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

Xiaoquan (Michael) ZHANG, Feng ZHU
presented by Junjun QUAN, Xiaobo YU

December 8, 2021

Introduction

WIKIPEDIA

The Free Encyclopedia

Literature Review on Private Provision of Public Goods

Alternative Theories

1. **Crowding Out Hypothesis** Andreoni (1988), Fries et al. (1991), Palfrey and Rosenthal (1984), Hindriks and Pancs (2002) ...
2. **"Warm Glow"** Andreoni (1989), Konow (2006), Ribar and Wilheim (2002) ...
3. **Social Effects** Manski (1993), Andreoni and Scholz (1998), Becker (1974), van Dijk and van Winder (1997), Rege (2004), Lerner and Tirole (2002) ...

Literature Review on Private Provision of Public Goods

Alternative Theories

1. **Crowding Out Hypothesis** Andreoni (1988), Fries et al. (1991), Palfrey and Rosenthal (1984), Hindriks and Pancs (2002) ...
2. **"Warm Glow"** Andreoni (1989), Konow (2006), Ribar and Wilheim (2002) ...
3. **Social Effects** Manski (1993), Andreoni and Scholz (1998), Becker (1974), van Dijk and van Winder (1997), Rege (2004), Lerner and Tirole (2002) ...

Different levels of Empirical Studeis

1. **Experimental Data** Sweeney (1973), Chamberlain (1978), Marwell and Ames (1979) ...
2. **Field Studies** Goetze et al. (1993), Brunner (1998) ...

Institutional Background

Six unblocks of Wikipedia (up to 2008)

