NET Institute*

www.NETinst.org

Working Paper #07-22

September 2007

Group Size and Incentive to Contribute: A Natural Experiment at Chinese Wikipedia

Michael Zhang
HKUST Business School

Feng Zhu
Harvard University

Group Size and Incentive to Contribute: A Natural Experiment at Chinese Wikipedia

Xiaoquan (Michael) Zhang and Feng Zhu*

Abstract

The literature of private provision of public goods suggests that incentive to contribute is inversely related to group size. This paper empirically tests this relationship using field data from Chinese Wikipedia, an online encyclopedia. We exploit an exogenous reduction in group size as a result of the blocking of Wikipedia in mainland China and examine whether individual contributions increase after the block as predicted in the literature. Our result indicates the opposite: individual contribution of unaffected contributors decreases by 42% on average as a result of the block. We attribute the cause to social effects: contributors care about the number of beneficiaries of their contributions. We build a simple model to illustrate how social effects and group size affect individual incentive to contribute. Consistent with our model prediction, we find that the more a contributor values social recognition, the greater the reduction in her contributions after the block. A series of robustness checks appear to support our explanation.

JEL: D85, H44, L14, L31, L86
Keywords: incentive to contribute; group size; public goods; social effects

*Hong Kong University of Science and Technology; Harvard University. Email: zhang@ust.hk and feng@hbs.edu. Financial support from the NET Institute is gratefully acknowledged. Preliminary draft; Comments welcome.

> Mencius asked [King Hsüan of Ch'î], "Which is the more pleasant — to perform music alone, or to perform it with others?" "To perform it with others," answered the King. "And which is the more pleasant — to perform music with a few, or to perform it with many?" "To perform it with many," was the answer.
>
> — The Works of Mencius (Circa 300 BC)

1 Introduction

A fundamental question in the private provision of public good concerns the relationship between group size and incentive to contribute. The major focus of the literature has been the "crowding out" hypothesis: when the population grows, individual contribution level will decline (e.g., Olson 1965; Andreoni 1988). The intuition is that individuals treat others' contributions as substitutes to their own. Consequently, as group size increases, they are more likely to free ride.

We empirically examine the causal effect of changes in group size on individual contributions. Identification hinges on an exogenous reduction in group size at Chinese Wikipedia, the Chinese language version of an online encyclopedia that relies entirely on voluntary contribution. Contributors to Chinese Wikipedia are composed of Chinese-speaking people in mainland China, Taiwan, Hong Kong, Singapore and other regions in the world. We exploit the block to Chinese Wikipedia in mainland China in October 2005. As a result of the block, while contributors from other parts of the world were not affected, mainland Chinese could not use or contribute to Chinese Wikipedia.

We empirically test whether the reduction in group size leads to increases in individual contributions as predicted in the literature. Our results indicate the opposite: individual contributions of unaffected contributors decrease by 42% on average as a result of the block. We attribute the cause to social effects: contributors care about the number of beneficiaries of their contributions. The shrinking group size weakened these social effects. As a result, even

though the incentive to free ride decreased after the block, when social effects are important, individual contributions could decrease. We build a simple framework to illustrate how social effects and group size affect individual incentive to contribute. Consistent with our model prediction, we find that the more a contributor values social recognition, the greater the reduction in her contribution after the block.

Our study contributes to the theoretical literature on private provision of public goods. The literature starts with Olson (1965)'s conjecture that public goods are less likely to be provided in larger groups as individuals have incentive to free ride other people's contributions. Andreoni (1988) and Fries et al. (1991) examine the conjecture in the context of continuous public goods where individual utility depends on private good consumption and total private provision of the public good. They show that as group size increases to infinity, individual contribution level falls to zero. Palfrey and Rosenthal (1984) and Hindriks and Pancs (2002) analyze discrete public goods[1] and find the same result. As Andreoni (1988) points out, these results fail to confirm empirical observations about charity and have limited predictive power. Therefore, it is necessary to generalize these models to include other motives for giving.

Our model is the first to incorporate group size into individuals' utility. In our model, for a given contribution, the larger the group, the more utility (e.g., social recognition) the contributor derives. The idea that contributors receive social recognition, or they experience a "warm glow" from contributing has been suggested numerous times in the literature (e.g., Becker 1974; Rege 2004; Rege and Telle 2004; Carman 2006). A number of recent studies on open source communities also find that contributors indeed derive utility from helping others and gaining respect (e.g., Lerner and Tirole 2002; Lakhani and von Hippel 2003). Surprisingly, theoretical studies related to group size have not considered these social effects. Even in models that explicitly incorporate warm-glow motives, warm glow is often assumed to be independent of group size (e.g., Ribar and Wilhelm 2002). We show that incorporat-

[1]That is, the public good is provided when a sufficient number of contributions are made. Otherwise, it is not provided.

ing group size could overturn the inverse relationship between group size and incentive to contribute found in prior theoretical models.

Our research also contributes to the empirical literature that examines the group size effects. The majority of the existing literature is based on experimental data. In general, these experiments find increased free-riding behavior in larger groups (e.g., Sweeney 1973; Chamberlin 1978), although some of the studies suggest that group size effects could be weak (e.g., Marwell and Ames 1979; Chamberlin 1978). One exception is Isaac et al. (1994), which finds that groups of 40 and 100 subjects provide the good more efficiently than groups of 4 and 10. In their study, however, the public good is divided among participants. As a result, an increase in group size reduces the benefit of the public good to each participant. Laboratory studies on this topic are necessarily handicapped because in comparison with the group sizes we often find in real world, groups used in the experiments are of very small sizes.

