Encouraging Online Knowledge Contributions - Evidence from a Field Experiment

Yikun Jiang *

November 26, 2024

Please Click Here for the Latest Version

Abstract

With the prominence of user-generated content platforms, online knowledge platforms have experienced substantial growth. However, it is unclear why individuals voluntarily contribute knowledge, and how platform strategies can facilitate individuals' contributions through non-monetary incentives. I develop a stylized model of individual knowledge contributions where individuals have social motivation to gain reputation and instrumental motivation to obtain functionality privileges on the platform. Based on the model, I design and implement a large-scale field experiment, involving 12,182 individuals on one of the largest online question-and-answer platforms. I sample and manually treat participating platform individuals daily over the course of four and a half months. The treatment gives one anonymous upvote to eligible answers from one of my own accounts, exogenously shifting individuals' social and instrumental motivations. I then track comprehensive data on individuals' subsequent behavior on the platform daily for four months. I find that the treatment significantly increases an individual's probability of contributing additional answers by around 15% of the baseline, and the difference between the control and treatment groups persists over time. The treatment effect is the most pronounced for individuals with low-to-moderate answering experience or reputation and is slightly stronger for those who are close to obtaining additional privileges after the treatment. The overall quality and effort of future answers remain stable. Using data from the field experiment, I structurally estimate the model of contribution decisions to quantify the relative importance of social and instrumental motivation. Simulation results suggest that social motivation is more important, and platform strategies that boost social motivation are more effective in encouraging contributions.

Keywords: Consumer Dynamics; Innovation; Field Experiment; User-Generated Contents; Online Knowledge Platforms; Online Reputation; Platform Design; Social Preferences

*University of California, Berkeley, Department of Economics. Email: jiangyk@berkeley.edu.

Previously titled, Dynamics in Online Knowledge Contributions - Evidence from a Field Experiment. I thank advisors, Stefano DellaVigna, Matthew Osborne, Nano Barahona, Benjamin Handel, and Olivia Natan. I appreciate helpful discussions with Ned Augenblick, Luisa Cefala, Nan Chen, Yan Chen, Bryan Chu, John Conlon, Avi Goldfarb, Ali Goli, David Holtz, Kei Kawai, Jura Liaukonyte, Emaad Manzoor, Nitin Mehta, Sarah Moshary, Ricardo Perez-Truglia, Carolyn Stein, Avner Strulov-Shlain, Steven Tadelis, Dmitry Taubinsky, J. Miguel Villas-Boas, Nathan Yang, participants at UC Berkeley seminars and ISMS Marketing Science Conference 2024. The experiment received IRB approval from the University of California, Berkeley, before implementation. The IRB approval number is 2023-04-16245. The pre-registration ID at AEA RCT Registry is AEARCTR-0011615. All errors are my own.

1 Introduction

In recent years, user-generated content platforms have gained significant prominence, with online knowledge platforms like Quora and Stack Overflow experiencing substantial growth in traffic. On these platforms, individuals contribute their knowledge as a public good for others to use, while the platforms generate revenue from this user-contributed content. Sustaining user contributions over the long term is crucial for the viability of these platforms. For instance, Yahoo Answers, the first question-and-answer platform launched in 2005, saw its growth slow down around 2011 and ultimately closed in May 2021 due to a decline in content contributions (Guo et al. 2023). Therefore, understanding how online knowledge platforms, and online content platforms in general, can sustain user contributions is essential.

A feature of several successful online knowledge platforms is the offering of non-monetary rewards to individuals to incentivize content contributions. These rewards generally fall into two categories: recognition from other users, such as reputation points and virtual badges, and usage privileges granted by the platforms. Non-monetary rewards are also widely used by various platforms that depend on user-generated content to drive engagement and revenue. Examples include online review platforms like TripAdvisor and Goodreads, online education websites such as Khan Academy, and social media platforms including Instagram, Twitter, TikTok, and Facebook.[1]

In this paper, I present evidence from a large-scale field experiment quantifying the relative importance of social motivation for gaining additional reputation - driven by factors such as social status, altruism, or warm glow - versus instrumental motivation for obtaining additional privileges to unlock additional platform functionalities[2] in knowledge contribu-

[1]On TripAdvisor, reviewers receive badges and earn higher levels by sharing their travel experiences, leading to increased influence and recognition. On Goodreads, book reviewers can earn likes, comments, and followers by contributing reviews, lists, and recommendations. Khan Academy also uses non-monetary rewards to encourage contributions in their course discussion forum - individuals who answer questions, contribute to discussions, or practice skills receive badges and energy points that build their online reputation. Social media platforms like Instagram, Twitter, TikTok, and Facebook let individuals click “like” on others’ posts as a way to encourage future posts.

