Motivating Contributions to Public Information Goods: A Field Experiment at Wikipedia - A Pre-Analysis Plan*

Yan Chen, Rosta Farzan, Robert Kraut, Iman YeckehZaare and Ark Fangzhou Zhang

April 20, 2018

Abstract

Wikipedia is among the most important information sources for the general public. Motivating domain experts to contribute to Wikipedia can improve the accuracy and completeness of its content. In a field experiment at Wikipedia, we examine individual motivations to contribute to public information goods. Using a 2-by-3 factorial design, we vary the expectation on the number of recipients along one dimension and the amount of private benefit along the other dimension. In the analysis, we will investigate how our interventions affect the experts' willingness to participate and contribution measured by both quantity and quality.

*Yan Chen: School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109 (e-mail: yanchen@umich.edu); Rosta Farzan: School of Information Sciences, University of Pittsburgh, 135 North Bellefield Avenue, Pittsburgh, PA 15260 (e-mail: rfarzan@pitt.edu); Robert Kraut: School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 (e-mail: robert.kraut@cmu.edu); Iman YeckehZaare: School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109 (e-mail: oneweb@umich.edu); Ark Zhang: School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109 (e-mail: arkzhang@umich.edu). We are grateful for the helpful comments from Raj Chetty, David Cooper, Matt Van Essen, Andrei Gomberg, Aaron Halfaker, Muriel Niederle, Axel Ockenfels, Paul Resnick, Tanya Rosenblat, Alvin Roth, Karl Schlag, Jean-Robert Tyran, David Yang, and seminar participants at the Centro de Investigación y Docencia Económicas (CIDE), Cologne, Facebook Core Data Science, ITAM, Lyon, Magdeburg, Michigan, Stanford and Vienna. We are grateful to Paul Hur for excellent research assistance. Research support from the National Science Foundation through grant number SES-1620319 to Chen and Kraut is gratefully acknowledged. The research was approved by the University of Michigan IRB (HUM00090577).

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.

These peer-produced public goods enabled by information technology, which we call *public information goods*, have distinct characteristics. Unlike textbook examples of pure public goods, such as national defense, where exclusion is technically difficult, public information goods are technically easy to exclude by requiring authentication. However, they are provided to community members or 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, an experimental economist working on coordination games might find it less costly to contribute a Wikipedia article on coordination games than an article on the business cycle. Because of the expertise she has developed over the years, her contribution quality on coordination games will be higher than it would be in a less well matched area. She might be motivated to contribute to this article as she cares about this subject matter being introduced correctly to the general public.

In this paper, we investigate individual motivations to contribute to public information goods. 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 contributions, such as being cited or publicly acknowledged. In addition, we investigate the effect of matching accuracy between the recommended public information goods and contributors' expertise, as well as the social distance between the contributor and the research team.

Research on charitable giving has identified two factors important for contribution, ask; and who does the asking. We add a third factor, what you ask people to do.

We conduct our study in the context of the English Wikipedia, the English language version of the online encyclopedia. The English Wikipedia was founded in the January of 2001 and is operated by the Wikimedia Foundation. It is one of the most popular sources for information goods about scientific entries (Thompson and Hanley, 2017). According to Alexa Internet, it ranks among the top five most popular websites globally, with over 262 million daily visits.[1] As of the February of 2018, the English Wikipedia has accumulated over 5.5 million articles, with over 3.1 billion words in all content pages. Aiming at becoming a free online encyclopedia, Wikipedia grants all Internet users open access to it. The defining feature of non-excludability distinguishes Wikipedia from the traditional repositories of scientific knowledge such as academic journals that requires subscription fees or institutional access.

