citation benefit does not appear to generate significantly more positive responses. Along the dimension of social impact, we find that the HighView condition with or without citation benefit does not significantly increase positive responses ($p =$ 0.959 for NoCite, 0.455 for Cite, and 0.542 for CiteAckn).

Proposition 1 predicts how our treatments might affect expert participation decisions, formulated below as Hypothesis 1.

Hypothesis 1. *The likelihood that experts express interests in participation follows the order of (a) AvgView < HighView, (b) NoCite < Cite, and (c) Cite < CiteAckn.*

In the actual implementation of the experiment, an expert can have three potential responses to our invitation email: positive (clicking "*Yes*"), negative (clicking "*No*"), or null response. To estimate the treatment effects on the experts' willingness to participate, we use the following multinomial regression framework:

$$
\begin{aligned}
R_i = \beta_0 &+ \beta_1 \times \text{HighView}_i + \beta_2 \times \text{Cite}_i + \beta_3 \times \text{CiteAckn}_i \\
&+ \beta_4 \times \text{HighView}_i \cdot \text{Cite}_i + \beta_5 \times \text{HighView}_i \cdot \text{CiteAckn}_i \\
&+ \mathbf{B_E} \times \text{expert-level controls}_i + \varepsilon_i,
\end{aligned}
$$

where the dependent variable R_i is an expert i' response, which can be positive (1), null (0) or negative (-1). The independent variables include the treatment dummies (HighView, Cite, and CiteAckn), the interactions among them, and expert-level controls such as the number of views one's abstracts received (as a proxy for the expert's reputation), whether an expert's primary institution is located in an English-speaking country, and whether an expert is in behavioral and experimental economics, which overlaps with the research areas of the research team as a proxy for social distance.

Table 3.3 reports the average marginal effects estimated from the multinomial logistic regression. Under the AvgView condition, the likelihood of negative response is reduced by 6.6 p.p. with citation benefits (p-value < 0.05), and by 4.5

Table 3.3: Average Marginal Effect on the First-stage Response

Dependent Variable:	Positive $\mathrm{P}(R=1)$ (1)	Null $\mathrm{P}(R=0)$ (2)	Negative $\mathrm{P}(R=-1)$ (3)	Positive $\mathrm{P}(R=1)$ (4)	Null $\mathrm{P}(R=0)$ (5)	Negative $\mathrm{P}(R=-1)$ (6)
HighView	0.002	0.021	-0.022	0.004	0.019	-0.023
	(0.030)	(0.026)	(0.027)	(0.030)	(0.026)	(0.027)
Cite	0.042	0.022	-0.064**	0.037	0.029	-0.066**
	(0.030)	(0.026)	(0.027)	(0.030)	(0.026)	(0.026)
CiteAckn	0.030	0.020	-0.050*	0.020	0.025	-0.045*
	(0.029)	(0.026)	(0.027)	(0.030)	(0.026)	(0.027)
HighView $\times$ Cite	0.021	-0.023	0.002	0.023	-0.028	0.005
	(0.042)	(0.037)	(0.037)	(0.042)	(0.037)	(0.037)
HighView $\times$ CiteAckn	0.017	-0.003	-0.013	0.022	-0.007	-0.014
	(0.042)	(0.037)	(0.038)	(0.042)	(0.037)	(0.038)
log(Abstract Views)				0.009	-0.039***	0.030***
				(0.009)	(0.008)	(0.008)
English Affiliation				-0.020	-0.037**	0.057***
				(0.018)	(0.015)	(0.015)
Overlapping Expertise				0.212***	-0.079***	-0.133***
				(0.034)	(0.028)	(0.025)
HighView + HighView $\times$ Cite	0.022	-0.002	-0.020	0.027	-0.009	-0.018
	(0.030)	(0.026)	(0.025)	(0.030)	(0.026)	(0.025)
Cite + HighView $\times$ Cite	0.063**	-0.001	-0.062**	0.060**	0.001	-0.061**
	(0.030)	(0.027)	(0.026)	(0.030)	(0.026)	(0.026)
HighView + HighView $\times$ CiteAckn	0.018	0.017	-0.036	0.025	0.012	-0.037
	(0.030)	(0.027)	(0.026)	(0.030)	(0.027)	(0.026)
CiteAckn + HighView $\times$ CiteAckn	0.047	0.016	-0.063**	0.041	0.018	-0.059**
	(0.030)	(0.027)	(0.027)	(0.030)	(0.027)	(0.027)
Model Specification		Multinomial Logistic			Multinomial Logistic	
Observations		3,346			3,301	

Notes. The dependent variable is the expert's response to the email in the first stage. Standard errors are provided in parentheses. Average marginal effects are calculated using the Delta method (Ai and Norton, 2003). *, ** and *** denote significance level at 10%, 5% and 1% level. Table D.1 in Appendix D provides robustness check using percentile measures of abstract view.

p.p. with citation and acknowledgement (p-value < 0.10). Similar results are obtained in the HighView condition: estimates for the average marginal effect is -6.1 p.p. for Cite + HighView $\times$ Cite (p-value < 0.05) and -5.9 p.p. for CiteAckn + HighView $\times$ CiteAckn (p-value < 0.05). We summarize the results below.

Result 1 (Treatment Effects on Participation)**.** *Under the HighView condition, mentioning citation benefit leads to a 6 p.p. increase in positive response rate, whereas under both the average and high view conditions, citation benefit leads to a 6 p.p. decrease in negative response rates.*

By Result 1, we reject the null in favor of Hypothesis 1(b), but fail to reject the null in favor of Hypothesis 1(a) or 1(c). Therefore, in the first stage, citation benefit significantly increases experts' participation interests, whereas social impact, at least between 426 and 1,000 views, does not.

Proposition 1 further suggests that the experts' participation decisions also depend on their opportunity cost and social distance with the askers. We use reputation as a proxy for an expert's opportunity cost of contribution. To examine this set of predictions, we measure an expert's reputation by one of three variables: 1) the number of views for her abstracts at *RePEc*, 2) whether her overall ranking is among the top-10 percentile of researchers at *RePEc*, and less obviously, 3) whether she is affiliated with an institution from an English speaking country. It turns out that all three measures are highly correlated. The Spearman's rank order tests indicate significant correlation between being ranked among top-10 percentile at *RePEc* and both of the other two measures (p-values < 0.01). Therefore, we use number of views for abstracts as a measure of reputation in our subsequent regression analysis, as it is a finer measure than Top 10%, which is binary. To measure social distance, we construct a dummy variable, Overlapping Expertise, which equals 1 if the expert has overlapping area of research with the research team and 0 otherwise. Hypothesis 2 formulates this set of predictions.

Hypothesis 2. *The likelihood that an expert is willing to participate increases (decreases) for those with overlapping expertise as the research team (a higher reputation).*

Columns 4 through 6 in Table 3.3 provides the average marginal effects from the multinomial logistic regression with expert-level controls. Note that the empirical distribution of Abstract Views is skewed toward zero (see Figure 3.2). To mitigate the potential impact from extreme values, we apply both log transformation and percentile ranking (Table D.1 in Appendix D) to Abstract Views in the regression. The effect of $\log$(Abstract Views) on negative response is 3 p.p. (p-value < 0.01). The experts who are affiliated with an institution from an English-speaking country is 5.7 p.p. more likely to decline the invitation (p-value < 0.01). Furthermore, the experts with overlapping expertise are 21.2 (13.5) p.p. more (less) likely to respond positively (negatively) than others (p-value < 0.01). We summarize the results below.

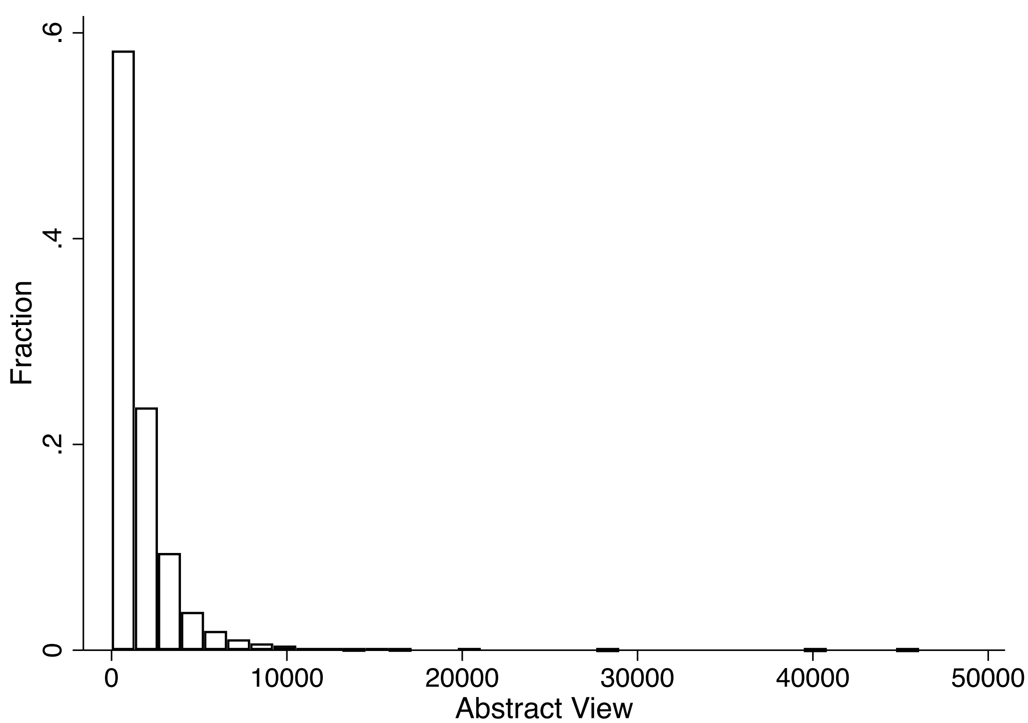


Figure 3.2: Empirical Distribution of Abstract View

Result 2 (Social Distance and Reputation Effects)**.** *An expert with overlapping expertise is 21 p.p. more likely to respond positively. In contrast, a unit increase in* *abstract views (affiliation with an institution from an English-speaking country) is associated with a 3 p.p. (6 p.p.) increase in negative response rate.*

By Result 2, we reject the null in favor of Hypothesis 2. The experts who enjoy a higher reputation (and thus associated with a higher opportunity cost) are more likely to respond negatively. From a back-of-the-envelope calculation, we find that a one standard deviation increase in $\log$(Abstract Views) is associated with a 25 p.p. increase in the likelihood of a negative response. This result is consistent with DellaVigna and Pope (2017), who invite academic scholars to predict the outcomes of a real-effort experiment and document that the completion rate is the highest among assistant professors and lowest among full professors. Furthermore, Result 2 supports the prediction in Proposition 1 that closer social distance is likely to yield more positive responses. To the extent that overlapping research expertise implies closer social distance (Akerlof, 1997; Castillo et al., 2014), our results indicate that in soliciting voluntary contribution to public information goods, an asker with closer social distance to the potential contributor is more likely to induce higher positive responses.

Results in this section reveal several interesting findings on increasing the interests of domain experts in contributing to public information goods. First, even the baseline positive response rate is as high as 44%, indicating the importance to ask and to personalize the ask. Second, mentioning citation benefit further increases experts' interests. Third, experts respond to peer solicitation. Those in the same research fields as the research team respond significantly more positively. Lastly, those with greater reputation (more views for author abstract) are more likely to say no.

Each expert who responded positively in the first stage receive a second email immediately, inviting them to comment on articles in their field of expertise within a month. Among the 1,603 experts who received a second email, 1,513 opened it (treated group). In this subsection, we restrict our analysis to this treated group. By the end of our experiment, 512 experts commented on at least one Wikipedia article and we received a total number of 1,188 comments. Figure 3.3 summarizes the number of participants in each stage of the experiment.

We evaluate both the quantity and quality of each expert's comments in our

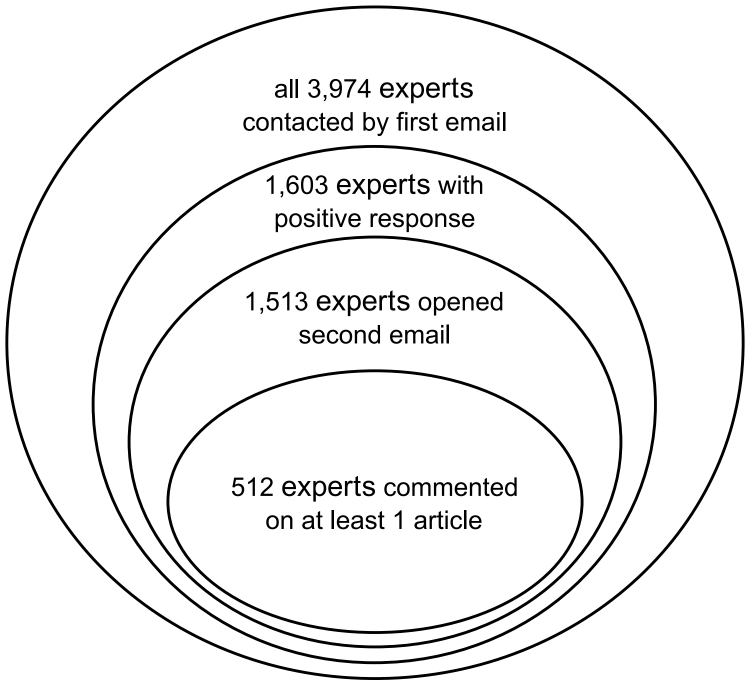


Figure 3.3: Experts' Responses in Each Stage of the Experiment

analysis. Contribution quantity is measured by the number of words in each comment. To measure contribution quality, we develop a rating protocol following the standard practice in content analysis (Krippendorff, 2003). Each comment is independently evaluated by three raters, who are expected to provide objective evaluations on the quality of the comments. In our rating procedure, raters first read the corresponding Wikipedia article. For each comment, raters start with a series of questions regarding various aspects of the comments prior to giving their overall ratings. Such a multi-item approach breaks down the global evaluation of the entire comment into concrete subcomponents and has been found to improve the inter-rate reliability for the overall quality rating (Strayhorn et al., 1993). The rating protocol and the corresponding summary statistics are provided in Appendix E.

We measure the quality of comments by the median of the three raters' responses to each of the three questions:

1. *Please rate the overall quality of the comment.* (1-7 Likert scale)

2. *Suppose you are to incorporate this comment. How helpful is it?* (1-7 Likert

scale)

3. *Suppose that you are to incorporate the expert's review of this Wikipedia article and you want to first break down the review into multiple comments. How many comments has the expert made to this Wikipedia article?* (non-negative integers)

A total of 68 raters participate in the evaluation of the experts' comments. These raters are recruited from junior/senior and graduate students at the University of Michigan who either major in economics or have completed the core requirements (including intermediate micro- and macroeconomics, as well as introduction to econometrics). All raters take part in a training session, which aims to reach a common understanding of the rating scale among them. In the training session, one research assistant first introduces the experiment to provide the raters with the background of the study. The research assistant then uses one piece of comment as an example and goes through the entire evaluation with the raters. For each rating question, the assistant discusses the rationales for the rating scale and provides clarifications for the rating instructions.

The raters conduct their evaluations through a web-based survey system, which requires Kerberose authentication. To guarantee that raters have background knowledge on the entries they evaluate, we assign the comments to those who have taken courses related to the associated Wikipedia articles.

Figure 3.4 presents the relationship between the two measures: the length of a comment, $\log(1 + \text{Word Count})$, is positively associated with the median rater's overall quality. Similar correlations hold between $\log(1 + \text{Word Count})$ and the helpfulness of a comment (Figure D.5) or the number of sub-comments in a comment (Figure D.6), both relegated to Appendix D. The Spearman's rank correlation between the quantity and the three quality measures varies between 0.663 and 0.682, and is statistically significant (p-value < 0.01). Similar positive associations between the quality and the quantity of experts' comments have been found in previous studies in the context of question-answering platforms, such as Yahoo! Answers (Adamic et al., 2008) and Google Answers (Chen et al., 2010c; Edelman, 2012).

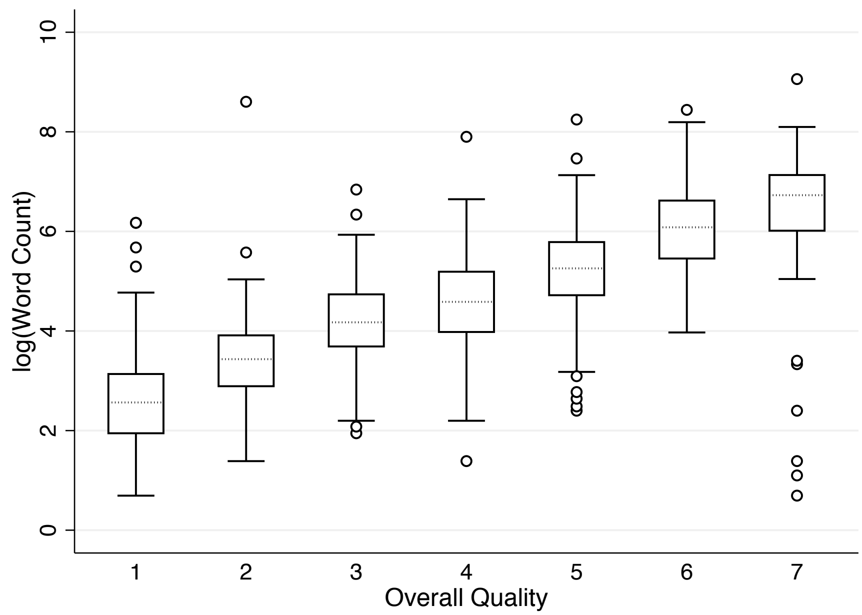


Figure 3.4: Word Count and Median Rater's Overall Quality Rating

Throughout the second-stage analysis, we specify the following statistical model:

$$\begin{aligned}Y_{i,k} = F\big(&\beta_0 + \beta_1 \times \text{HighView}_i + \beta_2 \times \text{Cite}_i + \beta_3 \times \text{CiteAckn}_i \\ &+ \beta_4 \times \text{HighView}_i \cdot \text{Cite}_i + \beta_5 \times \text{HighView}_i \cdot \text{CiteAckn}_i \\ &+ \mathbf{B_A} \times \text{article-level controls}_{i,k} + \mathbf{B_E} \times \text{expert-level controls}_i + \varepsilon_{i,k}\big),\end{aligned}$$

where i indexes the experts and k indexes the recommended Wikipedia articles. The dependent variable, $Y_{i,k}$, is the quantity or quality measure of an expert i's contribution to article k. HighView_i, Cite_i and CiteAckn_i are dummy variables representing the treatment status of expert i. The article-level controls include the article length, the quality class and the importance class. The expert-level controls include the number of views for one's abstracts, dummy variables for English affiliation and overlapping expertise with the research team.

