

Motivating Contributions to Public Information Goods: A Personalized Field Experiment on Wikipedia¹

Yan Chen

Rosta Farzan

Robert Kraut

Iman YeckehZaare

Ark Fangzhou Zhang

Abstract

We use a large-scale personalized field experiment on Wikipedia to examine the effect of motivation on the contributions of domain experts to public information goods. Experts are 13% more interested in contributing when we mention the private benefit of contribution, such as the likely citation of their work, together with the social impact of the public good. More importantly, we find that greater matching accuracy between a recommended Wikipedia article and an expert’s paper abstract increases both contribution quantity and quality. Our results show the potential of scalable personalized interventions using recommender systems to study drivers of prosocial behavior.

Keywords: public information goods, motivation, matching accuracy

JEL classification: C93, H41

¹Chen and YeckehZaare: School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109 (e-mail: yanchen@umich.edu, oneweb@umich.edu); Farzan: School of Computing and Information, University of Pittsburgh, 135 North Bellefield Avenue, Pittsburgh, PA 15260 (e-mail: rfarzan@pitt.edu); Kraut: School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 (e-mail: robert.kraut@cmu.edu); Zhang: Uber Intelligent Decision System, 1455 Market Street, San Francisco, CA 94103 (e-mail: arkzhang@umich.edu). We are grateful for the helpful comments from George Akerlof, Tim Besley, Raj Chetty, David Cooper, Matt Van Essen, Andrei Gombert, Uri Gneezy, Aaron Halfaker, Rachel Kranton, Shane Murphy, Muriel Niederle, Axel Ockenfels, Paul Resnick, Tanya Rosenblat, Alvin Roth, Sally Sadoff, Karl Schlag, Charlie Sprenger, Jean-Robert Tyran, David Yang, and seminar participants at CIDE, Cologne, Facebook Core Data Science, ITAM, Lyon, Magdeburg, Michigan, Michigan State, Stanford, Tel Aviv, Vienna, Windsor, the Wikimedia Foundation, the 2017 ESA World Meeting (San Diego, CA), 2018 ASSA Meeting (Philadelphia, PA), and the 2nd ERINN Conference (Oxford, UK). 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. This research is not funded by the Wikimedia Foundation. The research was approved by the University of Michigan IRB (HUM00090577), and registered at the AEA RCT Registry (AEARCTR-0002920).

1 Introduction

Online communities, social networking sites, and other online social environments have become popular mechanisms for creating public goods through member contributions of labor and resources. Dedicated to the provision of free information, the Wikipedia community has developed history’s most comprehensive encyclopedia (Lih, 2009). In the technology space, members of open source software development projects have created the software that runs the Internet and many other valuable software artifacts (Weber, 2004). In other contexts, question and answer sites such as Stack Overflow provide users with often highly specific advice about technical problems. Finally, a number of online communities have arisen to provide health-related public goods. Online health support groups, such as BreastCancer.org and the American Cancer Society’s Cancer Support Network, provide members dealing with serious illnesses with both informational and emotional support (Wang et al., 2012).

In each of these cases, the peer-produced public goods enabled by information technology, which we call *public information goods*, have distinct characteristics. They are information goods with free and open access to the general public. Unlike traditional examples of pure public goods, such as national defense, access to technology-based public information goods can easily be controlled through login restrictions. However, the free-by-design nature of these goods means that they are non-rivalrous by nature and non-excludable by choice. Second, public information goods are contributor-dependent in the sense that the right contributor can simultaneously improve the quality and lower the cost of a contribution. Furthermore, accurate matching can even invoke or burnish a contributor’s personal or professional identity, which can also motivate contributions. For example, a game theorist working on equilibrium selection might find it less costly to comment on the Wikipedia article on “Coordination game” than that on “Business cycle”. Her expertise in coordination games would also yield a higher quality contribution to the “Coordination game” article. Similarly, she would be more motivated to contribute to this article as she cares more about her own subject area being presented

accurately to the general public.² Moreover, if her contributions are recognized, they may enhance her reputation.

In this paper, we examine what motivates the willingness and effort of experts in deciding to contribute to public information goods. Specifically, we explore several possible motivations for contributing based on our theoretical model in Section 3. First, we examine whether individuals are motivated to contribute due to the social impact of the public goods (Andreoni, 2007; Zhang and Zhu, 2011). An expert might be more motivated to contribute if many recipients benefit from her contributions. Second, we examine whether individuals are motivated due to private benefits, such as being cited or publicly acknowledged.

In addition to examining motivations to contribute, we investigate the extent to which matching accuracy between the recommended task and the potential contributor’s expertise affects contribution quantity and quality. We use natural language processing techniques (Manning and Schütze, 1999) to determine matching accuracy. Doing so, we are able to identify that exactly what it is that individuals have been asked to contribute is critically important for the quality of their contributions. Our computation techniques to match expertise with contribution tasks are scalable to large communities and to any field with open content.

We conduct our field experiment in the context of the English language version of Wikipedia. The English Wikipedia was founded in January 2001 and is operated by the Wikimedia Foundation. Since its creation, it has become one of the most important information sources for the general public as measured by the number of daily visits.³ As of the February of 2018, the English Wikipedia provided over 5.5 million articles with open and free access to all Internet users. Recent field experiments demonstrate that Wikipedia not only reflects the state of scientific knowledge but also shapes science (Thompson and Hanley, 2017).

To investigate what motivates domain experts to contribute their expertise to public information goods, we design a field experiment where we exogenously vary the social impact of the public information goods and the potential private benefit

²We thank David Cooper for helpful discussions.

³According to Alexa Internet, Wikipedia ranks among the top five most popular websites globally, with more than 262 million daily visits. See <https://www.alexa.com/siteinfo/wikipedia.org>.

(in terms of reputation) that the contribution generates for the contributor using a 2×3 factorial design. Along the social impact dimension, we vary whether our experts are given information about the average number of article views or information that indicates more than twice the average number of views, which we use as a cutoff for all the Wikipedia articles recommended to experts in our sample. Along the private benefit dimension, we vary whether we mention the likelihood of an expert being cited with or without public acknowledgement of her contribution. Our study is registered at the AEA RCT Registry (AEARCTR-0002920).

We invite 3,974 academic economists with at least five research papers posted in a public research paper repository (*RePEc*) which we use for expertise matching. The baseline positive response rate to our initial email is 45%, much higher than the 2% positive response rate from a comparable field experiment inviting academic psychologists to review Wikipedia articles.⁴ Compared to the baseline, telling the economists that they would receive private benefit from their contribution in the form of citations and acknowledgements further increases the positive response rates by 13%. For those who respond positively, textual similarity between the experts' abstracts and the articles they were asked to comment upon has a substantial and significant impact on both contribution quantity and quality. Furthermore, mentioning citation benefit along with the fact that the expert will be publicly acknowledged significantly increases contribution quality. These findings suggest that accurate matching of volunteers to tasks is critically important in encouraging contributions to public information goods, and likely to public goods provision and volunteering in general.

Our study makes novel and important contributions to the experimental public goods literature (Ledyard, 1995; Vesterlund, 2015). First, domain experts are more interested in contributing when we mention the private benefit of contribution, such as the likely citation of their work. Second, our study shows the usefulness of natural language processing techniques to determine matching accuracy between volunteers and tasks, a characteristic which is a robust and significant predictor of

⁴In an unpublished field experiment, authors Farzan and Kraut emailed 9,532 members of the American Psychological Society (APS) inviting them to review Wikipedia articles, with a 2% positive response rate. They manipulated two main factors: identities of those who have done the work and identities of those who will benefit from the reviews provided by APS members.

both contribution quantity and quality. This technique can be extended to determining matching accuracy in other scholarly contexts as well as other types of volunteer activities where expertise matters. In addition to these scientific results, this research identifies public information goods as an increasingly important class of public goods and explores factors which encourage domain experts' contributions.

In addition, our study provides a methodological innovation that synthesizes the predictive accuracy of recommender systems with the causal inference of theory-guided field experiments (Kleinberg et al., 2015), representing a new wave of personalized intervention, analogous to the recent development of precision medicine (Collins and Varmus, 2015).

Finally, it is worth noting that our field experiment has generated valuable public information goods, i.e., 1,097 expert comments on Wikipedia articles in economics, all of which have been posted on the Talk Pages of the corresponding Wikipedia articles, where Wikipedians coordinate with each other in the production process. These comments help improve the quality of Wikipedia articles.

2 Literature Review

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

Economists have developed several perspectives to mitigate the incentive to under-contribute. The mechanism design perspective relies on incentive-compatible tax-subsidy schemes enforced by a central authority.⁵ Therefore, they cannot be di-

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

rectly applied to contexts where contribution is voluntary. In these contexts, a social norms and identity perspective applies insights from theories of social identity to the study of economic problems (Akerlof and Kranton, 2000, 2010). This body of research shows that when people feel a stronger sense of common identity with a group, they exert more effort and make more contributions to reach an efficient outcome (Chen and Chen, 2011; Eckel and Grossman, 2005). In addition to common identity, Bénabou and Tirole (2011) have shown that pro-social behavior can be motivated by the private benefits derived from public acknowledgement and awards. This finding is documented in other studies as well (Andreoni and Bernheim, 2009; Ariely et al., 2009; Rege and Telle, 2004).

In the context of Wikipedia, Algan et al. (2013) conduct a lab-in-the-field experiment among a diverse sample of 850 contributors. Using the public goods and the trust game, they find that reciprocal and altruistic participants are more cooperative when contributing to Wikipedia. In another study on Wikipedia, Kriplean et al. (2008) use naturally occurring data and find that editors who receive more barnstar awards are more likely to contribute.⁶ Following this line of research, Gallus (2016) uses a natural field experiment on the German language version of Wikipedia and finds that a purely symbolic award has a sizable and persistent impact on the retention of new editors. Our study extends this research by using a field experiment to examine how the incentives of being cited and being publicly acknowledged impact expert contributions.

Another potentially important factor that influences contributions to public goods is the social impact, or the number of beneficiaries of the public goods. In the linear public goods environment with voluntary contribution mechanisms, laboratory experiments find a positive effect of group size on total contribution levels with certain parameter configurations (Goeree et al., 2002; Isaac and Walker, 1988; Isaac et al., 1994). By comparison, in the non-linear public goods environment, where the production function is concave in the sum of players' contributions, Guttman (1986) finds evidence that increasing the group size leads to an increase in aggregate con-

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

tributions to the group, but a decrease in average contribution. More recently, Chen and Liang (2018) prove theoretically and find evidence in the lab that the effects of group size on public goods contributions depend on the complementarity of the production function. In the context of a congestable public good, Andreoni (2007) finds that although an increase in the number of recipients encourages a higher contribution, it does not lead to an equivalent increase in total contributions.⁷ The most closely related prior work on the effect of social impact on contributions to public information goods examines the natural experiment in which government blocking of the Chinese Wikipedia reduced the size of the readership and led to a 42.8% decrease in the level of contribution by overseas Wikipedia editors who were not blocked during that time (Zhang and Zhu, 2011). This paper indicates that a reduction in the social impact of the public information good discourages contributions.

Lastly, several studies have examined technology-based public information goods communities. For example, Cosley et al. (2007) deploy an intelligent task-routing agent, SuggestBot, that asks editors to improve articles similar to ones they have worked on before. Their findings show that personalized recommendations lead to nearly four times as many actual edits as random suggestions. While Cosley et al. (2007) utilize Wikipedia editors' existing editing history to recommend articles, we motivate domain experts who have never edited Wikipedia articles to contribute by recommending that they comment on Wikipedia articles similar to their publications and working papers. Our approach demonstrates the potential of developing personalized interventions in economics to promote prosocial behavior.

3 Theoretical Framework

In this section, we outline the theoretical framework that guides our field experiment on how motivation impacts the likelihood of contributing to public information goods. While our theoretical framework is closely related to the literature on voluntary contributions to public goods, we also incorporate features of public

⁷Note that in the standard laboratory experiment the contributors are the beneficiaries of the public good, whereas with Wikipedia editors generally the beneficiaries (readers) are distinct from contributors.

information goods production into our model to better represent the context of our field experiment.