[2]This is a general type of non-monetary rewards distributed by platforms. For instance, on Twitter, Community Notes contributors must rate a sufficient number of notes to unlock the ability to write notes themselves. Similarly,

tion decisions on a leading online question-and-answer platform, Stack Overflow. It is a popular platform for professional and enthusiast programmers, where individuals voluntarily contribute. Individuals can gain reputation points if their contributions are recognized by others using the platform, and with more reputation points, individuals can gain additional privileges on the platform once their reputation reaches certain thresholds. I experimentally vary the recognition given to a group of contributing individuals, providing an exogenous boost to their reputation and thereby shifting both their social and instrumental motivations.

To motivate the experiment design, I first introduce a model of online knowledge contribution, where individuals derive utility from gaining reputation and privileges on the platform. The model incorporates individuals forming expectations about reputation points to gain with each answer contribution and heterogeneous valuation of reputation and privileges. Then, I present the design of the field experiment with several distinctive features. These features are specifically tailored for the experiment to focus on new and less active participants, a group where platform strategies have the potential to yield significant increases in contributions.

From August 31st, 2023 to January 10th, 2024, I implement the large-scale field experiment on Stack Overflow, collecting and manually treating daily samples of individuals for a period of four and a half months to include 12,182 individuals in the experiment. The treatment gives one additional anonymous upvote from one of my own accounts to one of each individual's recent eligible answers and increases an individual's reputation by 10. I then track comprehensive data on individuals' subsequent daily behavior on the platform for more than four months up to July 1st, 2024.

Do exogenously assigned upvotes affect the primary outcome of interest—the decision to contribute online knowledge? I find that receiving one additional anonymous upvote substantially impacts individuals' decisions to contribute additional answers. At the extensive

on Wikipedia, more experienced contributors with a track record of valuable edits can ascend to administrator roles, which grant them the ability to delete pages, protect pages from being edited, block disruptive users, and access other administrative tools.

margin, the probability of individuals making additional contributions increases by approximately 15% of the baseline. At the intensive margin, receiving one more upvote on an answer can increase the number of answers contributed within three weeks by about 6%. These treatment effects on both margins persist for at least four months, indicating they are not driven by short-term intertemporal substitution. The effect at the extensive margin varies across individuals, with those having low-to-moderate answering experience and reputation showing the largest treatment effect. The overall quality and effort of future answers do not decrease with the treatment. There is suggestive evidence that although the treatment encourages contributions from individuals with lower past answer quality and effort, these individuals increase their effort and quality when they do contribute.

The experimental upvotes can affect an individual's future contribution decision at the extensive margin through two main channels - by shifting belief that relates to social motivation, and by giving additional reputation points, and thus changing an individual's distance to the next privilege threshold, that is, the instrumental part of the motivation. To assess the importance of the latter part, I estimate the treatment effect at the extensive margin separately for two groups of individuals: the first group, who move across the nearest reputation threshold due to the experimental upvote and is thus further from the next threshold than they were before. If instrumental motivation is critical, this could decrease future contributions. The second group, who get closer to the threshold by virtue of the experimental upvote, and thus should increase future contributions with increased instrumental incentives. I find that the impact of the treatment on the extensive margin is positive and significant in both groups, and the effect appears marginally larger for the group that is moved closer to the next privilege threshold. This suggests that the instrumental motivation to reach the next threshold is not the sole motivation.

Will the exogenously assigned upvotes affect individuals' activeness and other posting decisions on the platform? I find that the treatment does not significantly impact whether individuals login to the platform within 21 days since being sampled, but it slightly increases the number of days individuals login to the platform by 0.55, around 3.9% of the baseline. Further, the treatment does not significantly impact the probability of posting additional questions or the number of questions being posted within 21 days. The treatment significantly increases the probability of posting comments by 3.3 percentage points, and the number of comments within 21 days by around 5.8% of the control group.

How do the exogenously assigned upvotes affect how individuals evaluate others' posts? The treatment significantly impacts individuals' voting behavior within 21 days. For upvoting, the treatment significantly increases the probability by 7.30% of the baseline. The number of upvotes also increase by around 5% of the baseline. For downvoting, the treatment significantly increases the probability of making at least one downvote by around 20% of the baseline. The treatment also increases the number of downvotes by 2.1%.