Note that the data on contribution quantity features a semi-continuous distribution with a mass at the origin, as 86.5% articles recommendations receive zero contributions after the experts open the second-stage email. Such a large number

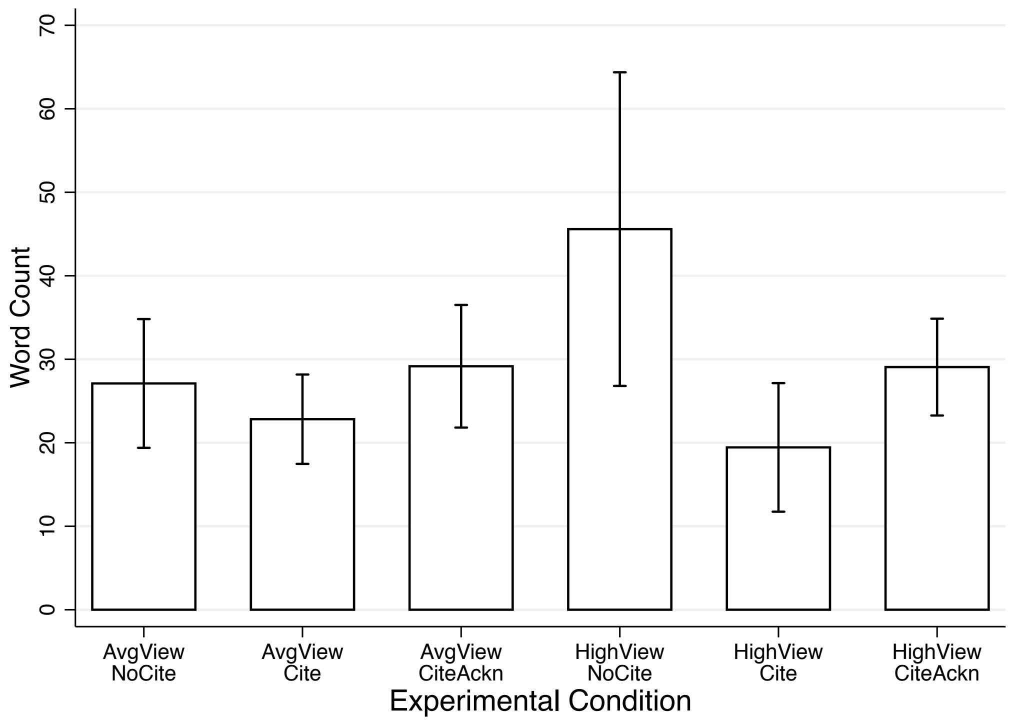


Figure 3.5: Average Word Count

of zeros make the common assumption of normality inappropriate and render the asymptotic inference problematic. To overcome this issue, we fit the data with the exponential dispersion model, which assumes that the variance of the outcome is a power function of the mean (Jorgensen, 1987; Zhang, 2013). Compared to other modeling choices which deal with an excessive number of zeros, the exponential dispersion model is applicable to continuous data rather than discrete ones.

We first explore how our experimental interventions can encourage experts' contribution. According to Proposition 2, an increase in the number of recipients or the private benefit yields more effort in contribution. This gives the following hypothesis.

Hypothesis 3. *Experts' comment quantity and quality follows the order of: (a) AvgView < HighView, and (b) NoCite < Cite < CiteAckn.*

Figure 3.5 plots average word count of the comments for each experimental

Table 3.4: Determinants of Contribution Quantity

Dependent Variable:	log(1 + Word Count)			
	(1)	(2)	(3)	(4)
HighView	-0.034	0.066	-0.051	0.029
	(0.100)	(0.214)	(0.101)	(0.216)
Cite	-0.070	-0.086	-0.085	-0.119
	(0.096)	(0.210)	(0.097)	(0.212)
CiteAckn	-0.069	-0.047	-0.086	-0.086
	(0.096)	(0.209)	(0.098)	(0.213)
HighView × Cite	-0.072	-0.202	-0.059	-0.177
	(0.137)	(0.299)	(0.138)	(0.302)
HighView × CiteAckn	0.131	0.147	0.149	0.173
	(0.138)	(0.295)	(0.139)	(0.299)
Cosine Similarity			1.768***	2.862***
			(0.166)	(0.359)
log(Article Length)			-0.040	-0.059
			(0.027)	(0.063)
log(Abstract View)			0.053**	0.083
			(0.032)	(0.069)
English Affiliation			0.095**	0.151
			(0.057)	(0.123)
Overlapping Expertise			0.373***	0.742***
			(0.099)	(0.194)
HighView + HighView × Cite	-0.105	-0.137	-0.110	-0.148
	(0.093)	(0.208)	(0.094)	(0.211)
Cite + HighView × Cite	-0.142	-0.289	-0.144	-0.296
	(0.097)	(0.212)	(0.098)	(0.215)
HighView + HighView × CiteAckn	0.098	0.213	0.097	0.202
	(0.095)	(0.203)	(0.096)	(0.207)
CiteAckn + HighView × CiteAckn	0.062	0.100	0.063	0.087
	(0.098)	(0.207)	(0.099)	(0.209)
Model Specification	OLS	Exp. Disp.	OLS	Exp. Disp.
Observations	8,819	8,819	8,635	8,635

Notes. The dependent variable is the log transformation of word count. Column (1) and (3) report the results from the OLS model and column (2) and (4) report the results from the exponential dispersion model. Quality class and importance class are controlled in all specifications. Fixed effects are included. Standard errors are reported in the parentheses. *, ** and *** denote significance level at 10%, 5% and 1% level. Table D.2 in Appendix D provides robustness check using percentile measure for article length and abstract view. The number of observations is the total number of recommended Wikipedia articles to experts who responded positively in the first stage.

Table 3.5: Determinants of Contribution Quality

Dependent Variable:	Overall Quality		Helpfulness		# of Sub-comments	
	(1)	(2)	(3)	(4)	(5)	(6)
HighView	0.870	0.899	0.846	0.868	0.885*	0.898*
	(0.161)	(0.168)	(0.157)	(0.163)	(0.056)	(0.058)
Cite	0.877	0.868	0.815	0.806	0.900*	0.894*
	(0.157)	(0.158)	(0.147)	(0.147)	(0.056)	(0.056)
CiteAckn	1.498**	1.565**	1.346	1.432**	1.094	1.119*
	(0.273)	(0.293)	(0.246)	(0.267)	(0.066)	(0.069)
HighView $\times$ Cite	1.403	1.429	1.642*	1.701**	1.122	1.139
	(0.375)	(0.386)	(0.439)	(0.460)	(0.105)	(0.107)
HighView $\times$ CiteAckn	1.058	1.020	1.239	1.152	1.045	1.008
	(0.275)	(0.241)	(0.322)	(0.306)	(0.092)	(0.090)
Cosine Similarity		11.904***		14.655***		3.421***
		(7.114)		(8.799)		(0.636)
log(Article Length)		1.062		1.084		1.074**
		(0.110)		(0.112)		(0.037)
log(Abstract View)		0.957		1.007		0.999
		(0.060)		(0.064)		(0.021)
English Affiliation		1.021		1.132		0.999
		(0.113)		(0.125)		(0.037)
Overlapping Expertise		1.381*		1.441**		1.108*
		(0.237)		(0.244)		(0.062)
HighView + HighView $\times$ Cite	1.220	1.285	1.388*	1.476**	0.993	1.022
	(0.235)	(0.253)	(0.267)	(0.290)	(0.068)	(0.070)
Cite + HighView $\times$ Cite	1.230	1.241	1.337	1.372	1.011	1.018
	(0.243)	(0.249)	(0.264)	(0.274)	(0.070)	(0.071)
HighView + HighView $\times$ CiteAckn	0.920	0.917	1.048	1.000	0.924	0.905
	(0.168)	(0.174)	(0.191)	(0.187)	(0.056)	(0.056)
CiteAckn + HighView $\times$ CiteAckn	1.584**	1.596**	1.668***	1.650***	1.143**	1.129*
	(0.295)	(0.302)	(0.310)	(0.311)	(0.072)	(0.072)
Model Specification	Ordered Logistic		Ordered Logistic		Poisson	
Observations	1,097	1,078	1,097	1,078	1,097	1,078

Notes. Columns (1)-(4) report odds ratio estimated from ordered logistic regressions. Columns (5)-(6) report incidence-rate ratio estimated from Poisson regressions. Quality class and importance class are controlled in all specifications. Fixed effects are included. Standard errors are reported in the parentheses. *, ** and *** denote significance level at 10%, 5% and 1% level. Table D.3 in Appendix D provides robustness check using percentile measure for article length and abstract view. Of the 1,188 comments provided by the experts, 1,097 remains after inappropriate comments are removed. The number of observations further drop to 1,078 after we remove experts without institutional affiliation information.

condition, with the error bars denoting one standard error. The average length of the experts' contribution ranges between 19.45 and 45.59, though it exhibits large variations. Along the dimension of article readership, the length of comments is not significantly different between the AvgView and HighView conditions (p-value = 0.65 for the NoCite condition, 0.10 for the Cite condition, 0.07 for the CiteAckn condition, using the Wilcoxon rank-sum test). Along the dimension of private benefit, we find that citation and acknowledgement does not lead to longer comments. Furthermore, given the HighView condition, it appears that the citation benefit reduces the quantity of experts' contribution (p-value < 0.05, using the Wilcoxon rank-sum test). The regression results in Table 3.4, which control for the expert fixed effect, also indicates no statistical evidence supporting treatment effects.

Result 3 (Treatment Effect on Quantity)**.** *Neither the social impact nor the private benefit variation leads to significantly longer comments.*

We next examine the treatment effect on contribution quality. The empirical distribution of median rating in Figure 3.6 demonstrates an increase in the quality of the comments along the dimension of private benefit. Nonparametric comparisons among the experimental conditions shows that the median overall quality is higher in the CiteAckn conditions than it is in the NoCite conditions (p-value < 0.05 for both the AvgView condition and the HighView condition, using the Wilcoxon rank-sum tests).

The regression analysis in Table 3.5 shows that the private benefit of acknowledgement consistently encourages high-quality contribution. The effect of CiteAckn on the proportional odds ratio for the ordered logistic model is significantly larger than 1. Put differently, the comments from the CiteAckn conditions are significantly more likely to receive a higher rating for overall quality. For example, the estimated marginal effect on the probability of be rated as 6 out of 7 is 3.38 p.p. in the AvgView condition (p-value < 0.01) and 3.32 p.p. in the HighView condition (p-value < 0.05) (see Table D.4 in Appendix D). Our results also speak to the quality measured by helpfulness (see column 3-4 in Table 3.5). Table D.5 in Appendix D shows that the average marginal effect of CiteAckn is significantly positive (negative) on the probability that the helpfulness of the comment is rated above (below)

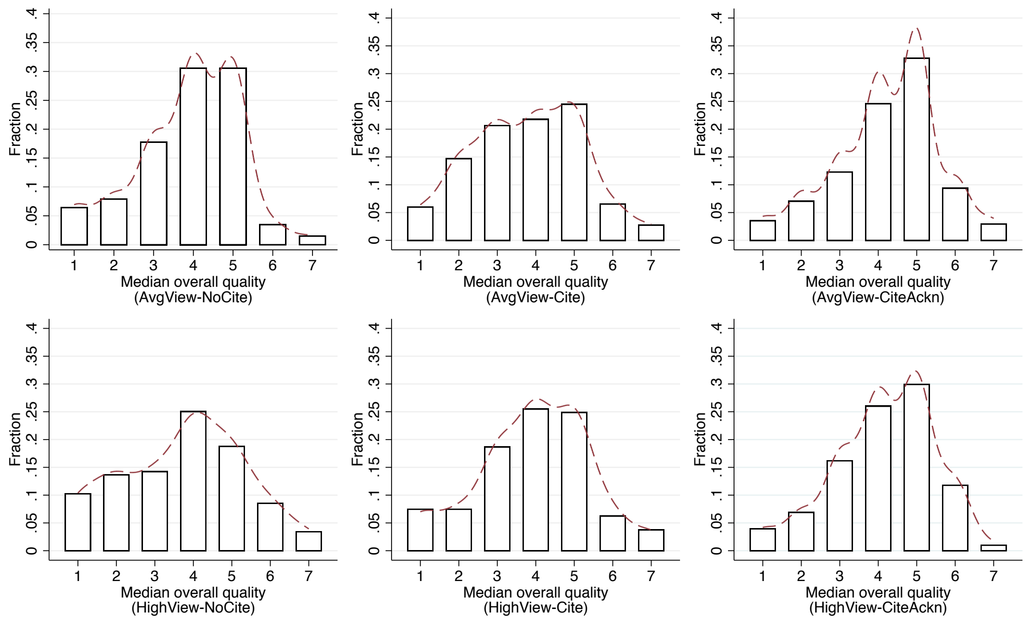


Figure 3.6: Distribution of Median Quality

4. The impact of CiteAckn on the number of sub-comments is positive but weakly significant.

Result 4 (Treatment Effect on Quality)**.** *Compared to the NoCite condition, the comments in CiteAckn condition receive significantly higher ratings regarding its overall quality and helpfulness.*

In sum, Results 3 and 4 offer mixed answers to our prediction in Hypothesis 3: while the private benefit of citation and acknowledgement is ineffective in generating longer comments, it consistently and remarkably improves the quality of the experts' contribution. Our result thus highlights the promise of non-monetary incentive, such as public acknowledgement, in inducing high-quality contributions.

We next investigate the prediction in Proposition 2 regarding how the proper matching between the expert's field of expertise and the Wikipedia articles can encourage contribution. To quantify the matching quality of the recommendations, we calculate the cosine similarity (Singhal, 2001) between the Wikipedia article k and expert i's research work. Cosine similarity is widely used in the area of informational retrieval as a measure of similarity between two documents. It starts by converting each of the two documents into a tokenizer (Huang et al., 2007) and processes them with a stemmer, which strips variants of the same word into its root (Airio, 2006). After this conversion process, the results are transformed into vectors by the tf-idf vectorizer, which weights by how specific the associated word is relative to the entire set of articles to be recommended (Leskovec et al., 2014). For example, the phrase "public goods" is given higher weight than the phrase "microeconomics", which in turn is given higher weight than a more generic word "economics". The similarity between a Wikipedia article and the expert' work is then given by the cosine value of the angle between the two word vectors.

Hypothesis 4. *The length and quality of a comment is higher if the cosine similarity between the Wikipedia article and the expert's abstract is higher.*

The cosine similarity between an expert's abstract and a Wikipedia article has an economically and statistically significant impact on the quantity of contribution. The parameter estimates on cosine similarity vary between 1.77 and 2.86 in columns 3 and 4 in Table 3.4 (p-value < 0.01). From a simple back-of-the-envelope calculation, we find that a unit increase in the cosine similarity implies 52.52 more words in an average comment.[13]

Consistent with the results on quantity measure, better matching between experts and Wikipedia articles remarkably improves the quality of contribution. Column 2 in Table 3.5 shows that a unit increase in the cosine similarity is associated with an increase of 11.90 in the odds ratio of overall quality. This represents, for example, an increase of 16 p.p. in the probability of being rated 6 (p-value < 0.01) and an increase of 7 p.p. in the probability of being rated 7 (p-value < 0.01). Similarly, columns 4 and 6 provide statistical evidence on the positive impact of cosine similarity on the helpfulness and number of sub-comments. The coefficient on the odds ratio of helpfulness is 14.66 (p-value < 0.01) and the coefficient on the incidence-rate ratio is 3.42 (p-value < 0.01). Result 5 summarizes our findings regarding cosine similarity.

[13]The change in character length is calculated as $\Delta(\text{Word Count}) = \hat{\beta}_x \cdot (1 + \text{Word Count}) \cdot \Delta x$, using $\hat{\beta}_x$ estimated in column 4 of Table 3.4.

Result 5 (Matching Accuracy and Contribution)**.** *An expert contributes longer and better comments to the Wikipedia articles with a higher cosine similarity to her research papers.*

Similar evidence is also provided by Edelman (2012) in an empirical study at Google Answers, who shows that the level of an answerer's specialization has a positive effect on the quality of her answers. Our result on cosine similarity highlights one feature distinguishing public information goods from the classic public goods context such as charitable giving - that the production cost exhibits substantial idiosyncrasy that depends on the matching quality between the specific public good and the contributor.

We next explore how opportunity cost affects contribution. Recall that we measure an experts' opportunity cost by 1) the number of views for her abstract and 2) whether she is affiliated with an institution from an English-speaking country. This yields the following hypothesis, again based on Proposition 2.

Hypothesis 5. *An expert with a higher number of abstract views or affiliated with an institution from an English-speaking country will contribute more.*

Kendall's rank correlation τ between Word Count and Abstract View is -0.013 and not significantly different from zero (p-value = 0.109). In the exponential dispersion model (column 6 in Table 3.4), the parameter estimates for $\log(\text{Abstract View})$ and English Affiliation is not significantly different from zero (p-value = 0.11 for both). Similar results also speak to the contribution quality. In Table 3.5, the coefficients on $\log(\text{Abstract View})$ and English Affiliation are not significantly different from zero in any of the three measure.

Result 6 (Opportunity Cost and Contribution)**.** *Conditional on being willing to participate, the length and quality of the contributions by experts who receive a higher number of views for abstract or who are from an English affiliation are not significantly different from those who receive a lower number of views or who are from an non-English affiliation.*

Finally, we investigate the prediction regarding social distance in Proposition 2.

Hypothesis 6. *The length and quality of a comment is higher if the expert's area of research overlaps that of the research team.*

A non-parametric comparison shows that the experts whose area of research overlaps that of the research team contribute comments that are significantly longer (p-value < 0.01 using the Wilcoxon rank-sum test). Returning to Table 3.4, the parameter estimates for Overlapping Expertise are 0.37 in the OLS model and 0.74 in the exponential dispersion model (p-value < 0.01 in both specifications). These estimates represent an increase of 2.83 and 11.09 more words in the comments, respectively.

The positive impact of social distance also applies to contribution quality. The estimates in columns 2, 4 and 6 in Table 3.5 indicate that overlapping expertise has a positive effect on the ratings for experts comments (though weakly significant for overall quality).