We design our field experiment to examine how individual contribution behavior is affect by the potential number of recipients of the public goods and the potential private benefit. The experimental manipulation is implemented through variations in both the content of invitation e-mail and the web interface. We randomly vary the amount of social impact and private benefit one expect to enjoy from her contribution using a 2-by-3 factorial design. Along one dimension, we provide the experts in the HighView condition with the information on the number of recipients of the public information goods, and the experts in the Avgview condition with only the information on the average readership across the entire English Wikipedia. Along the other dimension, we vary the expects' expectations on the amount of private benefit related to citation of their research work and acknowledgement of their contributions.

2 A Theoretical Framework

In this section, we outline a simple theoretical framework for contribution 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, $G \geq 0$. To simplify notation, we use a single public information good in the main text and generalize it to multiple public goods in the appendix. The set of potential contributors is I. In our context, we assume that the public good has an existing quantity and quality, represented by

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

$g_0 > 0$. The number of consumers of this public good is $n \geq 0$. Each agent, $i \in I$, selects a contribution level, $g_i \in [0, +\infty)$. Contribution leads to an improvement in both the quantity and the quality of the public good, $G = g_0 + \sum_{j \in I} g_j$. A contributor derives utility from the social impact of the public good, which is the product of G 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 the contributor's utility function is $v(n)G$, 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 editors who are not blocked (Zhang and Zhu, 2011).

The second type of benefit derived from contributing to a public information good 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. We use a specification that is general enough to encompass various types of private benefit. We capture the marginal private benefit by $w(n)$, where $w' \geq 0$. Thus, the private benefit of contribution is captured by $w(n)y_i$.

The charitable giving literature identifies two main reasons why people donate to charity. The first is that they have been asked. The second is that they have been asked by someone they care about (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 (Benabou and Tirole, 2007), or social pressure (DellaVigna et al., 2012). We capture this type of motives by a function, $s(g_i)$, with a positive and increasing function of g_i if contributor i cares about the askers, and zero otherwise.

A contributor's cost of contribution has two components. Choosing g_i entails a cost, $c(g_i)$, which is assumed to be convex in g_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 $r(T_i - g_i)$.

Crucially for public information goods, we can use information technology, such as a recommender system, to infer the expertise of a potential contributor, and recommend work which matches her expertise. Let $m_i \in [0, 1]$ be the matching quality between an agent'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 area of expertise. Second, matching an expert to work in her domain of expertise is a match in 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.

We consider a two-stage process, in a similar spirit as DellaVigna et al. (2012). In the first stage, we elicit the interests of the expert by manipulating the number of views (n) and the private benefit ($w(n)$).

Stage 1. Eliciting interests. In the first stage, the designer asks a potential contributor whether she would be interested to contribute to a public information good in her area of expertise, which implies an anticipated matching quality, $\bar{m}$. The contributor makes a participation decision by solving the following problem:

$$\max_{p_i \in \{0,1\}} v(n)(\bar{y}_0 + p_i\bar{y}_i + \sum \bar{y}_{-i}) + w(n)(p_i\bar{y}_i) + s(p_i\bar{y}_i) + r_i(T_i - p_i\bar{y}_i) - \frac{c(p_i\bar{y}_i)}{\bar{m}}. \quad (1)$$

Therefore, an agent will decide to participate ($p_i = 1$) if the following condition is satisfied:

$$[v(n) + w(n) + s_i - r_i]\bar{y}_i + s(\bar{y}_i) \geq \frac{c(\bar{y}_i)}{\bar{m}}. \quad (2)$$

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

1. *the public information good is consumed by more people,* $\frac{\partial p_i}{\partial n} > 0$*; or*
2. *the private benefit of contribution is higher,* $\frac{\partial p_i}{\partial w} > 0$*; or*
3. *her expertise overlaps with the askers,* $\frac{\partial p_i}{\partial s_i} > 0$*; or*
4. *her opportunity cost of time is lower,* $\frac{\partial p_i}{\partial r_i} > 0$*.*