With the above experimental results, it is clear that gaining recognition and obtaining privileges both motivate knowledge contribution decisions (i.e. contributing additional answers), the primary outcome of interests. However, it is unclear the extent to which social motivation of gaining reputation (due to social status, altruism, warm glow, etc.) versus instrumental motivation given by the platform (gaining more privileges with more reputation) determines the contribution decision. If social motivation is predominant, platforms might consider strategies to enhance it, such as increasing the likelihood for individuals to gain recognition. Conversely, if instrumental motivation is crucial, increasing the likelihood of individuals gaining recognition can backfire, as it makes it easier to reach privilege thresholds. Research on observational data suggests that individuals significantly reduce their contributions after gaining these platform incentives (Goes et al. 2016). In this case, instead, platforms can implement strategies to boost instrumental motivation, for example, by introducing additional privileges or adjusting the difficulty of obtaining them.

To quantify how both motivators affect the primary outcome of interest - the extensive margin of whether or not to contribute additional online knowledge, I structurally estimate the model of individual decisions with heterogeneous utility parameters to explain my experimental findings. The estimates suggest that the extent to which individuals value social motivation and instrumental platform privileges varies widely across individuals.

Using estimates from the structural model, I explore a set of counterfactuals. First, I simulate the scenario where social motivation is absent to determine its relative importance compared to instrumental motivations. The results indicate that the proportion of individuals willing to contribute drops to 26% of the baseline control group. Next, I simulate the removal of instrumental motivation. In this scenario, the proportion of contributing individuals remains at 84% of the baseline control group. The substantially larger reduction observed with the removal of social motivation suggests that social motivation plays a more critical role in influencing contribution decisions.

Therefore, I explore two platform strategies designed to enhance social motivation: amplifying the number of upvotes for answers that already have at least one upvote by highlighting those answers, and giving an additional upvote to high-quality answers. My findings indicate that tripling the number of upvotes for answers with at least one upvote can increase the proportion of contributing individuals by 15% compared to the baseline. Additionally, awarding one extra upvote to high-quality answers results in a 25% increase in contributions relative to the baseline.

This paper contributes to the broader discussion about social preferences and contributions to public goods (e.g., Lerner and Tirole (2002), Andreoni (2007)). On Chinese Wikipedia, Zhang and Zhu (2011) finds that an exogenous reduction in the group size of users decreases contribution levels of non-blocked contributors substantially due to reduced social benefits. Similarly, Wang et al. (2019) shows an exogenous expansion in the user population of a major online review platform causes individuals to write more reviews with higher quality. Gallus (2017) finds that purely symbolic awards have a sizeable effect on volunteer retention on Wikipedia. Chen et al. (2023) shows that experts contribute longer and better comments to Wikipedia when the actual match between a recommended article and an expert's expertise was better. Within the context of public good contribution, this paper contributes to the discussion about incentivizing open source and innovation (e.g., Athey and Ellison (2014)). Conti et al. (2023) uses a difference-in-differences approach and shows that receiving an extrinsic monetary reward may crowd out developers' social motivation, diverting their effort away from community and service-oriented activities on open source[3]. In terms of research context, this paper is related to Xu et al. (2020), which studies a group of particularly active Stack Overflow users who switch jobs. Those individuals were invited to apply for jobs posted on Stack Overflow Jobs and the study shows career incentives partially drive individuals' decisions to contribute knowledge using a difference-in-differences design. This paper complements their study by focusing on the vast majority of contributing users who newly registered and are much less active, and investigates the role of online reputation and platform instrumental privileges in driving users' contribution behavior[4]. The experiment is implemented during a period when Stack Overflow Jobs was disabled, when the career incentives are minimal[5].

Second, the findings add to the general understanding of motivations behind the creation of user-generated content online[6]. Several studies have documented how monetary incentives can bias buyer reviews (Cabral and Li 2015) and how combining monetary incentives and social aspects can help (Sun et al. (2017), Burtch et al. (2018)). There are also studies focusing on the role of non-monetary incentives in motivating user-generated content. Several studies shed light on this with observational data. Jin et al. (2015) presents correlational evidence from a Chinese online question-and-answer community that peer recognition is positively

[3]Incentivizing open-source contributions can be more important in the era of AI. A study by Burtch et al. (2023) shows that ChatGPT has led to large, significant declines in Stack Overflow questions, with larger effects manifesting for topics where ChatGPT is more likely to excel, based on the volume of public, online data that would have been available for training.