Result 7 (Social Distance and Contribution)**.** *An expert with overlapping expertise as that of the research team contributes longer and better comments.*

3.6 Conclusion

Public information goods, such as Wikipedia, have the potential of giving everyone "free access to the sum of all human knowledge" (Miller, 2004). To reach that potential, we need contributions from not only enthusiasts, but also domain experts in various fields. This study explores factors that motivates domain experts to contribute to public information goods.

Using a field experiment designed to explore both the private benefit and the social impact of contributions, we find that private benefits, such as the likelihood of citation of one's own work, significantly increases experts' interests to participate. Shorter social distance, such as overlapping domain of expertise with the askers, also increases the likelihood of participation. Conditional on willingness to contribute, public acknowledgement of one's contributions increases the quality of contributions. These finding affirms the motivating effects of non-pecuniary private benefits in giving. From the large research literature on the charitable giving, we know that people give to charity because they are asked to do so, and that they are more likely to give if they are asked by someone they care about. We confirm both of these findings in the public information goods context.

Furthermore, we uncover a third factor in motivating contributions to public information goods, that is, *what* you ask people to do is crucially important. Accurate matching between expertise and the task significantly increases both contribution quality and quantity. Specifically, in our experiment, conditional on participation, the accuracy of matching a participant's expertise and the tasks, measured by the cosine similarity between an expert's abstract and the corresponding Wikipedia article, is the single most significant predictor of both contribution quantity and quality. This result highlights the potential of utilizing information technology, such as recommender systems, in promoting pro-social behavior.

Beyond public information goods, we expect that matching accuracy between a contributor's expertise and tasks will improve public goods contributions and volunteering more generally. Using machine learning to accurately match volunteers with tasks is a new and promising tool for social scientists. We expect that there will be more work along this direction.

CHAPTER IV

Group Membership and Contributions to Wikipedia: The Case of WikiProject

Abstract

We investigate the effects of group identity on contribution behavior on the English Wikipedia, the largest online encyclopedia that gives free access to the public. Using an instrumental variable approach that exploits the variations in one's exposure to WikiProject, we find that joining a WikiProject has a significant impact on one's level of contribution, with an average increase of 79 revisions or 8,672 character per month. To uncover the potential mechanism underlying the treatment effect, we use the size of home page for WikiProject as a proxy for the number of recommendations from a project. The results show that the users who join a WikiProject with more recommendations significantly increase their contribution to articles under the joined project, but not to articles under other projects.

4.1 Introduction

Teams, groups and organizations are prevalent in the online economy to motivate and organize the production of public information goods. Classical examples include Github teams at the software development community[1] and Kiva lending teams at the microfinance site[2]. The rise of many Internet platforms – whose content relies entirely on voluntary contributions – has provided a rich set of real world examples to examine the question on motivating the provision of public goods by team-based approach. Would team membership induce higher contributions to pub-

[1]See `https://help.github.com/articles/about-teams/`.

[2]See `https://www.kiva.org/teams`.

lic information goods from individuals? What is the mechanism through which teams exerts its impact on motivating private contributions?

In this paper, we present an empirical study that investigates the effect of team membership on the provision of public information goods at Wikipedia. Wikipedia was created to provide web-based, free-content encyclopedia to the public. Ever since its establishment in 2001, Wikipedia has developed the most comprehensive encyclopedia in history. As of May 2018, it has accumulated more than 5.6 million articles, with over 838 million revisions from 33 million registered users. According to statistics from Alexa Internet, Wikipedia is ranked among the top five most popular websites globally[3].

The content of Wikipedia relies entirely on the voluntary contribution. To motivate users' efforts, Wikipedia introduced WikiProject - a collaboration platform through which users create and join groups in 2001. A WikiProject is typically focused on a specific topic and has a set of relevant Wikipedia articles that fall in its scope. It provides guidance for its members in various ways, such as offering advices for users and keeping track of articles of interests. Thus, the introduction of WikiProjects offers an opportunity to explore how team-based approach can motivate voluntary contributions to public information goods.

Evaluating the causal impact of WikiProject on user contribution is complicated by the potential bias due to selection on unobservables. For example, the Wikipedia users who are WikiProject members can be inherently more active than those who are not, introducing systematic bias in cross-sectional comparison. We deal with this challenge by exploiting plausibly exogenous variation in how intensive users are exposed to WikiProjects when they contribute to Wikipedia articles. Becaused Wikipedia articles are manually classified to WikiProjects by users, the quasi-randomness in the association of articles to WikiProjects offers a natural source of variation in the level of exposure. Using an instrumental variable approach, we find that joining a WikiProject has a substantial effect on user contribution, measured by both the number and size of revisions.

To uncover the potential mechanisms that drive our main results, we examine how the increases in user contribution relate to the characteristics of WikiProjects.

[3]See https://www.alexa.com/siteinfo/wikipedia.org

Using the character size of project homepage as a proxy for the number of recommendations, we find that the more suggestions a WikiProject offers, the greater the increase in the contribution level of the users who join. This result suggests that team coordination, which lowers the search cost one incurs to identify the Wikipedia articles that she is interested in contributing, could be the factor explaining the effect of team membership.

4.2 Literature Review

Motivating voluntary contribution to public goods is of major interest for economics since the seminal work of Samuelson (1954) and Bergstrom et al. (1986). To address the problem of under-contribution, the mechanism-design approach focuses on incentive-compatible schemes with tax-subsidies that internalize the externalities from one's contribution (Chen, 2008; Groves and Ledyard, 1987). However, these schemes are typically enforced by a central authority and therefore can hardly be implemented in domains that rely on voluntary participation such as Wikipedia.

The social identity approach, in comparison, focuses on inducing pro-social behavior by fostering a common identity among individuals. Akerlof and Kranton (2000b) offers a systematic theoretical framework that incorporates social identity into standard economic analysis and shows that deviation from the expected behavior prescribed by the group leads to disutility. The identity-based model has since been applied to a large variety of economic contexts. For example, various laboratory evidence shows that participants with a salient group identity exert more effort and achieve a more efficient outcome in the context of minimum-effort game (Bornstein et al., 2002; Chen et al., 2010b; Weber, 2006) and public goods provision (Charness et al., 2014; Croson et al., 2008; Eckel and Grossman, 2005).

The potential of identity-based approach to induce economically efficient outcome has also been examined in the field. For example, Chen et al. (2017) and Ai et al. (2016) examine the impact of team membership on pro-social lending behavior at Kiva. Using both naturally-occuring data and field experiments, they provide evidence demonstrating that users are substantially more engaged in lending behavior after joining a team and attribute to the effect to goal-setting and team coordination.

Similar evidence is also provided by Erev et al. (1993), who show that competition between teams leads to higher performance in orange-picking tasks. Our study contributes to understanding identity-based approach in fostering provision of public information goods and indicating that coordination through task suggestion could be one underlying mechanism.

A large body of empirical works dedicated to increasing user engagement at Wikipedia. Various factors have been shown to influence user motivations, such as size of audience (Zhang and Zhu, 2011a), number of collaborators (Solomon and Wash, 2014), private benefit (Gallus, 2016; Kriplean et al., 2008). More recently, a number of studies provide both quantitative and qualitive evidence on how user contribution at Wikipedia is related to a variety dimensions of WikiProjects, including experience composition of members (Chen et al., 2010a; Solomon and Wash, 2014; Zhu et al., 2012), division of labor (Kriplean et al., 2008; Morgan et al., 2014), coordination (Kittur and Kraut, 2008; Kittur et al., 2009; Morgan et al., 2013; Ransbotham and Kane, 2011). For example, Solomon and Wash (2014) documents the observations from over 1,000 WikiProjects and find that the edits provided by peripheral editors have a positive effect of the activeness of the project. Chen et al. (2010a) explores how the age composition of WikiProject members can encourage or frustrate newcomers. They find that WikiProjects with the age diversity that is either extremely low (the experience of WikiProject members are evenly distributed) or extremely high (most WikiProject members are newcomers or veterans) increases the likelihood that new editors leave. Based on the longitudinal history of 2,065 featured articles at Wikipedia, Ransbotham and Kane (2011) investigates the impact of editors' coordination measured by membership turnover on the success of collaboration. Their results indicate that the inclusion of new members helps an article be promoted to the featured article status. Therefore, the amount of coordination appears to have a curvilinear relationship with article outcomes and coordination among editors can be especially important during knowledge retention stage. While all these studies identify various characteristics of WikiProjects that correlate with user contribution, one common limitation is that they do not deliver a causal interpretation.

4.3 WikiProject: An Overview

In September 2001, the Wikimedia Foundation launched WikiProject, a collaboration platform that allow editors with similar interests connect to each other and coordinate their efforts. Both the number and size of WikiProject have expanded since then and as of May 2018, there exist over 2,300 projects in Wikipedia, with varying degrees in size and activity. These projects are related to either an area of topics (e.g., WikiProject Economics) or a specific type of maintanence task that applies to the entire Wikipedia (e.g., WikiProject Article for creation). The task-based WikiProjects typically consists of administrators, who are typically granted the access to perform certain actions at Wikipedia, such as block user accounts and deleting pages. Because the purpose of this study is to examine how WikiProject impacts the contribution behavior of ordinary users, we will only focus on topic-based projects.

According to Wikipedia, a WikiProject is "a group of contributors who want to work together as a team to improve Wikipedia". They typically focus on a specific topic and has a mission statement articulating its scope, goal and vision on the homepage. For each WikiProject, there is an associated talk page, which functions as a forum that enables open discussions. In addition, many WikiProjects have its own project userbox - an emblem that members can put on their userpage to demonstrate their membership. For example, the userbox for WikiProject Economics is a supply-demand diagram sided with the note "This user is a member of WikiProject". These components constitute a basis for the project membership, which can foster project identity.

Wikipedia allows any registered user to initiate and join a WikiProject. To initiate a WikiProject, one needs to create a homepage that follows a standard template, which includes pre-defined sections such as introduction to scope of the project and a list of participants. When a WikiProject is created, a Wiki link for the project, which allow other Wikipedia pages to be directed to it, is automatically registered.[4].

[4]Edits at Wikipedia are made through the Wiki markup, a coding language that is designed to facilitate the presentation of text. A Wiki link is a string of Wiki markup text specially designed to enable hyperlink across pages at Wikipedia. For a WikiProject, its Wiki link is represented by its name enclosed with double brackets, such as {{`WikiProject Economics`}}.

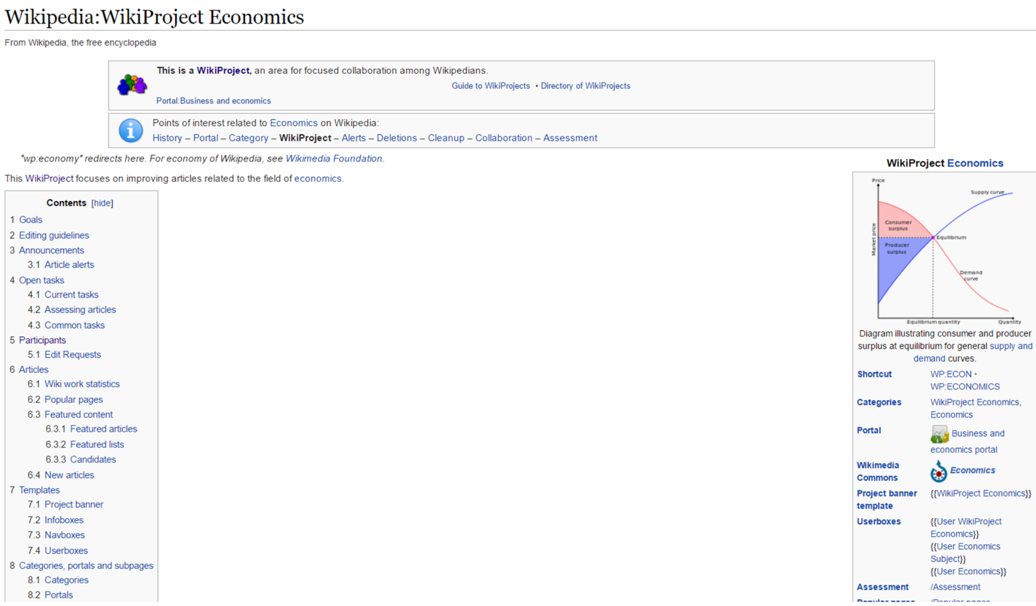
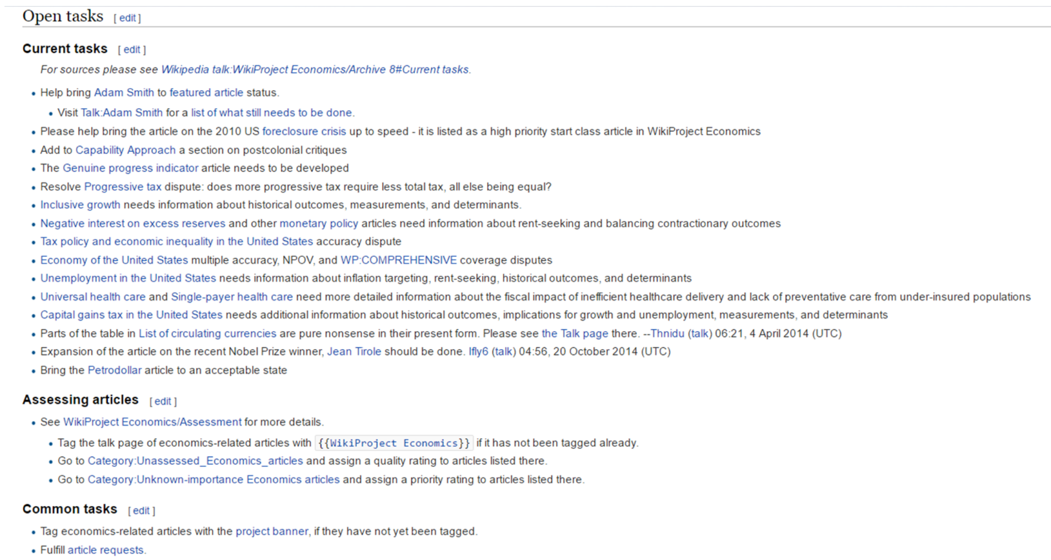


Figure 4.1: Screenshot for the Homepage of WikiProject Economics

Wiki link has been extensively used by user to join a WikiProject and associate articles with a WikiProject. To join a WikiProject, a user can either directly add the Wiki link of her user page to the list of members of the project, or incorporate the Wiki link of the project to her user page. To associate articles with a WikiProject, one can simply add to the talk page of the article the Wiki link of the corresponding WikiProject.

To increase user engagement, WikiProjects use a variety of methods to motivate and organize members' efforts. For example, several WikiProjects designate one or two articles to create or improve during a collaboration period. Once the target article for collaboration is chosen, it is in the project homepage and members will get notified. Zhu et al. (2012) find that during the collaboration period, WikiProject members increase their contributions by 5 times. Other examples include list of open tasks and monthly newsletters. These recommendations are usually presented on the homepage in a salient manner and the Wiki links of the associated articles are typically embedded. These suggest that article recommendation could be a potential mechanism through which WikiProjects motivate user contribution.



Open tasks [edit]

Current tasks [edit]

For sources please see Wikipedia talk:WikiProject Economics/Archive 8#Current tasks.

- Help bring Adam Smith to featured article status.
  - Visit Talk:Adam Smith for a list of what still needs to be done.
- Please help bring the article on the 2010 US foreclosure crisis up to speed - it is listed as a high priority start class article in WikiProject Economics
- Add to Capability Approach a section on postcolonial critiques
- The Genuine progress indicator article needs to be developed
- Resolve Progressive tax dispute: does more progressive tax require less total tax, all else being equal?
- Inclusive growth needs information about historical outcomes, measurements, and determinants.
- Negative interest on excess reserves and other monetary policy articles need information about rent-seeking and balancing contractionary outcomes
- Tax policy and economic inequality in the United States accuracy dispute
- Economy of the United States multiple accuracy, NPOV, and WP:COMPREHENSIVE coverage disputes
- Unemployment in the United States needs information about inflation targeting, rent-seeking, historical outcomes, and determinants
- Universal health care and Single-payer health care need more detailed information about the fiscal impact of inefficient healthcare delivery and lack of preventative care from under-insured populations
- Capital gains tax in the United States needs additional information about historical outcomes, implications for growth and unemployment, measurements, and determinants
- Parts of the table in List of circulating currencies are pure nonsense in their present form. Please see the Talk page there. --Thnidu (talk) 06:21, 4 April 2014 (UTC)
- Expansion of the article on the recent Nobel Prize winner, Jean Tirole should be done. Ifly6 (talk) 04:56, 20 October 2014 (UTC)
- Bring the Petrodollar article to an acceptable state

Assessing articles [edit]

- See WikiProject Economics/Assessment for more details.
  - Tag the talk page of economics-related articles with `{{WikiProject Economics}}` if it has not been tagged already.
  - Go to Category:Unassessed_Economics_articles and assign a quality rating to articles listed there.
  - Go to Category:Unknown-importance Economics articles and assign a priority rating to articles listed there.

Common tasks [edit]

- Tag economics-related articles with the project banner, if they have not yet been tagged.
- Fulfill article requests.

Figure 4.2: Open Tasks Listed at WikiProject Economics

4.4 Data and Summary Statistics

Our original data come from the top 10,000 editors on the English Wikipedia ranked by number of edits. Wikipedia allows a user to contribute anonymously and the edits will be recorded with only the IP address. Since joining a WikiProject requires an editor to have a username, our analysis focuses only on the editors with a registered account. We also exclude bot editors which are automatic tools developed to perform programmed tasks. The bot editors are usually involved in maintaining the Wikipedia community, such as organizing syntax of text and fighting vandelism, but not editing articles with real topics.

We eventually identify 9,183 editors with registered accounts. The data of these editors' contributions are collected through the Wikipedia data dump that archive all editors' complete editing histories on Wikipedia. Our data contains 98,584,912 edits over a time period spanning from February 2001 and April 2015. Each edit includes the editor's username, the size (i.e. number of characters changed), timestamp, editor's comment (i.e. description of the edit such as adding links) and the title of the page on which the edit is made.

Because editors' contributions to Wikipedia are sporadic, analyzing the behavior on a daily level can suffer from sparsity of the data. Therefore, we aggregate the editors' edits by at the monthly level. Our final dataset contains 821,192 editor-month observations, which constitute our unit of analysis. For each observation, the data contains the editor's username, the corresponding month, the size and number of edits contributed during that month both as aggregate and breaking down into additions and deletions.