A couple of studies have used field data to test the group size effects. Goetze et al. (1993) examine contributions to public television stations and find that average contributions per contributor decrease with group size. Brunner (1998) studies contributions to public radio stations and finds that the proportion of listeners contributing to public radio decreases as the number of listeners increases. In addition, he finds that as group size increases, average contributions per contributor remain constant. Our empirical study differs from these studies in two aspects. First, we measure changes in contributions at the individual level, while the two studies use aggregate data. Second, we exploit an exogenous change of group size to establish the causal relationship between group size and incentive to contribute. As many factors could change group size and incentive to contribute at the same time, the relationships found in prior studies could be spurious.

The rest of the paper is organized as follows. Section 2 presents the theoretical model. Section 3 presents the empirical results and various robustness checks. Section 4 concludes.

2 The Model

Assume there are n people in the economy who benefit from the provision of a public good. Each person can choose whether to contribute to the public good. If she contributes, she incurs a fixed cost, c. If not, she could free ride the public good. The n people differ in their willingness to contribute, which is denoted by θ. Assume that for each person i, her type θ_i is independently and identically drawn from a probability distribution function $F(\theta)$ with the corresponding density function $f(\theta)$. Assume the person with type θ^* is indifferent between contributing and not contributing. That is, any one with $\theta \geq \theta^*$ will contribute. The number of contributors in the economy is $n[1 - F(\theta^*)]$.

2.1 Baseline Case

More contributors lead to better quality and hence greater utility from using the public good. Let the utility from the public good be $U\big(n[1 - F(\theta^*)]\big)$, where $U(\cdot)$ is an increasing and concave function, and $U(0) = 0$. Denote the warm-glow motive by w (Andreoni 1989). Let $k^* \equiv 1 - F(\theta^*)$ be the fraction of population that contributes to the public good.

Specifically, k^* is implicitly defined by

$$w + U'(nk^*) = c, \tag{1}$$

To be consistent with the literature, we define warm glow as the pure joy of contributing to a good cause to distinguish it from social effects. Therefore, warm glow here does not depend on the number of potential beneficiaries and the number of contributors. We assume $w < c$ so that not everyone contributes.

It is straightforward to show, from (1), that warm glow can partially alleviate the free-riding problem.

Proposition 1. *If there exists warm-glow motive, a higher percentage of population contributes to the public good.*

Proof. Let the proportion of contributors be $\hat{k}$, and $\hat{k}$ satisfies $U'(n\hat{k}) = c$. Since U is a concave function, $U'(nk^*) = c - w < c = U'(n\hat{k})$, we have $k^* > \hat{k}$. □

When someone makes a contribution, the quality of the public good increases. In addition, she could enjoy more utility from the product than from free-riding. $U'(nk^*)$ captures this additional utility. Therefore, the left hand side of (1) measures the marginal benefit from contributing to the public good. The right hand side is the marginal cost of contributing.

We can derive the familiar free-riding result as below.

Proposition 2. *The fraction of people free-riding increases with the group size.*

Proof. Since $U(\cdot)$ is well-behaved, we can write

$$k^* = \frac{1}{n} U'^{-1}(c - w), \tag{2}$$

where $U'^{-1}(\cdot)$ is the inverse function of $U'(\cdot)$. When n goes to infinity, the proportion of contributors goes to zero:

$$\lim_{n\to\infty} k^* = \lim_{n\to\infty} \frac{1}{n} U'^{-1}(c - w) = 0$$

□

Proposition 2 indicates that when the population size increases, the proportion of contributors declines. In terms of θ^*, this result suggests that when the group size is very large, only those with very high willingness-to-contribute will contribute.

Note that the result in Proposition 2 holds regardless of the inclusion of the warm-glow motive. This observation suggests that warm-glow motive cannot be an explanation of the provision of public goods in very large groups.

2.2 Social Effects

We now extend the baseline model to incorporate social effects. When a person contributes to the public good, she may enjoy additional utility when she knows that her contribution is valued by others even if there is no direct communication with the potential beneficiaries.

Let $\tilde{k}$ be the fraction of population that contributes in this case.[2] Denote the utility derived from social effects by $V(n, \tilde{k})$ for all contributors. $V(0, \cdot) = 0$. We assume $\frac{\partial V(n,\tilde{k})}{\partial n} > 0$, and $\frac{\partial V(n,\tilde{k})}{\partial k} > 0$, so that the utility of social value increases with both the population and the number of contributors. Note that $V(n, \tilde{k})$ is a function of both n and $\tilde{k}$. If the contributors only value social interaction with other contributors, then $V(n, \tilde{k}) = V(n\tilde{k})$. By choosing this more general form, we are able to consider more flexible interpretations of social effects.

Similar to (1), the marginal contributor has

$$w + U'(n\tilde{k}) + V(n, \tilde{k}) = c. \tag{3}$$

Equation (3) suggests that for the marginal contributor, the cost of contributing equals the total utility from warm glow, additional benefit from the improved public good, and social benefits. Comparing (3) with (1) gives the following result.

Proposition 3. *For any given group size, the proportion of population that contributes is higher when contributors receive additional utility from social interactions.*

Proof. From equations (1) and (3), we have $U'(n\tilde{k}) = c - w - V(n, \tilde{k}) < c - w = U'(nk^*)$. As $U(\cdot)$ is concave, $n\tilde{k} > nk^*$, and hence $\tilde{k} > k^*$. □

We now examine how $\tilde{k}$ changes with n with the inclusion of social effects. Let $B \equiv w + U'(n\tilde{k}) + V(n, \tilde{k})$ be the marginal benefit from contributing.

$$\frac{dB}{dn} = \tilde{k}U''(n\tilde{k}) + \frac{\partial V(n, \tilde{k})}{\partial n}.$$

[2]This $\tilde{k}$ is uniquely associated with a type $\tilde{\theta}$ that maps the type of the marginal contributor when social interactions are valued by the contributors.