Our study centers around the question of how potential contributors choose to contribute to a public information good, $y \geq 0$. To simplify notation, we use a single public information good. It is straightforward to generalize the results to multiple public goods. To begin, we first let the set of potential contributors, or agents, be I , and the number of consumers of this public good be $n \geq 0$. We then specify that each agent, $i \in I$, selects a contribution level, $y_i \in [0, T_i]$, where $T_i > 0$ is the total resources available to agent i . The quantity of the public information good is obtained as the sum of all individual contributions, $y = \sum_{j \in I} y_j$.

A contributor's utility function is comprised of several components. Let the social impact of the public good be the product of the individual valuation of the public good, $f_i(y)$, and the value derived from the number of consumers, $v_i(n)$, where both $v_i(\cdot)$ and $f_i(\cdot)$ are concave. Thus, the first component of a contributor's utility function is $v_i(n)f_i(y)$, which we call the social impact of the public information good. Incorporating the social impact of contributions is supported by the effects of the exogenous blocking of the Chinese Wikipedia on the contribution behavior of editors who were not blocked (Zhang and Zhu, 2011).

The second component is the private benefit from the act of contribution. Previous research has shown that individuals choose to contribute to public goods due to the warm glow from contributing (Andreoni, 1989, 1990), or increased visibility of the contributor's own work, which should be an increasing function of the number of consumers of the good. Note that our specification allows us to capture various types of private benefits, $w_i(n)$, where $w_i(\cdot)$ is again concave. Thus, the private benefit of contribution is captured by $w_i(n)y_i$.

In comparison, a contributor's cost of contribution has two components. Contributing $y_i \geq 0$ entails a cost, $c_i(y_i)$, which is assumed to be convex in y_i . Let $r_i \geq 0$ be the contributor's marginal opportunity cost. Here, we assume that contributing to the public information good takes time away from other activities, such as one's own research or paid work, that would yield a private benefit of $r_i(T - y_i)$. In our experiment, we measure the marginal opportunity cost, r_i , by the number of views of expert i 's abstracts in a public working paper repository, which serves as

a proxy for the expert's reputation.

In addition to a marginal opportunity cost, a potential contributor faces a cost in terms of how much time and effort is required for a contribution. In our study, we determine an individual's domain expertise through her prior work and use this expertise to identify matched tasks. We represent this by letting $m_i \in (0, 1]$ be the matching accuracy between an expert's domain of expertise and the public information good. Tasks that are matched with domain expertise reduce the cost of contribution as the individual already has the required information at her disposal.⁸ Matching accuracy is primarily determined by the state of art of the recommender system. Let $G(m_i)$ be the cumulative distribution function of matching accuracy. We assume that experts share the same common prior with regard to the distribution of matching accuracy.

After specifying the benefits and costs of individual contributions to the public information good, we now model the process of contribution. To do so, we consider a two-stage process, participation and contribution, in a similar spirit as DellaVigna et al. (2012).

The first stage: participation. In the first stage, we model the expert's interest in contributing to a public information good in her area of expertise. In this stage, matching accuracy is not realized. In deciding to participate, the expert forms an expectation of the matching accuracy, and chooses to participate if the expected utility from participation dominates that of nonparticipation. Those who express interests in participation move to the second stage.

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

⁸Matching an expert to tasks in her domain of expertise might also invoke her professional identity, which could also increase the value she places on the public good. For simplicity, we focus on the former and omit the latter.

$$\max_{y_i \in [0, T_i]} v_i(n) f_i(y) + w_i(n) y_i + r_i(T_i - y_i) - \frac{c_i(y_i)}{m_i}. \quad (1)$$

Using backward induction, we solve expert i 's optimal contribution level in the second stage, y_i^* , and then solve the participation decision in the first stage. The respective proofs are relegated to Appendix A. Solving the optimization problem (1), we first obtain the following comparative statics for the contribution stage.

Proposition 1 (Contribution). *After an expert agrees to participate, she will contribute more if*

- (a) *more people consume the public information good, $\frac{\partial y_i^*}{\partial n} \geq 0$; or*
- (b) *the private benefit of contribution is more salient, $\frac{\partial y_i^*}{\partial w_i} \geq 0$; or*
- (c) *the matching accuracy between the public information good and her expertise is higher, $\frac{\partial y_i^*}{\partial m_i} \geq 0$; or*
- (d) *her opportunity cost of time is lower, $\frac{\partial y_i^*}{\partial r_i} \leq 0$.*

Going back to the first stage when the expert does not know the matching quality, we define expert i 's expected utility as EU_i . We next solve the participation problem and obtain the following comparative statics.

Proposition 2 (Participation). *Ceteris paribus, an expert is more likely to participate if*

- (a) *more people consumer the public information good, $\frac{\partial EU_i}{\partial n} \geq 0$; or*
- (b) *the private benefit of contribution is more salient, $\frac{\partial EU_i}{\partial w_i} \geq 0$; or*
- (c) *her opportunity cost of time is lower, $\frac{\partial EU_i}{\partial r_i} \leq 0$.*

Together, our propositions provide guidance to our experimental design and form the basis for our subsequent hypotheses.

4 Experimental Design

We translate these theoretically derived propositions into a field experiment to explore factors that motivate domain experts to contribute to public information goods. We choose the English language version of Wikipedia as the research site as it is one of the best known and most widely used general public information resources. We choose academic economists as participants, as we know the subject area well. In addition, it is a field with a large public repository of economic research. In what follows, we present our sample selection strategies, design of treatments and experimental procedures.

4.1 Sample Selection: Experts and Articles

The experts whom we invite to contribute to Wikipedia are academic economists registered on *Research Papers in Economics (RePEc)*.⁹ *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 information about her research, such as fields of expertise and a list of publications and working papers. To determine a match between an expert’s domain and a proposed Wikipedia contribution task, we identify her most recent field of expertise based on her most recent publications and working papers. Appendix B provides more details on the algorithm we use in this process.

A power analysis, based on the positive response rate from a pilot experiment conducted in the summer of 2015 ($N = 142$), suggests we would need at least 636 participants per experimental condition (or 3,816 participants for six experimental conditions) to detect a 10% change between two treatments holding one factor constant, with $\alpha = 0.05$ and $\beta = 0.10$.¹⁰ We randomly selected a sample of 3,974 experts from the *RePEc* database who maintained a research profile at *RePEc*, that included both an email address and research specialization and who had at least five research articles in English archived in *RePEc*. The latter requirement for the

⁹See <https://ideas.repec.org>.

¹⁰Our pre-analysis plan contains a more detailed explanation of our sample size calculation (AEARCTR-0002920).

recommendation algorithm was needed to produce accurate matches between their expertise and the Wikipedia articles they would be asked to work on.

The Wikipedia articles recommended to an expert are selected according to their relevance to her research. For each expert, for each of her five or six most recent papers, we first used the Google custom search API to retrieve a list of Wikipedia articles that were most relevant to the keywords in the expert’s research paper. Among these articles, we filtered out those with fewer than 1,500 characters. We further eliminated articles viewed less than 1,000 times in the past 30 days. Therefore, all articles in our sample have a minimum amount of content for experts to comment on, with more than twice as many views as the average Wikipedia article at the time of our experiment, which was 426 views. The average number of views is computed using a Wikipedia data dump the month before the launch of our experiment. We then took the superset of all Wikipedia articles for each expert, ranked them based on the number of repetitions, and recommended the top five or six articles to each expert. Multiple experts could receive some of the same article on which to comment.

In sum, our dataset contains 3,974 experts and 3,304 unique Wikipedia articles. For each expert, the dataset includes the number of times the abstracts for her research papers on *RePEc* had been viewed in 2016, whether she was ranked within the top 10 percentile at *RePEc*, and the affiliated institution.¹¹ For each Wikipedia article, our dataset includes the quality and importance class assessed by Wikipedia, the number of characters comprising the article, the number of revisions, and the number of times it has been viewed over the past 30 days.¹²

¹¹*RePEc* assigns a percentile ranking for each expert based on her number of publications and citations, and lists the top 10 percentile in its public database.

¹²The quality scale at Wikipedia contains the following six classes in increasing order: *Stub*, *Start*, *C*, *B*, *Good Article* and *Featured Article*. The criteria 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 contains four classes: *Low*, *Mid*, *High* and *Top*. The criteria range from “not particularly notable or significant even within its field of study” for the *Low* class to “extremely important, even crucial, to its specific field” for the *Top* class. See information at https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Wikipedia/Assessment.

4.2 Experimental Procedure and Treatments

Our experiment consists of two stages. In the first stage, we sent experts an initial email inquiring whether they were willing to provide comments on Wikipedia articles related to their expertise. This email implements one of our six experimental treatments described below.

We implement a 2×3 between-subject factorial design in which we vary two factors in the emails inviting experts to contribute to Wikipedia (see Table 1).

Table 1: Features of Experimental Conditions

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

Social Impact. To assess the effect of social impact on motivation to contribute, we vary experts’ expectations about the number of times articles are likely to be viewed. In the Average View (AvgView) condition, we tell experts that a typical Wikipedia article received 426 views per month. This information sets a baseline impact expectation. In the High View (HighView) condition, we provide an expert with the additional information that we will only recommend articles which have been viewed at least 1,000 times in the past month. Recall that every Wikipedia article in our sample has been viewed at least 1,000 times per month.

Private benefit. Along the private benefit dimension, we vary experts’ expectation about the private benefit they might receive from their contribution, either giving them no information about citations (NoCite baseline), suggesting contributing would increase the likelihood others would cite their work (Cite), and a third condition in which they were also told that their contributions would be acknowledged on a WikiProject Economics Page used by Wikipedians who curate the economics articles (Citation & Acknowledgement, shortened as CiteAckn).

The treatments are operationalized through the personalized invitation emails

we send the experts. For each condition, we send one of six personalized email messages. The subject line of the email contains the expert’s area of expertise as identified by Algorithm 1 in Appendix B. Each email consists of three sections. The first section is common to all treatments (with words in square brackets personalized for each expert), starting with a brief introduction of Wikipedia and mentioning the average number of views a typical Wikipedia article receives:

Dear Dr. [Chen],

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

Depending on the experimental condition, the second section manipulates social impact by providing information about the readership of the articles to be recommended to the expert and/or the private benefits she can expect to receive. In the HighView condition, we mention that we select articles with over 1,000 views. In the Cite condition, we mention that the articles recommended to the experts are likely to cite their research, by randomly inserting one of the following three messages: “*may include some of your publications in their references*”, “*might refer to some of your research*”, or “*are likely to cite your research*”. Results from χ^2 tests show that the null hypothesis of independence between the actual realization of the email messages and the experts’ first-stage responses cannot be rejected for the Cite condition (p -value = 0.564) or the CiteAckn condition (p -value = 0.435). The following is an example excerpt from a HighView-Cite email message:¹³

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

¹³The order of the HighView and Cite messages was randomized.

1,000 views in the past month, so that your feedback will benefit many Wikipedia readers.

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

The CiteAckn condition strengthens the private benefit by mentioning public acknowledgement as an additional private benefit. In this condition, the experts are told in the email message that their contributions will be acknowledged on the WikiProject Economics page at Wikipedia (see Figure C.4).¹⁴ WikiProject Economics is a group of Wikipedia editors who sign up to improve articles related to economics. Being acknowledged for one's contribution in the WikiProject Economics page thus serves as an additional private benefit beyond that of citation. To avoid potential confound due to the (likely) asynchronous timing of experts' contributions, we froze during the main experiment the acknowledgement page to include only contributions from our pilot phase. Thus, the acknowledgement page seen by the experts does not vary.

The last section of the email asks whether the expert is willing to contribute by commenting on the recommended Wikipedia articles. The experts are provided with two options: *"Yes, please send some Wikipedia articles to comment on."* and *"No, I am not interested."* Authors Chen and Kraut sign the email with their respective titles and institutional affiliations. A screen shot of an example email in the HighView-Cite condition is included in Appendix C as Figure C.1.

Experts who respond positively (i.e., clicking "Yes") to the first-stage email were then sent a second email immediately thanking them and listing the articles recommended to them for comments. As described in more detail below, for experts in the HighView condition, the list also shows the actual number of views each recommended article has received in the past month (Figure C.2). For each article, there was a hyperlink directing the experts to a webpage in which to put comments.