Stage 2. Contribution decision. In the second stage, expert i decides whether and how much to contribute to the public information good, after seeing the recommended work, i.e., the quality of the match, m_i. The extent that the recommended work matches her expertise is characterized by m_i, which determines the contribution cost, $c(y_i)/m_i$. Therefore, the closer the match is, the lower the contribution cost will be. Agent i solves the following optimization problem:

$$\max_{y_i \in [0,\infty)} v(n)(y_0 + y_i + \sum y_{-i}) + w(n)(y_i) + s_i(y_i) + r_i(T - y_i) - \frac{c(y_i)}{m_i}. \quad (3)$$

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., $c'(0) < [v(n) + w(n) + s_i - r_i]m_i$, the unique interior solution is characterized by the following first-order condition:

$$c_i'(y_i) = [v(n)y + w(n) + s_i - r_i]m_i. \quad (4)$$

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

$$y_i^* = k((v(n) + w(n) + s_i - r_i)m_i). \tag{5}$$

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

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.

Proposition 2 (Contribution)**.** *After an agent agrees to participate, the following comparative statics hold:*

1. *An increase in the number of consumers of the public information good leads to an increased level of contribution, i.e.,* $\frac{\partial y^*}{\partial n} = g'[v'(n) + w'(n)]m_i > 0$.
2. *An increase in the private benefit of contributions leads to an increased level of contributions, i.e.,* $\frac{\partial y^*}{\partial w} = k'm_i > 0$.
3. *An expert with a higher reputation will contribute less, i.e.,* $\frac{\partial y_i^*}{\partial r_i} = -k'm_i < 0$. *This is because the opportunity cost of contribution is higher.*
4. *Better matching between the content of the public information good and the agent's expertise leads to an increased level of contributions, i.e.,* $\frac{\partial y_i^*}{\partial m_i} = k'[v(n) + w(n) + s_i - r_i] \geq 0$ *if and only if* $v(n)y + w(n) + s_i \geq r_i$.

We now examine the welfare implications of having a better matching technology. An omniscient social planner uses a Samuelsonian social welfare function which considers both consumers and producers of the public information good, and where everyone has equal weight, $W = \sum_{i \in C} v_i + \sum_{i \in E} u_i$. We obtain the following result.

Proposition 3 (Technology and welfare)**.** *While total equilibrium contribution to the public information good is lower than the social optimum, an improvement in matching quality increases equilibrium contributions,* $\dfrac{\partial g_i^*}{\partial m_i} > 0$, *and total welfare,* $\dfrac{\partial W}{\partial m_i} > 0$, $\forall\, i$.

All proofs are relegated to Appendix A.

3 Experimental Design

3.1 Expert and article selection: recommender systems

The experts whom we invite to contribute are the academic scholars registered on *Research Papers in Economics* (*RePEc*).[2] *RePEc* is a public repository of working papers and journal articles in the field of economics. It maintains a profile for each registered economist, including the 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 and research works. For each expert, we identify her most recent field of expertise according to the area on which her recent research works are mostly focused.

The initial sample consists of 31,670 economists who maintain a research profile at *RePEc*. From this sample, there are 13,261 economists with either email address or research specialization not provided. We exclude these experts and reach a sample of 18,409 economists. To guarantee the accuracy of recommendation, we further restrict the experiment to the 3,974 experts with at least five research papers on *RePEc*.

The Wikipedia articles recommended to an expert are selected according to how relevant they are 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 the articles that is either classified as a "Stub" (see footnote 3) or less than 1,500 characters long. We further keep the ones which are viewed for at least 1,000 times in the past 30 days. The main purpose of this restriction is to control the potential difference in the actual readership across our experimental conditions.

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 top 10% at *RePEc* and the affiliated institution. *RePEc* assigns an index for each expert based on her number of publications and citations, and list the top experts who rank 10% among all. For each Wikipedia article, our dataset includes the quality class 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.[3]

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

[3] The quality scale at Wikipedia contains seven classes: *Stub*, *Start*, *C*, *B*, *A*, *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 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

3.2 Power calculation

We perform the power calculation based on the results from the pilot study. The pilot study was conducted between November 2015 and January 2016 and used a 2-by-2 factorial design (CiteAckn condition was not included).