As $U''(n\tilde{k}) < 0$ and $\frac{\partial V(n,\tilde{k})}{\partial n} > 0$, $\frac{dB}{dn}$ can be positive or negative. An increase in group size now has two effects. On one hand, it induces more people to free ride as now the marginal benefit from the improved public good becomes smaller. On the other hand, as more people are contributing, the benefit from social effects increases. Hence, the net effect of group size on the total marginal benefit from contributing depends on the relative strength of each effect.

When $\frac{dB}{dn} > 0$ $\left(\text{i.e., } \frac{\partial V(n,\tilde{k})}{\partial n} > -\tilde{k}U''(n\tilde{k})\right)$, the increase in the utility from social value is greater than the decrease in utility from the public good as n increases, so that the overall benefit from contributing increases. As the marginal cost does not change, the indifferent person will prefer contributing to the public good when n increases. Hence, a larger proportion of population will contribute when n increases.

When $\frac{dB}{dn} < 0$ $\left(\text{i.e., } \frac{\partial V(n,\tilde{k})}{\partial n} < -\tilde{k}U''(n\tilde{k})\right)$, the utility from social effects is small so that as n increases, the indifferent person will prefer free-riding. Hence, a smaller proportion of population will contribute in this case.

Proposition 4. *When contributors value social interactions, as group size increases, the percentage of population that contributes increases if* $\frac{\partial V(n,\tilde{k})}{\partial n} > -\tilde{k}U''(n\tilde{k})$*, and decreases if* $\frac{\partial V(n,\tilde{k})}{\partial n} < -\tilde{k}U''(n\tilde{k})$.

The result suggests a tension between the incentive to free ride and the desire to enjoy increased social benefits when the group size increases. We find that when the increase in the social benefits is relatively small, individual contribution will decrease as the group size increases. When such increase is large, individual contributions may increase.

3 Empirical Analysis

3.1 Background

Wikipedia is a Web-based free encyclopedia project operated by the Wikimedia Foundation. Its goal is to "give every single person in the world free access to the sum of all human knowledge."[3] Wikipedia's articles have been written collaboratively by volunteers around the world and can be edited by anyone with access to the Internet. Chinese Wikipedia started in October 2002, and is the Chinese language edition of Wikipedia.

Due to political concerns, the Chinese government has established the "great firewall" to censor the access of mainland Chinese to various information sources, among which are all Wikipedia's sites. Access to Chinese Wikipedia has been blocked and unblocked in mainland China five times.[4] The first block took place on June 2, 2004. All Wikipedia sites[5] were blocked in mainland China. In response to the block, two administrators of the Chinese Wikipedia site, 'Shizhao' and 'Mountain,' contacted their respective ISPs, and drafted an appeal, which was submitted on June 15, 2004. During the period between June 17 and June 21, 2004, all the Wikipedia sites were unblocked.

The second block, which lasted from September 23 to September 27, 2004, was not universal. While some users in mainland China reported that during this period access to Wikipedia had been erratic or unavailable, many users in mainland China were never affected.

The third block began on October 19, 2005. Shizhao once again submitted an appeal to his ISP on October 21. Given the experience with the first two blocks, many people expected the block to be lifted soon. The appeal got no response. However, on the morning of October 31, 2005, contributors from mainland China began to report that they could

[3] http://en.wikipedia.org/wiki/User:Jimbo_Wales, accessed August 2007.

[4] Information about the five blocks can be found at http://en.wikipedia.org/wiki/Blocking_of_Wikipedia_in_mainland_China

[5] In addition to Wikipedia in various languages, the Wikimedia Foundation manages other sites such as a multi-language dictionary and thesaurus named Wiktionary and an encyclopedia of quotations named Wikiquote.

access Wikipedia. It turned out that this "unblocking" was linked to a server upgrade in the Korean server cluster. A change of the IP address of Wikipedia sites for users in China circumvented the block. Within a few hours, Wikipedia was once again blocked. The prompt block on October 31 made it abundantly clear to the Chinese Wikipedia community that the block was going to be permanent. Since then, no more plans for appeals have been made. This block was widely discussed in both mainland China and other places.

Nearly a year later, the block was partially lifted. Beginning from October 10, 2006, some parts of mainland China could access Wikipedia. On November 10, 2006, Chinese Wikipedia appeared to have been fully unblocked but was reblocked on November 17.

The fifth block took place in July 2007. On June 15, 2007, the block to Wikipedia sites was lifted, with the exception of several sensitive articles and Chinese Wikipedia. On July 25, 2007, Chinese Wikipedia was unblocked as well, only to be blocked again after several hours.

Chinese Wikipedia offers an ideal empirical setting to study the relationship between group size and incentive to contribute for several reasons. First, individual contribution can be easily measured in Wikipedia. Wikipedia keeps the complete editing history of all articles, and each edit can be traced to an ID that uniquely identifies the contributor. As a result, we are able to accurately measure the contributions from each contributor over time.

Second, changes in individual contribution could correlate with group size for various reasons. Merely associating the group size with the incentive of contribution may be misleading in assessing their relationship. For example, as group size in Wikipedia increases over time, a reduction in contribution of the existing contributors can be observed. While this observation is consistent with the crowding-out hypothesis in the literature, it might be a result of the contributors having less to contribute over time, or might simply be a case of their slacking off after a "honeymoon" period. To establish a causal relationship, it is critical to look for factors that change group size exogenously. The blocks provide natural experiments to examine the impact of group size. The blocks are exogenous as for each block there

was no warning beforehand nor explanations afterwards. As contributors outside mainland China were unaffected, we could examine changes in their contribution levels and study how their incentive to contribute changes as a result of the change in group size.