To make this process easier, that experts could comment on an article without having to learn Wikipedia's markup language or how to edit a wiki page, the commenting page consists of a mirror image of the Wikipedia article on the right side

¹⁴See detailed information at https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Economics/ExpertIdeas.

of the screen and a dashboard with a textbox for comments on the left. The interface displayed the article and the text box side by side so that the experts can input their comments without switching between browser pages. The interface disabled all the hyperlinks in the article that could direct the expert away from the article (Figure C.5). As soon as the experts submitted a comment, they were sent a thank-you email (Figure C.3) and their comments were posted on the talk page associated with the corresponding Wikipedia article by our bot, the ExpertIdeas Bot.¹⁵

The experiment started on May 6, 2016 and ended on December 22, 2016. The emails were sent between 6:00 AM and 7:00 PM on weekdays based on the local time of an expert’s primary institutional affiliation. To avoid the emails being filtered as spam, we sent no more than 10 emails within a four-hour period. Throughout the experiment, we used a tracking tool to monitor whether emails sent to experts were opened. If an expert did not respond after two weeks, we sent up to four reminder emails. If the expert declined in any stage, they received no additional email from the experiment. All emails were sent from the first author’s University of Michigan email address.

5 Results

We first investigate the treatment effects on experts’ participation decisions in the first stage. We then present the results for those who elect to participate, exploring the impact of treatment condition, matching accuracy, opportunity cost of time and social proximity on the experts’ contribution behavior in the second stage.

5.1 First Stage: Participation

We first investigate whether our randomization across experimental conditions works. Table 2 reports the summary statistics for our pre-treatment characteristics, broken down into the six experimental conditions. Panels A and B present the characteristics of the experts and recommended Wikipedia articles, respectively. Columns (1) through (6) report average values as well as standard deviations. We

¹⁵See <https://en.wikipedia.org/wiki/User:ExpertIdeasBot>.

Table 2: Characteristics of Experts and Recommended Wikipedia Articles, by Experimental Conditions

	Average View			High View			
	NoCite (1)	Cite (2)	CiteAckn (3)	NoCite (4)	Cite (5)	CiteAckn (6)	p-value (7)
Panel A: Characteristics of Experts							
Abstract Views	1,610 (1,763)	1,633 (1,875)	1,764 (2,637)	1,697 (2,106)	1,810 (2,652)	1,644 (1,764)	0.493
Top 10%	0.360 (0.480)	0.378 (0.485)	0.358 (0.480)	0.347 (0.476)	0.371 (0.483)	0.386 (0.487)	0.712
English Affiliation	0.417 (0.493)	0.457 (0.499)	0.434 (0.496)	0.452 (0.498)	0.477 (0.500)	0.407 (0.492)	0.103
Observations	678	669	671	637	661	658	
Panel B: Characteristics of Article Recommendations							
Article Length	34,266 (33,552)	33,973 (33,194)	34,579 (34,269)	36,269 (36,399)	35,000 (34,875)	34,150 (33,582)	0.044
Number of Edits	725 (997)	725 (1,081)	708 (1,000)	754 (1,066)	750 (1,102)	712 (1,036)	0.273
Views in Past Month	14,409 (17,086)	14,023 (19,842)	14,013 (19,956)	14,348 (18,108)	14,471 (19,955)	13,934 (21,391)	0.732
Quality:							
Featured Article	0.054 (0.227)	0.050 (0.217)	0.046 (0.210)	0.058 (0.235)	0.047 (0.211)	0.048 (0.213)	0.095
Good Article	0.216 (0.412)	0.211 (0.408)	0.215 (0.411)	0.226 (0.418)	0.205 (0.404)	0.201 (0.401)	0.120
B	0.594 (0.491)	0.604 (0.489)	0.601 (0.490)	0.581 (0.493)	0.613 (0.487)	0.613 (0.487)	0.037
C	0.127 (0.333)	0.125 (0.331)	0.126 (0.332)	0.123 (0.328)	0.122 (0.328)	0.127 (0.333)	0.978
Start & Stub	0.009 (0.094)	0.010 (0.099)	0.011 (0.106)	0.012 (0.109)	0.013 (0.113)	0.011 (0.103)	0.582
Importance:							
Top	0.168 (0.374)	0.160 (0.367)	0.158 (0.365)	0.173 (0.378)	0.152 (0.359)	0.153 (0.360)	0.077
High	0.350 (0.477)	0.339 (0.474)	0.353 (0.478)	0.347 (0.476)	0.358 (0.480)	0.348 (0.476)	0.630
Mid	0.255 (0.436)	0.270 (0.444)	0.256 (0.437)	0.245 (0.430)	0.264 (0.441)	0.263 (0.440)	0.192
Low	0.064 (0.245)	0.073 (0.260)	0.070 (0.256)	0.067 (0.251)	0.067 (0.251)	0.071 (0.257)	0.664
Observations	3,924	3,872	3,845	3,693	3,779	3,794	

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

perform χ^2 tests on joint orthogonality across the treatments and report the associated p -values in column (7). The statistics in Table 2 show that the randomization yields balanced experimental groups along most characteristics. One exception is that the recommended Wikipedia articles in the HighView-NoCite condition are longer and of higher quality compared to those in the other conditions.

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

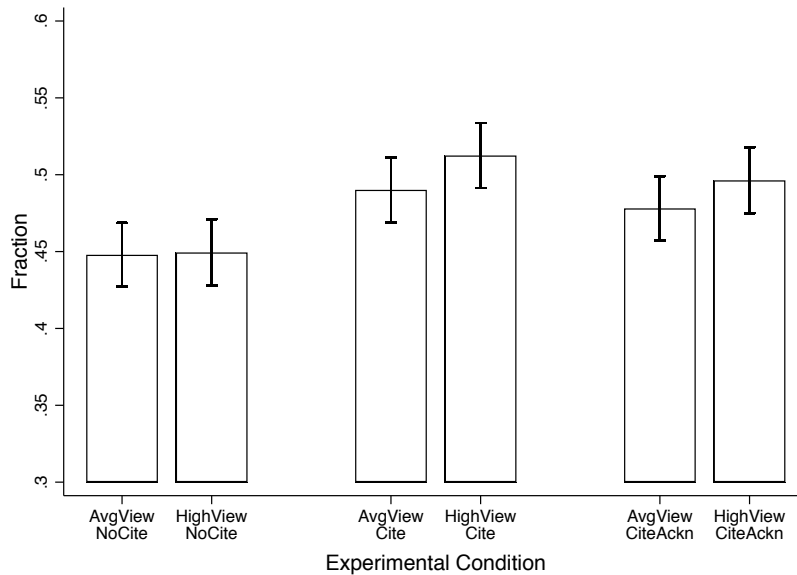


Figure 1: Proportion of positive responses in the first stage: Error bars denote one standard error of the mean.

Figure 1 presents the proportion of positive responses from the treated experts, with the error bars denoting one standard error above and below the mean. Figure 1 indicates that, over all treatment conditions, the baseline willingness to participate is surprisingly high. In the baseline condition mentioning only the average num-

ber of article views (NoCite-AvgView), 44.8% experts respond positively to our invitation, much higher than the 2% positive response rate from a comparable field experiment.¹⁶ While our study and theirs differ in experimental design, we interpret our finding as an indication that our reference to domain expertise in the subject line and first paragraph may have piqued our experts’ interest in responding.

Proposition 2 predicts how our treatments might affect expert participation decisions, formulated below as Hypothesis 1.

Hypothesis 1. The likelihood that experts express interest in participating after receiving the initial email request for contributions follows the order of (a) AvgView < HighView, (b) NoCite < Cite, and (c) Cite < CiteAckn.

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

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

where the dependent variable R_i is an expert i ’ response, which can be positive (1), null (0) or negative (-1). The independent variables include the treatment dummies (HighView, Cite, and CiteAckn), the interactions among these treatment variables, and expert-level control variables including the number of views an expert’s abstracts receive (as a proxy for the expert’s reputation), whether the expert’s primary institution is located in an English-speaking country, and whether the expert is in behavioral and experimental economics, the email senders’ domain expertise (as a proxy for social proximity).

¹⁶In an unpublished field experiment, authors Farzan and Kraut emailed 9,532 members of the American Psychological Society inviting them to review Wikipedia articles, and obtained a 2% positive response rate.

Table 3: Average Marginal Effect on the First-stage Response

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

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

Table 3 reports the results for the average marginal effects estimated from the multinomial logistic regression. Under the average view condition, the likelihood of a negative response is reduced by 6.6 p.p. with citation benefits (p -value < 0.05). Similar results are obtained in the high view condition: estimates for the average marginal effect is 6.0 p.p. for Cite + HighView $\times$ Cite (p -value < 0.05), corresponding to a 13% increase over the baseline response rate of 45%, and -5.9 p.p. for CiteAckn + HighView $\times$ CiteAckn (p -value < 0.05). We summarize the results below.

Result 1 (Treatment Effects on Participation). Under the high view condition, mentioning a citation benefit leads to a 13% increase in the positive response rate, whereas under both the average and high view conditions, mentioning a citation benefit leads to a significant decrease in the negative response rate.

By Result 1, we reject the null in favor of Hypothesis 1(b), but fail to reject the null in favor of Hypothesis 1(a) or 1(c). Overall, we find that mentioning a citation benefit together with a social impact, significantly increases experts' participation interest, whereas mentioning a social impact in terms of number of article views, at least between 426 and 1,000 views, has no effect on participation interest by itself.

Recall that Proposition 2 further suggests that willingness to participate depends on the opportunity cost of doing so. To measure opportunity cost, we use expert reputation, as determined by one of three variables: 1) the number of views for her abstracts at *RePEc*, 2) whether her overall ranking is among the top-10 percentile of researchers at *RePEc*, and 3) whether she is affiliated with an institution from an English-speaking country. A Spearman's rank order test indicates significant correlation between being ranked among the top-10 percentile at *RePEc* and both of the other two measures (p -values < 0.01). Therefore, we use number of abstract views as our measure of expert reputation in our subsequent regression analysis, as it is a finer measure than Top 10%, which is binary. To control for social proximity between the recipient and the requester, we construct a dummy variable, Overlapping Expertise, which equals 1 if the expert has an overlapping area of research with the research team and 0 otherwise. Hypothesis 2 formulates this set of predictions:

Hypothesis 2. The likelihood that an expert is willing to participate decreases (increases) for those who have a higher reputation (more similar expertise to the research team).

Columns (4) through (6) in Table 3 provide the results for the average marginal effects from the multinomial logistic regression including expert-level controls. Note that the empirical distribution of Abstract Views is skewed toward zero (see Figure 2). To mitigate any potential effect of extreme values, we apply both log transformation and percentile ranking (Table D.1 in Appendix D) to Abstract Views in the regression. Doing so, we find that the effect of $\log(\text{Abstract Views})$ on negative response is 3 p.p. ($p\text{-value} < 0.01$). Similarly, we find that experts affiliated with an institution from an English-speaking country are 5.7 p.p. more likely to decline the invitation ($p\text{-value} < 0.01$). Finally, we find that experts with overlapping expertise as the requesters are 21.2 (13.5) p.p. more (less) likely to respond positively (negatively) than others ($p\text{-value} < 0.01$). We summarize these findings in Result 2.

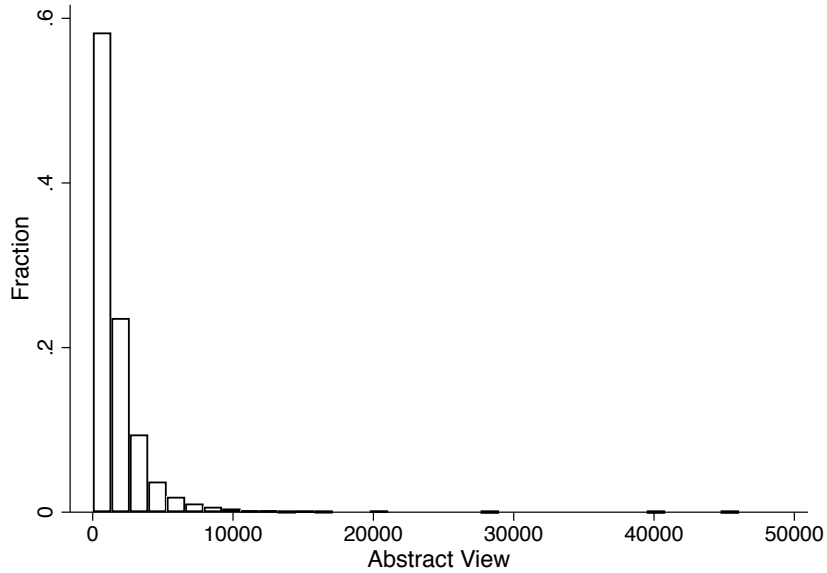


Figure 2: Empirical Distribution of Abstract View

Result 2 (Reputation and Social Proximity). A unit increase in abstract views (affiliation with an institution from an English-speaking country) is associated with a 3 p.p. (6 p.p.) increase in negative response rate. In contrast, an expert whose field overlaps with those requesting her participation is 21 p.p. more likely to respond positively.