Suppose that there are two treatments X and Y. There are n_x experts exposed to treatment X and n_y experts exposed to treatment Y. Each subject has a binary outcome, a positive response and a negative response, denoted by 0 and 1, respectively. The probability that an expert responds positively is p for treatment X and q for treatment Y. Let x and y denote the number of experts responding postively in each of the two treatments. Therefore, we have

$$x \sim Bi(n_x, p), \quad y \sim Bi(n_y, q)$$

We want to test whether there is a treatment effect on the probability of a positive response. The main null hypothesis is

$$H_0 : p = q, \quad H_1 : p > q$$

Because the joint distribution of (x, y) belongs to an exponential family, the uniformly most powerful unbiased test with level α takes the form of rejecting H_0 if

$$x > c_\alpha(x + y),$$

where $c(\cdot)$ is a function satisfying that

$$P_{H_0}[y > c(u)|x + y = u] \leq \alpha.$$

Under H_0, it can be shown that $P_{H_0}[y|x + y = u]$ follows a hypergeometric distribution $HG(u, n_y, n_x)$. Given that p and q is the true parameter that generates x and y, we want to determine n_x and n_y such that the test above gives power $1 - \alpha$. The probability of rejection H_0 given p and q (i.e., power) is:

$$\mathbb{E}[\mathbb{P}_{(p,q)}[y > c(u)|x + y = u]]$$

Practically, we would like to reject the null hypothesis of no treatment with level $\alpha = 0.05$ and power $1 - \beta = 0.9$. We also assume that $n_x = n_y$. The empirical $\hat{p}$ and $\hat{q}$ that we will use come from the pilot results:

We calculate the sample size required for testing the treatment effect on positive response for each factor. For example, given that $p = 0.412$ and $q = 0.490$ (or given NoCite condition, to test HighView), we need: $n_x = n_y = 636$. The following table provide the sample size needed to detect a 10 percentage point difference for different combinations of p and q.

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 1: Positive Response Rate from the Pilot Study

	No Citation	Citation
Average View	0.412 $(N = 51)$	0.389 $(N = 18)$
High View	0.490 $(N = 49)$	0.500 $(N = 24)$

Table 2: Sample Size for Various Combinations of p and q

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.2	232								
0.3	74	338							
0.4	39	97	408						
0.5	24	47	111	445					
0.6	17	30	53	116	445				
0.7	12	18	31	53	111	408			
0.8	10	12	18	30	47	97	338		
0.9	9	10	12	17	25	39	74	232	
1.0	5	6	8	9	12	15	25	38	78

3.3 Treatments

We implement a 2-by-3 between-experts factorial design in which we vary the email content inviting experts' contribution to Wikipedia articles (see Table 3). Along one dimension, we vary the experts' expectation on the number of recipients of the public information goods which they contribute to. In the average view condition, we provide the experts with only the average number of views a typical Wikipedia article received in the past 30 days. This information serves to set the experts' expectation on the readership of a typical Wikipedia article. In the high view condition, we provide the experts with the additional information on the number of views the recommended articles received in the past 30 days.