Third, contributors at Wikipedia are motivated mostly by altruism and joy (Nov forthcoming). In many other public good provision circumstances such as open source software projects, contributors may be motivated by monetary payment and potential career advancement (see, for example, Lerner and Tirole 2002), and as a result, their incentive to contribute is difficult to study.

Finally, Wikipedia is widely regarded as a promising model for knowledge sharing. Many other online communities are starting to adopt similar models. Our results thus have important implications for future knowledge-sharing communities.

Our empirical analysis focuses on the third block which took place in October 2005 and lasted for nearly one year. Figure 1 shows the number of new contributors in Chinese Wikipedia over time. We find that before this block, the number of contributors grew exponentially over time. The number dropped significantly as a result of the block. Shortly after the block, the number continued to grow at an even faster rate, most likely due to the great deal of publicity Chinese Wikipedia received as a result of the block. We choose to focus on the third block as it is the longest block among the five blocks. In the other four instances, blocking and unblocking often happened within several days and their impact is limited. In addition, as we are interested in individual-level contribution, a short time period does not allow us to accurately measure the changes in contribution. Finally, the third block was widely publicized. In response to this block, the home page of Chinese Wikipedia added a link at the top directing any mainland China user to a status page on October 20, 2005. Unaffected contributors could easily learn about the block. We empirically investigate the changes in contribution of these unaffected contributors before October 19, 2005 and after October 31, 2005.

3.2 Data

We obtain our data set from the Chinese Wikipedia Web site (http://zh.wikipedia.org/). In our analysis, we analyze the full text of all Wikipedia articles and their complete editing histories. These articles were posted and edited between October 2002 and February 2007. We have 196,130 articles in our data set. Contributors are identified by their IDs if they have registered. Otherwise, they are identified by their network IP addresses at the time of connection. As the same IP address can map to multiple contributors and a contributor may not always use the same IP address when contributing, we focus on registered contributors in our analysis. Since administrators and robots have different editing patterns, we also exclude them from the analysis. For each article, we record the revision time, contributor ID, and number of characters[6] added and deleted in each revision. We then use this information to generate contribution history of each contributor in each week.

In addition to article pages on Wikipedia, we also obtain the contribution history of each contributor in user pages and user-talk pages. Each contributor can have his own user page or user-talk page. Many contributors add information about themselves such as contact information, photographs, and information about their areas of expertise and interest on their user pages or user-talk pages. Some contributors also use these pages to discuss their opinions about Wikipedia with others. Generally these user pages and user-talk pages, like Wikipedia articles, can also be modified by anyone.

3.3 Analysis

We first identify the contributors who are unaffected by the block. Wikipedia does not reveal geographic information about individual contributors. As a result, we rely on the contribution history of each contributor to infer whether she was affected. We consider a contributor as unaffected if she joins Chinese Wikipedia before the beginning of the block

[6]In Chinese, characters form the basic unit of meaning. Not all characters can stand alone as a word but most Chinese words are formed by two or three characters.

(i.e., October 19, 2005) and contributes at least once during the blocked period (i.e., between November 1, 2005 and October 10, 2006). 6042 contributors joined Wikipedia before the block and 1377 of them are classified as unaffected contributors. It is possible that our classification scheme excludes those unaffected contributors who made no contribution during the blocked period. Our results would be strengthened had these contributors been included.

Figure 2 shows the weekly total contribution measured by the total number of characters added and deleted for unaffected contributors over time. We find that for both measures the level of contribution drops after the block. The drop is more significant for the number of characters added.

We now test the change in contributions in a regression framework for each individual contributor. We employ the following specification:

$$Contributions = \beta_0 + \beta_1 AfterBlock + ControlVars + \epsilon \tag{4}$$

The dependent variable, *Contributions*, is the weekly contributions of each unaffected contributor to Wikipedia articles.

We use logarithms of the number of characters added and deleted as measures for individual contributions.[7] We consider addition and deletion as different forms of editing as the amount of effort involved may be different. $AfterBlock$ is a dummy which equals 1 if the time period is after the block and 0 otherwise. We also include age, which is measured as the number of weeks since the contributor joins Wikipedia, as a control variable, and the square of age to control for possible curvilinear effects.

Table 1 reports our regression results. We use the logarithm of the weekly total characters added, deleted and their sums as the dependent variables in Models 1, 2 and 3 respectively. We repeat the analysis controlling for individual fixed effects in Models 4, 5 and 6. All results suggest the same pattern: while contribution in general decreases at a decreasing rate with age, the block has significantly reduced incentive to contribute. In addition, the

[7]We add 1 to these measures before taking logarithms as some numbers can be zero.

number of characters added drops more than the number of characters deleted. A back-of-the-envelope calculation based on Model 3 suggests that weekly individual contribution dropped by approximately 42% as a result of the block (Halvorsen and Palmquist 1980; Kennedy 1981).

We then turn to examine the impact of social effects. We measure the amount of social interactions each unaffected contributor had had before the block using data on her participation in user pages and user-talk pages on Wikipedia. Active participation in these pages suggests that these contributors are more likely to desire a high level social recognition. Therefore, if social effects provide important motivation for contribution, we expect these socially active contributors to be affected more by the block. We consider the following differences-in-differences specification:

$$\begin{aligned} Contributions &= \beta_0 + \beta_1 AfterBlock + \beta_2 SociallyActive \times AfterBlock \\ &\quad + \beta_3 SociallyActive + ControlVars + \epsilon \end{aligned}$$

where $SociallyActive$ is the logarithm of the sum of total addition and total deletion in user pages or user-talk pages by each contributor before the block.