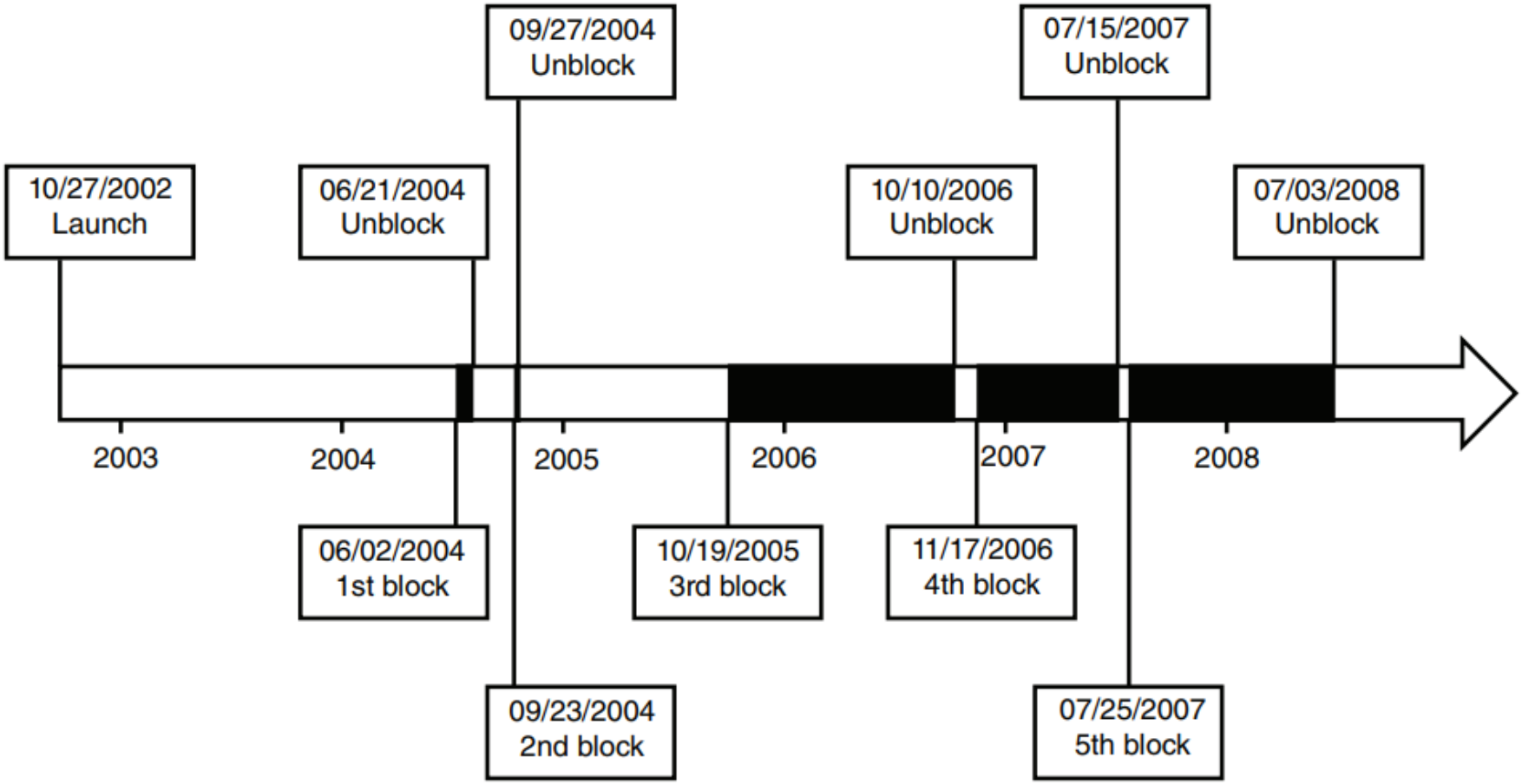


FIGURE 1. TIMELINE OF CHINESE WIKIPEDIA

Why Wikipedia and why the third block?

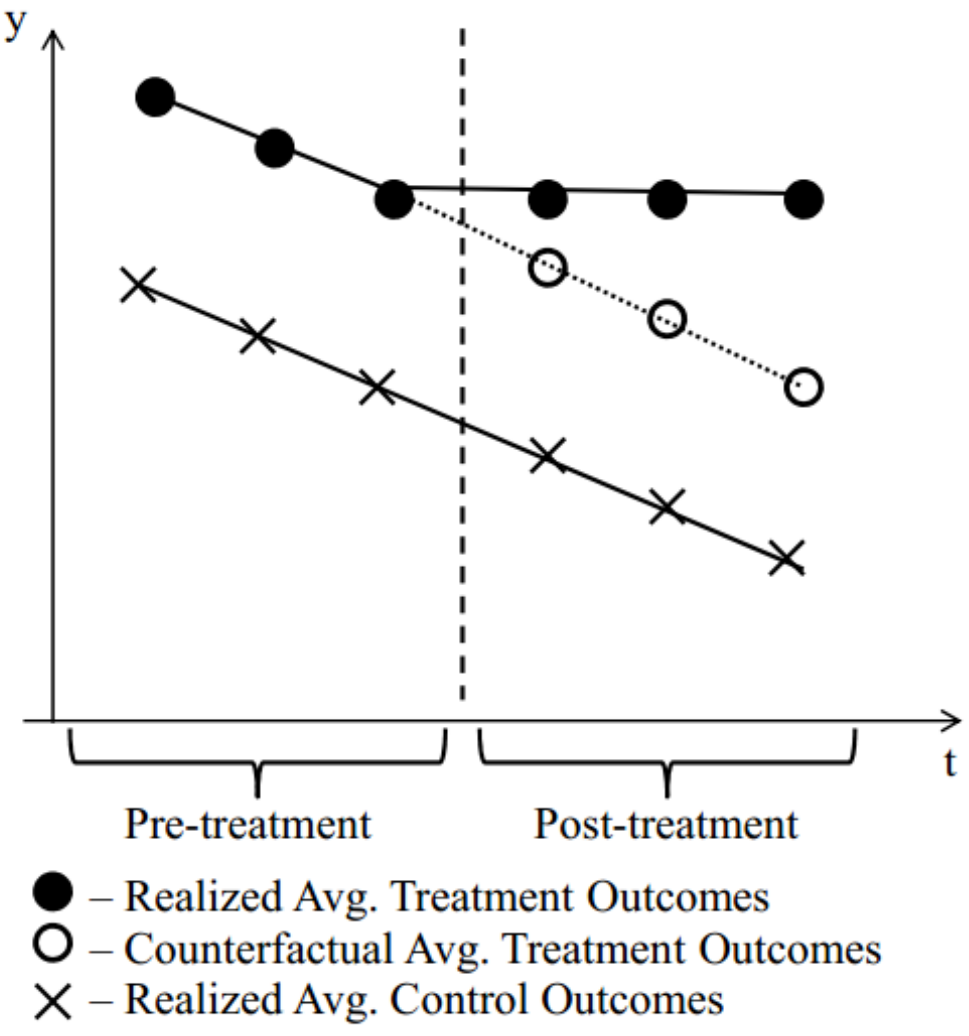
Why Wikipedia?