By Result 2, we reject the null in favor of Hypothesis 2. As predicted, we find that experts who enjoy a higher reputation (and thus have a higher opportunity cost related to participation) are more likely to decline to participate. From a back-of-the-envelope calculation, we find that a one standard deviation increase in $\log(\text{Abstract Views})$ is associated with a 25 p.p. increase in the likelihood of a negative response. This result is consistent with that of DellaVigna and Pope (2017) who find that assistant professors are more likely to accept an invitation to predict the outcomes of a real-effort experiment and to complete the task than are full professors. Furthermore, Result 2 provides support for the prediction that closer social proximity is likely to yield more positive responses. To the extent that overlapping research expertise implies closer social proximity (Akerlof, 1997; Castillo et al., 2014), our results indicate that the similarity between the recipient and the requester yields greater success in contribution requests for public information goods.

Overall, the results from the first stage of our experiment reveal several interesting findings regarding expert willingness to contribute to public information goods. First, our baseline positive response rate of 45% indicates that even a simple request yields a positive response, especially when the request is tailored to the expert's field. Our results also show that experts are 13% more likely to respond favorably when both the private benefit of citation and the social impact of the public information good are mentioned. Finally, our results show that it is more difficult to get those experts with a higher reputation to respond favorably to a contribution request.

5.2 Second Stage: Contribution Quantity and Quality

Of the 1,603 experts who responded favorably to our initial request, 1,513 opened the second email we sent providing recommendations for articles to com-

ment on. We consider this group to be our treated group for our second-stage analysis. From this treated group, 512 experts commented on at least one Wikipedia article and we received a total number of 1,188 comments, 1,097 of which have been posted on the talk pages of the corresponding Wikipedia articles. Figure 3 summarizes the number of participants at each stage of the experiment.

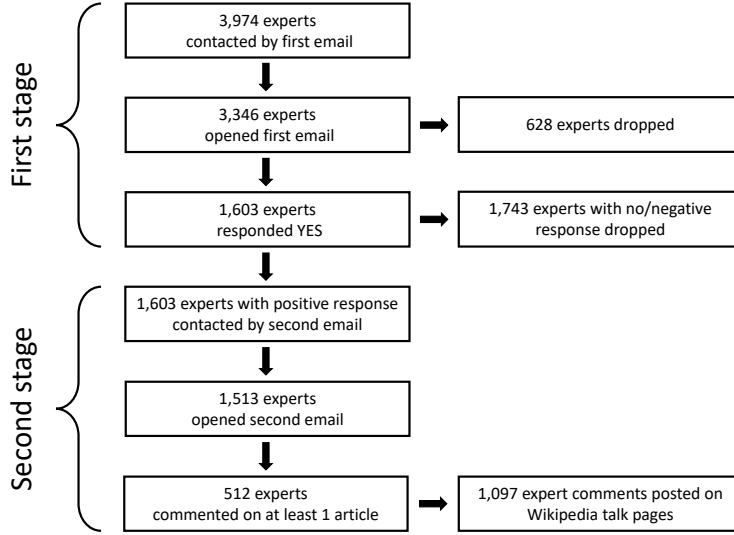


Figure 3: Experts’ Responses in Each Stage of the Experiment

We evaluate both the quantity and quality of each expert’s comments in our analysis. To measure the quantity of an expert’s contribution, we count the number of words in each comment. To measure the quality of an expert’s contribution, we develop a rating protocol following standard content analysis practices (Krippendorff, 2003). Using this protocol, each comment is independently evaluated by three raters who are trained to provide objective evaluations on the quality of the comments. In our rating procedure, raters first read the corresponding Wikipedia article. For each comment, raters start with a series of questions regarding various aspects of the comments prior to giving their overall ratings. This multi-item approach breaks down the comment evaluation task into discrete concrete subcomponents. Doing so has been shown to improve inter-rater reliability for the overall quality rating (Jr. et al., 1993). The rating protocol and the corresponding summary

statistics are provided in Appendix E.

Specifically, we measure the quality of a given comment as the median of the three raters' responses to each of the three following questions:

1. *Please rate the overall quality of the comment.* (1-7 Likert scale)
2. *Suppose you are to incorporate this comment. How helpful is it?* (1-7 Likert scale)
3. *Suppose that you are to incorporate the expert's review of this Wikipedia article and you want to first break down the review into multiple comments. How many comments has the expert made to this Wikipedia article?* (non-negative integers)

Our raters are 68 junior/senior and graduate students at the University of Michigan who either major in economics or have completed the core requirements (including intermediate micro- and macroeconomics, as well as introduction to econometrics). All raters first take part in a training session designed to build a common understanding of the rating scale. In the training session, one research assistant first introduces the experiment to provide the raters with the background of the study. The research assistant then uses one piece of comment as an example and goes through the entire evaluation with the raters as a full group. For each rating question, the assistant discusses the rationale for the rating scale and provides clarification for the rating instructions. The raters then individually practice with the rating scale, with a full group discussion of ratings on the practice comments.

After receiving their training, our raters conduct their evaluations through a web-based survey system which requires Kerberos authentication. To guarantee that raters have background knowledge on the entries they evaluate, we identify courses taken by the raters and then assign the comments to those who have taken courses related to the associated Wikipedia articles.

We assess inter-rater reliability of our raters' evaluation using the intra-class correlation coefficient (ICC[1,3]),¹⁷ which generalizes Cohen's Kappa into the multi-rater case. The reliability statistics for the three responses that we use to measure

¹⁷There are six main cases of intra-class correlation coefficient, distinguished by the numbers in the parentheses following the letters ICC (Shrout and Fleiss, 1979). The first number indicates the

quality is 0.66 for overall quality and helpfulness, and 0.86 for number of sub-comments. In general, values above 0.75 indicate excellent reliability and values between 0.40 and 0.75 indicate fair to good reliability. Therefore, our raters on average provide reliable ratings on the quality of the comments.

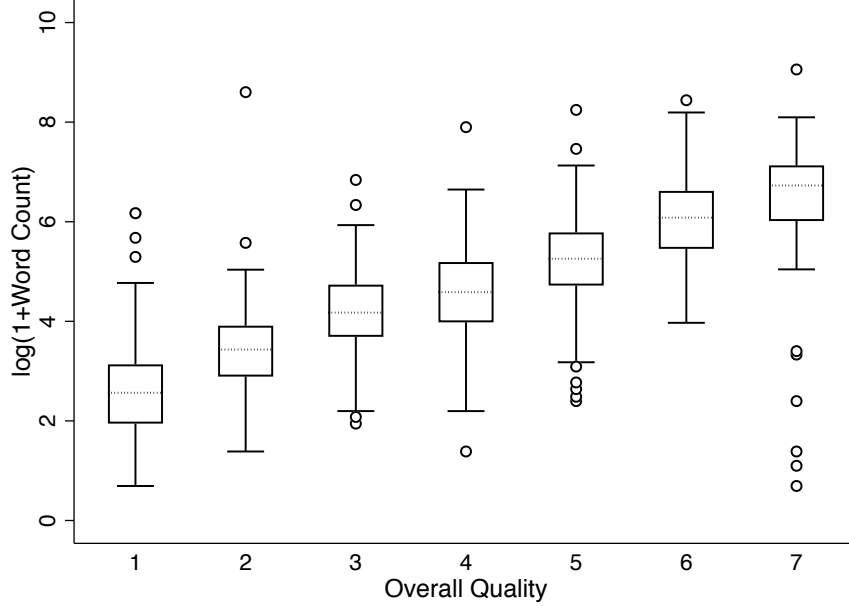


Figure 4: Word Count and Median Rater’s Overall Quality Rating

Figure 4 presents the relationship between our measures of contribution quantity and quality. From Figure 4, we see that quantity, measured as the length of a comment, $\log(1 + \text{Word Count})$, is positively associated with the median rater’s overall quality. Similar correlations hold between $\log(1 + \text{Word Count})$ and the rated helpfulness of a comment (upper panel of Figure D.2) as well as the number of sub-comments contained in a comment (lower panel of Figure D.2). The Spearman’s rank correlation between the quantity and the three quality measures varies between 0.663 and 0.682, and is statistically significant ($p\text{-value} < 0.01$). Similar positive associations between the quality and the quantity of experts’ comments have been found in previous studies in the context of question-answering platforms,

model specification and the second number indicates the number of raters. In our study, we use Case 1 model, which does not require each rating target to be evaluated by a fixed set of raters.

such as Yahoo! Answers (Adamic et al., 2008) and Google Answers (Chen et al., 2010; Edelman, 2012).

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

$$\begin{aligned}
Y_{i,k} = & \beta_0 + \beta_1 \times \text{HighView}_i + \beta_2 \times \text{Cite}_i + \beta_3 \times \text{CiteAckn}_i + \beta_4 \times \text{HighView}_i \cdot \text{Cite}_i \\
& + \beta_5 \times \text{HighView}_i \cdot \text{CiteAckn}_i + \beta_6 \times \text{MatchingAccuracy}_{i,k} \\
& + \mathbf{B}_A \times \text{article-level controls}_i + \mathbf{B}_E \times \text{expert-level controls}_i + \varepsilon_{i,k},
\end{aligned}$$

where i indexes the experts and k indexes the recommended Wikipedia articles. The dependent variable, $Y_{i,k}$, is the quantity or quality measure of expert i 's contribution to article k . HighView_i , Cite_i and CiteAckn_i are dummy variables representing the respective treatment status of expert i , and $\text{MatchingAccuracy}_{i,k}$ measures the quality of matching between expert i 's expertise and the recommended article k . In addition, we include in our regression article-level controls for article length, quality class, and importance class. We also include the same expert-level controls as in our earlier analyses: number of abstract views, English-speaking institution affiliation, and similar expertise as the requesting research team.

Contribution quantity. Note that the data on contribution quantity features a semi-continuous distribution with a mass at the origin, as 86.5% articles recommendations received no comments after the experts opened the second-stage email. Such a large number of zeros would make the common assumption of normality inappropriate and render the asymptotic inference problematic. To overcome this issue, we fit the data with an exponential dispersion model that assumes that the variance of the outcome is a power function of the mean (Jorgensen, 1987; Zhang, 2013). Compared to other models that address a disproportionate number of zeros in the data, the exponential dispersion model is applicable to continuous data rather than discrete ones.

In our analysis of the impact of expert motivation on contribution effort, Proposition 1 suggests that an increase in the number of viewers or the private benefit to the contributor yields more effort in contribution. This leads to Hypothesis 3.

Hypothesis 3. Experts' comment quantity and quality follows the order of: (a)

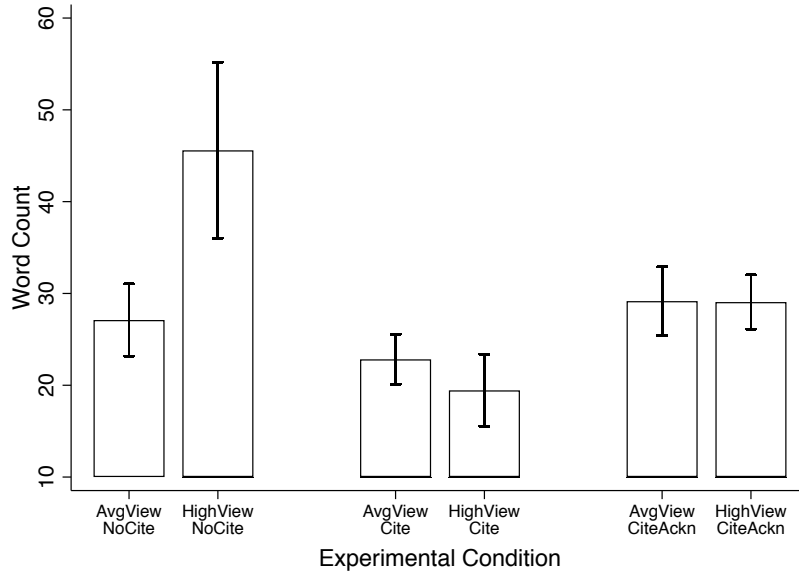


Figure 5: Average word count by experimental condition: Error bars denote one standard error of the mean

AvgView < HighView, and (b) NoCite < Cite < CiteAckn.