Table 4.1 presents the summary statistics for our dataset. Among the 9,133 editors in our sample, there are 6,215 who join a WikiProject with a total of 576,132 editor-month observations. The main outcome variables that we use in the analysis include the number and size of edits, additions and deletions. The size of edits is measured by the number of characters modified by the editor. For each measure, Table 4.1 gives the average value both for entire sample and for the treated and non-treated editors separately. The 9,183 editors contribute a average of 116 edits in every month, which amounts to 44,114 characters. The treated editors who join a WikiProject contribute on average 114 edits and 44,345 characters in each month. Prior to joining a WikiProject, the treated editors contribute to 64 edits and 22,300 characters. After that, they contribute to 139 edits and 55,172 characters on average. That is, after joining a WikiProject, the average contribution level measured by the number and size of edits more than doubles.

Table 4.1: Summary Statistics of Editors

		Treated Editors			
	All Editors	Before	After	Combined	Non-treated Editors
Number of Edits	116.244	63.908	138.991	114.260	120.906
	(255.267)	(191.380)	(270.028)	(249.414)	(268.468)
Number of Additions	71.796	42.164	84.260	70.395	75.090
	(171.809)	(134.900)	(181.505)	(168.756)	(178.738)
Number of Deletions	36.810	17.860	45.772	36.579	37.353
	(98.430)	(63.083)	(103.026)	(92.742)	(110.653)
Size of Edits	44.114	22.300	55.172	44.345	43.571
(1,000 characters)	(284.032)	(153.677)	(302.122)	(263.117)	(327.993)
Size of Additions	30.888	16.170	38.367	31.056	30.495
(1,000 characters)	(207.251)	(120.918)	(223.431)	(195.968)	(231.623)
Size of Deletions	13.226	6.131	16.805	13.289	13.096
(1,000 characters)	(122.263)	(57.892)	(136.930)	(117.060)	(133.701)
# of Editors	9,133		6,215		2,968
Observations	821,192	189,764	386,368	576,132	245,060

4.5 Results

In this section, we present our results investigating the impact of WikiProject on users' contribution behavior. We first explore how joining a WikiProject affect one's contribution measured by the number and size of revisions. Then, we explore the mechanism underlying its impact.

4.5.1 Proof of Treatment

As introduced in Section 3, WikiProject induces a common identity among users who share similar interests. How does such a common identity affect individual contribution behavior? The group-contingent social preference framework (Basu, 2006; Chen and Li, 2009; Mcleish and J.Oxoby, n.d.) models a potential contributor as maximizing a weighted sum of her own and others payoff, and predicts that those who share the same group identity tend to behave more pro-socially. Evidence from both laboratory (Chen and Chen, 2011; Eckel and Grossman, 2005) and field experiments (Ai et al., 2016; Erev et al., 1993) provide supports to this prediction. Therefore, we expect users who join a WikiProject increase their level of contribution.

Hypothesis 1. *Compared to those who are not a WikiProject member, joining a WikiProject increases one's level of contribution.*

Before reporting the regression results, we first provide a graphical illustration on the editors' contribution behavior before and after joining a WikiProject. In Figure 4.3, we recenter the month in which the editors join a WikiProject to zero and plots the average number and size of edits, additions and deletions in six months before and after. The solid lines represent the average value in each month, with the error bars denoting one standard error.

We first note in the figure that prior to the treated editors joining a WikiProject, there is an increase in the number of edits. For example, the average contribution level starts from 132 edits and 44,675 characters in six month before joining a WikiProject, and reaches 240 edits with 106,694 characters in one month before. Both the average number and size of edits reach their peaks in the month when

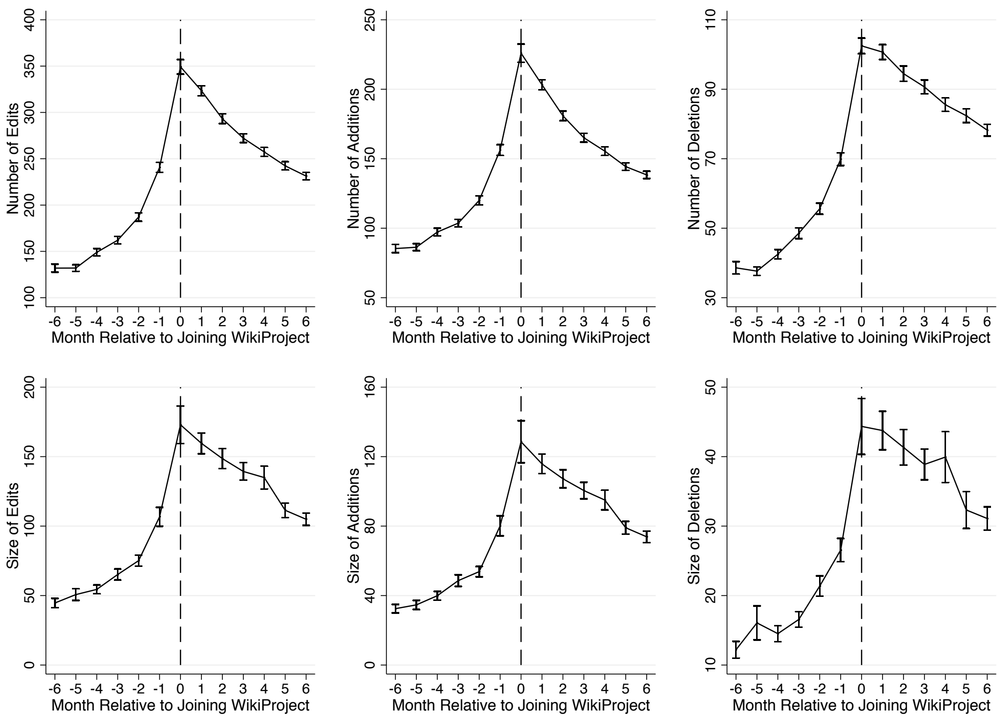


Figure 4.3: Contribution level before and after WikiProject membership

the editors join the WikiProject. While the contribution level decreases gradually afterwards, it remains higher than it is before the editor join the WikiProject for at least five months. The plots for additions and deletions show a similar pattern that editors become increasingly active and decline in their contribution level after joining a WikiProject. These observations indicate the possibility that there may be pre-existing trends in the behavior of the treated editors before they join a WikiProject. More importantly, given the gradual increase in the contribution level prior to WikiProject membership, the decisions to join a WikiProject might be driven by editors' unobserved characteristics. Therefore, a direct before-after comparison of the treated editors' behavior may suffer from a self-selection bias - the editors who join a WikiProject are intrinsically more active than those who do not.

The visual evidence presented in Figure 4.3 shows that the treated editors who join a WikiProject make remarkably more contributions than before. However, it does not rule out the possibility that some unobservable factors drives the difference. To identify and estimate the causal effect of WikiProject membership, we adopt an instrumental variable approach that exploits exogenous variations correlated with joining a WikiProject. We start with the following regression framework:

$$Y_{i,t} = \beta_0 + \beta_1 \times \text{Membership}_{i,t} + \varepsilon_{i,t},$$

where i indexes the editors and t indexes months. The dependent variable, $Y_{i,t}$, is the quantity measure of an editor's contribution in month t. $\text{Membership}_{i,t}$ is a dummy variable representing the membership status of editor i.

Our instrumental variable approach is inspired by how the editors are exposed to WikiProjects. For each Wikipedia article, the WikiProjects to which it belong are listed on its talk page (see Figure 4.4 for the example of the article "Public Good"). Depending on how much the editors visit articles' talk pages, the lists of the associated WikiProjects give the editors various levels of exposure to the WikiProjects. The lists of associated Wikirojects are maintained in a decentralized manner by individual editors. To categorize an article under a WikiProject, any registered editor can add the Wiki link of that WikiProject to the talk page of the article. To remove an article from a WikiProject, any registered editor can delete

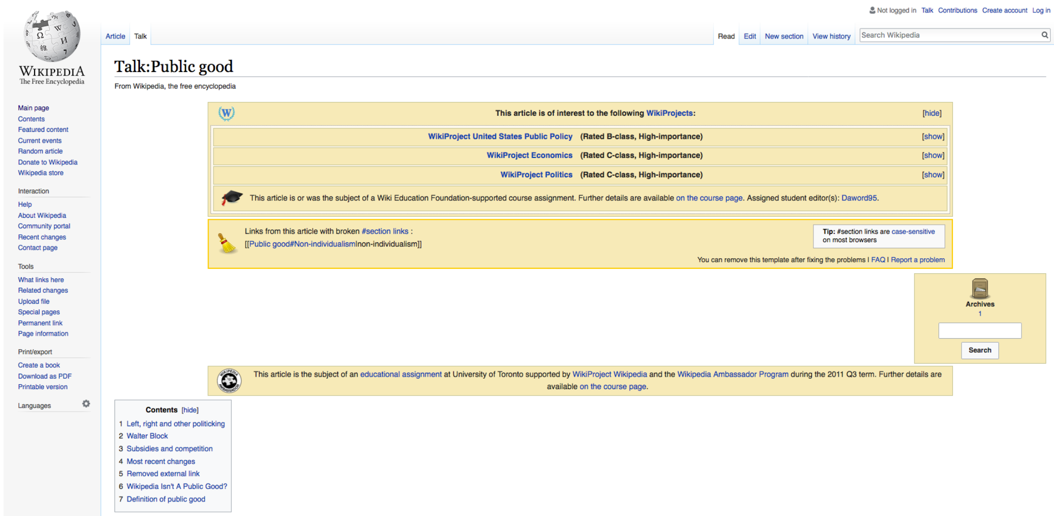


Figure 4.4: Screenshot for the Talk Page of Article “Public Good”

the Wiki link of the WikiProject from the talk page of the article.

The decentralized maintenance of the list of associated WikiProjects introduces a natural source of variation in the editors’ exposure to WikiProjects, which will arguably affects whether one joins a WikiProject. Consider a simple example where an article “public goods” is categorized to WikiProject Economics on time t. Editors who visit it talk page before time t will not see the Wiki link of the project and therefore be less exposed to the project, while those who visit after time t will.

To construct our instrumental variable, we define the intensity of exposure to WikiProject as below. For each editor i in month t, we focus on the set of articles whose talk pages are revised by her in that month. Among these articles, let $T^*_{i,t}$ denote the ones that are assigned a WikiProject by month t and $T_{i,t}$ denote the ones that are eventually associated with a WikiProject by the October of 2016. For example, consider an editor i who revises the talk page of four articles - “public goods”, “public economics”, “microeconomics” and “Wikipedia” in month t. Suppose that “public goods”, “public economics” and “microeconomics” would all belong to WikiProject Economics, but the first two are associated with the project in month $t_1 < t$ and the third one is associated in month $t_2 > t$. The article “Wikipedia” does not belong to any WikiProject. For editor i, $T_{i,t}$ contains the three articles that eventually belong to WikiProject Economics "public goods", "public economics", "microeconomics" but $T^*_{i,t}$ contains only the first two articles that have been associated by month t. The measure for WikiProject exposure is calculated by taking the ratio of the number of articles contained in these two sets, $|T^*_{i,t}|/|T_{i,t}|$.

Our choice of instrument is valid if it affects contribution behavior only through its impact on one's decision to join a WikiProject. When an editor revises an article's talk page, whether an associated WikiProject is listed is determined by the editors who edited it previously. Hence, whether that article belongs to a WikiProject at the time is exogenous to that editor. One may be concerned that the instrument merely captures how active an editor is. However, we normalize the measure of exposure to WikiProject by $|T_{i,t}|$ so that the variation in the instrument is driven by the level of exposure to WikiProject relative to one's activeness. Another potential source of endogeneity is that editors who are extremely interested in specific topics are more likely to edit talk pages of articles that belong to relevant WikiProjects. In that case, the instrument would reflect the editors' interests. However, we only focus on articles that would eventually belong to a WikiProjects rather than all articles.

We present the estimation for the treatment effect in Table 4.2. We use number and size (count of characters) of revisions and additions as measures for one's contribution. For each measure, we report both the estimation results using simple OLS regressions and the two-stage least square IV regressions. Editor-level and month-level fixed effects are also controlled. We also include the first-stage results for the IV regressions. There exists a strong and positive correlation between the constructed measure of exposure to WikiProject and whether one joins a WikiProject. The associated F statistics for the first stage is 137.85, indicating that our estimation results do not suffer from the bias due to weak instrument.

The regression results in Table 4.2 indicate that joining a WikiProject yields economically and statistically significant impact on individual contributions. Column 2 shows that WikiProject membership leads to an average increase of 79 revisions per month (p-values < 0.01). Compared with the average level of 64 revisions per month before the treatment, joining a WikiProject more than doubles an editor's contribution. Similar evidence is obtained for the size of revisions measured by the number of characters: joining a WikiProject yields 8,672 more characters per month, which is 1.5 times more than the level before.

Columns 6 and 8 present the estimate for the impact of WikiProject membership on the additions. After joining a WikiProject, the treatment leads to 39 additions and 6,199 characters per month (p-values < 0.01). Combining with the results obtained in columns 2 and 4, we find an interesting pattern that additions account for less than 50% of the average effect when measured by the number of revisions and more than 60% when measured by the size. One explanation for the discrepancy is that addition differs from other types of revisions (such as deletion and uploading files) in that addition is usually involved with more effort.

Result 1. *Compared to those who are not a WikiProject member, joining a WikiProject leads to a significantly increase in one's contribution level by*

1. *79 more number of revisions or 8,672 more characters per month,*

2. *39 more number of additions or 6,199 more characters per month.*

We also provide in Table 4.2 the estimation results obtained from the ordinary least square regressions. The OLS estimates for the impact from joining a WikiProject are consistently higher than the IV estimates. For example, this indicates that the results from simple OLS regression might overestimate the impact of WikiProject membership.

Table 4.2: Impact of WikiProject Membership on Contribution

	Number of Revisions		Size of Revisions		Number of Additions		Size of Additions	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WikiProject	81.364***	79.117***	10.147***	8.722***	45.059***	39.089***	6.937***	6.244***
	(0.734)	(1.340)	(0.217)	(0.396)	(0.496)	(0.905)	(0.159)	(0.289)
First-stage coefficient		0.573***		0.573***		0.573***		0.573***
		(0.001)		(0.001)		(0.001)		(0.001)
F statistics		137.92		137.92		137.92		137.85
Specification	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Observations	821,944	821,944	821,944	821,944	821,944	821,944	821,944	821,944

Notes. The dependent variable is a user's contribution in monthly level. Size of revisions and size of additions (columns 3, 4, 7 and 8) are denoted in 1,000 bytes. Standard errors are provided in parentheses. *, ** and *** denote significance level at 10%, 5% and 1% level.

4.5.2 Mechanism for WikiProject Membership

One potential mechanism underlying the impact of WikiProject is the reduction in search cost. WikiProjects provide editors with relevant articles to improve through various ways, such as open tasks and collaborations of the week or month (Zhu et al., 2012). These recommendations lowers the cost of identifying the articles one is interested in and make editors' contribution more centralized (Kittur and Kraut, 2008; Kittur et al., 2009; Morgan et al., 2013; Ransbotham and Kane, 2011).

To examine whether our result can be explained by this mechanism, the most straightforward way is to explore how the change in editors' behavior relates to the number of recommendations provided by the WikiProjects. However, the WikiProjects vary in the manner in which the recommendations are presented and it is difficult to measure the number of recommendations uniformly. We overcome this problem by using the character size of WikiProject home page as a proxy for the number of recommendations provided. The rationale behind this measure is that various devices that provide recommendation are usually presented at the home page of a WikiProject. For example, in the template for WikiProject home page, there are pre-defined sections that aim to provide recommendations such as open tasks, new articles to be created and assessment. Thus, we can use the size of WikiProject home page as an alternative measure for the variation in the number of recommendations. If coordination through article recommendation drives the increase in editors' contribution, we would expect that WikiProjects of larger sizes (and hence more recommendations) encourages more contribution. This gives the following hypothesis regarding the size of WikiProject home page and increase in contribution behavior.

Hypothesis 2. *The users who join a WikiProject with a home page of larger size increase their level of contribution more than those who join a WikiProject with a home page of smaller size do.*

Figure 4.5 provides a visual presentation on the association between size of WikiProject and change in editors' contribution, measured by the number and size of revisions and additions. Each point in the figures corresponds an editor who is a member of a WikiProject and the dotted line is a simple OLS fit. The vertical axis represents the increase in the contribution level in the month in which she joins relative to the month immediately before. The horizontal axis represents the size of WikiProject. To mitigate the impact from outliers, we apply log transformations to both measures.

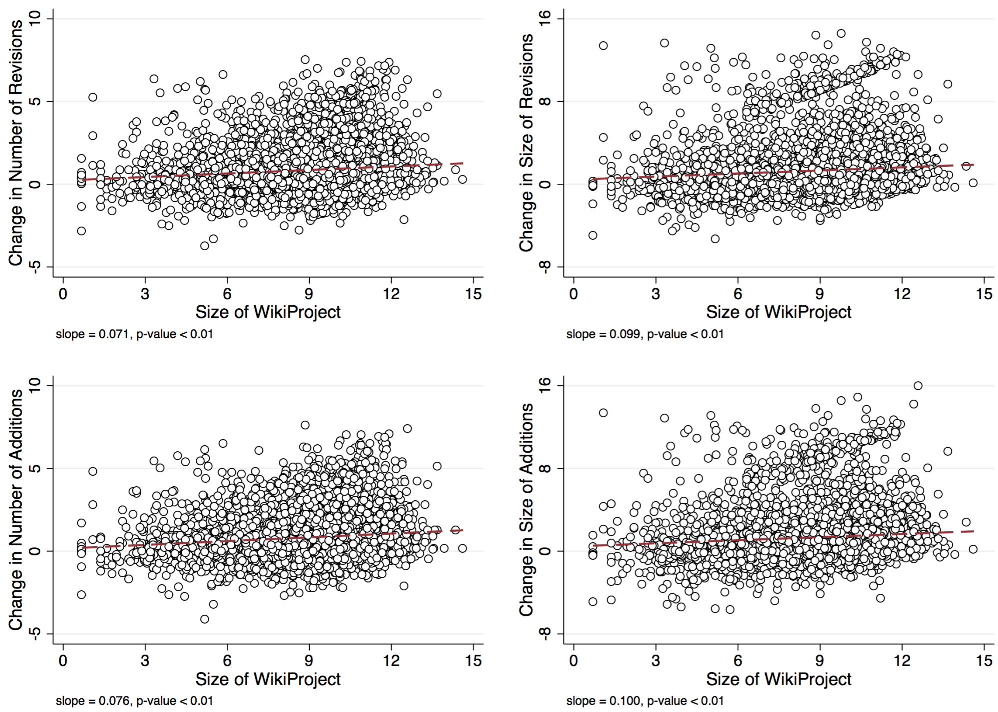


Figure 4.5: Change in Editors' Contribution and Size of WikiProjects

The slopes of the fitted line are significantly larger than 0 for all the four measures. That is, there is a positive relationship between the size of a WikiProject and the change in the level of contributions by editors' who becomes a member. Given this, we can further investigate the quantitative relationship by estimating the following equation:

$$\log(y_i) = \alpha_0 + \alpha_1 \times \log(\text{Project size}_i) + \alpha_2 \times \log(Ly_i) + \varepsilon_i,$$

where y_i is the measure of an editor i's contribution in the month in which she becomes a WikiProject member and Project Size$_i$ is the character size of the WikiProject in which an editor joins. $\log Ly_i$ is the one-month period lagged term and controls for an editor's contribution level prior to WikiProject membership.