[4]Among the 1,301 individuals studied by Xu et al. (2020), the average number of total answers per individual is 255.59, while among the 12,182 individuals sampled in this study, the average is 18.15, and the median is 4. In fact, 99.02% of the sampled individuals have contributed less than 255.59 answers at the time of being sampled.

[5]Stack Overflow Jobs was disabled on March 31, 2022, one year and a half before the experiment started, and the announcement can be retrieved from https://meta.stackoverflow.com/questions/415293/sunsetting-jobs-developer-story. The platform introduced a significantly different version of Stack Overflow Jobs on May 8th, 2024, a few months after the field experiment in this paper concluded, and the announcement can be retrieved from https://meta.stackexchange.com/questions/399440/testing-a-new-version-of-stack-overflow-jobs.

[6]For a comprehensive discussion, please refer to Chen (forthcoming).

associated with knowledge contribution behavior. Ahn et al. (2016) develop and estimate a dynamic rational expectations equilibrium model with data on an Internet forum site. Goes et al. (2016) finds that glory-based incentives can have unintended consequences, where they encourage users to contribute before certain goals are reached, but user contributions decrease significantly after that[7]. More recently, Deolankar et al. (2023) finds that receiving negative peer feedback can also increase a user's subsequent commenting activity on Reddit. Paridar et al. (2024) finds that peer rewards and platform rewards can have opposite effects on users' posting behavior on a board game platform.

Beyond studies using observational data on user-generated content, a rapidly growing body of literature in recent years has explored non-monetary incentives through field experiments. Chen et al. (2010) finds that social comparisons can significantly lift contribution levels by below-median users on an online movie rating platform. Burtch et al. (2022) shows that anonymous peer awards on Reddit motivate individuals to generate more content on Reddit. Toubia and Stephen (2013) documents that increasing followers can have a heterogeneous treatment effect on individuals' decisions to post on Twitter. Eckles et al. (2016) finds that providing feedback increases individuals' posting on Facebook. Jiménez Durán (2021) shows that reporting hatred posts does not reduce future posts. Huang et al. (2022) shows that exogenous variation in attention and recognition on an image-sharing social network makes creators produce and share different content, than the ones that received attention and recognition. More recently, Mummalaneni et al. (2023), Srinivasan (2023), Zeng et al. (2023) and Zhang and Luo (2024) find that individuals increase content generation with more engagement or social interactions with other users.

This paper contributes to the above two streams of literature by conducting a large-scale field experiment that exogenously varies the recognition individuals receive on a leading online knowledge-sharing platform, with a focus on new and less active individuals. New

[7] Lacetera and Macis (2010) has similar findings in blood donation with a nonlinear incentive scheme, where donors' donation frequency only significantly increases right before reaching the thresholds for which the publicly-announced symbolic rewards are given.

and less active individuals comprise a large proportion of the platform's user base, making platform strategies crucial for engaging them. This study tracks sampled individuals over four months to provide evidence of the long-term treatment effects. Additionally, it models the dynamics of individual contributions, examining why individuals decide to contribute and quantifying the role of social motivation and instrumental motivation. Through counterfactual simulations, the study also explores platform strategies to encourage long-term contributions.

In terms of novelty, this is among the first field experiments on online knowledge platforms and, so far, the largest in scale. The work extends the literature on online knowledge contributions with a randomized experiment to establish causality in a more credible manner than prior studies using observational data. The field experiment is the only one that was implemented after the introduction of AIs significantly reduced individual online knowledge contributions (Burtch et al. 2023). Yet, the study shows that anonymous recognition can still substantially increase contributions, which has important implications for online knowledge platforms in the post-AI era. In addition, I present the first field experimental evidence in content contribution decisions that quantifies the relative importance of social motivation versus instrumental motivation with a structural model. Findings from this study have broad implications beyond online question-and-answer platforms, extending to other contexts where user-generated content is crucial, such as online review platforms, gaming forums, educational discussion forums, and social media sites.

The remaining parts of this paper are organized as follows. Section 2 introduces the data I use and institutional details of Stack Overflow. Section 3 presents a model of knowledge contribution. Section 4 presents the design of the field experiment. Section 5 presents the experimental results. Section 6 provides a discussion of identification and estimation strategy. Section 7 presents structural estimates. Section 8 presents simulated counterfactuals. Section 9 concludes.