1. Easy to measure contribution
2. Easy to address exogeneity
3. No Monetary and Career Incentive Concern
4. Promising model for Knowledge Sharing

Why the third block?

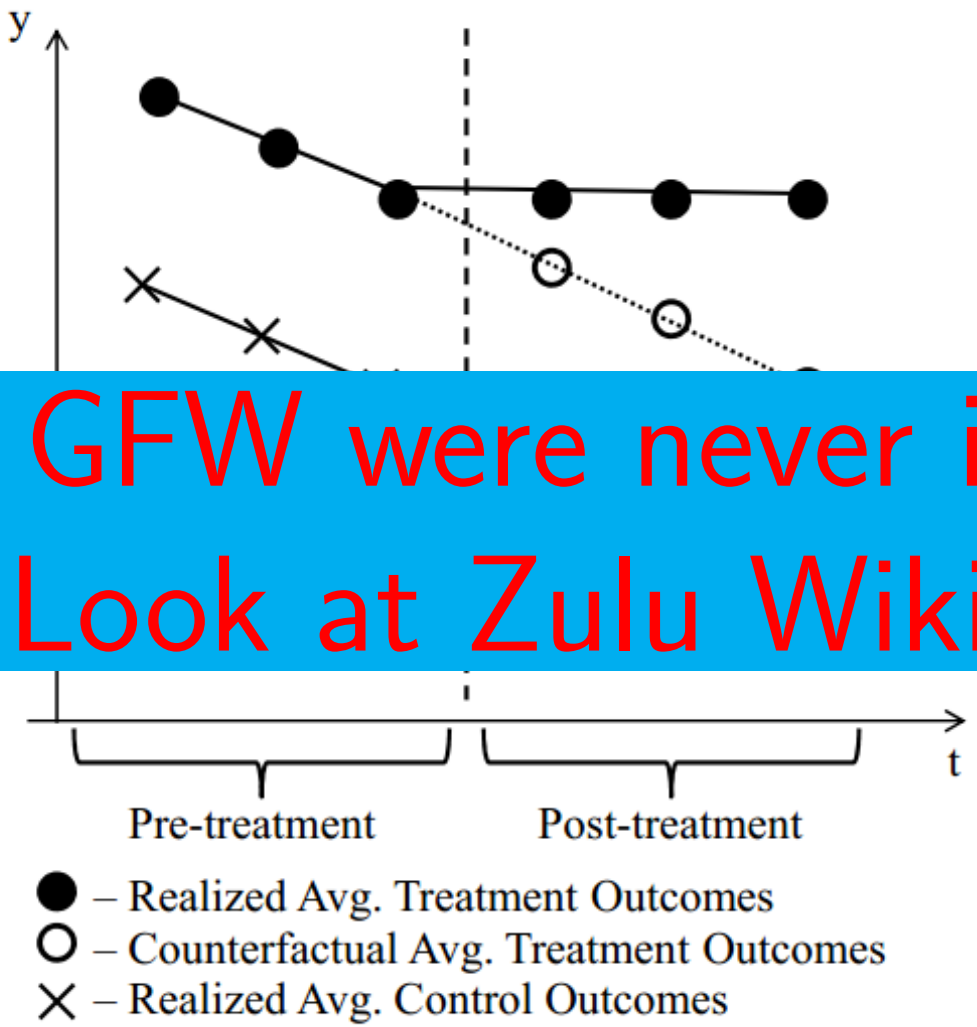
1. Longest among all
2. Publicization and Contributor awareness
3. Durable Goods and forward looking writers
4. Confounding Events (Olympics)

Methodology: Diff-in-Diff Recap



This figure is clipped from the lecture notes of Thierry Magnac.