Figure 5 plots average word count of the comments for each experimental condition, with the error bars denoting one standard error. The average length of the experts' contribution ranges between 19.45 and 45.59, though it exhibits large variations. The social impact manipulation does not influence the length of comments between the AvgView and HighView conditions (p -value = 0.65 for the NoCite condition, 0.10 for the Cite condition, 0.07 for the CiteAckn condition, using the Wilcoxon rank-sum test). In terms of private benefit, neither the citation nor the citation & acknowledgement condition lead to longer comments than the non-citation baseline. Furthermore, within the HighView condition, it appears that mentioning the citation benefit actually reduces the quantity of experts' contribution (p -value < 0.05, using the Wilcoxon rank-sum test). The regression results in Table 4, which control for the expert fixed effect, also indicates no statistical evidence supporting treatment effects on contribution quantity. We summarize this discussion in Result 3.

Table 4: Determinants of Contribution Quantity

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

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

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

Contribution quality. We next examine the treatment effect on contribution quality. Figure 6 plots the average overall quality of the comments for each experimental condition, with the error bars denoting one standard error. From Figure 6, we see that there is an increase in the quality of the comments along the dimension of private benefit. Nonparametric comparisons among the experimental conditions shows that the average overall quality is higher in the CiteAckn conditions than it is in the NoCite conditions (p -value < 0.05 for both the AvgView condition and the HighView condition, using the Wilcoxon rank-sum tests).

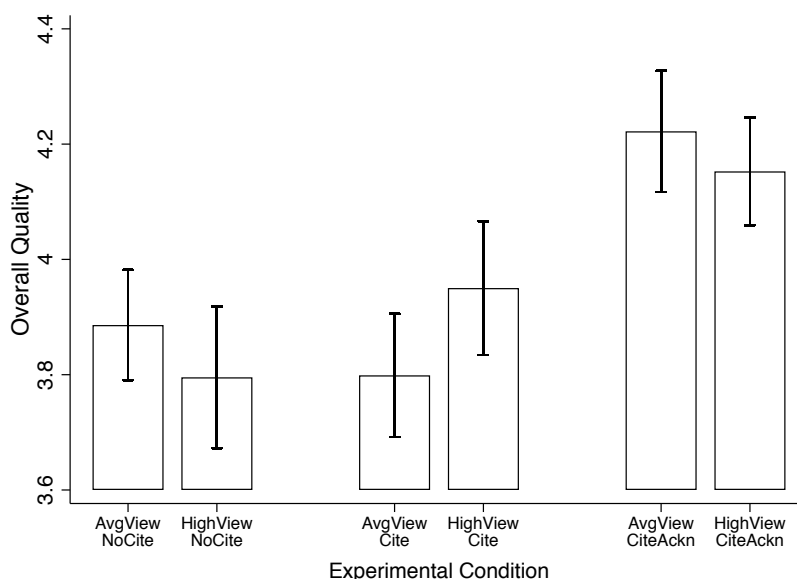


Figure 6: Average overall quality by experimental condition: Error bars denote one standard error of the mean

Table 5 presents the results of our second-stage regression analysis on contribution quality. These results show that mentioning the private benefit of acknowledgment encourages higher quality contributions. Specifically, we find that the effect of CiteAckn on the proportional odds ratio for the ordered logistic model is significantly larger than 1. Put differently, the comments from the CiteAckn conditions

Table 5: Determinants of Contribution Quality

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

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

are significantly more likely to receive a higher rating for overall quality than the no citation base rate. Tables in Appendix D.2 show that results reported in Table 5 are robust when we use the percentile of article length and abstract view (Table D.3), and when we report the ordered logit model for each quality category for overall quality (Table D.4), helpfulness (Table D.5) and the number of sub-comments contained in a contribution (Table D.6). We summarize these findings in Result 4.

Result 4 (Treatment Effect on Quality). Compared to expert contributions in the baseline no citation condition, those in the citation & acknowledgement conditions receive significantly higher ratings in terms of overall quality and helpfulness.

Taken together, Results 3 and 4 offer a mixed answer to our predictions in Hypothesis 3. While we find that mentioning citation and acknowledgement benefits does not yield longer contributions from our expert participants, we find that doing so does improve the quality of the contributions. Thus, we conclude that our results highlight the potential usefulness of a non-monetary incentive, such as public acknowledgement, in inducing high-quality contributions.

Matching accuracy. We next examine how matching accuracy impacts the quantity and quality of expert contributions to the public information good, predicted in Proposition 1. To quantify the matching quality of the recommendations, we calculate the cosine similarity (Singhal et al., 2001) between the Wikipedia article k and expert i 's research work. Cosine similarity is widely used in the area of informational retrieval as a measure of the extent of similarity between two documents. To compute the cosine similarity between two documents, we convert each document into a vector of words, and compute the cosine value of the angle between the two word vectors. For example, a one unit increase in the cosine similarity measure translates into a match with no overlapping words ($\cos \theta = 0$) and a perfect match ($\cos \theta = 1$). Appendix F provides a detailed description and an example on the calculation of cosine similarity. Using the above similarity measure, we can now state Hypothesis 4.

Hypothesis 4. Experts will contribute longer and better-quality comments when they are assigned to tasks that match their expertise more accurately.

Examining the results of our analysis, we find that the cosine similarity between a recommended Wikipedia article and an expert’s abstract has an economically and statistically significant impact on the quantity of contribution. Specifically, the results in Table 4, columns (3) and (4) yield parameter estimates on cosine similarity between 1.77 and 2.86 (p -value < 0.01). Using a simple back-of-the-envelope calculation, we find that a one unit increase in cosine similarity, which translates into a match with no overlapping words and a perfect match, implies 52.52 more words in an average comment.¹⁸

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

Result 5 (Matching Accuracy and Contribution). An expert contributes longer and better quality comments to Wikipedia articles that exhibit a higher cosine similarity to the abstracts of her research papers.

By Result 5, we reject the null in favor of Hypothesis 4. Our result indicates that contributions to public information goods in particular depend on the matching quality between the specific public good and the contributors’ attributions. This finding reinforces prior results in Edelman (2012), who shows that the specialization level of a Google Answers contributor has a positive effect on the quality of her answers.

We next explore how opportunity cost affects the quantity and quality of a contribution. Recall that we measure an expert’s opportunity cost by the number of

¹⁸The change in character length is calculated as $\Delta(\text{Word Count}) = \hat{\beta}_x \cdot (1 + \text{Word Count}) \cdot \Delta x$, using the $\hat{\beta}_x$ estimated in column (4) of Table 4.

views for her research paper abstract. This yields the following hypothesis, again based on Proposition 1.

Hypothesis 5. Experts with a higher reputation (more abstract views) will contribute less.

The results of our analysis yield an insignificant Kendall’s rank correlation τ between Word Count and Abstract View of -0.013 (p -value = 0.109). Using an exponential dispersion model, the results in column (6) of Table 4 yield insignificant parameter estimates for $\log(\text{Abstract View})$ (p -value = 0.11). Finally, the results in Table 5 for the effect of opportunity cost on contribution quality show insignificant coefficients for $\log(\text{Abstract View})$. We summarize this discussion in Result 6.

Result 6 (Opportunity Cost and Contribution). Conditional on being willing to participate, the length and quality of the contributions by experts who receive a higher number of views for their research paper abstracts are not significantly different from those of their counterparts.

By Result 6, we fail to reject the null in favor of Hypothesis 5. Finally, we investigate the effect of social proximity on contribution length and quality, as outlined below in Hypothesis 6.

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

A non-parametric comparison shows that the experts whose area of research overlaps that of the research team contribute comments that are significantly longer (p -value < 0.01 using the Wilcoxon rank-sum test). Returning to Table 4, the parameter estimates for Overlapping Expertise are 0.37 in the OLS model and 0.74 in the exponential dispersion model (p -value < 0.01 in both specifications). These estimates represent an increase of 2.83 and 11.09 more words in the comments, respectively.

The positive impact of social proximity also applies to contribution quality. The estimates in columns (2), (4) and (6) in Table 5 indicate that overlapping expertise has a positive effect on the ratings for experts’ comments (though weakly significant for overall quality).

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

By Result 7, we reject the null in favor of Hypothesis 6. Although we label this result as one due to the social proximity between the potential contributor and the contribution requesters (Castillo et al., 2014), there could be a number of reasons at work, such as social image concern (Bénabou and Tirole, 2011), or social pressure (DellaVigna et al., 2012). We remain agnostic as to the underlying mechanism for the positive effect of this factor.

In sum, our personalized field experiment uncovers several interesting factors in encouraging domain experts to contribute to public information goods. First, in the elicitation stage, experts are more willing to participate when we mention the private benefit of contribution, such as the likely citation of their work, together with the higher social impact of the public information good. In the contribution stage, we find that greater matching accuracy between a recommended Wikipedia article and an expert’s paper abstract increases both contribution quantity and quality, whereas mentioning public acknowledgement leads to better but not longer comments. Furthermore, experts with a higher reputation are less likely to participate. However, conditional on agreeing to contribute, their contribution quantity and quality are no different from others. Lastly, experts with overlapping expertise with the research team are both more likely to participate, and contribute longer and better quality comments.

6 Conclusion

Public information goods, such as the articles provided by Wikipedia, have the potential of giving everyone “free access to the sum of all human knowledge” (Miller, 2004). However, to realize this potential, they require the input of experts who have other demands on their time and energy. One way to increase expert contributions is to understand what motivates these individuals to contribute. This study explores factors that encourage domain experts to contribute to public information goods.

Using a field experiment designed to explore both the private benefit and the social impact of contributions, we find that private benefits, such as the likelihood that a Wikipedia article would cite one's own work, significantly increases experts' interest in contributing. Closer social proximity, such as overlapping domain of expertise with the requesters, also increases the likelihood of participation. In the second stage of our experiment, we examine the impact of motivation on the quantity and quality of contributions. Here, we find that mentioning public acknowledgement of one's contributions increases the quality of contributions. These findings indicate that non-pecuniary private benefits can be used as motivators in improving contribution efforts. However surprisingly the likely social impact of one's contributions does not increase contribution by itself, whereas social impact in combination with private benefit does.

Furthermore, in the case of contributions to public information goods as opposed to money, the *nature* of what people are being asked to contribute is crucially important. Accurate matching between expertise and the task significantly increases both contribution quantity and quality. Specifically, conditional on expressed willingness to contribute, the accuracy of matching of participants' expertise and their tasks, measured by the cosine similarity between the text in their abstracts and the Wikipedia articles to which they are assigned, is the single most significant predictor of both contribution quantity and quality. This result highlights the potential of utilizing information technology, such as recommender systems, in promoting pro-social behavior. Although psychologists have stressed the matching of volunteer tasks with volunteers' motivations to contribute (Stukas et al., 2009), our research shows that matching on task expertise is also crucially important.

References

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

Ai, Chunrong and Edward C. Norton, “Interaction Terms in Logit and Probit Models,” *Economics Letters*, 2003, 80 (1), 123–129.

Airio, Eija, “Word Normalization and Decompounding in Mono- and Bilingual IR,” *Information Retrieval*, 2006, 9 (3), 249–271.

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

___ and **Rachel E. Kranton**, “Economics and Identity,” *Quarterly Journal of Economics*, 2000, 115 (3), 715–753.

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

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

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

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

___, “Giving Gifts to Groups: How Altruism Depends on the Number of Recipients,” *Journal of Public Economics*, 2007, 91 (9), 1731–1749.