If the hypothesized mechanism explains the treatment, We expect that the editors who join a project of larger size makes more contributions to articles related to it. Table 4.3 reports the results from the above regression. In the alternative specifications, we also include the square term of $\log(\text{Project size})$ to control for the potential U-shape pattern of the project size. It turns out that the goodness-of-fit is not improved as the adjusted R^2 is lower. Using the likelihood ratio test, we cannot reject the null hypothesis that including the squared term of $\log(\text{Project size})$ gives a better fit for the data. Furthermore, the regression results show that the dependent variables are increasing in $\log(\text{Project size})$ within the range of values in our dataset. We relegate the results from this alternative specification to Table D.7 in Appendix Appendix D.

We see that the variations in the size of WikiProject have a significantly positive impacts on the level of contribution. The estimated coefficients for $\log(\text{Project size})$ are significantly positive for both the number and size of revisions (column 1 and 2) and additions (column 3 and 4). Based on a simple back-of-the-envelope calculation, a one standard deviation increase in $\log(\text{Project size})$ yields more than a 8% increase in the number of revisions and nearly 14% increase in the size of revisions. Repeating the analysis for contribution level measured by additions gives estimates that are similar in both magnitude and significance level. The results suggest that a WikiProject with more information on its homepage tend to induce higher increases in the contributions from its membership, indicating that article recommendation could be the mechanism underlying the impact of WikiProjects.

Result 2. *The increase in a user's level of contribution to the articles under the joined project is significantly larger if the home page of the project is larger. Specifically, a one standard deviation increase in* $\log(\textit{Project size})$ *leads to*

1. *8.157% increase in the number of revisions or 13.821% increase in number of characters,*

2. *7.437% increase in the number of additions or 15.299% increase in number of characters added.*

Table 4.3: Project Size and Contribution to the Articles under the Joined Project

Dep. var.:	Revision		Addition	
Measure:	Number	Size	Number	Size
	(1)	(2)	(3)	(4)
log (Project size)	0.043***	0.070***	0.040***	0.074***
	(0.011)	(0.021)	(0.011)	(0.024)
Lagged dep. var.	0.554***	0.395***	0.571***	0.416***
	(0.012)	(0.011)	(0.012)	(0.017)
Δ% of dep. var. in response to				
unit s.d. change in log (Project size)	8.157***	13.821***	7.437***	15.299***
	(2.292)	(4.865)	(2.290)	(5.279)
p-values for LR Test	0.830	0.913	0.735	0.848
R^2	0.393	0.281	0.382	0.282
Observations	3,500	3,500	3,500	3,500

Notes. The dependent variable is a user's contribution after joining a WikiProject. Standard errors are provided in parentheses. *** denotes significance level at 1% level.

Previous works exploring the effect of social identity on pro-social behavior provide evidence that when individuals tend to exhibit less other-regarding behavior to groups that they identify less with (Bernhard et al., 2006; Chen and Li, 2009; Goette et al., 2006). Given the large influence of the number of recommendations provided by a project on an editor's behavior, it is interesting to explore whether there is any spillover to articles that do not belong to the project. To the extent that WikiProjects constitute a social group, we expect that the change in members' contribution behavior to articles under other WikiProjects are not affected through this channel. This gives the following hypothesis.

Hypothesis 3. *The size of the home page of the WikiProject which a user join does not affect her contribution to articles that do not belong to the joined WikiProject.*

Using the same specifications in Table 4.3, Table 4.4 report the link between project size and change in contributions to articles outside the project that an editor join. Unlike the impacts on articles under the joined project, the estimates for those outside are closer to zero and statistically insignificant. This is consistent with our proposed mechanism of recommendation: the articles that are not associated with the project that one joins are less likely to be listed on the home page, and therefore the size of the project has smaller impact on editors' contribution on them.

Result 3. *The increase in a user's level of contribution to the articles that do not belong to the joined project is not significantly affected by the size of the home page for the project.*

To further test our proposed mechanism, we extend the analysis by estimating the following equation:

$$\log (y_{i,k}) = \alpha_0 + \alpha_1 \times \text{Jaccard similarity}_{i,k} + \alpha_2 \times \log (Ly_{i,k}) + \varepsilon_{i,k}.$$

Here, $y_{i,k}$ is editor i's contribution to a project k after joining a WikiProject. Jaccard similarity$_{i,k}$ is a continuous measure for how similar the project k is to the project that editor i joins. Jaccard similarity is an widely used in the area of information retrieval to quantify how close or similar two sets are (Leskovec et al., 2014). Formally, Jaccard similarity between two finite sets is calculated by the size of their intersection divided by the size of their union. For example, a Jaccard similarity of 1 implies two identical sets, where as a Jaccard similarity of 0 implies two disjoint sets. In our context, we follow Platt and Romero (2018) and define the Jaccard similarity between two WikiProjects as the number of articles associated with both of the two projects and the number of articles associated with either one. The results in Table 4.5 further supports the proposed mechanism. The parameter estimates for Jaccard similarity are significantly positive in columns 1 through 4. The estimates show that a one percentile increase in the Jaccard similarity is associated with a 0.64% increase in the number of revisions and a 2.06% increase in the size. Taken together, the above results provide evidence indicating that the recommendations of articles can explain the influence of WikiProject membership.

It is noteworthy that although the results presented in Table 4.3 give support to the effect of article recommendation, it does not rule out the possibility that WikiProjects encourage more user contribution through other channels. For example, users can be more intrinsically motivated by the common identity that they share with the WikiProjects that they join (Solomon and Wash, 2014). In that sense, WikiProjects with a stronger group identity (e.g., with a project user-box)

Table 4.4: Project Size and Contributions to Articles outside the Joined Project

Dep. var.:	Revision		Additions	
Measure:	Number	Size	Number	Size
	(1)	(2)	(3)	(4)
$\log$ (Project size)	0.005	-0.014	0.006	-0.021
	(0.007)	(0.013)	(0.008)	(0.014)
Lagged dep. var.	0.537***	0.391***	0.539***	0.393***
	(0.011)	(0.013)	(0.014)	(0.015)
$\Delta\%$ of dep. var. in response to unit s.d. change in $\log$ (Project size)	1.126	-2.706	1.223	-3.886
	(1.634)	(2.476)	(1.622)	(2.505)
R^2	0.431	0.306	0.414	0.285
Observations	3,500	3,500	3,500	3,500

Notes. The dependent variable is a user's contribution after joining a WikiProject. Standard errors are provided in parentheses. *** denote significance level at 1% level.

Table 4.5: Jaccard Similarity and Contribution Level

Dep. var.:	Revision		Additions	
Measure:	Number	Size	Number	Size
	(1)	(2)	(3)	(4)
Jaccard similarity	0.638***	2.038***	0.490***	1.865***
	(0.011)	(0.031)	(0.010)	(0.032)
Lagged dep. var.	0.312***	-0.015	0.417***	0.121***
	(0.008)	(0.050)	(0.007)	(0.008)
$\Delta\%$ of dep. var. in response to percentile change in Jaccard Similarity	0.640***	2.059***	0.491***	1.882***
	(0.011)	(0.031)	(0.010)	(0.033)
R^2	0.130	0.030	0.172	0.040
Observations	636,340	636,340	636,340	636,340

Notes. The dependent variable is a user's contribution after joining a WikiProject. Standard errors are provided in parentheses. *** denote significance level at 1% level.

might drive their members to contribute more than those without one. We expect to more future work will be dedicated to fully disentangle the mechanisms of article recommendation versus project identity.

4.6 Conclusion

In this paper, we present field evidence on the causal effects of group membership on encouraging individual contribution to public information goods. We analyze how joining WikiProjects - a collaboration platform designed to motivate and coordinate user effort - can lead to more edits to Wikipedia articles from individual users. Using an instrumental variable approach that exploits variations in the intensity of exposure to WikiProjects, we find that joining a WikiProject results in a substantial increase in user contribution, namely 79 more revisions and 34,686 more characters per month. Compared with the user behavior prior to WikiProject membership, the magnitude of the estimated treatment effect amounts to more than 1.5 times than the contribution level before.

To understand the underlying mechanism behind the influence of WikiProject, we examine how the characteristics of WikiProjects relate to the increase in user contribution. Using the size of project homepage as a proxy measure for the number of recommendation devices, our analysis shows that a project with more recommendations yields significantly larger change in the level of user contribution. Specifically, the Wikipedia users who join a large This indicates that coordination through article recommendation could be the potential mechanism that fosters user engagement.

The findings in this paper contribute to the research on motivating pro-social behavior with identity-based approach in various field settings. While previous studies have confirmed the potential of group identity to induce pro-social behavior through coordination and competition, the evidence in this paper suggests the group membership can encourage contribution to public information goods through the provision of recommendations to users.

Lastly, our analysis provide design implications on Wikipedia and many other online platforms that rely on voluntary contribution from individual users. While much effort has been dedicated to encouraging user engagement enhancing members' identity, our results suggest that recommending tasks efficiently through group coordination can be important considerations.

Appendix A Proofs

In this appendix, we will present the proofs for the two propositions in section 3.3. We use backward induction to solve the second stage optimization problem first.

Proof of Proposition 2: After seeing the realization of the matching accuracy, m_i, expert i solves the following optimization problem:

$$\max_{y_i \in [0,T]} v(n)\Big(y_0 + y_i + \sum y_{-i}\Big) + w(n, s_i)y_i + r_i(T - y_i) - \frac{c(y_i)}{m_i}. \tag{1}$$

For agents whose marginal contribution cost at zero is flat relative to the net marginal benefit and opportunity cost adjusted by match quality, i.e., $c'(0) < (v(n) + w(n, s_i) - r_i) \cdot m_i$, the unique interior solution is characterized by the following first-order condition:

$$c'(y) = \big(v(n) + w(n, s_i) - r_i\big) \cdot m_i.$$

Let $k(\cdot) \equiv (c')^{-1}(\cdot)$, which is the inverse of the marginal cost function. The interior solution is characterized by

$$y_i = k\big((v(n) + w(n, s_i) - r_i) \cdot m_i\big). \tag{2}$$

For agents whose marginal contribution cost at zero is sufficiently steep relative to the net marginals adjusted by match quality, i.e., $c'(0) \geq (v(n) + w(n, s_i) - r_i) \cdot m_i$, the optimal contribution level is a corner solution, $y_i = 0$. Taken together, the optimal contribution level in the second stage is

$$y^* = \max\Big\{0, k\big((v(n) + w(n, s_i) - r_i) \cdot m_i\big)\Big\} \tag{3}$$

For the interior solution, it is straightforward to obtain the following comparative statics, which serve as the benchmark for our experimental design and data analysis.

(a) An increase in the number of consumers of the public information good leads to an increased level of contribution, i.e.,

$$\frac{\partial y^*}{\partial n} = k' \cdot \big(v'(n) + w_n(n, s_i)\big) \cdot m_i > 0.$$

(b) An increase in the private benefit of contributions leads to an increased level of contributions, i.e.,

$$\frac{\partial y^*}{\partial w} = k' \cdot m_i > 0.$$

(c) An expert with a higher reputation will contribute less, i.e.,

$$\frac{\partial y^*}{\partial r} = -k' \cdot m_i < 0.$$

(d) Better matching between the content of the public information good and the agent's expertise leads to an increased level of contributions, i.e.,

$$\frac{\partial y^*}{\partial m_i} = k' \cdot \big(v(n) + w(n, s_i) - r_i\big).$$

If the interior solution exists, we have $v(n) + w(n, s_i) - r_i \geq 0$. Hence,

$$\frac{\partial y^*}{\partial m_i} \geq 0.$$

Q.E.D.

Proof of Proposition 1: We now offer the proof for proposition 1. In the first stage, an expert does not see the realization of the match accuracy, m_i, but knows its distribution $f(m)$. Therefore, she forms her expectations for the matching accuracy m_i. Given a matching accuracy m_i, the optimal contribution level in the second stage is $y^*(m_i)$. If the expert chooses to participate, her expected utility is given by

$$\int_0^1 \left(v(n)\Big(y^* + \sum y_{-i}\Big) + w(n, s_i)y^* + r_i(T - y^*) - \frac{c(y^*)}{m}\right) f(m)dm.$$

If the expert chooses not to participate, her expected utility is

$$v(n) \cdot \sum y_{-i} + r_i T.$$

Therefore, one chooses to participate if

$$G(n, w, s_i, r_i) = \int_0^1 \left(v(n)y^* + w(n, s_i)y^* - r_i y^* - \frac{c(y^*)}{m} \right) f(m)dm \geq 0. \quad (4)$$

It now suffices to show that $G(n, w, s_i, r_i)$ is increasing in n and w, and decreasing s_i and r_i.

Let $\underline{m}(n, w, s_i, r_i)$ be implicitly defined by $y^*(n, w, s_i, r_i, \underline{m}) = 0$. That is, $\underline{m}$ is the minimum matching accuracy required to make an expert contribute. By definition, we have

$$y^* = \begin{cases} k\big((v(n) + w(n, s_i) - r_i) \cdot m\big), & \text{if } m > \underline{m} \\ 0, & \text{if } m \leq \underline{m} \end{cases}$$

This implies that

$$G(n, w, s_i, r_i) = \int_{\underline{m}(n)}^1 \left(v(n)y^* + w(n, s_i)y^* - r_i y^* - \frac{c(y^*)}{m} \right) f(m)dm.$$

Hence,

$$\begin{aligned} \frac{\partial G}{\partial n} &= \frac{\partial}{\partial n} \int_{\underline{m}}^1 \left(v(n)y^* + w(n, s_i)y^* - r_i y^* - \frac{c(y^*)}{m} \right) f(m)dm \\ &= \underline{m}'(n) \cdot \left(v(n)y^*(n, \underline{m}) + w(n, s_i)y^*(n, \underline{m}) - r_i y^*(n, \underline{m}) - \frac{c(y^*(n, \underline{m})}{\underline{m}} \right) f(\underline{m}) \\ &\quad + \int_{\underline{m}}^1 \frac{\partial}{\partial n} \left(v(n)y^*(n, m) + w(n, s_i)y^*(n, m) - r_i y^*(n, m) - \frac{c(y^*(n, m))}{m} \right) f(m)dm \\ &= \int_{\underline{m}}^1 \frac{\partial}{\partial n} \left(v(n)y^*(n, m) + w(n, s_i)y^*(n, m) - r_i y^*(n, m) - \frac{c(y^*(n, m))}{m} \right) f(m)dm \end{aligned}$$

When $m > \underline{m}$, an interior solution exists for optimization problem 1. Applying the envelope theorem yields

$$\frac{\partial}{\partial n}\left(v(n)y^* + w(n,s_i)y^* - r_i y^* - \frac{c(y^*)}{m}\right) = (v_n + w_n)y^* \geq 0,$$

which indicates that $\partial G/\partial n \geq 0$, i.e., a potential contributor is more likely to participate if the number of consumers increases.

Similarly, we have

$$\frac{\partial G}{\partial w} = \int_{\underline{m}}^{1} y^*(n,w,s_i,r_i)f(m)dm \geq 0.$$
$$\frac{\partial G}{\partial s_i} = \int_{\underline{m}}^{1} w_s y^*(n,w,s_i,r_i)f(m)dm \leq 0.$$
$$\frac{\partial G}{\partial r_i} = -\int_{\underline{m}}^{1} y^*(n,w,s_i,r_i)f(m)dm \leq 0.$$

Therefore, a potential contributor is more likely to participate if the private benefit increases, or if her expertise overlaps with the askers, or her opportunity cost of time is lower. Q.E.D.

Appendix B Recommendation algorithm

We first describe the method we use to identify experts' domain of expertise. We develop a filtering algorithm which is based on the experts' recent research papers archived at *New Economics Papers* (*NEP*). *NEP* is an announcement service aiming to provide up-to-date information about research literature. It disseminates and archives new research papers in 97 research areas.[1] For each expert, we refer to *NEP* her recent research works as well as the research fields where each work is classified. Then, we select the research field in which her research works are classified most and use that one as the most recent domain of expertise. The pseudocode for the filtering algorithm used to identify the most recent domain of expertise is presented below.

```
foreach Expert
    ResearchList ← Expert's research works at NEP
    foreach Work in ResearchList
        ResearchFields ← research fields of Work at NEP
        foreach Field in ResearchFields
            NumFields[Field] ← NumFields[Field] + 1
DomainExpertise ← max(NumFields)
```

In what follows, we present details about our selection criteria for Wikipedia articles. For each of an expert's research papers listed at *NEP*, the recommendation algorithm submits a search query containing the keywords in the paper through Google Custom Engine API. The search result returned from Google contains Wikipedia articles that are potentially relevant enough for recommendation. After we iterate over all research papers by an experts, we reach a list of Wikipedia articles indicated as relevant to the experts' recent research focus. We further restrict this list using the following criteria: 1) The article must be under the namespace 0 (i.e., main articles);[2] 2) The article is not edit protected;[3] 3) The character

[1] See `http://nep.repec.org/`.

[2] Wikipedia uses namespace to categorize webpages according to their functions. All encyclopedia articles at Wikipedia are under namespace 0. Webpages under other namespaces include talk pages and user pages. See `https://en.wikipedia.org/wiki/Wikipedia:Namespace` for a detailed explanation of namespace at Wikipedia.