We report the results in Models 1, 2 and 3 of Table 2. The significant positive coefficients of $SociallyActive$ indicate that in general, contributors deriving more social utility contribute more. The negative coefficients of the interaction variable suggest that these contributors are affected significantly more by the block.

To ensure that the results are not driven by outliers, in Models 5, 6 and 7, we exclude contributors whose weekly contribution exceeds four standard deviations from the mean. Forty five contributors are excluded and we obtain similar results.

3.4 Robustness Checks

3.4.1 Seasonality

A natural concern is whether the results above are a consequence of time-specific effects. It could be that contributors contribute less in November than in September and early October in any year. We therefore replicate the analysis in Table 1, and the first three models in Table 2, for the same time period in years 2003 and 2004, and report the results in Table 3. We find that while socially active contributors contribute more during these periods, we detect no significant difference in contribution before October 19 and after October 31 in 2003 and 2004.

3.4.2 Controversial Articles

We are also concerned that the decrease in contribution may result from less disagreement in editing after the block. Contributors in mainland China may hold different political views from those in other regions such as Taiwan.[8] After the block, we would expect to see less disagreement in editing Wikipedia articles and hence less editing. We conduct two tests. First, we focus our analysis on new articles. If the decrease is caused entirely by less disagreement rather than social factors, we would expect no decline in the level of effort in creating new articles.[9] We count the number of characters in each new article and use the logarithm of weekly contribution in new articles by each contributor as the dependent variable. Table 4 reports the result. We find that after the block contributors create fewer new articles, and the decrease is mostly associated with socially active contributors.

In the second test, we examine category information for each article. When editing articles, a contributor can map these articles to a list of categories in the database. New categories can also be easily created. For example, an article about "auction theory" would

[8]Political sensitive articles are often edited more frequently. For example, the six most edited articles as reported by Wikipedia are Republic of China, China, People's Republic of China, Mao Zedong, Chiang Kai-shek, and Hong Kong.

[9]In fact, the level of effort could increase since contributors could spend more time creating articles after the block.

be mapped to "applied mathematics," "economics and finance," and "game theory." We compile a list of 31,871 categories from all articles, and manually go through the list.[10] In the end, we identify 2,500 categories as contentious categories. We then exclude contributions to articles in these categories and repeat the analysis. Table 5 reports the results. The results are similar.

3.4.3 Fewer New Articles for Revision

Our third concern is that the decrease is caused by fewer new articles created. As fewer new articles are created after the block, contributors may have less content to improve. We recompute individual contribution, this time only including the contribution to articles created before the time window (i.e., four weeks before October 19, 2005). We repeat the analysis and obtain similar results (Table 6).

3.4.4 Proxy Server

Finally, technically-adept contributors in mainland China might use proxy servers to circumvent the block. As proxy servers are slow, they might contribute less than before. Our classification scheme would classify them as unaffected contributors. As a result, we may observe a decrease in contribution. This concern is alleviated by a Wikipedia policy[11] that prevents contributors from editing using open and anonymous proxies. In addition, the Chinese Wikipedia administrators forbid open proxies in a way such that even registered contributors cannot use open proxies to edit articles. It is, however, still possible that not all open proxies are blocked. We conduct our robustness check using the encoding of each article and each edit. Like characters in many non-English languages, Chinese characters are encoded in two major standards in order to be correctly displayed. Due to historical reasons, Taiwan, Hong Kong, and Macau adopted Traditional Chinese characters (encoded

[10] The complete list of categories can be found at http://stats.wikimedia.org/EN/CategoryOverview_ZH_Complete.htm.

[11] See Wikipedia: No open proxies at http://en.wikipedia.org/wiki/Wikipedia:No_open_proxies for details.

in BIG5) while those from mainland China, Singapore, and Malaysia use Simplified Chinese characters (encoded in GB2312) to edit articles.[12] Therefore, those who use Traditional Chinese to edit articles are likely to be outside mainland China and thus are unaffected by the block. We analyze characters added by each contributor and consider those who added at least one Traditional Chinese character as unaffected. We repeat the analysis for these unaffected contributors and obtain similar results (Table 7).

4 Concluding Remarks

In this paper, we utilize an exogenous shock on the group size of Wikipedia participants to study how incentive to contribute reacts to a change in group size. In contrast to prior studies, we find that when the group size is exogenously reduced, contribution from the remaining population reduced by approximately 42%.

To explain this phenomenon, we provide a simple theoretical framework to demonstrate that the observation may be attributed to social effects. In various empirical specifications, we consistently find that socially more active contributors react to the exogenous shock more strongly than those who are less active.

Our results suggest that when people contribute to public goods, they care about not only the consequence of their contributions (provision/quality of the good), but also about the social recognition of their contributions. Awareness of the impact of social effects may help governments, non-profit organizations, and other public goods providers to introduce this important dimension into their planning and campaigns.

References

Andreoni, James, "Privately Provided Public Goods in a Large Economy: The Limits of Altru-

[12]See http://en.wikipedia.org/wiki/Big5 and http://en.wikipedia.org/wiki/GB2312, respectively.

ism," *Journal of Public Economics*, 1988, *35* (1), 57–73.

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

Becker, Gary S., "A Theory of Social Interactions," *Journal of Political Economy*, Nov.-Dec. 1974, *82* (6), 1063–1093.

Brunner, Eric J, "Free Riders or Easy Riders?: An Examination of the Voluntary Provision of Public Radio," *Public Choice*, December 1998, *97* (4), 587–604.