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

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

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

Bergstrom, Theodore, Lawrence Blume, and Hal Varian, “On the Private Provision of Public Goods,” *Journal of Public Economics*, 1986, 29 (1), 25–49.

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.

Chen, Roy and Yan Chen, “The Potential of Social Identity for Equilibrium Selection,” *American Economic Review*, October 2011, 101 (6), 2562–2589.

Chen, Yan, “Incentive-Compatible Mechanisms for Pure Public Goods: A Survey of Experimental Research,” in Charles Plott and Vernon Smith, eds., *The Handbook of Experimental Economics Results*, Vol. 1, Amsterdam: North-Holland, 2008, pp. 625 – 643.

___ and **Yingzhi Liang**, “Optimal Team Size in Public Goods Provision: Theory and Experiments,” 2018. University of Michigan Working Paper.

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

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

Cosley, Dan, Dan Frankowski, Loren Terveen, and John Riedl, “SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia,” in “Proceedings of the 12th International Conference on Intelligent User Interfaces” ACM 2007, pp. 32–41.

DellaVigna, Stefano and Devin Pope, “What motivates effort? Evidence and expert forecasts,” *Review of Economic Studies*, 2017, 85 (2), 1029–1069.

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

Eckel, Catherine C. and Phillip Grossman, “Managing Diversity by Creating Team Identity,” *Journal of Economic Behavior & Organization*, 2005, 59 (1), 17–45.

Edelman, Benjamin, “Earnings and Ratings at Google Answers,” *Economic Inquiry*, 2012, 50 (2), 309–320.

Gallus, Jana, “Fostering Public Good Contributions with Symbolic Awards: A Large-scale Natural Field Experiment at Wikipedia,” *Management Science*, 2016, 63 (12), 3999–4015.

Goeree, Jacob K., Charles A. Holt, and Susan K. Laury, “Private Costs and

Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior,” *Journal of Public Economics*, 2002, 83 (2), 255–276.

Groves, Theodore and John O. Ledyard, “Incentive Compatibility since 1972,” in Theodore Groves, Roy Radner, and Stanley Reiter, eds., *Information, Incentives and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*, Minneapolis: University of Minnesota Press, 1987, pp. 48–111.

Guttman, Joel M., “Matching Behavior and Collective Action: Some Experimental Evidence,” *Journal of Economic Behavior & Organization*, 1986, 7 (2), 171–198.

Huang, Chu-Ren, Petr Šimon, Shu-Kai Hsieh, and Laurent Prévot, “Rethinking Chinese Word Segmentation: Tokenization, Character Classification, or Wordbreak Identification,” in “Proceedings of the 45th Annual meeting of the ACL on Interactive Poster and Demonstration Sessions” Association for Computational Linguistics 2007, pp. 69–72.

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

___, ___, and **Arlington W. Williams**, “Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups,” *Journal of Public Economics*, 1994, 54 (1), 1–36.

Jorgensen, Bent, “Exponential Dispersion Models,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 1987, pp. 127–162.

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

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

Kriplean, Travis, Ivan Beschastnikh, and David W. McDonald, “Articulations of Wikiwork: Uncovering Valued Work in Wikipedia through Barnstars,” in “Proceedings of the 2008 ACM conference on Computer

Supported Cooperative Work” ACM 2008, pp. 47–56.

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

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

Leskovec, Jure, Anand Rajaraman, and Jeffrey D. Ullman, *Mining of Massive Datasets*, Cambridge: Cambridge University Press, 2014.

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

Manning, Christopher D. and Hinrich Schütze, *Foundations of Statistical Natural Language Processing*, MIT press, 1999.

Miller, Rob, “Wikipedia Founder Jimmy Wales Responds,” *Slashdot*, July 28 2004.

Rege, Mari and Kjetil Telle, “The Impact of Social Approval and Framing on Cooperation in Public Good Situations,” *Journal of Public Economics*, 2004, 88 (7), 1625–1644.

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

Shrout, Patrick E. and Joseph L. Fleiss, “Intraclass Correlations: Uses in Assessing Rater Reliability,” *Psychological Bulletin*, 1979, 86 (2), 420–428.

Singhal, Amit et al., “Modern Information Retrieval: A Brief Overview,” *IEEE Data Eng. Bull.*, 2001, 24 (4), 35–43.

Stukas, Arthur A, Keilah A Worth, E Gil Clary, and Mark Snyder, “The matching of motivations to affordances in the volunteer environment: An index for assessing the impact of multiple matches on volunteer outcomes,” *Nonprofit and Voluntary Sector Quarterly*, 2009, 38 (1), 5–28.

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

Vesterlund, Lise, “Using experimental methods to understand why and how we give to charity,” in John H. Kagel and Alvin E. Roth, eds., *The Handbook of Experimental Economics*, Vol. 2, Princeton, New Jersey: Princeton University

Press, 2015.

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

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

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

Zhang, Yanwei, “Likelihood-based and Bayesian Methods for Tweedie Compound Poisson Linear Mixed Models,” *Statistics and Computing*, 2013, *23* (6), 743–757.

For Online Publication Only

Appendices for

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

Yan Chen Rosta Farzan Robert Kraut
Iman YeckehZaare Ark Fangzhou Zhang

List of Appendices

A: Proofs	42
B: Sample selection: Experts and articles	45
C: Screen shots	48
D: Robustness checks	52
E: Rating protocol	60
F: Calculating cosine similarity	62

Appendix A Proofs

In this appendix, we present the proofs for the two propositions in Section 3. We use backward induction to solve the second stage optimization problem first.

Proof of Proposition 1: In the second stage, upon observing the realized matching accuracy, m_i , expert i solves the following optimization problem:

$$\max_{y_i \in [0, T]} v_i(n) f\left(\sum y_{-i} + y_i\right) + w_i(n) y_i + r_i(T_i - y_i) - \frac{c_i(y_i)}{m_i}. \quad (2)$$

Let y_i^* be expert i 's optimal contribution level. The first order condition requires:

$$v_i(n) f'\left(\sum y_{-i} + y_i^*\right) + w_i(n) - r_i - \frac{c'(y_i^*)}{m_i} = 0. \quad (3)$$

Because the valuation function for the public good, $f_i(y)$, is concave and the cost function, $c_i(y_i)$, is convex, the second order condition is satisfied:

$$v_i(n) f''\left(\sum y_{-i} + y_i^*\right) - \frac{c''(y_i^*)}{m_i} \leq 0. \quad (4)$$

In what follows, we proceed to show that y_i^* is increasing in n , w_i , m_i and decreasing in r_i .

- (a) An increase in the number of consumers of the public information good leads to an increased level of contribution. Taking the derivative of Equation (3) with respect to n , we obtain:

$$\left[v_i(n) f''\left(\sum y_{-i} + y_i^*\right) - \frac{c''(y_i^*)}{m_i} \right] \frac{\partial y_i^*}{\partial n} = -v'(n) f'\left(\sum y_{-i} + y_i^*\right) - w'_i(n).$$

Because $w'_i(n) \geq 0$, $v'(n) \geq 0$, $f'(y) \geq 0$ and (4), we have:

$$\frac{\partial y_i^*}{\partial n} \geq 0.$$

- (b) An increase in the private benefit of contributions leads to an increased level of contributions. Taking the derivative of Equation (3) with respect to w_i , we obtain:

$$\left[v_i(n) f''\left(\sum y_{-i} + y_i^*\right) - \frac{c''(y_i^*)}{m_i} \right] \frac{\partial y_i^*}{\partial w_i} = -1.$$

Because of the second order condition (4), we have:

$$\frac{\partial y_i^*}{\partial w_i} \geq 0.$$

- (c) Better matching between the content of the public information good and the agent's expertise leads to an increased level of contributions. Taking the derivative of Equation (3) with respect to m_i , we obtain:

$$\left[v_i(n) f'' \left(\sum y_{-i} + y_i^* \right) - \frac{c''(y_i^*)}{m_i} \right] \frac{\partial y_i^*}{\partial m_i} = -\frac{c'(y_i^*)}{m_i^2}.$$

Because $c'(y_i^*) \geq 0$ and (4), we have:

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

- (d) An expert with a higher reputation will contribute less. Taking the derivative of Equation (3) with respect to r_i , we obtain:

$$\left[v_i(n) f'' \left(\sum y_{-i} + y_i^* \right) - \frac{c''(y_i^*)}{m_i} \right] \frac{\partial y_i^*}{\partial r_i} = 1.$$

Because of the second order condition (4), we have

$$\frac{\partial y_i^*}{\partial r_i} \leq 0.$$

Q.E.D.

Proof of Proposition 2: In the first stage, an expert does not see the realization of the match accuracy, m_i , but knows its distribution $G(m_i)$. Therefore, she forms her expectations for the matching accuracy m_i .

Let $V_i(n, w_i, r_i, m_i)$ be the value function for the optimization problem in (2) at optimal solution y_i^* :

$$V_i(n, w_i, r_i, m_i) = v_i(n) f \left(\sum y_{-i} + y_i^* \right) + w_i(n) y_i^* + r_i(T_i - y_i^*) - \frac{c_i(y_i^*)}{m_i}.$$

By the envelope theorem, we have

$$\begin{aligned}
\frac{\partial V_i}{\partial n} &= v'(n)f\left(\sum y_{-i} + y_i^*\right) + w'_i(n)y_i^* \geq 0 \\
\frac{\partial V_i}{\partial w_i} &= y_i^* \geq 0 \\
\frac{\partial V_i}{\partial r_i} &= -y_i^* \leq 0 \\
\frac{\partial V_i}{\partial m_i} &= \frac{c_i(y_i^*)}{m_i^2} \geq 0
\end{aligned}$$

In the first stage, expert i does not observe the realization of matching quality, but knows its distribution $G(m_i)$, which is assumed to have a continuous density function. If expert i chooses to participate, her expected utility is

$$EU_i(n, w_i, r_i) = \int_0^1 V_i(n, w_i, r_i, m) dG(m_i). \quad (5)$$

Otherwise, her utility is $v_i(n)f\left(\sum y_{-i}\right) + r_i \cdot T_i$. An expert participates if $EU_i(n, w_i, r_i) \geq v_i(n)f\left(\sum y_{-i}\right) + r_i \cdot T$.

To prove the comparative statics in Proposition 2, we want to show that $EU_i(n, w_i, r_i)$ is increasing in n , w_i and decreasing r_i .

- Taking derivative of EU_i with respect to n :

$$\frac{\partial EU_i}{\partial n} = \frac{\partial}{\partial n} \int_0^1 V_i(n, w_i, r_i, m) dG(m_i) = \int_0^1 \frac{\partial}{\partial n} V_i(n, w_i, r_i, m) dG(m_i) \geq 0$$

- Taking derivative of EU_i with respect to w_i :

$$\frac{\partial EU_i}{\partial w_i} = \frac{\partial}{\partial w_i} \int_0^1 V_i(n, w_i, r_i, m) dG(m_i) = \int_0^1 \frac{\partial}{\partial w_i} V_i(n, w_i, r_i, m) dG(m_i) \geq 0$$

- Taking derivative of EU_i with respect to r_i :

$$\frac{\partial EU_i}{\partial r_i} = \frac{\partial}{\partial r_i} \int_0^1 V_i(n, w_i, r_i, m) dG(m_i) = \int_0^1 \frac{\partial}{\partial r_i} V_i(n, w_i, r_i, m) dG(m_i) \leq 0$$

Q.E.D.

Appendix B Recommendation algorithms

In this appendix, we describe methods used to identify experts' domains of expertise as well as those used to identify the most relevant Wikipedia articles for each expert.