[3] Edit protection restriction a Wikipedia article from being edited by users. It is usually applied to

length of article is not less than 1,500 characters; 4) The article is viewed at least 1,000 times in the past 30 days (dynamically updated) prior to exposure to the intervention.[4] Finally, we choose the six Wikipedia articles that appear most frequently in the search results by Google Custom Engine for our recommendation.

---

foreach `Expert`
    foreach `Work` **in** `ResearchList`
        `WPArticles` ← Results from Google searching keyword in `Work`
        foreach `Article` **in** `WPArticles`
            `CountWP[Article]` ← `CountWP[Article]` + 1
foreach `Article` in `CountWP`
    if `Article` is not namespace 0 **or**
        `Article` is edit protected **or**
        `Article` has fewer than 1,500 characters **or**
        `Article` is viewed for fewer than 1,000 times over the past 30 days
            remove `Article` **from** `CountWP`
`DomainExpertise` ← **max**(`CountFields`)

---

articles that is subject to content disputes or the risk of vandalism. The decision to apply or remove edit protection is made by administrators at Wikipedia. See `https://en.wikipedia.org/wiki/Wikipedia:Protection_policy` for a detailed explanation of the edit protection policy at Wikipedia.

[4]This restriction guarantees that articles recommended in the AvgView condition are balanced regarding the number of views versus those recommended in the HighView condition. The experimental manipulation is merely the mentioning of number of views, but not any inherent difference in viewship.

Appendix C Screen shots

In this section, we provide screen shots of interface design for our field experiments. Figure C.1 presents our webpage where experts enter their comments. The interface is designed to minimize entry cost. An expert does not need to know how to edit a wiki. In the split screen design, the right side is the corresponding Wikipedia article that the expert can scroll up or down. The left side has a quality rating and a text box for the expert to enter comments. As long as she can use Word, she will be able to enter comments.

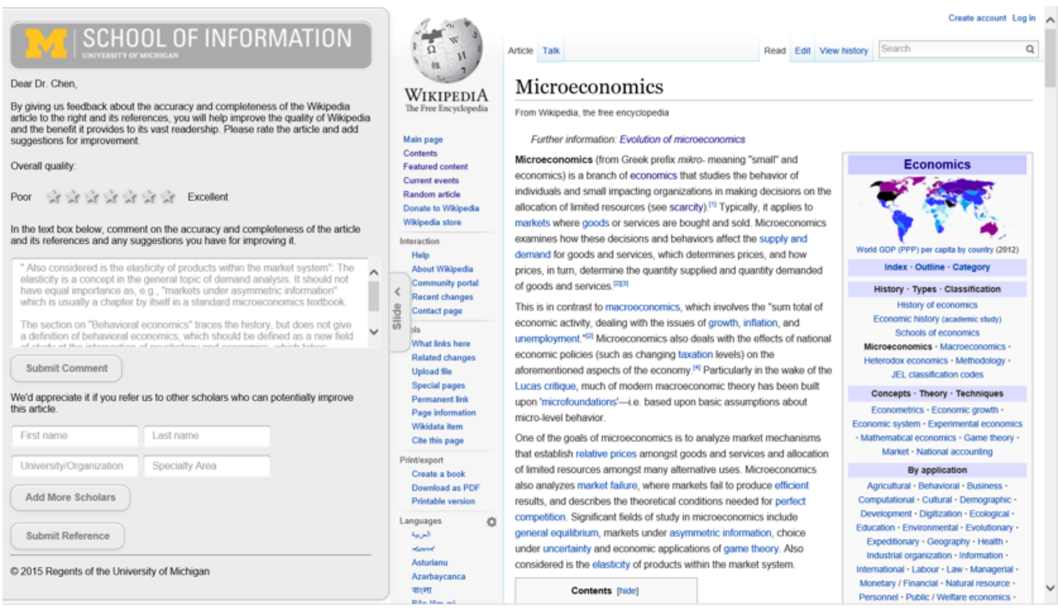
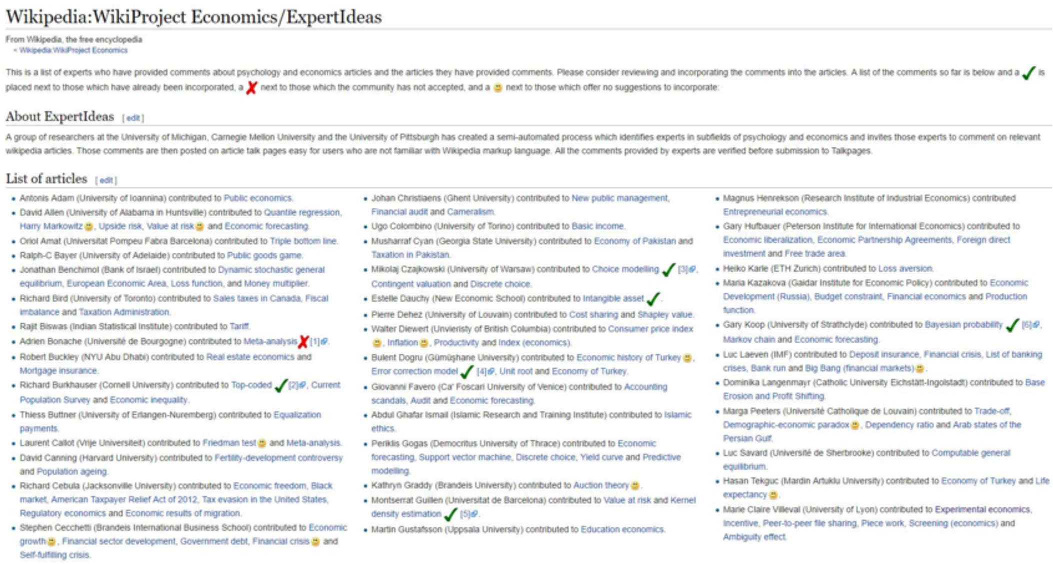


Figure C.1: Webpage of Comment

Figure C.2 presents our public acknowledgement of expert contributions to Wikipedia articles. This page was assembled by a Wikipedia editor who was a doctoral student in Economics at the University of Lancaster. The economists on this list contributed to our project during its pilot phase.



Wikipedia:WikiProject Economics/ExpertIdeas

From Wikipedia, the free encyclopedia
< Wikipedia:WikiProject Economics

This is a list of experts who have provided comments about psychology and economics articles and the articles they have provided comments. Please consider reviewing and incorporating the comments into the articles. A list of the comments so far is below and a ✓ is placed next to those which have already been incorporated, a ✗ next to those which the community has not accepted, and a 😐 next to those which offer no suggestions to incorporate:

About ExpertIdeas [edit]

A group of researchers at the University of Michigan, Carnegie Mellon University and the University of Pittsburgh has created a semi-automated process which identifies experts in subfields of psychology and economics and invites those experts to comment on relevant wikipedia articles. Those comments are then posted on article talk pages easy for users who are not familiar with Wikipedia markup language. All the comments provided by experts are verified before submission to Talkpages.

List of articles [edit]

- Antonis Adam (University of Ioannina) contributed to Public economics.
- David Allen (University of Alabama in Huntsville) contributed to Quantile regression, Harry Markowitz 😐, Upside risk, Value at risk 😐 and Economic forecasting.
- Oriol Amat (Universitat Pompeu Fabra Barcelona) contributed to Triple bottom line.
- Ralph-C Bayer (University of Adelaide) contributed to Public goods game.
- Jonathan Benchimol (Bank of Israel) contributed to Dynamic stochastic general equilibrium, European Economic Area, Loss function, and Money multiplier.
- Richard Bird (University of Toronto) contributed to Sales taxes in Canada, Fiscal imbalance and Taxation Administration.
- Rajit Biswas (Indian Statistical Institute) contributed to Tariff.
- Adrien Bonache (Université de Bourgogne) contributed to Meta-analysis ✗ [1].
- Robert Buckley (NYU Abu Dhabi) contributed to Real estate economics and Mortgage insurance.
- Richard Burkhauser (Cornell University) contributed to Top-coded ✓ [2], Current Population Survey and Economic inequality.
- Thiess Buttner (University of Erlangen-Nuremberg) contributed to Equalization payments.
- Laurent Callot (Vrije Universiteit) contributed to Friedman test 😐 and Meta-analysis.
- David Canning (Harvard University) contributed to Fertility-development controversy and Population ageing.
- Richard Cebula (Jacksonville University) contributed to Economic freedom, Black market, American Taxpayer Relief Act of 2012, Tax evasion in the United States, Regulatory economics and Economic results of migration.
- Stephen Cecchetti (Brandeis International Business School) contributed to Economic growth 😐, Financial sector development, Government debt, Financial crisis 😐 and Self-fulfilling crisis.
- Johan Christiaens (Ghent University) contributed to New public management, Financial audit and Cameralism.
- Ugo Colombino (University of Torino) contributed to Basic income.
- Musharraf Cyan (Georgia State University) contributed to Economy of Pakistan and Taxation in Pakistan.
- Mikolaj Czajkowski (University of Warsaw) contributed to Choice modelling ✓ [3], Contingent valuation and Discrete choice.
- Estelle Dauchy (New Economic School) contributed to Intangible asset ✓.
- Pierre Dehez (University of Louvain) contributed to Cost sharing and Shapley value.
- Walter Diewert (Unvieristy of British Columbia) contributed to Consumer price index 😐, Inflation 😐, Productivity and Index (economics).
- Bulent Dogru (Gümüşhane University) contributed to Economic history of Turkey 😐, Error correction model ✓ [4], Unit root and Economy of Turkey.
- Giovanni Favero (Ca' Foscari University of Venice) contributed to Accounting scandals, Audit and Economic forecasting.
- Abdul Ghafar Ismail (Islamic Research and Training Institute) contributed to Islamic ethics.
- Periklis Gogas (Democritus University of Thrace) contributed to Economic forecasting, Support vector machine, Discrete choice, Yield curve and Predictive modelling.
- Kathryn Graddy (Brandeis University) contributed to Auction theory 😐.
- Montserrat Guillen (Universitat de Barcelona) contributed to Value at risk and Kernel density estimation ✓ [5].
- Martin Gustafsson (Uppsala University) contributed to Education economics.
- Magnus Henrekson (Research Institute of Industrial Economics) contributed Entrepreneurial economics.
- Gary Hufbauer (Peterson Institute for International Economics) contributed to Economic liberalization, Economic Partnership Agreements, Foreign direct investment and Free trade area.
- Heiko Karle (ETH Zurich) contributed to Loss aversion.
- Maria Kazakova (Gaidar Institute for Economic Policy) contributed to Economic Development (Russia), Budget constraint, Financial economics and Production function.
- Gary Koop (University of Strathclyde) contributed to Bayesian probability ✓ [6], Markov chain and Economic forecasting.
- Luc Laeven (IMF) contributed to Deposit insurance, Financial crisis, List of banking crises, Bank run and Big Bang (financial markets) 😐.
- Dominika Langenmayr (Catholic University Eichstätt-Ingolstadt) contributed to Base Erosion and Profit Shifting.
- Marga Peeters (Université Catholique de Louvain) contributed to Trade-off, Demographic-economic paradox 😐, Dependency ratio and Arab states of the Persian Gulf.
- Luc Savard (Université de Sherbrooke) contributed to Computable general equilibrium.
- Hasan Tekguc (Mardin Artuklu University) contributed to Economy of Turkey and Life expectancy 😐.
- Marie Claire Villeval (University of Lyon) contributed to Experimental economics, Incentive, Peer-to-peer file sharing, Piece work, Screening (economics) and Ambiguity effect.

Figure C.2: Public Acknowledgement Hosted on a WikiProject Economics Page

Dear Dr. Chen,

Would you be willing to spend 10 – 20 minutes providing feedback on a few Wikipedia articles related to behavioral and experimental economics? Wikipedia is among the most important information sources the general public uses to find out about a wide range of topics. A Wikipedia article is viewed on average 426 times each month. While many Wikipedia articles are useful, articles written by enthusiasts instead of experts can be inaccurate, incomplete, or out of date.

If you are willing to help, we will send you links to a few Wikipedia articles in your area of expertise. We will select only articles, with over 1,000 views in the past month, so that your feedback will benefit many Wikipedia readers.

These articles may include some of your publications in their references.

Please click one of the following links to continue:

Yes, please send me some Wikipedia articles to comment on.

No, I am not interested.

Thank you for your attention.

Sincerely,

Yan Chen, Daniel Kahneman Collegiate Professor of Information, University of Michigan

Robert Kraut, Herbert A. Simon Professor of Human-Computer Interaction, Carnegie Mellon University

Figure C.3: First-stage Email

Dear Dr. Chen,

Thank you for your willingness to provide feedback on the quality of Wikipedia articles. The following articles are suggested by our algorithm as related to law & economics.

Please comment on the articles most relevant to your research. Your feedback can significantly improve these articles' accuracy and completeness, and the comments and the references that you provide will be incorporated therein. These articles might refer to some of your research. We would appreciate receiving your comments by Jan 14, 2017. Thank you very much for your help.

Wikipedia Article Title	Number of views in the past month	Link to review the article
Shareholder value	6,298	Click here
Corporate governance	38,351	Click here
Managerial economics	17,771	Click here
Economic nationalism	8,931	Click here
University of Delaware	17,123	Click here
Corporatocracy	10,479	Click here

Sincerely,

Yan Chen, Daniel Kahneman Collegiate Professor of Information, University of Michigan

Robert Kraut, Herbert A. Simon Professor of Human-Computer Interaction, Carnegie Mellon University

Figure C.4: Second-stage Email

Appendix D Robustness checks

Table D.1: Average Marginal Effect on the First-stage Response.

Dependent Variable:	Positive $\mathrm{P}(R=1)$ (1)	Null $\mathrm{P}(R=0)$ (2)	Negative $\mathrm{P}(R=-1)$ (3)	Positive $\mathrm{P}(R=1)$ (4)	Null $\mathrm{P}(R=0)$ (5)	Negative $\mathrm{P}(R=-1)$ (6)
HighView	0.002	0.021	-0.022	0.004	0.018	-0.022
	(0.030)	(0.026)	(0.027)	(0.030)	(0.026)	(0.027)
Cite	0.042	0.022	-0.064**	0.037	0.030	-0.067**
	(0.030)	(0.026)	(0.027)	(0.030)	(0.026)	(0.026)
CiteAckn	0.030	0.020	-0.050*	0.020	0.024	-0.044*
	(0.029)	(0.026)	(0.027)	(0.030)	(0.026)	(0.027)
HighView × Cite	0.021	-0.023	0.002	0.023	-0.028	0.005
	(0.042)	(0.037)	(0.037)	(0.042)	(0.037)	(0.037)
HighView × CiteAckn	0.017	-0.003	-0.013	0.021	-0.005	-0.016
	(0.042)	(0.037)	(0.038)	(0.042)	(0.037)	(0.038)
Percentile of Abstract Views				0.029	-0.039***	0.030***
				(0.030)	(0.008)	(0.008)
English Affiliation				-0.020	-0.037**	0.057***
				(0.018)	(0.015)	(0.015)
Overlap				0.212***	-0.079***	-0.133***
				(0.034)	(0.028)	(0.025)
HighView + HighView × Cite	0.022	-0.002	-0.020	0.027	-0.010	-0.017
	(0.030)	(0.026)	(0.025)	(0.030)	(0.026)	(0.025)
Cite + HighView × Cite	0.063**	-0.001	-0.062**	0.060**	0.002	-0.062**
	(0.030)	(0.027)	(0.026)	(0.030)	(0.026)	(0.026)
HighView + HighView × CiteAckn	0.018	0.017	-0.036	0.025	0.013	-0.038
	(0.030)	(0.027)	(0.026)	(0.030)	(0.027)	(0.026)
CiteAckn + HighView × CiteAckn	0.047	0.016	-0.063**	0.041	0.019	-0.060**
	(0.030)	(0.027)	(0.027)	(0.030)	(0.027)	(0.027)
Model Specification	Multinomial Logistic			Multinomial Logistic		
Observations	3,346			3,301		

Notes. The dependent variable is the expert's response to the email in the first stage. Standard errors are provided in parentheses. Average marginal effects are calculated using the Delta method. *, ** and *** denote significance level at 10%, 5% and 1% level.

Table D.2: Determinants of Contribution Quantity

Dependent Variable:	log(1 + Word Count)			
	(1)	(2)	(3)	(4)
HighView	-0.034	0.066	-0.051	0.030
	(0.100)	(0.214)	(0.101)	(0.216)
Cite	-0.070	-0.086	-0.086	-0.119
	(0.096)	(0.210)	(0.097)	(0.213)
CiteAckn	-0.069	-0.047	-0.085	-0.086
	(0.096)	(0.209)	(0.098)	(0.213)
HighView $\times$ Cite	-0.072	-0.202	-0.058	-0.176
	(0.137)	(0.299)	(0.138)	(0.302)
HighView $\times$ CiteAckn	0.131	0.147	0.147	0.175
	(0.138)	(0.295)	(0.139)	(0.299)
Cosine Similarity			1.768***	2.861***
			(0.166)	(0.360)
Percentile of Article Length			-0.116*	-0.166
			(0.080)	(0.186)
Percentile of Abstract View			0.154*	0.213
			(0.099)	(0.217)
English Affiliation			0.097**	0.155
			(0.057)	0.123
Overlap			0.373***	0.741***
			(0.099)	(0.194)
HighView + HighView $\times$ Cite	-0.105	-0.137	-0.108	-0.145
	(0.093)	(0.208)	(0.094)	(0.211)
Cite + HighView $\times$ Cite	-0.142	-0.289	-0.144	-0.295*
	(0.097)	(0.212)	(0.098)	(0.215)
HighView + HighView $\times$ CiteAckn	0.098	0.213	0.097	0.205
	(0.095)	(0.203)	(0.096)	(0.207)
CiteAckn + HighView $\times$ CiteAckn	0.062	0.100	0.063	0.089
	(0.098)	(0.207)	(0.099)	(0.209)
Model Specification	OLS	Exp. Disp.	OLS	Exp. Disp.
Observations	8,819	8,819	8,635	8,635

Notes. The dependent variable is the log transformation of word count. Columns (1) and (3) report the results from the OLS model, and columns (2) and (4) reports the results from the exponential dispersion model. Quality class and importance class are controlled in all specifications. Fixed effects are controlled at the expert level. Standard errors are reported in the parentheses. *, ** and *** denote significance level at 10%, 5% and 1% level.