Methodology: Diff-in-Diff Recap



What if GFW were never invented?
Look at Zulu Wiki!

This figure is clipped from the lecture notes of Thierry Magnac.

Data

1. Source: Chinese Wikipedia Website
2. Time Frame: four weeks before Oct. 19, 2005, and four weeks after Oct. 31, 2005.
    1. 9,048 new articles initiated; 53,519 revisions made
    2. 10,436,966 characters added; 4,321,112 characters deleted

Data

1. Source: Chinese Wikipedia Website
2. Time Frame: four weeks before Oct. 19, 2005, and four weeks after Oct. 31, 2005.
    1. 9,048 new articles initiated; 53,519 revisions made
    2. 10,436,966 characters added; 4,321,112 characters deleted
3. Sample selection: registered contributors
    1. exclude unregistered contributors, administrators, and robots

Data

1. Source: Chinese Wikipedia Website
2. Time Frame: four weeks before Oct. 19, 2005, and four weeks after Oct. 31, 2005.
    1. 9,048 new articles initiated; 53,519 revisions made
    2. 10,436,966 characters added; 4,321,112 characters deleted
3. Sample selection: registered contributors
    1. exclude unregistered contributors, administrators, and robots
4. Additional Data: editing history of user page and user-talk page
    1. user page: geographic, contact, photographs, expertise, and interest
    2. user-talk page: leave comments and communicate

Data

1. Source: Chinese Wikipedia Website
2. Time Frame: four weeks before Oct. 19, 2005, and four weeks after Oct. 31, 2005.
    1. 9,048 new articles initiated; 53,519 revisions made
    2. 10,436,966 characters added; 4,321,112 characters deleted
3. Sample selection: registered contributors
    1. exclude unregistered contributors, administrators, and robots
4. Additional Data: editing history of user page and user-talk page
    1. user page: geographic, contact, photographs, expertise, and interest
    2. user-talk page: leave comments and communicate
5. Identify Unblocked Contributors (1,707 in total)
    1. time criterion: joined before the beginning of the third block (6,062) and contributed at least once after the beginning (1,623)
    2. encoding criterion: more than 50 percent of additions in traditional Chinese (1,207) and 118 of them joined before the block

Summary Statistics

TABLE 1—SUMMARY STATISTICS FOR CONTRIBUTIONS BY INDIVIDUAL CONTRIBUTORS

	Pre-block		Post-block		Paired *t*-test
	Mean	Standard error	Mean	Standard error	*t*-stats
Panel A. Contributions from all contributors					
Addition	1,105.53	225.67	538.01	117.00	2.89***
Deletion	528.96	167.25	174.57	38.21	2.32**
Total	1,634.49	389.77	712.58	153.13	2.69***
Panel B. Contributions from nonblocked contributors					
Addition	3,599.79	797.50	1,910.62	413.73	2.43**
Deletion	1,806.45	592.76	619.95	135.13	2.19**
Total	5,406.24	1,379.49	2,530.57	541.45	2.37**

Notes: We examine the contributions made by unique contributors before and after the block. Contributions are measured by the number of characters they added and deleted. We report the contributions made during the four weeks before October 19, 2005, in the Pre-Block column, and the contributions made during the four weeks after October 31, 2005, in the Post-Block column. In the last column, we report results from paired *t*-tests.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

Baseline Analysis

1. The change in contributions of each individual contributor

$$Contributions_{it} = \beta_0 + \beta_1 AfterBlock_t + ControlVars_{it} + \epsilon_{it} \quad (1)$$

   i indexes the contributors and t indexes weeks.
2. Contributions are measured by logarithms of weekly total characters added, total characters deleted, and their sums.
3. Control variables include account age and age square