We first describe the method we use to identify our experts' respective domains of expertise. To do so, we develop a filtering algorithm which is based on the experts' recent research papers archived in *New Economics Papers (NEP)*. *NEP* is an announcement service that disseminates and archives new research papers in 97 research areas.¹⁹ For each expert, we refer to *NEP* to obtain her recent research papers as well as the research fields where each work is classified. Then, we select the research field in which her research papers are classified most often and use that one as the most recent domain of expertise. The pseudo-code for the filtering algorithm that identifies an expert's most recent domain of expertise is presented as Algorithm 1 below.

```
foreach expert do
  ResearchList  $\leftarrow$  expert's research papers at NEP.
  foreach research paper do
    Retrieve the list of NEP categories the research paper belongs to.
    foreach category do
      specDict[category] += 1
      if specDict[category] == 7 then
        Result: Return the list of the expert's research papers under
                  this category as his or her recent research papers
                  and the category as his or her recent field of interest.
      end
    end
  end
  Data: maxSpec := the specialization in specDict with maximum # of
                publications.
  Result: Return the list of the expert's research papers under this
            category as his or her recent research papers and the category
            as his or her recent field of interest.
end
```

Algorithm 1: The algorithm for identifying an expert's most recent domain of expertise.

¹⁹See <http://nep.repec.org/>.

In what follows, we present the details for our selection criteria for Wikipedia articles. For each of an expert’s research papers listed in *NEP*, the recommendation algorithm submits a search query containing the keywords in the paper through Google Custom Engine API. The search result returned from Google contains Wikipedia articles that are potentially relevant enough for recommendation. After we iterate over all research papers by an expert, we obtain a list of Wikipedia articles indicated as relevant to the expert’s recent research focus. We further restrict this list using the following criteria: 1) The article must be under the namespace 0 (i.e., main articles);²⁰ 2) The article is not edit protected;²¹ 3) The length of the article is not less than 1,500 characters; 4) The article is viewed at least 1,000 times in the past 30 days (dynamically updated) prior to exposure to the intervention.²² Finally, we choose the five to six Wikipedia articles that appear most frequently in the search results by Google Custom Engine for our recommendation. The pseudo-code for the algorithm that identifies the most relevant articles for each expert is presented as Algorithm 2.

For both algorithms, our code can be accessed from GitHub through the following URL: <https://github.com/ImanYZ/ExpertIdeas>. The back-end uses Python (Django framework) and MySQL Database, whereas the front-end uses HTML, CSS3 and JavaScript (jQuery).

²⁰Wikipedia uses namespace to categorize webpages according to their functions. All encyclopedia articles at Wikipedia are under namespace 0. Webpages under other namespaces include talk pages and user pages. See <https://en.wikipedia.org/wiki/Wikipedia:Namespace> for a detailed explanation of namespace at Wikipedia.

²¹The edit protection restricts a Wikipedia article from being edited by users. It is usually applied to articles that are subject to content disputes or the risk of vandalism. The decision to apply or remove edit protection is made by administrators at Wikipedia. See https://en.wikipedia.org/wiki/Wikipedia:Protection_policy for a detailed explanation.

²²This restriction guarantees that articles recommended in the AvgView condition are similar to those recommended in the HighView condition in terms of the number of views.

```

foreach expert do
  Data: RecommendationsDict := empty dictionary of recommendations
    and their # of repetition.
  foreach publication by the author do
    Data: keyword := the first keyword listed in the RePEc profile of
      the publication.
    recommendations = Retrieved Google search Engine API results
      searching (“econ+” + keyword);
    if |recommendations| ≠ 0 then
      foreach recommendation in recommendations do
        if recommendation is under the namespace 0 (Main/Article)
          ∧
          recommendation is not edit protected ∧ recommendation
            is not a “Stub” ∧
          the character length of recommendation is not less than
            1,500 characters ∧
          recommendation has not been viewed less than 1,000
            times over the past 30 days then
          Result: Save recommendation as one of the
            recommendations for publication.
          Increment # of repetition of recommendation in
            RecommendationsDict.
        end
      end
    end
  end
  foreach publication by the author do
    Result: Save the most repeated recommendation as the
      recommendation for publication.
    end
  end

```

Algorithm 2: Algorithm for matching and recommending Wikipedia articles with an expert’s most recent publications.

Appendix C Screen shots

In this section, we provide screen shots of the interface design for our field experiments, starting with examples of the three emails we sent to the experts.

Our first email implements the treatments. Below is an example in the High-View & Citation treatment. Note that the order of the HighView and the Citation paragraphs was randomized for each expert before the email was sent out. In all three examples, we replace the expert's real last name by the first author's last name.

Dear Dr. Chen,

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

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

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

Please click one of the following links to continue:

[Yes, please send me some Wikipedia articles to comment on.](#)

[No, I am not interested.](#)

Thank you for your attention.

Sincerely,

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

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

Figure C.1: First-stage email: An example in the HighView & Citation treatment.

Dear Dr. Chen,

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

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

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

Sincerely,

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

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

Figure C.2: Second-stage email: An example in the HighView & Citation treatment

Dear Dr. Chen,

Thank you for providing feedback on Wikipedia articles. We have posted your comments to the following article talk page(s), which is where Wikipedia editors discuss changes to articles. You can see the original article or your comments, by clicking on the appropriate links below.

Wikipedia Article	Your Comment
Shareholder value	Your comment on the Talk Page
Corporate governance	Your comment on the Talk Page
Managerial economics	Your comment on the Talk Page
Economic nationalism	Your comment on the Talk Page

Thank you again for your contribution to Wikipedia!

Sincerely,

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

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

Figure C.3: Thank-you Email

Figure C.4 presents our public acknowledgement of expert contributions to Wikipedia articles. This page was assembled by a Wikipedian, Shane Murphy, who was a doctoral student in Economics at the University of Lancaster. The economists on this list contributed to our project during its pilot phase. The list was kept constant during our experiment.

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 

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 

- 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-O. Bayer (University of Adelaide) contributed to Public goods game. - Jonathan Bencimot (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  . - Robert Buckley (NYU Abu Dhabi) contributed to Real estate economics and Mortgage insurance. - Richard Burkhauser (Cornell University) contributed to Top-coded   Current Population Survey and Economic inequality. - Theres Butner (University of Erlangen-Nuremberg) contributed to Equalization payments. - Laurent Calais (Vrije Universiteit) contributed to Friedman test  and Meta-analysis. - David Canning (Harvard University) contributed to Fertility-development controversy and Population ageing. - Richard Cebuta (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. - Musharrat Cyan (Georgia State University) contributed to Economy of Pakistan and Taxation in Pakistan. - Mikołaj Czajkowski (University of Warsaw) contributed to Choice modelling  . - 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 (University 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  , 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. - Petris 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 . - Martin Gustafsson (Uppsala University) contributed to Education economics.	- Magnus Henrekson (Research Institute of Industrial Economics) contributed to 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 (Caird 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  , 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) . - Dominka Langenmayr (Catholic University Eichstätt-Ingolstadt) contributed to Base Erosion and Profit Shifting. - Marga Peeters (Université Catholique de Louvain) contributed to Trade-off, Demographic-economic paradox , Dependency ratio and Arab states of the Persian Gulf. - Luc Savard (Université de Sherbrooke) contributed to Computable general equilibrium. - Hasan Tekguc (Mardin Artuklu University) contributed to Economy of Turkey and Life expectancy . - Marie Claire Villeval (University of Lyon) contributed to Experimental economics, Incentive, Peer-to-peer file sharing, Piece work, Screening (economics) and Ambiguity effect.
--	--	--

Figure C.4: Public Acknowledgement Hosted on a WikiProject Economics Page

A larger version of this page hosted on Wikipedia can be accessed through the following URL: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Economics/ExpertIdeas.

Figure C.5 presents our webpage where experts enter their comments. The interface is designed to minimize entry cost. An expert does not need to know how to edit a wiki. In the split screen design, the right side is the corresponding Wikipedia article that the expert can scroll up or down. The left side has a quality rating and a text box for the expert to enter comments. Thus, the process only requires knowledge of Word.

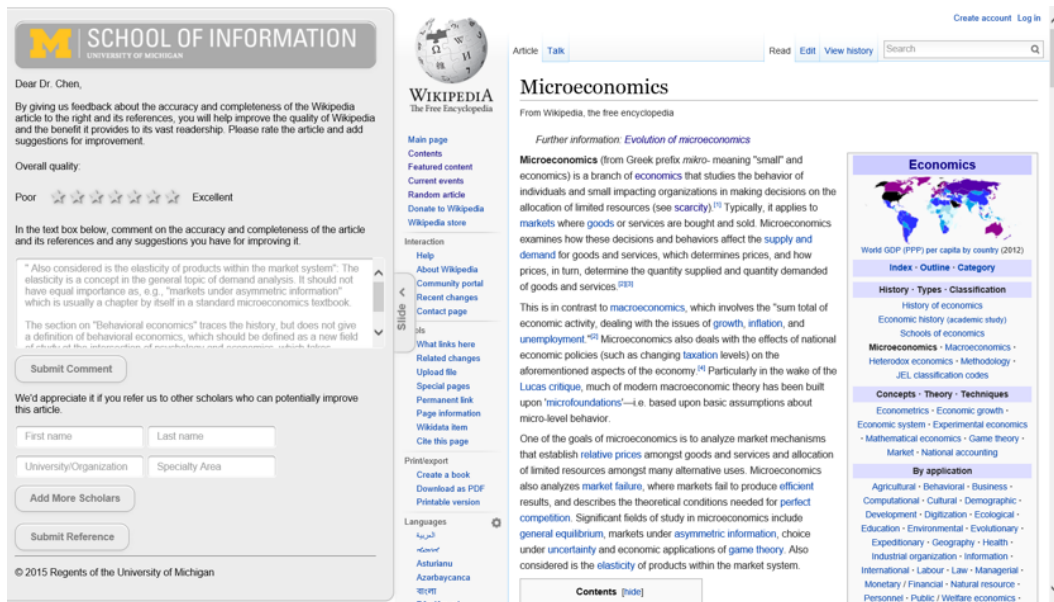


Figure C.5: Web interface for experts to enter comments

Appendix D Robustness checks

D.1 First-stage response and contribution quantity

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

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

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

Table D.2: Determinants of Contribution Quantity

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

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

D.2 Robustness check: Contribution quality

This subsection contains robustness checks for contribution quality. Table D.3 provides robustness checks for Table 5 when we replace the logarithmic transformation by percentile of article length and abstract view. We find that the estimated marginal effect on the probability of be rated as 6 out of 7 is 3.38 p.p. in the AvgView condition (p -value < 0.01) and 3.32 p.p. in the HighView condition (p -value < 0.05). Tables D.5 and D.6 provide a complete ordered probit analysis providing robustness checks for helpfulness (see column 3-4 in Table 5) and the number of sub-comments, respectively. We find that the average marginal effect of CiteAckn is significantly positive (negative) on the probability that the helpfulness of the comment is rated above (below) 4, whereas the impact of CiteAckn on the number of sub-comments is positive but weakly significant.