2 Data and Setting

I collect user data from Stack Overflow, a leading question-and-answer website for professional and enthusiast programmers. As of April 2023, Stack Overflow had 20 million registered users, 24 million questions, and 35 million answers. 69% of all questions are answered. Each day, there are 5.9 million visits and around 3,900 questions posted[8]. According to the 2023 Developer Survey by Stack Overflow, 63% of respondents spend more than 30 minutes daily searching for answers or solutions on the platform, with 25% spending over 60 minutes per day[9].

The site and similar programming question-and-answer sites have globally mostly replaced programming books for day-to-day programming reference in the 2000s, and today are an important part of computer programming. Based on the type of tags assigned to questions, the top ten most discussed topics on the site are JavaScript, Python, Java, C#, PHP, Android, HTML, jQuery, C++, and CSS[10].

Figure 1a shows a screenshot of Stack Overflow. Similar to other online question-and-answer communities, individuals can post questions on Stack Overflow, and other individuals voluntarily post answers to the questions. When individuals log in to the platform, as shown in Figure 1a, questions are ranked and displayed in the order they were posted or updated. Questions that have most recently been posted, answered, or commented on appear at the top[11]. If individuals find either the question or the answer helpful, they can upvote it[12].

The platform designs and maintains a user reputation system. Figure 1b shows a sample question and Figure 1c shows a sample answer on Stack Overflow, where each individual's

[8]The number of registered users, questions, answers, visits, questions per day, and the percentage of answered questions are retrieved from https://stackexchange.com/sites?view=list#users, on April 14th, 2023.

[9]The information can be retrieved from, https://survey.stackoverflow.co/2023/section-productivity-impacts-daily-time-spent-searching-for-answers-solutions.

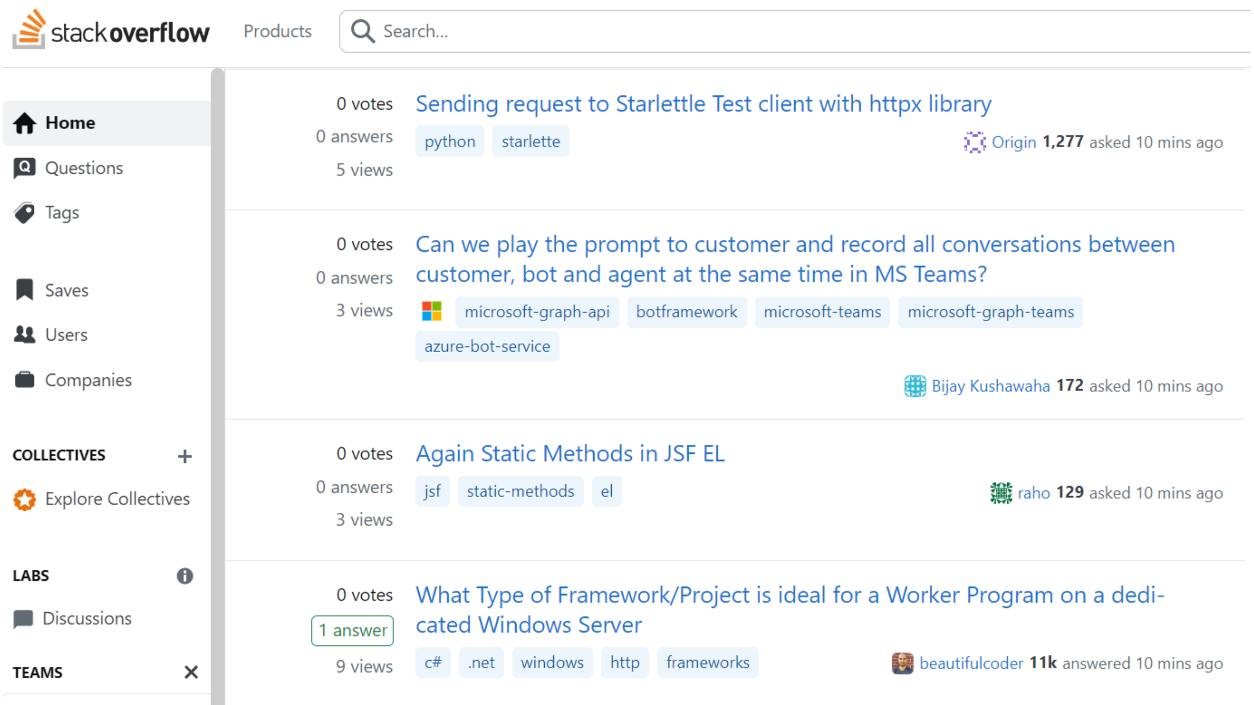
[10]This information is obtained from, https://stackoverflow.com/tags, retrieved on April 14th, 2023.