Carman, Katherine Grace, "Social Influences and the Private Provision of Public Goods: Evidence from Charitable Contributions in the Workplace," Dec. 2006. Working Paper, Tilburg University.

Chamberlin, John R., "The logic of collective action: Some experimental results," *Behavioral Science*, 1978, *23* (5), 441–445.

Fries, Timothy L., Edward Golding, and Richard Romano, "Private Provision of Public Goods and the Failure of the Neutrality Property in Large Finite Economies," *International Economic Review*, 1991, *32* (1), 147–157.

Goetze, Linda, T. F. Glover, and B. Biswas, "The Effects of Group Size and Income on Contributions to the Corporation for Public Broadcasting," *Public Choice*, 1993, *77* (2), 407–14.

Halvorsen, Robert and Raymond Palmquist, "The Interpretation of Dummy Variables in Semilogarithmic Equations," *American Economic Review*, June 1980, *70* (3), 474–75.

Hindriks, Jean and Romans Pancs, "Free Riding on Altruism and Group Size," *Journal of Public Economic Theory*, 2002, *4* (3), 335–46.

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*, 1–36.

Kennedy, Peter E, "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations," *American Economic Review*, September 1981, *71* (4), 801.

Lakhani, Karim R. and Eric von Hippel, "How Open Source Software Works: "Free" User-to-User Assistance," *Research Policy*, 2003, *32* (6), 923–943.

Lerner, Josh and Jean Tirole, "Some Simple Economics of Open Source," *Journal of Industrial Economics*, 2002, *50* (2), 197–234.

Marwell, Gerald and Ruth E. Ames, "Experiments on the Provision of Public Goods. I. Resources, Interest, Group Size, and the Free-Rider Problem," *American Journal of Sociology*, 1979, *84* (6), 1335–1360.

Nov, Oded, "What Motivates Wikipedians, or how to Increase User-generated Content Contribution," *Communications of the ACM*, forthcoming.

Olson, Mancur, *Logic of Collective Action: Public Goods and the Theory of Groups*, Harvard University Press, 1965.

Palfrey, Thomas R. and Howard Rosenthal, "Participation and the Provision of Discrete Public Goods: A Strategic Analysis," *Journal of Public Economics*, 1984, *24*, 171–193.

Rege, Mari, "Social Norms and Private Provision of Public Goods," *Journal of Public Economic Theory*, 2004, *6* (1), 65–77.

—— **and Kjetil Telle**, "The impact of social approval and framing on cooperation in public good situations," *Journal of Public Economics*, 2004, *88* (7-8), 1625–44.

Ribar, David and Mark O. Wilhelm, "Altruistic and Joy-of-Giving Motivations in Charitable Behavior," *Journal of Political Economy*, 2002, *110* (2), 425–457.

Sweeney, John W., "An Experimental Investigation of the Free-Rider Problem," *Economic Journal*, 1973, *2* (3), 277–292.

Figure 1. Number of New Contributors over Time

The vertical line indicates the start of the third block.

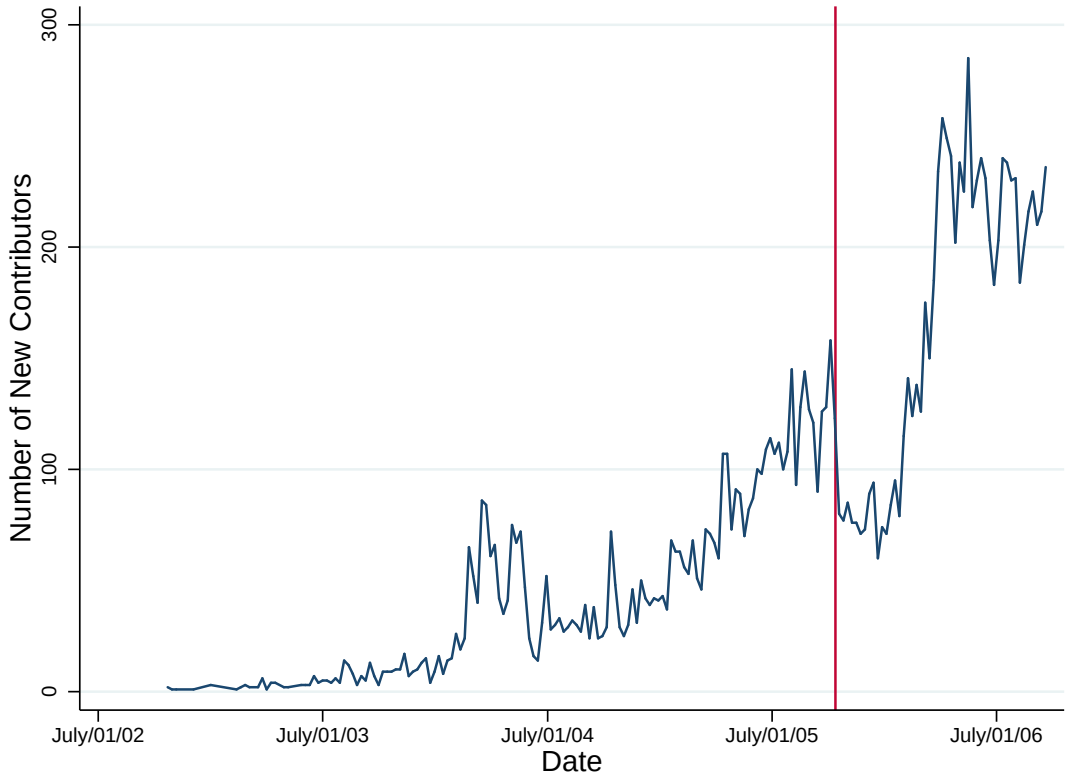

Figure 2. Weekly Contribution of Unaffected Contributors over Time

The vertical line indicates the start of the third block.