Table D.3: Determinants of Contribution Quality

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

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

Table D.4: Average Marginal Effect on Overall Quality

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

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

Table D.5: Average Marginal Effect on Helpfulness

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

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

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

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

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

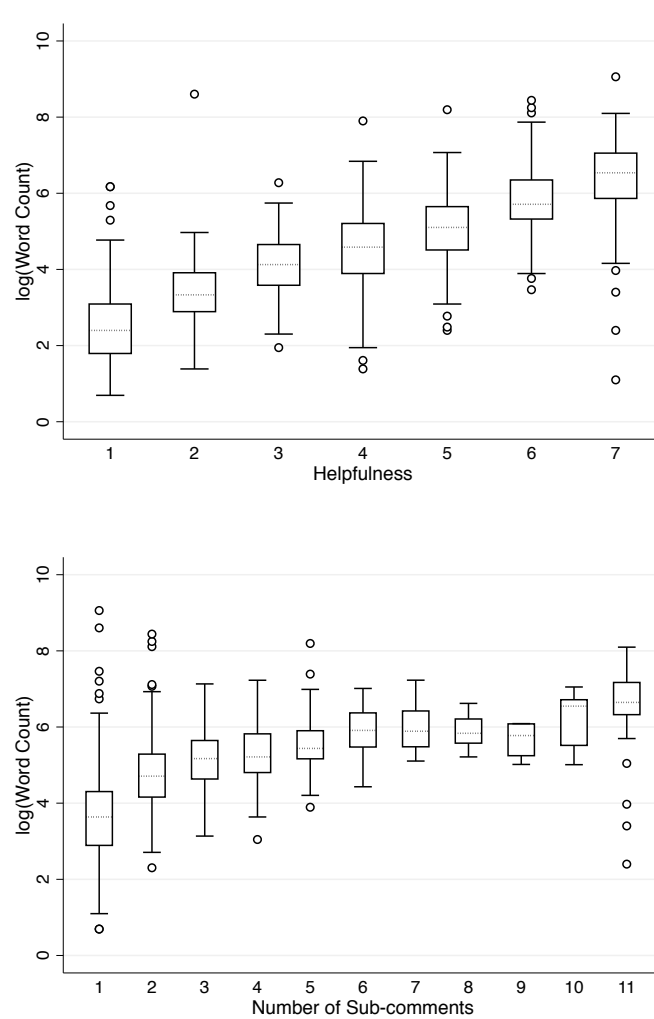


Figure D.6: Word count and median helpfulness (upper panel); Word count and median number of subcomments within a comment (lower panel)

Appendix E Rating protocol

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

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

- Suppose that you are to incorporate the expert's review of this Wikipedia article and you want to break down the review into multiple pieces of comments. How many pieces of comments has the expert made to this Wikipedia article? (mean: 2.711, median: 2, standard deviation: 0.069)
- According to the expert, this Wikipedia article has
 - _____ errors (mean: 1.444, median: 0, standard deviation: 0.912)
 - _____ missing points (mean: 1.098, median: 1, standard deviation: 0.040)
 - _____ missing references (mean: 0.626, median: 0, standard deviation: 0.049)
 - _____ outdated information (mean: 0.043, median: 0, standard deviation: 0.007)
 - _____ outdated references (mean: 0.010, median: 0, standard deviation: 0.003)
 - _____ irrelevant information (mean: 0.134, median: 0, standard deviation: 0.013)
 - _____ irrelevant references (mean: 0.016, median: 0, standard deviation: 0.005)
 - _____ other issues. (mean: 0.238, median: 0, standard deviation: 0.019)
Please specify: _____
- How many references does the expert provide for the Wikipedia article? _____ (mean: 1.508, median: 0, standard deviation: 0.074)
- How many self-cited references does the expert provide for the Wikipedia article? _____ (mean: 0.374, median: 0, standard deviation: 0.032)

- Rate the amount of effort needed to address the experts' comments. (1 = cut and paste; 7 = rewrite the entire article) (mean: 3.621, median: 4, standard deviation: 0.057)
- Rate the amount of expertise needed to address the experts' comments. (1 = high school AP economics classes; 7 = PhD in economics) (mean: 3.887, median: 4, standard deviation: 0.057)
- How easily can the issues raised in the comment be located in the Wikipedia article? (1 = unclear where to modify in the Wikipedia article; 7 = can be identified at the sentence level) (mean: 4.572, median: 5, standard deviation: 0.061)
- Suppose you are to incorporate this expert's comments. How helpful are they? (1 = not helpful at all; 7 = very helpful) (mean: 4.121, median: 4, standard deviation: 0.045)
- Please rate the overall quality of the comment. (1 = not helpful at all; 7 = extremely helpful) (mean: 3.968, median: 4, standard deviation: 0.044)

Appendix F Cosine similarity

In this appendix, we describe the process used to compute the cosine similarity between two documents, an expert’s abstract and a Wikipedia article. Cosine similarity of two documents measures the similarity between them in terms of overlapping vocabulary.

1. Retrieving two pieces of text:

- (a) Document a is the abstract of Akerlof and Kranton (2000):

“This paper considers how identity, a person’s sense of self, affects economic outcomes. We incorporate the psychology and sociology of identity into an economic model of behavior. In the utility function we propose, identity is associated with different social categories and how people in these categories should behave. We then construct a simple game-theoretic model showing how identity can affect individual interactions. The paper adapts these models to gender discrimination in the workplace, the economics of poverty and social exclusion, and the household division of labor. In each case, the inclusion of identity substantively changes conclusions of previous economic analysis.”

- (b) Document b is the Wikipedia article on Identity Economics (https://en.wikipedia.org/wiki/Identity_economics), with only the text part of the article retrieved from the MediaWiki API on December 2, 2018.

“Identity economics Identity economics captures the idea that people make economic choices based on both monetary incentives and their identity: holding monetary incentives constant, people avoid actions that conflict with their concept of self. The fundamentals of identity economics was first formulated by Nobel Prize–winning economist George Akerlof and Rachel Kranton in their article “Economics and Identity,” [1] published in Quarterly Journal of Economics. This article provides a framework for incorporating social identities into standard economics models, expanding the standard utility function to include both pecuniary payoffs and identity utility. The authors demonstrate the importance of identity in economics by showing how predictions of the classic

principal-agent problem change when the identity of the agent is considered. Akerlof and Kranton provide an overview of their work in the book “Identity Economics,” [2] published in 2010. In the book, they provide a layman’s approach to Identity Economics and apply the concept to workplace organization, gender roles, and educational choice, summarizing several previous papers on the applications of Identity Economics. [3][4][5] While this macro-economic theory deals exclusively with already well established categories of social identity, Laszlo Garai when applied the concept of social identity in economic psychology [6] takes into consideration identities in *statu nascendi* (i.e. in the course of being formed and developed). [7][8] This theory that is referred to the macro-processes based on a “large-scale production” later gets applied to the individual creativity’s psychology: Garai derived it from the principal’s and, resp., agent’s “identity elaboration”. A further special feature of Garai’s theory on social identity is that it resolved the contradiction between the inter-individual phenomena studied by the social identity theories and the intraindividual mechanisms studied by the brain theories: L. Garai presented [9] a theory on an inter-individual mechanism acting in the world of social identity. The theory that was referred in the beginning to the macro-processes based on a large-scale production later has been applied by Garai to the micro-processes of individual creativity. [10] Following papers have used social identity to examine a variety of subjects within economics. Moses Shayo uses the concept of social identity to explain why countries with similar economic characteristics might choose substantially different levels of redistribution. [11] The paper won the 2009 Michael Wallerstein Award, given to the best article published in the area of political economy. Daniel Benjamin, James Choi, and Joshua Strickland examine the effect of social identity, focusing on ethnic identity, on a wide range of economic behavior. [12] For a review of papers that study economics and identity, see articles by Claire Hill (2007) and John Davis (2004). [13][14]”

2. Filtering the text: remove all the non-alphabetic characters from Documents *a* and *b*. Document *a* becomes:

“This paper considers how identity a person s sense of self affects economic outcomes We incorporate the psychology and sociology of identity into an economic model of behavior In the utility function we propose identity is associated with different social categories and how people in these categories should behave We then construct a simple game theoretic model showing how identity can affect individual interactions The paper adapts these models to gender discrimination in the workplace the economics of poverty and social exclusion and the household division of labor In each case the inclusion of identity substantively changes conclusions of previous economic analysis”

3. Tokenizing: enter both text files into a tokenizer (Huang et al., 2007), which divides text into a sequence of tokens, which roughly correspond to words. Document *a* becomes the following list of tokens:

[‘This’, ‘paper’, ‘considers’, ‘how’, ‘identity’, ‘a’, ‘person’, ‘s’, ‘sense’, ‘of’, ‘self’, ‘affects’, ‘economic’, ‘outcomes’, ‘We’, ‘incorporate’, ‘the’, ‘psychology’, ‘and’, ‘sociology’, ‘of’, ‘identity’, ‘into’, ‘an’, ‘economic’, ‘model’, ‘of’, ‘behavior’, ‘In’, ‘the’, ‘utility’, ‘function’, ‘we’, ‘propose’, ‘identity’, ‘is’, ‘associated’, ‘with’, ‘different’, ‘social’, ‘categories’, ‘and’, ‘how’, ‘people’, ‘in’, ‘these’, ‘categories’, ‘should’, ‘behave’, ‘We’, ‘then’, ‘construct’, ‘a’, ‘simple’, ‘game’, ‘theoretic’, ‘model’, ‘showing’, ‘how’, ‘identity’, ‘can’, ‘affect’, ‘individual’, ‘interactions’, ‘The’, ‘paper’, ‘adapts’, ‘these’, ‘models’, ‘to’, ‘gender’, ‘discrimination’, ‘in’, ‘the’, ‘workplace’, ‘the’, ‘economics’, ‘of’, ‘poverty’, ‘and’, ‘social’, ‘exclusion’, ‘and’, ‘the’, ‘household’, ‘division’, ‘of’, ‘labor’, ‘In’, ‘each’, ‘case’, ‘the’, ‘inclusion’, ‘of’, ‘identity’, ‘substantively’, ‘changes’, ‘conclusions’, ‘of’, ‘previous’, ‘economic’, ‘analysis’]

4. Removing stop words: make all the characters lower-case and remove all the stop words. Document *a* becomes:

[‘paper’, ‘considers’, ‘identity’, ‘person’, ‘sense’, ‘self’, ‘affects’, ‘economic’, ‘outcomes’, ‘incorporate’, ‘psychology’, ‘sociology’, ‘identity’, ‘economic’, ‘model’, ‘behavior’, ‘utility’, ‘function’, ‘propose’, ‘identity’, ‘associated’, ‘different’, ‘social’, ‘categories’, ‘people’, ‘categories’, ‘behave’, ‘construct’, ‘simple’, ‘game’, ‘theoretic’, ‘model’, ‘showing’, ‘identity’, ‘affect’, ‘individual’, ‘interactions’, ‘paper’, ‘adapts’, ‘models’, ‘gender’, ‘discrimination’,

'workplace', 'economics', 'poverty', 'social', 'exclusion', 'household', 'division', 'labor', 'case', 'inclusion', 'identity', 'substantively', 'changes', 'conclusions', 'previous', 'economic', 'analysis']

5. Stemming: convert each token to its corresponding stem, which strips variants of the same word into the word's root (Airio, 2006). Document *a* becomes:

['paper', 'consid', 'ident', 'person', 'sens', 'self', 'affect', 'econom', 'outcom', 'incorpor', 'psycholog', 'sociolog', 'ident', 'econom', 'model', 'behavior', 'util', 'function', 'propos', 'ident', 'associ', 'differ', 'social', 'categori', 'peopl', 'categori', 'behav', 'construct', 'simpl', 'game', 'theoret', 'model', 'show', 'ident', 'affect', 'individu', 'interact', 'paper', 'adapt', 'model', 'gender', 'discrimin', 'workplac', 'econom', 'poverti', 'social', 'exclus', 'household', 'divis', 'labor', 'case', 'includ', 'ident', 'substant', 'chang', 'conclus', 'previou', 'econom', 'analysi']

6. Defining the stemmed corpus: take the union of the two stemmed documents, where each unique stemmed token is defined as a dimension.

stemmed-corpus = ['paper', 'consid', 'ident', 'person', 'sens', 'self', 'affect', 'econom', 'outcom', 'incorpor', 'psycholog', 'sociolog', 'model', 'behavior', 'util', 'function', 'propos', 'associ', 'differ', 'social', 'categori', 'peopl', 'behav', 'construct', 'simpl', 'game', 'theoret', 'show', 'individu', 'interact', 'adapt', 'gender', 'discrimin', 'workplac', 'poverti', 'exclus', 'household', 'divis', 'labor', 'case', 'includ', 'substant', 'chang', 'conclus', 'previou', 'analysi', 'captur', 'idea', 'make', 'choic', 'base', 'monetari', 'incent', 'hold', 'constant', 'avoid', 'action', 'conflict', 'concept', 'fundament', 'first', 'formul', 'nobel', 'prize', 'win', 'economist', 'georg', 'akerlof', 'rachel', 'kranton', 'articl', 'publish', 'quarterli', 'journal', 'provid', 'framework', 'standard', 'expand', 'includ', 'pecuniari', 'payoff', 'author', 'demonstr', 'import', 'predict', 'classic', 'princip', 'agent', 'problem', 'overview', 'work', 'book', 'layman', 'approach', 'appli', 'organ', 'role', 'educ', 'summar', 'sever', 'applic', 'macro', 'theori', 'deal', 'alreadi', 'well', 'establish', 'laszlo', 'garai', 'take', 'consider', 'statu', 'nascendi', 'e', 'cours', 'form', 'develop', 'refer', 'process', 'larg', 'scale', 'product', 'later', 'get', 'creativ', 'deriv', 'resp', 'elabor', 'special', 'featur', 'resolv', 'contradict', 'inter', 'phenomena', 'studi',