Along the second dimension, we vary the experts' expectation on the amount of private benefit they receive from their contribution. We include three conditions: a baseline condition, a citation condition and an acknowledgement condition. The baseline condition serves as a control and no private benefit is mentioned in the email. In the citation condition, we mention that the articles recommended to the experts are likely to cite their research. The acknowledgement condition strengthens the private benefit by including acknowledgement as an additional benefit. The

Table 3: Features of Experimental Conditions

	No Citation	Citation	Citation & Acknowledge
Average View	AvgView-NoCite ($N = 678$)	AvgView-Cite ($N = 669$)	AvgView-CiteAckn ($N = 671$)
High View	HighView-NoCite ($N = 637$)	HighView-Cite ($N = 661$)	HighView-CiteAckn ($N = 658$)

experts are told in the email message that their contributions will be addressed on a WikiProject Economics page at Wikipedia (see Figure 3).[4] WikiProject Economics is a collection of editors who work together as a team to improve articles related to economics. Being acknowledged for one's contribution in the WikiProject Economics thus serves as a private benefit in additional to the citation benefit. To avoid potential confound due to the experts' sequential contribution, we only post the acknowledgement to the contributions from the experts in our pilot stage and keep the acknowledgement page frozen at the through the main experiment.

Table 4 reports the summary statistics for the pre-treatment characteristics, broken down into the six experimental conditions. Panel A presents the characteristics of the experts and panel B gives the characteristics of the Wikipedia articles recommended to the experts. Columns (1) through (6) report average values as well as standard errors. 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, the articles which are recommended in the HighView-NoCite condition are longer and assigned higher quality class, compared to those in the other conditions.

3.4 Experimental Procedure

Our experiment consists of two stage. In the first stage, we send an initial email inquiring whether the 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 the method described above. The content starts with a brief introduction of Wikipedia, including the average readership of a typical article. Depending on the experimental condition, the second paragraph provides information regarding the readership of the articles to be recommended to the expert and the private benefits she expects to receive. To avoid the misinter-

[4]See https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Economics/ExpertIdeas.

Table 4: Characteristics of Experts and Recommended Articles, by Conditions[a]

	AvgView NoCite (1)	AvgView Cite (2)	AvgView CiteAckn (3)	HighView NoCite (4)	HighView Cite (5)	HighView CiteAckn (6)	p-value (7)
Panel A: Characteristics of Experts							
Abstract Views	1,610	1,633	1,764	1,697	1,810	1,644	0.493
	(68)	(73)	(103)	(84)	(104)	(69)	
Top 10%	0.360	0.378	0.358	0.347	0.371	0.386	0.712
	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	
English Affiliation	0.417	0.457	0.434	0.452	0.477	0.407	0.103
	(0.019)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)	
Observations	678	669	672	636	660	658	
Panel B: Characteristics of Recommended Articles							
Article Length	34,266	33,973	34,579	36,269	35,000	34,150	0.044
	(536)	(533)	(553)	(599)	(567)	(546)	
Number of Edits	725	725	708	754	750	712	0.273
	(16)	(17)	(16)	(18)	(18)	(17)	
Views in Past Month	14,409	14,023	14,013	14,348	14,471	13,934	0.732
	(273)	(319)	(322)	(298)	(325)	(348)	
Quality:							
Featured Article	0.054	0.050	0.046	0.058	0.047	0.048	0.095
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	
Good Article	0.216	0.211	0.215	0.226	0.205	0.201	0.120
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	
B	0.594	0.604	0.601	0.581	0.613	0.613	0.037
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
C	0.127	0.125	0.126	0.123	0.122	0.127	0.978
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	
Start & Stub	0.009	0.010	0.011	0.012	0.013	0.011	0.582
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
Importance:							
Top	0.168	0.160	0.158	0.173	0.152	0.153	0.077
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	
High	0.350	0.339	0.353	0.347	0.358	0.348	0.630
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Mid	0.255	0.270	0.256	0.245	0.264	0.263	0.192
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	
Low	0.064	0.073	0.070	0.067	0.067	0.071	0.664
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	
Observations	3,924	3,872	3,845	3,693	3,779	3,794	

Note. Columns (1) through (6) report average values and 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.

pretation of particular phrases, we randomly select one of the three following ways to deliver the HighView condition: "*especially popular*", "*highly visible*", "*highly popular*". Similarly, one of the following three messages is randomly chosen to deliver the Cite and the CiteAckn condition: "*may include some of your publications in their references*", "*might refer to some of your research*", or "*are likely to cite your research*". 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.*"