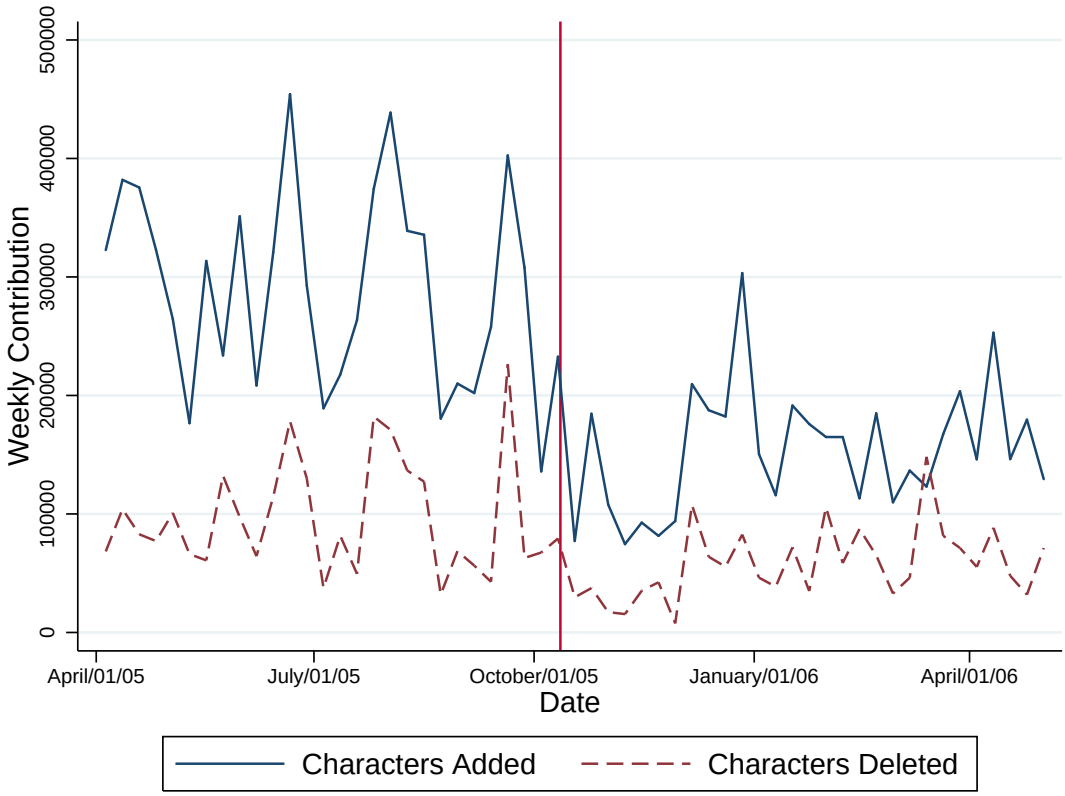

Table 1. Regression Results

Model	1	2	3	4	5	6
Dependent Variable	Addition	Deletion	Total	Addition	Deletion	Total
Age	−0.023***	−0.013***	−0.024***	−0.068***	−0.025*	−0.067***
	[0.003]	[0.003]	[0.003]	[0.018]	[0.014]	[0.018]
Age^2	0.000***	0.000***	0.000***	0.001***	0.000***	0.001***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
AfterBlock	−0.500***	−0.346***	−0.543***	−0.472***	−0.419***	−0.538***
	[0.052]	[0.039]	[0.054]	[0.100]	[0.077]	[0.104]
Observations	10657	10657	10657	10657	10657	10657
R^2	0.02	0.01	0.02	0.03	0.03	0.04
Number of ID				1377	1377	1377
Specification	OLS	OLS	OLS	FE	FE	FE

Table 2. Regression Results with Social Effects

Model	1	2	3	4	5	6
Dependent Variable	Addition	Deletion	Total	Addition	Deletion	Total
Age	−0.033***	−0.020***	−0.034***	−0.028***	−0.016***	−0.030***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.002]	[0.003]
Age^2	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
AfterBlock	−0.240***	−0.118***	−0.265***	−0.244***	−0.123***	−0.267***
	[0.049]	[0.036]	[0.052]	[0.046]	[0.032]	[0.049]
SociallyActive	0.484***	0.364***	0.515***	0.434***	0.323***	0.464***
	[0.020]	[0.017]	[0.021]	[0.020]	[0.017]	[0.021]
SociallyActive ×	−0.157***	−0.141***	−0.169***	−0.160***	−0.144***	−0.173***
AfterBlock	[0.027]	[0.022]	[0.028]	[0.026]	[0.022]	[0.027]
Observations	10657	10657	10657	10301	10301	10301
R^2	0.14	0.12	0.14	0.13	0.12	0.13
Specification	OLS	OLS	OLS	OLS	OLS	OLS

Table 3. Robustness Check: Previous Years

Model	1	2	3	4	5	6
Dependent Variable	Addition	Deletion	Total	Addition	Deletion	Total
Age	−0.074***	−0.048***	−0.078***	−0.186***	−0.089***	−0.192***
	[0.007]	[0.005]	[0.007]	[0.035]	[0.026]	[0.035]
Age^2	0.001***	0.001***	0.001***	0.004***	0.003***	0.005***
	[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
AfterBlock	0.029	0.027	0.03	−0.079	−0.018	−0.019
	[0.064]	[0.043]	[0.067]	[0.197]	[0.136]	[0.204]
SociallyActive	0.280***	0.165***	0.289***	0.479***	0.325***	0.509***
	[0.032]	[0.022]	[0.032]	[0.079]	[0.073]	[0.081]
SociallyActive ×	−0.052	0.004	−0.035	−0.13	−0.078	−0.155
AfterBlock	[0.042]	[0.032]	[0.044]	[0.115]	[0.101]	[0.119]
Observations	3658	3658	3658	565	565	565
R^2	0.13	0.11	0.13	0.29	0.28	0.29
Specification	OLS	OLS	OLS	OLS	OLS	OLS
Year	2004	2004	2004	2003	2003	2003