Baseline Analysis

TABLE 2—DETECTING THE CHANGE IN INDIVIDUAL CONTRIBUTION LEVELS AFTER THE BLOCK

Dependent variable	Total	Addition	Delection	Total	Addition	Deletion
Model	(1)	(2)	(3)	(4)	(5)	(6)
AfterBlock	−0.463***	−0.430***	−0.293***	−0.555***	−0.498***	−0.403***
	[0.044]	[0.042]	[0.032]	[0.082]	[0.079]	[0.060]
Age	−0.025***	−0.024***	−0.013***	−0.062***	−0.063***	−0.026**
	[0.002]	[0.002]	[0.002]	[0.016]	[0.015]	[0.011]
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]
Observations	13,376	13,376	13,376	13,376	13,376	13,376
R^2	0.03	0.03	0.02	0.04	0.04	0.03
Specification	OLS	OLS	OLS	FE	FE	FE

Notes: Addition(Deletion) refers to the logarithm of the number of characters added to (deleted from) the articles. Total is the logarithm of the sum of Addition and Deletion. *Age* is the number of weeks since the contributor joins Wikipedia. *AfterBlock* is a dummy variable that takes the value one if the time period is after the block and zero otherwise. Heteroskedasticity-adjusted standard errors in brackets.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

Impact of Social Effects

1. Contributors who derive more social benefits from contributing are affected more after the block
2. Contributors' Participation in User Pages and User-Talk Pages

$$\begin{aligned} Contributions_{it} &= \beta_0 + \beta_1 AfterBlock_t + \beta_2 SocialParticipation_i \\ &+ \beta_3 SocialParticipation_i \times AfterBlock_t + ControlVars_{it} + \epsilon_{it} \end{aligned} \tag{2}$$

3. Collaboration Network Affected by the Block

$$\begin{aligned} Contributions_{it} &= \beta_0 + \beta_1 AfterBlock_t + \beta_2 PercentageBlocked_i \\ &+ \beta_3 PercentageBlocked_i \times AfterBlock_t + ControlVars_{it} + \epsilon_{it} \end{aligned} \tag{3}$$

Impact of Social Effects

TABLE 3—DIFFERENCE-IN-DIFFERENCES ESTIMATIONS OF THE IMPACT OF THE BLOCK ON CONTRIBUTORS WITH DIFFERENT LEVELS OF SOCIAL PARTICIPATION

Dependent variable	Total	Addition	Deletion	Total	Addition	Deletion
Model	(1)	(2)	(3)	(4)	(5)	(6)
AfterBlock	−0.105**	−0.091**	−0.013	−0.212**	−0.170**	−0.131**
	[0.041]	[0.039]	[0.028]	[0.086]	[0.082]	[0.062]
SocialParticipation × *AfterBlock*	−0.195***	−0.186***	−0.153***	−0.190***	−0.182***	−0.150***
	[0.024]	[0.023]	[0.019]	[0.024]	[0.023]	[0.020]
SocialParticipation	0.539***	0.510***	0.372***			
	[0.018]	[0.018]	[0.015]			
Age	−0.023***	−0.022***	−0.012***	−0.062***	−0.062***	−0.026**
	[0.002]	[0.002]	[0.002]	[0.015]	[0.015]	[0.011]
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]
Observations	13,376	13,376	13,376	13,376	13,376	13,376
R^2	0.18	0.17	0.15	0.06	0.05	0.05
Specification	OLS	OLS	OLS	FE	FE	FE

Notes: SocialParticipation is the logarithm of the weekly average of total addition and total deletion in user pages or user-talk pages by each contributor before the block. The variable *SocialParticipation* drops in the fixed-effect specifications as its value is fixed for each contributor. Heteroskedasticity-adjusted standard errors in brackets.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

Impact of Social Effects

TABLE 4—DIFFERENCE-IN-DIFFERENCES ESTIMATIONS OF THE IMPACT OF THE BLOCK ON CONTRIBUTORS WITH DIFFERENT PERCENTAGES OF COLLABORATORS BLOCKED