[11]This aspect of the platform is well-suited for the experiment, as the experimental treatment of one additional upvote is unlikely to directly influence the ranking of questions on the front page. This minimizes concerns regarding potential interference from platform-driven recommendations in the question ranking process.

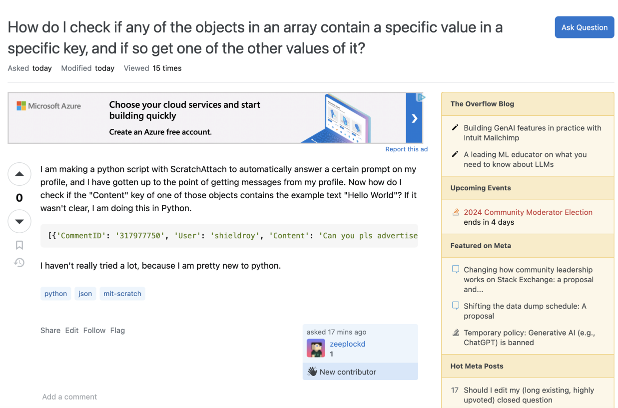
[12]In addition, the individual who asks questions can mark one of the answers to the question received as "accepted". Accepting does not necessarily mean it is the best answer, it just means that it worked for the individual who asked.

Figure 1: Stack Overflow and Sample Content

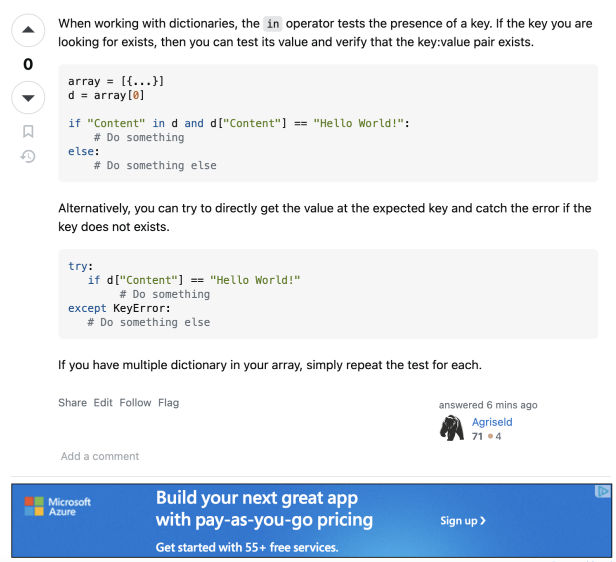
(a) Stack Overflow Front Page



(b) Stack Overflow - Sample Question



(c) Stack Overflow - Sample Answer



username and reputation points are displayed at the end of their question or answer. An individual's reputation score increases when others vote up their questions or answers. Individuals will receive 10 reputation points when a question or answer they posted receives one upvote[13]. Upvoting others' questions or answers will not give additional reputation points to those who upvoted. If individuals deem questions or answers unhelpful, they have the option to downvote them. In such cases, both the users who receive downvotes and those who issue them will incur a penalty of two reputation points each. In this way, the platform encourages individuals to exert more caution when downvoting others. In addition, advertisements are shown at the beginning and the end of the page, which is how the platform profits from user-generated content.

The platform allows individuals to unlock new platform privileges as they reach reputation thresholds. For instance, as shown in Figure 2a, individuals need at least 10 reputation points to remove new user restrictions and create wiki posts, the privilege to upvote other individuals' answers is granted with at least 15 reputation points, only individuals with more than 50 reputation points are allowed to comment everywhere, and the highest privilege threshold is 25,000, which enable individuals to access to internal and Google site analytics of Stack Overflow. The privilege thresholds are more concentrated at lower reputation levels, and more sparsely located at higher reputation levels. There are 12 privilege thresholds at or below 200 reputation points. Between 250 and 25,000 reputation points, there are 13 additional privilege thresholds.