Table 4. Robustness Check: New Articles

Model	1	2	3
Dependent Variable	NewArticlesAddition	NewArticlesAddition	NewArticlesAddition
Age	−0.015***	−0.023*	−0.020***
	[0.003]	[0.013]	[0.003]
Age^2	0.000***	0.000***	0.000***
	[0.000]	[0.000]	[0.000]
AfterBlock	−0.148***	−0.192***	−0.023
	[0.035]	[0.074]	[0.032]
SociallyActive			0.226***
			[0.016]
SociallyActive × AfterBlock			−0.075***
			[0.021]
Observations	10657	10657	10657
R^2	0.01	0.01	0.07
Number of ID		1377	
Specification	OLS	FE	OLS

Table 5. Robustness Check: Non-contentious Categories

Model	1	2	3	4	5	6	7	8	9
Dependent Variable	Addition	Deletion	Total	Addition	Deletion	Total	Addition	Deletion	Total
Age	−0.019***	−0.011***	−0.020***	−0.062***	−0.02	−0.064***	−0.026***	−0.015***	−0.027***
	[0.003]	[0.002]	[0.003]	[0.019]	[0.015]	[0.020]	[0.003]	[0.002]	[0.003]
Age^2	0.000***	0.000***	0.000***	0.001***	0.001***	0.001***	0.000***	0.000***	0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
AfterBlock	−0.513***	−0.345***	−0.558***	−0.517***	−0.458***	−0.576***	−0.250***	−0.112***	−0.269***
	[0.054]	[0.041]	[0.057]	[0.109]	[0.083]	[0.114]	[0.053]	[0.038]	[0.055]
SociallyActive							0.432***	0.323***	0.463***
							[0.020]	[0.017]	[0.021]
SociallyActive × AfterBlock							−0.147***	−0.133***	−0.162***
							[0.027]	[0.022]	[0.028]
Observations	9063	9063	9063	9063	9063	9063	9063	9063	9063
R^2	0.02	0.02	0.02	0.04	0.03	0.04	0.13	0.12	0.14
Number of ID				1158	1158	1158			
Specification	OLS	OLS	OLS	FE	FE	FE	OLS	OLS	OLS

Table 6. Robustness Check: Old Articles

Model	1	2	3	4	5	6	7	8	9
Dependent Variable	Addition	Deletion	Total	Addition	Deletion	Total	Addition	Deletion	Total
Age	−0.016***	−0.008***	−0.017***	−0.072***	−0.028**	−0.070***	−0.025***	−0.015***	−0.027***
	[0.003]	[0.002]	[0.003]	[0.017]	[0.014]	[0.018]	[0.003]	[0.002]	[0.003]
Age^2	0.000***	0.000**	0.000***	0.001***	0.000***	0.001***	0.000***	0.000***	0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
AfterBlock	−0.519***	−0.375***	−0.565***	−0.408***	−0.380***	−0.476***	−0.220***	−0.111***	−0.244***
	[0.049]	[0.039]	[0.052]	[0.099]	[0.078]	[0.103]	[0.047]	[0.035]	[0.050]
SociallyActive							0.451***	0.350***	0.488***
							[0.019]	[0.017]	[0.020]
SociallyActive × AfterBlock							−0.181***	−0.163***	−0.195***
							[0.025]	[0.022]	[0.026]
Observations	9733	9733	9733	9733	9733	9733	9733	9733	9733
R^2	0.02	0.01	0.02	0.04	0.03	0.04	0.14	0.13	0.15
Number of ID				1158	1158	1158			
Specification	OLS	OLS	OLS	FE	FE	FE	OLS	OLS	OLS

Table 7. Robustness Check: Non-mainland Contributors

Model	1	2	3	4	5	6	7	8	9
Dependent Variable	Addition	Deletion	Total	Addition	Deletion	Total	Addition	Deletion	Total
Age	−0.031***	−0.017***	−0.032***	−0.050**	−0.018	−0.047**	−0.043***	−0.026***	−0.045***
	[0.003]	[0.002]	[0.003]	[0.021]	[0.016]	[0.022]	[0.003]	[0.002]	[0.003]
Age^2	0.000***	0.000***	0.000***	0.001***	0.001***	0.001***	0.000***	0.000***	0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
AfterBlock	−0.556***	−0.372***	−0.605***	−0.633***	−0.527***	−0.710***	−0.351***	−0.167***	−0.383***
	[0.060]	[0.046]	[0.062]	[0.118]	[0.091]	[0.122]	[0.060]	[0.044]	[0.063]
SociallyActive							0.446***	0.346***	0.476***
							[0.020]	[0.018]	[0.021]
SociallyActive × AfterBlock							−0.113***	−0.115***	−0.122***
							[0.028]	[0.023]	[0.029]
Observations	8865	8865	8865	8865	8865	8865	8865	8865	8865
R^2	0.03	0.02	0.03	0.03	0.03	0.04	0.14	0.12	0.14
Number of ID				1240	1240	1240			
Specification	OLS	OLS	OLS	FE	FE	FE	OLS	OLS	OLS