Table D.3: Determinants of Contribution Quality

Dependent Variable:	Overall Quality		Helpfulness		# of Sub-comments	
	(1)	(2)	(3)	(4)	(5)	(6)
HighView	0.870	0.901	0.846	0.870	0.885*	0.898*
	(0.161)	(0.169)	(0.157)	(0.164)	(0.056)	(0.058)
Cite	0.877	0.869	0.815	0.806	0.900*	0.893*
	(0.157)	(0.158)	(0.147)	(0.147)	(0.056)	(0.056)
CiteAckn	1.498**	1.575**	1.346	1.447**	1.094	1.125*
	(0.273)	(0.295)	(0.246)	(0.271)	(0.066)	(0.069)
HighView $\times$ Cite	1.403	1.428	1.642*	1.703**	1.122	1.141
	(0.375)	(0.386)	(0.439)	(0.461)	(0.105)	(0.107)
HighView $\times$ CiteAckn	1.058	1.008	1.239	1.139	1.045	1.003
	(0.275)	(0.268)	(0.322)	(0.302)	(0.092)	(0.089)
Cosine Similarity		12.221***		15.085***		3.422***
		(7.302)		(9.056)		(0.635)
Percentile of Article Length		1.078		1.168		1.232**
		(0.326)		(0.354)		(0.125)
Percentile of Abstract View		0.930		1.105		1.049
		(0.184)		(0.220)		(0.070)
English Affiliation		1.018		1.130		0.998
		(0.112)		(0.125)		(0.037)
Overlap		1.387*		1.451**		1.112*
		(0.238)		(0.246)		(0.063)
HighView + HighView $\times$ Cite	1.220	1.286	1.388*	1.481**	0.993	1.025
	(0.235)	(0.253)	(0.267)	(0.291)	(0.068)	(0.071)
Cite + HighView $\times$ Cite	1.230	1.241	1.337	1.372	1.011	1.019
	(0.243)	(0.248)	(0.264)	(0.275)	(0.070)	(0.071)
HighView + HighView $\times$ CiteAckn	0.920	0.908	1.048	0.991	0.924	0.901*
	(0.168)	(0.172)	(0.191)	(0.186)	(0.056)	(0.055)
CiteAckn + HighView $\times$ CiteAckn	1.584**	1.588**	1.668***	1.648***	1.143**	1.129**
	(0.295)	(0.301)	(0.310)	(0.311)	(0.072)	(0.072)
Model Specification	Ordered Logistic		Ordered Logistic		Poisson	
Observations	1,097	1,078	1,097	1,078	1,097	1,078

Notes. Columns (1)-(4) report odds ratio estimated from ordered logistic regressions. Column (5) and (6) report incidence-rate ratio estimated from Poisson regressions. Quality class and importance class are controlled in all specifications. Standard errors are reported in the parentheses. *, ** and *** denote significance level at 10%, 5% and 1% level.

Table D.4: Average Marginal Effect on Overall Quality

Dependent Variable:	Overall Quality						
	$\mathrm{P}(Y=1)$	$\mathrm{P}(Y=2)$	$\mathrm{P}(Y=3)$	$\mathrm{P}(Y=4)$	$\mathrm{P}(Y=5)$	$\mathrm{P}(Y=6)$	$\mathrm{P}(Y=7)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HighView	0.007	0.010	0.008	-0.000	-0.014	-0.007	-0.003
	(0.013)	(0.013)	(0.015)	(0.001)	(0.025)	(0.012)	(0.005)
Cite	0.010	0.015	0.011	-0.001	-0.019	-0.009	-0.004
	(0.013)	(0.013)	(0.014)	(0.002)	(0.025)	(0.011)	(0.005)
CiteAckn	-0.024**	-0.022*	-0.038**	-0.013*	0.057**	0.034**	0.015**
	(0.010)	(0.011)	(0.016)	(0.007)	(0.024)	(0.015)	(0.007)
HighView $\times$ Cite	-0.024*	-0.036*	-0.029	0.000	0.048	0.023	0.010
	(0.018)	(0.018)	(0.022)	(0.003)	(0.037)	(0.017)	(0.007)
HighView $\times$ CiteAckn	-0.003	-0.010	-0.001	0.004	0.004	-0.001	-0.001
	(0.015)	(0.016)	(0.022)	(0.009)	(0.034)	(0.020)	(0.009)
Cosine Similarity	-0.149***	-0.169***	-0.203***	-0.037**	0.322***	0.174***	0.078***
	(0.040)	(0.041)	(0.049)	(0.018)	(0.076)	(0.044)	(0.023)
log(Article Length)	-0.004	-0.005	-0.005	-0.001	0.008	0.004	0.002
	(0.006)	(0.007)	(0.008)	(0.002)	(0.013)	(0.007)	(0.003)
log(Abstract View)	0.003	-0.000	-0.004	0.001	-0.006	-0.003	-0.001
	(0.004)	(0.004)	(0.005)	(0.001)	(0.008)	(0.004)	(0.002)
English Affiliation	-0.001	-0.008	-0.002	-0.000	0.003	0.001	0.001
	(0.007)	(0.007)	(0.009)	(0.002)	(0.014)	(0.008)	(0.003)
Overlap	-0.019*	-0.023**	-0.026*	-0.005	0.042*	0.023*	0.010*
	(0.011)	(0.011)	(0.014)	(0.003)	(0.022)	(0.012)	(0.006)
HighView $+$ HighView $\times$ Cite	-0.017	-0.026**	-0.020	-0.000	0.034	0.016	0.007
	(0.013)	(0.013)	(0.016)	(0.003)	(0.027)	(0.013)	(0.006)
Cite $+$ HighView $\times$ Cite	-0.014	-0.021	-0.018	-0.001	0.029	0.014	0.006
	(0.013)	(0.013)	(0.016)	(0.003)	(0.027)	(0.013)	(0.006)
HighView $+$ HighView $\times$ CiteAckn	0.004	0.000	0.007	0.004	-0.010	-0.007	-0.003
	(0.009)	(0.010)	(0.016)	(0.009)	(0.022)	(0.016)	(0.008)
CiteAckn $+$ HighView $\times$ CiteAckn	-0.027**	-0.031**	-0.039**	-0.008	0.061**	0.033**	0.015**
	(0.012)	(0.012)	(0.016)	(0.006)	(0.025)	(0.014)	(0.006)
Model Specification	Ordered Logistic						
Observations	1078						

Notes. Columns (1)-(7) report the average marginal effects on the probability that median overall quality receives the corresponding score. Quality class and importance class are controlled in all specifications. Standard errors are reported in the parentheses. *, ** and *** denote significance level at 10%, 5% and 1% level.

Table D.5: Average Marginal Effect on Helpfulness

Dependent Variable:	Helpfulness						
	$\mathrm{P}(Y=1)$	$\mathrm{P}(Y=2)$	$\mathrm{P}(Y=3)$	$\mathrm{P}(Y=4)$	$\mathrm{P}(Y=5)$	$\mathrm{P}(Y=6)$	$\mathrm{P}(Y=7)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HighView	0.010	0.010	0.011	0.003	-0.017	-0.011	-0.005
	(0.014)	(0.013)	(0.014)	(0.004)	(0.023)	(0.015)	(0.007)
Cite	0.016	0.015	0.016	0.004	-0.026	-0.017	-0.007
	(0.014)	(0.013)	(0.013)	(0.004)	(0.022)	(0.014)	(0.006)
CiteAckn	-0.021*	-0.022*	-0.027*	-0.017*	0.038*	0.033*	0.016*
	(0.011)	(0.011)	(0.014)	(0.010)	(0.020)	(0.018)	(0.009)
HighView $\times$ Cite	-0.038*	-0.036*	-0.040*	-0.013	0.063*	0.043*	0.019*
	(0.020)	(0.018)	(0.020)	(0.008)	(0.032)	(0.022)	(0.010)
HighView $\times$ CiteAckn	-0.010	-0.010	-0.011	-0.003	0.017	0.011	0.005
	(0.017)	(0.016)	(0.020)	(0.013)	(0.029)	(0.024)	(0.012)
Cosine Similarity	-0.175***	-0.169***	-0.200***	-0.098***	0.295***	0.235***	0.111***
	(0.043)	(0.041)	(0.045)	(0.026)	(0.066)	(0.054)	(0.029)
log(Article Length)	-0.005	-0.005	-0.006**	-0.003	0.009	0.007	0.003
	(0.007)	(0.007)	(0.008)	(0.004)	(0.011)	(0.009)	(0.004)
log(Abstract View)	-0.000	-0.000	-0.001	-0.000	0.001	0.001	0.000
	(0.004)	(0.004)	(0.005)	(0.002)	(0.007)	(0.006)	(0.003)
English Affiliation	-0.008	-0.008	-0.009	-0.005	0.014	0.011	0.005
	(0.007)	(0.007)	(0.008)	(0.004)	(0.012)	(0.010)	(0.005)
Overlap	-0.024**	-0.023**	-0.027**	-0.013**	0.040**	0.032**	0.015**
	(0.011)	(0.011)	(0.013)	(0.006)	(0.019)	(0.015)	(0.007)
HighView + HighView $\times$ Cite	-0.027*	-0.026**	-0.029**	-0.010	0.046**	0.032*	0.014*
	(0.014)	(0.013)	(0.015)	(0.007)	(0.023)	(0.017)	(0.008)
Cite + HighView $\times$ Cite	-0.022	-0.021	-0.024	-0.010	0.037	0.027	0.012
	(0.014)	(0.013)	(0.015)	(0.007)	(0.023)	(0.017)	(0.008)
HighView + HighView $\times$ CiteAckn	0.000	0.000	0.000	0.000	-0.000	-0.000	-0.000
	(0.010)	(0.010)	(0.014)	(0.012)	(0.017)	(0.019)	(0.010)
CiteAckn + HighView $\times$ CiteAckn	-0.032**	-0.031**	-0.038***	-0.020**	0.055***	0.044***	0.021**
	(0.013)	(0.012)	(0.014)	(0.008)	(0.021)	(0.017)	(0.008)
Model Specification	Ordered Logistic						
Observations	1078						

Notes. Columns (1)-(7) report the average marginal effects on the probability that median helpfulness receives the corresponding score. Quality class and importance class are controlled in all specifications. Standard errors are reported in the parentheses. *, ** and *** denote significance level at 10%, 5% and 1% level.

Table D.6: Average Marginal Effect on # of Sub-comments

Dependent Variable:	# of Sub-comments
HighView	-0.288*
	(0.170)
Cite	-0.297*
	(0.167)
CiteAckn	0.335*
	(0.183)
HighView × Cite	0.343
	(0.243)
HighView × CiteAckn	-0.010
	(0.251)
Cosine Similarity	3.364***
	(0.512)
log(Article Length)	0.195**
	(0.0095)
log(Abstract View)	-0.002
	(0.058)
English Affiliation	-0.002
	(0.102)
Overlap	0.279*
	(0.154)
HighView + HighView × Cite	0.056
	(0.175)
Cite + HighView × Cite	0.046
	(0.177)
HighView + HighView × CiteAckn	-0.298
	(0.185)
CiteAckn + HighView × CiteAckn	0.324*
	(0.171)
Model Specification	Poisson
Observations	1078

Notes. Columns (1)-(7) report the average marginal effects on the number of subcomments receives the corresponding score. Quality class and importance class are controlled in all specifications. Standard errors are reported in the parentheses. *, ** and *** denote significance level at 10%, 5% and 1% level.

Table D.7: Project Size and Contribution to the Articles under the Joined Project

Dep. var.:	Revision		Addition	
Measure:	Number	Size	Number	Size
	(1)	(2)	(3)	(4)
$\log(\text{Project size})$	0.050	0.083	0.044	0.087
	(0.011)	(0.059)	(0.034)	(0.062)
$\log(\text{Project size})^2$	-0.001	-0.002	-0.000	-0.001
	(0.003)	(0.004)	(0.003)	(0.005)
Lagged dep. var.	0.554***	0.426***	0.571***	0.383***
	(0.012)	(0.011)	(0.012)	(0.010)
R^2	0.393	0.281	0.382	0.282
Observations	3,500	3,500	3,500	3,500

Notes. The dependent variable is a user's contribution after joining a WikiProject. Standard errors are provided in parentheses. *** denotes significance level at 1% level.

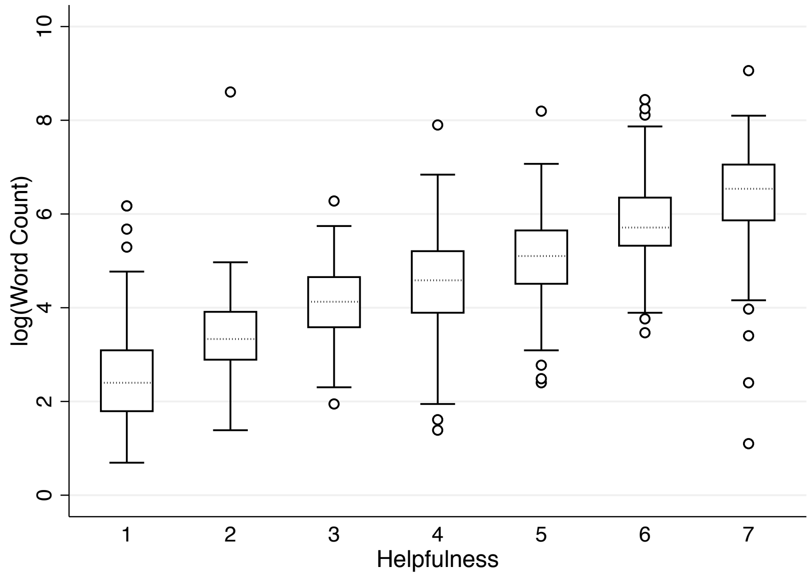


Figure D.5: Word count and median quality

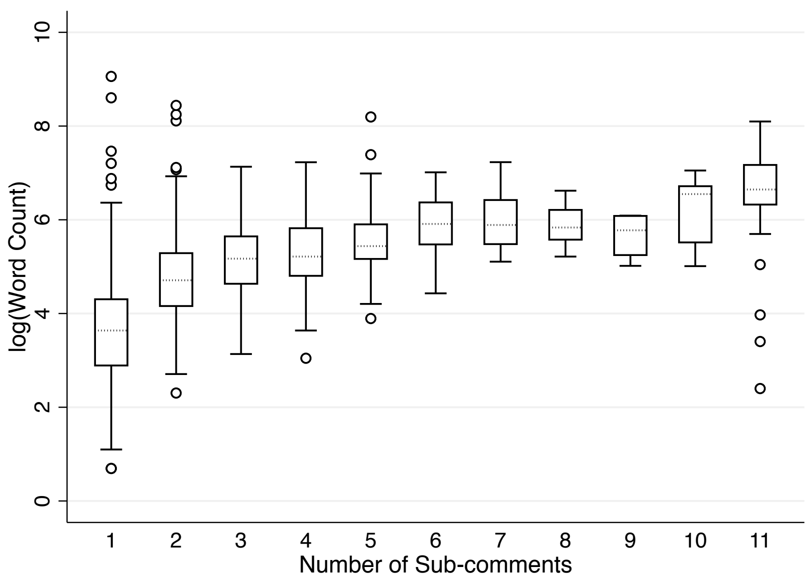


Figure D.6: Word count and # of sub-comments

Appendix E Rating Protocol

Below we provide the rating protocol. For each of the rating question, we also provide the mean, median and standard error.

Welcome to this rating session. Before you rate each comment, please read the associated Wikipedia article first.

- Suppose that you are to incorporate the expert's review of this Wikipedia article and you want to break down the review into multiple pieces of comments. How many pieces of comments has the expert made to this Wikipedia article? (mean: 2.711, median: 2, standard error: 0.069)

- According to the expert, this Wikipedia article has

  ____ errors (mean: 1.444, median: 0, standard error: 0.912)

  ____ missing points (mean: 1.098, median: 1, standard error: 0.040)

  ____ missing references (mean: 0.626, median: 0, standard error: 0.049)

  ____ outdated information (mean: 0.043, median: 0, standard error: 0.007)

  ____ outdated references (mean: 0.010, median: 0, standard error: 0.003)

  ____ irrelevant information (mean: 0.134, median: 0, standard error: 0.013)

  ____ irrelevant references (mean: 0.016, median: 0, standard error: 0.005)

  ____ other issues. (mean: 0.238, median: 0, standard error: 0.019) Please specify: ______________

- How many references does the expert provide for the Wikipedia article?____ (mean: 1.508, median: 0, standard error: 0.074)

- How many self-cited references does the expert provide for the Wikipedia article? ____ (mean: 0.374, median: 0, standard error: 0.032)

- Rate the amount of effort needed to address the experts' comments. (1 = cut and paste; 7 = rewrite the entire article) (mean: 3.621, median: 4, standard error: 0.057)

- Rate the amount of expertise needed to address the experts' comments. (1 = high school AP economics classes; 7 = PhD in economics) (mean: 3.887, median: 4, standard error: 0.057)

- How easily can the issues raised in the comment be located in the Wikipedia article? (1 = unclear where to modify in the Wikipedia article; 7 = can be identified at the sentence level) (mean: 4.572, median: 5, standard error: 0.061)

- Suppose you are to incorporate this expert's comments. How helpful are they? (1 = not helpful at all; 7 = very helpful) (mean: 4.121, median: 4, standard error: 0.045)

- Please rate the overall quality of the comment. (1 = not helpful at all; 7 = extremely helpful) (mean: 3.968, median: 4, standard error: 0.044)

References

Adamic, Lada A., Jun Zhang, Eytan Bakshy, and Mark S. Ackerman, "Knowledge Sharing and Yahoo Answers: Everyone Knows Something," in "Proceedings of the 17th International Conference on World Wide Web" ACM 2008, pp. 665–674.

Ai, Chunrong and Edward C Norton, "Interaction terms in logit and probit models," *Economics letters*, 2003, *80* (1), 123–129.

Ai, Wei, Roy Chen, Yan Chen, Qiaozhu Mei, and Webb Phillips, "Recommending Teams Promotes Prosocial Lending in Online Microfinance," *Proceedings of the National Academy of Sciences*, 2016, *113* (52), 14944–14948.

Airio, Eija, "Word normalization and decompounding in mono-and bilingual IR," *Information Retrieval*, 2006, *9* (3), 249–271.

Akerlof, George A., "Social Distance and Social Decisions," *Econometrica*, 1997, *65* (5), 1005–1027.

__ **and Rachel E. Kranton**, "Economics and Identity," *Quarterly Journal of Economics*, August 2000, *115* (3), 715–753.