The experts who respond positively (i.e., clicking "Yes") to the first stage email are then sent a second email automatically. This email starts with a thank-you message addressing the expert's willingness to contribute. It then presents a table listing the articles recommended to the expert. If the expert is assigned into the high view condition, the table also shows the 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 2 in appendix C). The webpage consists of a mirror image of the article on the right 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 the expert can leave her comment on the article. We place display the mirror image of the article and the text box are placed side by side so that the experts can input their comments without switching between browser pages. After the expert submit her comment, a thank-you email is automatically sent to her and the comments are posted on the talk pages associated with the corresponding Wikipedia articles.

The emails are sent between 6:00 AM and 7:00 PM on weekdays based on the local time of the expert's affiliation. To avoid the emails being filtered as spam, we send up to 10 emails every four hours. Throughout the field experiment, we use a tracking tool to monitor whether the emails sent to the 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 the invitation in any phase, no more emails will be sent to her.

4 Analsis Plan

The analysis will start with the impact of the experimental conditions on the experts' participation decisions in the first stage. We then examine the impact of opportunity cost and cosine similarity. Given the experts being willing to participate, we next explore the theoretical predictions in section 2 regarding the impact of treatment, matching quality, opportunity cost and social distance on the experts' behavior in the second stage.

4.1 The first stage: participation decision

According to proposition 1, we formulate the predictions on how our interventions affect experts' participation decisions in hypothesis 1 and 2.

Hypothesis 1. Experts' interests in participation follows the order of

1. AvgView < HighView,
2. NoCite < Cite < CiteAckn.

Hypothesis 2. An expert is more likely to be willing to participate if

1. she has a higher number of views for her research work,
2. she is affiliated with an English affiliation
3. she has overlapping areas of expertise with the research team.

To estimate the treatment effects on the experts' willingness to participate, we use the following 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 response (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, whether an expert's primary institution is located in an English-speaking country, and whether an expert is in behavioral and experimental economics.

4.2 The second stage: contribution quantity and quality

According to proposition 2 in the theoretical framework, we develop the following set of hypotheses regarding the experts' contribution in the second stage.

Hypothesis 3. The length and median quality rating of a comment in different conditions follows the order of

1. AvgView < HighView,
2. NoCite < Cite < CiteAckn.

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

Hypothesis 5. The length and quality of a comment is higher if the expert 1) has a higher number of views for her abstract and 2) is affiliated with an institution from an English-speaking country.

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

We evaluate an expert's comment in two aspects: contribution quantity and contribution quality. The contribution quantity is measured straightforwardly by the number of words contained. To measure the quality of the experts' contribution, we develop a rating protocol based on the guidance from the literature on examining peer review of manuscripts. The raters are expected to provide objective evaluations on the quality of the comments written by the experts. In our rating procedure, raters first read the associated Wikipedia article. For each piece of 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 is provided in the appendix.

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

1. *Please rate the overall quality of the comment.*
2. *Suppose you are to incorporate this comment. How helpful is it?*
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?*

Throughout the analysis on the second stage result, 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/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, the importance class. The expert-level controls include the number of views for one's abstract, dummies variables for English affiliation and overlap in research areas.

References

Andreoni, James, "Giving gifts to groups: How altruism depends on the number of recipients," *Journal of public Economics*, 2007, *91* (9), 1731–1749.

Benabou, Roland and Jean Tirole, "Identity, Dignity and Taboos: Beliefs as Assets," 2007. CEPR Discussion Paper 6123.

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.

DellaVigna, Stefano, John A List, and Ulrike Malmendier, "Testing for altruism and social pressure in charitable giving," *The quarterly journal of economics*, 2012, *127* (1), 1–56.

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*," *The American Journal of Psychiatry*, 1993, *150* (6), 947–952.