Dependent variable	Total	Addition	Deletion	Total	Addition	Deletion
Model	(1)	(2)	(3)	(4)	(5)	(6)
AfterBlock	−0.131*	−0.125*	−0.011	−0.163	−0.134	−0.076
	[0.068]	[0.064]	[0.048]	[0.100]	[0.096]	[0.074]
PercentageBlocked × *AfterBlock*	−1.717***	−1.568***	−1.524***	−2.506***	−2.327***	−2.090***
	[0.448]	[0.425]	[0.336]	[0.398]	[0.386]	[0.335]
PercentageBlocked	7.552***	7.058***	5.253***			
	[0.359]	[0.341]	[0.281]			
Age	−0.034***	−0.032***	−0.019***	−0.079***	−0.078***	−0.040***
	[0.002]	[0.002]	[0.002]	[0.016]	[0.015]	[0.012]
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]
Observations	13,376	13,376	13,376	13,376	13,376	13,376
R^2	0.11	0.11	0.09	0.05	0.04	0.04
Specification	OLS	OLS	OLS	FE	FE	FE

Notes: *PercentageBlocked* is the percentage of collaborators blocked after the third block for each contributor. The variable *PercentageBlocked* drops in the fixed-effect specifications as its value is fixed for each contributor. Heteroskedasticity-adjusted standard errors in brackets.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Robustness Checks

1. Seasonal Effects: placebo regression during the same eight-week period in 2003 and 2004

Robustness Checks

1. Seasonal Effects: placebo regression during the same eight-week period in 2003 and 2004
2. Less Disagreement in Editing: mainland China vs other regions
    1. creation of new articles
    2. exclude potentially contentious categories

Robustness Checks

1. Seasonal Effects: placebo regression during the same eight-week period in 2003 and 2004
2. Less Disagreement in Editing: mainland China vs other regions
    1. creation of new articles
    2. exclude potentially contentious categories
3. Self-Serving Edits
    1. promoting a firm or a product, advocating own views, etc.
    2. Wikipedia disallows self-serving edits
    3. method: exclude contributors who have been banned (37 in 1,707)

Robustness Checks

1. Seasonal Effects: placebo regression during the same eight-week period in 2003 and 2004
2. Less Disagreement in Editing: mainland China vs other regions
    1. creation of new articles
    2. exclude potentially contentious categories
3. Self-Serving Edits
    1. promoting a firm or a product, advocating own views, etc.
    2. Wikipedia disallows self-serving edits
    3. method: exclude contributors who have been banned (37 in 1,707)
4. Less Need for Revision: placebo regression on articles created before the block

Robustness Checks

1. Seasonal Effects: placebo regression during the same eight-week period in 2003 and 2004
2. Less Disagreement in Editing: mainland China vs other regions
    1. creation of new articles
    2. exclude potentially contentious categories
3. Self-Serving Edits
    1. promoting a firm or a product, advocating own views, etc.
    2. Wikipedia disallows self-serving edits
    3. method: exclude contributors who have been banned (37 in 1,707)
4. Less Need for Revision: placebo regression on articles created before the block
5. Proxy Server: Wikipedia policy prevents using open and anonymous proxies while editing (closed proxy servers could work by channeling their traffic through computers outside mainland China)
    1. exclude dedicated and technically adept Wikipedia contributors

Conclusions

1. Major Findings
    1. Positive relationship between group size and contributors' contribution level
    2. Contributors who are likely to care more about social benefits react to the change more strongly than those who value them less
2. Link to Existing Evidence and Literature
    1. helps explain the existence of many public goods with a large number of contributors
    2. helps explain the empirical evidence that many contributors prefer to contribute to large online communities

Limitations

1. External Validity
   1. only one particular kind of public goods: contributors are at the same time users (open source software, YouTube, etc.)
   2. online public good
2. Motivation behind Social Effects
   1. broad definitions of social effects
   2. different social motivations could lead to similar behavioral patterns
3. Treatment Effect Heterogeneity
   1. allow treatment effect to depend on group size
   2. random effects