__ **and** __ , "Economics and Identity," *Quarterly Journal of Economics*, 2000, *115* (3), 715–753.

__ **and** __ , *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*, Princeton, New Jersey: Princeton University Press, 2010.

Algan, Yann, Yochai Benkler, Mayo Fuster Morell, and Jérôme Hergueux, "Cooperation in a Peer Production Economy Experimental Evidence from Wikipedia," in "Workshop on Information Systems and Economics, Milan, Italy" 2013, pp. 1–31.

Andreoni, James, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, 1989, *97* (6), 1447–1458.

__ , "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *Economic Journal*, 1990, *100* (401), 464–477.

__ , "Giving Gifts to groups: How Altruism Depends on the Number of Recipients," *Journal of Public Economics*, 2007, *91* (9), 1731–1749.

__ **and B. Douglas Bernheim**, "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica*, 2009, *77* (5), 1607–1636.

Anya, Obinna, "Bridge the Gap!: What Can Work Design in Crowdwork Learn from Work Design Theories?," in "Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing" ACM 2015, pp. 612–627.

Arazy, Ofer, Lisa Yeo, and Oded Nov, "Stay on the Wikipedia Task: When Task-related Disagreements Slip into Personal and Procedural Conflicts," *Journal of the American Society for Information Science and Technology*, 2013, *64* (8), 1634–1648.

__ **, Oded Nov, Raymond Patterson, and Lisa Yeo**, "Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict," *Journal of Management Information Systems*, 2011, *27* (4), 71–98.

__ **, Wayne Morgan, and Raymond Patterson**, "Wisdom of the Crowds: Decentralized Knowledge Construction in Wikipedia," in "16th Annual Workshop on Information Technologies & Systems (WITS) Paper" 2006.

Argote, Linda, Marlene E. Turner, and Mark Fichman, "To Centralize or Not to Centralize: The Effects of Uncertainty and Threat on Group Structure and Performance," *Organizational Behavior and Human Decision Processes*, 1989, *43* (1), 58–74.

Ariely, Dan, Anat Bracha, and Stephan Meier, "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," *American Economic Review*, 2009, *99* (1), 544–555.

Baldwin, Carliss Y. and Eric A. von Hippel, "Modeling a Paradigm Shift: From Producer Innovation to User and Open Collaborative Innovation," *Organization Science*, 2011, *22* (6), 1399–1417.

Basu, Kaushik, "Identity, Trust and Altruism: Sociological Clues to Economic Development," 2006. CAE Working Paper #06-06.

Bénabou, Roland and Jean Tirole, "Identity, Morals, and Taboos: Beliefs as Assets," *Quarterly Journal of Economics*, 2011, *126* (2), 805–855.

Bergstrom, Theodore, Lawrence Blume, and Hal Varian, "On the Private

Provision of Public Goods," *Journal of Public Economics*, 1986, *29* (1), 25–49.
Bernhard, Helen, Ernst Fehr, and Urs Fischbacher, "Group Affiliation and Altruistic Norm Enforcement," *Amer. Econ. Rev.*, May 2006, *96* (2), 217–221.
Bornstein, Gary, Uri Gneezy, and Rosmarie Nagel, "The Effect of Intergroup Competition on Group Coordination: An Experimental Study," *Games and Economic Behavior*, October 2002, *41*, 1 – 25.
Castillo, Marco, Ragan Petrie, and Clarence Wardell, "Fundraising through Online Social Networks: A Field Experiment on Peer-to-peer Solicitation," *Journal of Public Economics*, 2014, *114*, 29–35.
Charness, Gary, Ramón Cobo-Reyes, and Natalia Jiménez, "Identities, Selection, and Contributions in a Public-goods Game," *Games and Economic Behavior*, 2014, *87*, 322–338.
Chen, Jilin, Yuqing Ren, and John Riedl, "The Effects of Diversity on Group Productivity and Member Withdrawal in Online Volunteer Groups," in "Proceedings of the SIGCHI Conference on Human Factors in Computing Systems" ACM 2010, pp. 821–830.
Chen, Roy and Yan Chen, "The Potential of Social Identity for Equilibrium Selection," *American Economic Review*, 2011, *101* (6), 2562–2589.
__ , __ , Yang Liu, and Qiaozhu Mei, "Does Team Competition Increase Pro-social Lending? Evidence from Online Microfinance," *Games and Economic Behavior*, 2017, *101*, 311–333.
Chen, Yan, "Incentive-Compatible Mechanisms for Pure Public Goods: A Survey of Experimental Research," in Charles Plott and Vernon Smith, eds., *The Handbook of Experimental Economics Results*, Vol. 1, Amsterdam: North-Holland, 2008, pp. 625 – 643.
__ and Sherry Xin Li, "Group Identity and Social Preferences," *The American Economic Review*, March 2009, *99* (1), 431–457.
__ and Yingzhi Liang, "Optimal Team Size in Public Goods Provision: Theory and Experiments," 2018. University of Michigan Working Paper.
__ , F. Maxwell Harper, Joseph Konstan, and Sherry Xin Li, "Social Comparisons and Contributions to Online Communities: A Field Experiment on Movielens," *American Economic Review*, 2010, *100* (4), 1358–98.

__ , Teck-Hua Ho, and Yong-Mi Kim, "Knowledge Market Design: A Field Experiment at Google Answers," *Journal of Public Economic Theory*, 2010, *12* (4), 641–664.

Cohendet, Patrick and Laurent Simon, "Always Playable: Recombining Routines for Creative Efficiency at Ubisoft Montreal's Video Game Studio," *Organization Science*, 2016, *27* (3), 614–632.

Collins, Francis S. and Harold Varmus, "A New Initiative on Precision Medicine," *New England Journal of Medicine*, 2015, *372* (9), 793–795.

Cosley, Dan, Dan Frankowski, Loren Terveen, and John Riedl, "SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia," in "Proceedings of the 12th international conference on Intelligent user interfaces" ACM 2007, pp. 32–41.

Croson, Rachel T. A., Melanie B. Marks, and Jessica Snyder, "Groups Work for Women: Gender and Group Identity in the Provision of Public Goods," *Negotiation Journal*, October 2008, *24* (4), 411–427.

Cummings, Jonathon N., J. Alberto Espinosa, and Cynthia K. Pickering, "Crossing Spatial and Temporal Boundaries in Globally Distributed Projects: A Relational Model of Coordination Eelay," *Information Systems Research*, 2009, *20* (3), 420–439.

DellaVigna, Stefano and Devin Pope, "What Motivates Effort? Evidence and Expert Forecasts," *Review of Economic Studies*, 2017, *85* (2), 1029–1069.

__ , John A. List, and Ulrike Malmendier, "Testing for Altruism and Social Pressure in Charitable Giving," *Quarterly Journal of Economics*, 2012, *127* (1), 1–56.

Dorfman, Robert, "A Formula for the Gini Coefficient," *The Review of Economics and Statistics*, 1979, pp. 146–149.

Driskell, James E. and Eduardo Salas, "Group Decision Making under Stress," *Journal of Applied Psychology*, 1991, *76* (3), 473.

Dutton, Jane E., Monica C. Worline, Peter J. Frost, and Jacoba Lilius, "Explaining Compassion Organizing," *Administrative Science Quarterly*, 2006, *51* (1), 59–96.

Eckel, Catherine C. and Philip J. Grossman, "Managing Diversity by

Creating Team Identity," *Journal of Economic Behavior & Organization*, 2005, *58* (3), 371–392.
Edelman, Benjamin, "Earnings and Ratings at Google Answers," *Economic Inquiry*, 2012, *50* (2), 309–320.
Erev, Ido, Gary Bornstein, and Rachely Galili, "Constructive Intergroup Competition as a Solution to the Free Rider Problem: A Field Experiment," *Journal of Experimental Social Psychology*, 1993, *29* (6), 463–478.
Filippova, Anna and Hichang Cho, "The Effects and Antecedents of Conflict in Free and Open Source Software Development," in "Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work & Social Computing" ACM 2016, pp. 705–716.
Gallus, Jana, "Fostering Public Good Contributions with Symbolic Awards: A Large-scale Natural Field Experiment at Wikipedia," *Management Science*, 2016, *63* (12), 3999–4015.
Gladstein, Deborah L. and Nora P. Reilly, "Group Decision Making under Threat: The Tycoon Game," *Academy of Management Journal*, 1985, *28* (3), 613–627.
Goeree, Jacob K., Charles A. Holt, and Susan K. Laury, "Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior," *Journal of Public Economics*, 2002, *83* (2), 255–276.
Goette, Lorenz, David Huffman, and Stephan Meier, "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups," *Amer. Econ. Rev.*, May 2006, *96* (2), 212–216.
Groves, Theodore and John O. Ledyard, "Incentive Compatibility since 1972," in Theodore Groves, Roy Radner, and Stanley Reiter, eds., *Information, Incentives and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*, Minneapolis: University of Minnesota Press, 1987, pp. 48–111.
Guttman, Joel M., "Matching behavior and collective action: Some experimental evidence," *Journal of Economic Behavior & Organization*, 1986, *7* (2), 171 – 198.
Harrington, Robert, David Lemak, and K.W. Kendall, "The Threat-rigidity

Thesis in Newly Formed Teams: An Empirical Test," *Journal of Business and Management*, 2002, *8* (2), 127.

Hinds, Pamela J. and Diane E. Bailey, "Out of Sight, Out of Sync: Understanding Conflict in Distributed Teams," *Organization science*, 2003, *14* (6), 615–632.

Huang, Chu-Ren, Petr Šimon, Shu-Kai Hsieh, and Laurent Prévot, "Rethinking Chinese word segmentation: tokenization, character classification, or wordbreak identification," in "Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions" Association for Computational Linguistics 2007, pp. 69–72.

Isaac, R Mark and James M Walker, "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism," *Quarterly Journal of Economics*, 1988, *103* (1), 179–199.

Isaac, R. Mark, James M. Walker, and Arlington W. Williams, "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups," *Journal of Public Economics*, 1994, *54* (1), 1–36.

Jackson, Susan and Jane Dutton, "Discerning Threats and Opportunities," *Administrative Science Quarterly*, 1988, pp. 370–387.

Jordan, MI and TM Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, 2015, *349* (6245), 255–260.

Jorgensen, Bent, "Exponential dispersion models," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1987, pp. 127–162.

Jr., Joseph Strayhorn, John F. McDermott, and Peter Tanguay, "An Intervention to Improve the Reliability of Manuscript Reviews for the *Journal of the American Academy of Child and Adolescent Psychiatry*," *American Journal of Psychiatry*, 1993, *150* (6), 947–952.

Kamphuis, Wim, Anthony Gaillard, and A. Vogelaar, "The Effects of Physical Threat on Team Processes during Complex Task Performance," *Small Group Research*, 2011, *42* (6), 700–729.

Karau, Steven J and Kipling D Williams, "Social loafing: A meta-analytic review and theoretical integration.," *Journal of personality and social psychology*, 1993, *65* (4), 681.

Keegan, Brian, Darren Gergle, and Noshir Contractor, "Do Editors or Articles Drive Collaboration?: Multilevel Statistical Network Analysis of Wikipedia Coauthorship," in "Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work & Social Computing" ACM 2012, pp. 427–436.

Kittur, Aniket and Robert E. Kraut, "Harnessing the Wisdom of Crowds in Wikipedia: Quality through Coordination," in "Proceedings of the 11th ACM Conference on Computer Supported Cooperative Work & Social Computing" ACM 2008, pp. 37–46.

__ and __ , "Beyond Wikipedia: Coordination and Conflict in Online Production Groups," in "Proceedings of the 13th ACM Conference on Computer Supported Cooperative Work & Social Computing" ACM 2010, pp. 215–224.

__ , Bryan Pendleton, and Robert E. Kraut, "Herding the Cats: The Influence of Groups in Coordinating Peer Production," in "Proceedings of the 5th International Symposium on Wikis and Open Collaboration" ACM 2009, p. 7.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer, "Prediction Policy Problems," *American Economic Review*, 2015, *105* (5), 491–95.

Kriplean, Travis, Ivan Beschastnikh, and David W. McDonald, "Articulations of Wikiwork: Uncovering Valued Work in Wikipedia through Barnstars," in "Proceedings of the 11th ACM Conference on Computer Supported Cooperative Work & Social Computing" ACM 2008, pp. 47–56.

Krippendorff, Klaus, *Content Analysis: An Introduction to Its Methodology*, 2nd ed., Thousand Oaks, CA: Sage Publications, 2003.

Ledyard, John, "Public goods: A Survey of Experimental Research," in John H. Kagel and Alvin E. Roth, eds., *The Handbook of Experimental Economics*, Vol. 1, Princeton, New Jersey: Princeton University Press, 1995.

Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman, *Mining of massive datasets*, Cambridge university press, 2014.

Lih, Andrew, *The Wikipedia Revolution: How a Bunch of Nobodies Created the World's Greatest Encyclopedia*, London, UK: Aurum Press, 2009.

MacKinnon, David P., Amanda J. Fairchild, and Matthew S. Fritz,

"Mediation Analysis," *Annual Review of Psychology*, 2007, *58*, 593.
Manning, Christopher D., Christopher D. Manning, and Hinrich Schütze, *Foundations of Statistical Natural Language Processing*, MIT press, 1999.
Mcleish, Kendra N. and Robert J.Oxoby, "Identity,Cooperation, and Punishment." IZA Discussion Paper No.2572.
Miller, Rob, "Wikipedia Founder Jimmy Wales Responds," *Slashdot*, July 28 2004.
Morgan, Jonathan T., Michael Gilbert, David W. McDonald, and Mark Zachry, "Editing beyond Articles: Diversity & Dynamics of Teamwork in Open Collaborations," in "Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing" ACM 2014, pp. 550–563.
__ , Siko Bouterse, Heather Walls, and Sarah Stierch, "Tea and Sympathy: Crafting Positive New User Experiences on Wikipedia," in "Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work & Social Computing" ACM 2013, pp. 839–848.
Ocasio, William, "Attention to Attention," *Organization Science*, 2011, *22* (5), 1286–1296.
Platt, Edward L. and Daniel M. Romero, "Network Structure, Efficiency, and Performance in WikiProjects," *arXiv preprint arXiv:1804.03763*, 2018.
Ransbotham, Sam and Gerald C. Kane, "Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia," *MIS Quarterly*, 2011, *35* (3), 613–627.
Rege, Mari and Kjetil Telle, "The Impact of Social Approval and Framing on Cooperation in Public Good Situations," *Journal of Public Economics*, 2004, *88* (7), 1625–1644.
Robert, Lionel P. and Daniel M. Romero, "Crowd Size, Diversity and Performance," in "Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems" ACM 2015, pp. 1379–1382.
__ and __ , "The Influence of Diversity and Experience on the Effects of Crowd Size," *Journal of the Association for Information Science and Technology*, 2017, *68* (2), 321–332.
Romero, Daniel M., Dan Huttenlocher, and Jon Kleinberg, "Coordination

and Efficiency in Decentralized Collaboration," in "Proceedings of the Ninth International AAAI Conference on Web and Social Media" pp. 367–376.

Samuelson, Paul A., "The Pure Theory of Public Expenditure," *Review of Economics and Statistics*, 1954, *36* (4), 387–389.

Singhal, Amit, "Modern Information Retrieval: A Brief Overview," *IEEE Data Engineering Bulletin*, 2001, *24* (4), 35–43.

Solomon, Jacob and Rick Wash, "Critical Mass of What? Exploring Community Growth in WikiProjects," in "Proceedings of the 8th International Conference on Web and Social Media" AAAI 2014, pp. 476–484.

Staw, Barry M., Lance E. Sandelands, and Jane E. Dutton, "Threat rigidity effects in organizational behavior: A multilevel analysis," *Administrative science quarterly*, 1981, pp. 501–524.

Sutcliffe, Kathleen M. and Timothy J. Vogus, "Organizing for Resilience," *Positive Organizational Scholarship: Foundations of a New Discipline*, 2003, *94*, 110.

Thompson, Neil and Douglas Hanley, "Science is Shaped by Wikipedia: Evidence From a Randomized Control Trial," 2017.

Tsvetkova, Milena, Ruth García-Gavilanes, and Taha Yasseri, "Dynamics of Disagreement: Large-Scale Temporal Network Analysis Reveals Negative Interactions in Online Collaboration," *Scientific Reports*, 2016, *6*.

Vesterlund, Lise, "Using Experimental Methods to Understand Why and How We Give to Charity," in John H. Kagel and Alvin E. Roth, eds., *The Handbook of Experimental Economics*, Vol. 2, Princeton, New Jersey: Princeton University Press, 2015.

Viegas, Fernanda B., Martin Wattenberg, Jesse Kriss, and Frank Van Ham, "Talk before you type: Coordination in Wikipedia," in "System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on" IEEE 2007, pp. 78–78.

Wang, Yi-Chia, Robert Kraut, and John M. Levine, "To Stay or Leave?: The Relationship of Emotional and Informational Support to Commitment in Online Health Support Groups," in "Proceedings of the 15th ACM Conference on Computer Supported Cooperative Work & Social Computing" ACM 2012, pp. 833–842.

Weber, Roberto A., "Managing Growth to Achieve Efficient Coordination in Large Groups," *American Economic Review*, 2006, *96* (1), 114–126.

Weber, Steven, *The Success of Open Source*, Cambridge, MA: Harvard University Press, 2004.

Windeler, Jaime B., Likoebe M. Maruping, Lionel P. Robert, and Cynthia K. Riemenschneider, "E-profiles, Conflict, and Shared Understanding in Distributed Teams," *Journal of the Association for Information Systems*, 2015, *16* (7), 608.

Yu, Lixiu Lisa and Jeffrey V. Nickerson, "Generating Creative Ideas through Crowds: An Experimental Study of Combination," in "Thirty Second International Conference on Information Systems" 2011.

Zhang, Xiaoquan and Feng Zhu, "Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia," *American Economic Review*, 2011, *101* (4), 1601–1615.

Zhang, Xiaoquan Michael and Feng Zhu, "Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia," *American Economic Review*, 2011, *101* (4), 1601–15.

Zhang, Yanwei, "Likelihood-based and bayesian methods for tweedie compound poisson linear mixed models," *Statistics and Computing*, 2013, *23* (6), 743–757.

Zhu, Haiyi, Robert Kraut, and Aniket Kittur, "Effectiveness of Shared Leadership in Online Communities," in "Proceedings of the 15th ACM Conference on Computer Supported Cooperative Work & Social Computing" ACM 2012, pp. 407–416.