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

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.

For Online Publication Only: Appendices

Appendix A Expert Selection

RePEc.org[5] has classified authors based on their domains of expertise at "Authors at RePEc" webpage[6]. However, authors usually belong to more than one category and these categories do not represent the experts' most recent area of focus. To capture the most up-to-date description of the experts' fields of expertise, we developed a filtering algorithm to identify their most recent domain of expertise and mentioned that domain name in the subject lines of our emails. For this purpose, we used "NEP reports on IDEAS"[7] and the author profiles to identify most repeated genre among their recent publications in NEP (New Economics Papers) reports, and picked that category as the experts' most recent area of focus (from here on noted as "domain of expertise").

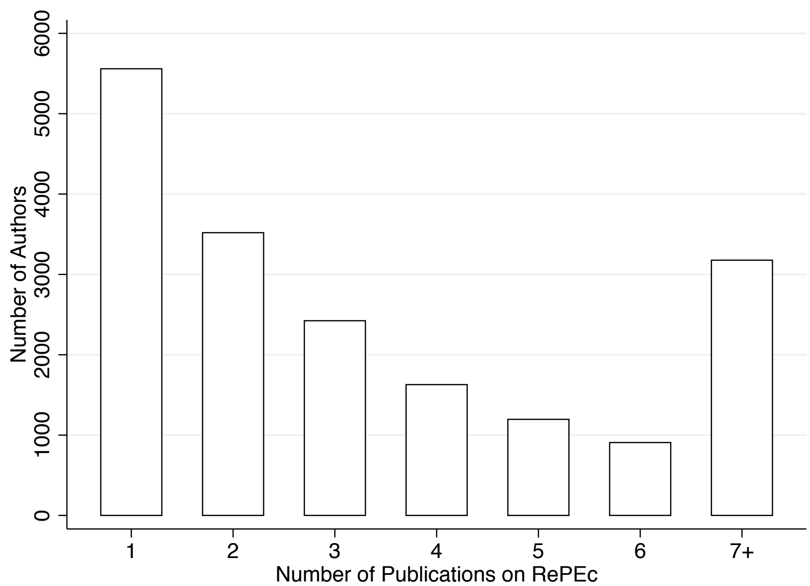

Figure 1: Distribution of Number of Publications on RePEc

[5]https://repec.org/
[6]https://ideas.repec.org/i/e.html
[7]https://ideas.repec.org/n/

Appendix B Wikipedia Articles' Selection

For each of the 6 or 7 publications authored by the economists in the dataset, our system retrieves and recommends a Wikipedia article related to that publication. For this purpose, we use "Google Custom Engine API"[8] to narrow down the list of possible recommended Wikipedia articles to the most relevant ones for each publication of each of the economists. We chose 6 recommended Wikiedia articles for each economist to increase the likelihood of having accurately matched articles in the list of recommendations. All the recommended Wikipedia articles satisfy the following criteria:

- The article must be under the namespace 0 (Main/Article)
- The article is not edit protected[9];
- The article is not a "Stub"[10];
- The character length of article is not less than 1,500 characters;
- The article is viewed at least 1,000 times in the past 30 days (dynamically updated) prior to exposure to the intervention[11].

To this end, in this experiment there are 3,304 unique Wikipedia articles recommended to the economists.

Appendix C Screen shots

In this section, we provide screen shots of interface design for our field experiments.

[8] https://cse.google.com/cse/all

[9] Edit protected Wikipedia articles are not appropriate for the purpose of recommending to economists. A comprehensive explanation of Wikipedia protection policy is available at: Wikipedia:Protection policy - Wikipedia.

[10] "Stub" Wikipedia articles are not appropriate for the purpose of recommending to economists. However, a number of economists asked us to provide them with the commenting interface on specific Wikipedia articles classified as Stub. So, there are few Stubs included in our dataset. A comprehensive explanation of Stub Wikipedia articles is available at: Wikipedia:Stub - Wikipedia.

[11] This restriction is due to the "high-view" (public benefit) condition in the design of the experiment. In order to prevent sample selection bias, all restrictions with less than 1,000 views over the past 30 days have been excluded from the study.

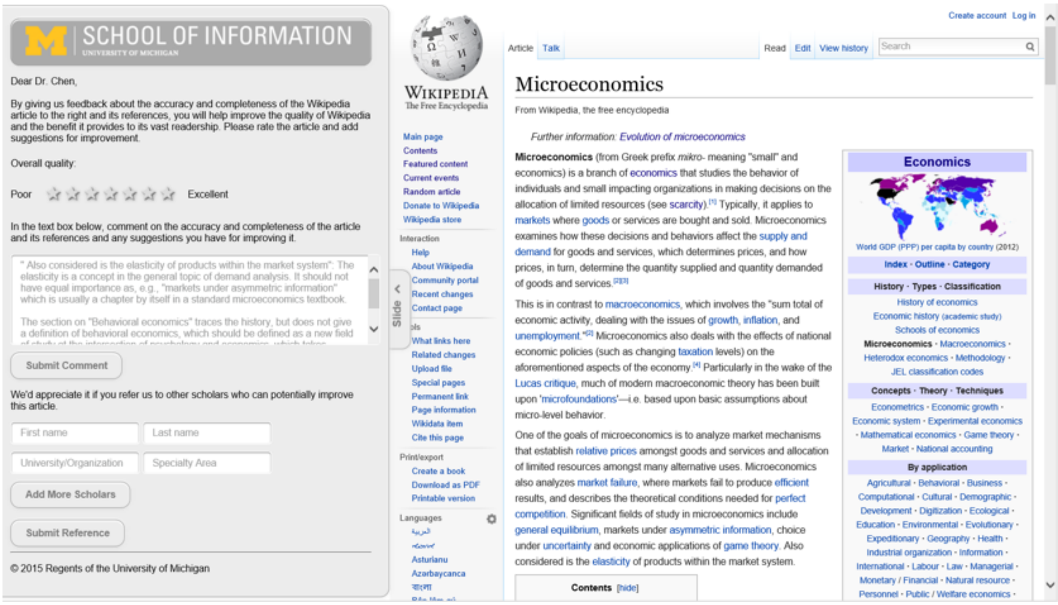

Figure 2: Webpage of Comment

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 3: Webpage of Comment

Appendix D Email messages

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 4: Stage 1 Email

Dear Dr. Bebchuk,

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 5: Stage 2 Email

Appendix E Rating Protocol

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?

- According to the expert, this Wikipedia article has

  _____ errors

  _____ missing points

  _____ missing references

  _____ outdated information

____ outdated references

____ irrelevant information

____ irrelevant references

____ other issues. Please specify: _______________

- According to the expert, this Wikipedia article:
  - ☐ should be restructured.
  - ☐ should contain new sections(s).
  - ☐ contains section(s) that should be removed.
  - ☐ does not need change in structure.
- How many references does the expert provide for the Wikipedia article? ____
- How many self-cited references does the expert provide for the Wikipedia article? ____
- Rate the amount of effort needed to address the experts' comments. (1 = cut and paste; 7 = rewrite the entire article)
- Rate the amount of expertise needed to address the experts' comments. (1 = high school AP economics classes; 7 = PhD in economics)
- 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)
- Suppose you are to incorporate this expert's comments. How helpful are they? (1 = not helpful at all; 7 = very helpful)
- Suppose you have incorporated all of this expert's comments. According to the quality scale of Wikipedia (`https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Wikipedia/Assessment`), this Wikipedia article will become:
  1. Unsure
  2. Stub

3. Start
4. C
5. B
6. Good Article
7. Featured Article

- Please rate the overall quality of the comment. (1 = not helpful at all; 7 = extremely helpful)