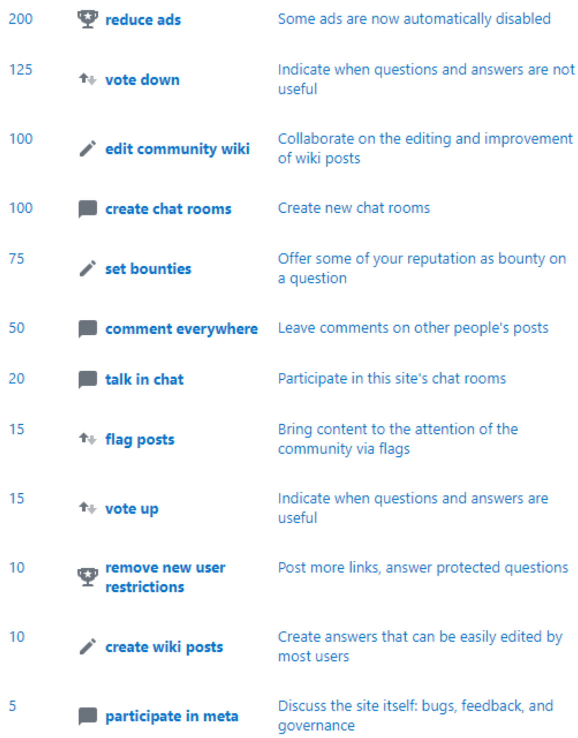
As individuals reach privilege thresholds, the platform prominently displays the remaining distance to the next threshold at the top of their profile page. This feature is designed to motivate users to contribute more actively. Figure 2d illustrates examples of two individuals, each with a distinct reputation level. The green bar at the bottom of each profile indicates the next privilege to be gained and the distance to the next threshold.

[13]Individuals will gain 15 more reputation points if one of their answers is accepted by the individual who asked the question they answered.

Figure 2: Privileges and Distances

(a) Thresholds for Different Privileges

(i) Below 200



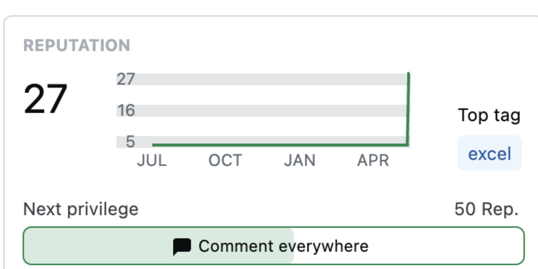
200	reduce ads	Some ads are now automatically disabled
125	vote down	Indicate when questions and answers are not useful
100	edit community wiki	Collaborate on the editing and improvement of wiki posts
100	create chat rooms	Create new chat rooms
75	set bounties	Offer some of your reputation as bounty on a question
50	comment everywhere	Leave comments on other people's posts
20	talk in chat	Participate in this site's chat rooms
15	flag posts	Bring content to the attention of the community via flags
15	vote up	Indicate when questions and answers are useful
10	remove new user restrictions	Post more links, answer protected questions
10	create wiki posts	Create answers that can be easily edited by most users
5	participate in meta	Discuss the site itself: bugs, feedback, and governance

(ii) Above 200

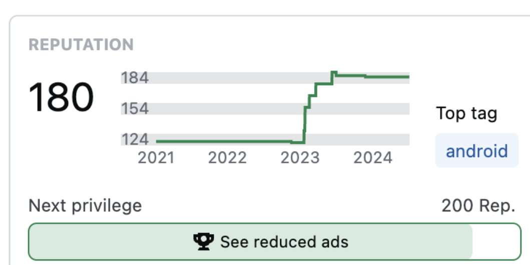
25,000	access to site analytics	Access to internal and Google site analytics
20,000	trusted user	Expanded editing, deletion and undeletion privileges
15,000	protect questions	Mark questions as protected
10,000	access to moderator tools	Access reports, delete questions, review reviews
5,000	approve tag wiki edits	Approve edits to tag wikis made by regular users
3,000	cast close and reopen votes	Help decide whether posts are off-topic or duplicates
2,500	create tag synonyms	Decide which tags have the same meaning as others
2,000	edit questions and answers	Edits to any question or answer are applied immediately
1,500	create tags	Add new tags to the site
1,000	established user	You've been around for a while; see vote counts
1,000	create gallery chat rooms	Create chat rooms where only specific users may talk
500	access review queues	Access the First posts and Late answers review queues
250	view close votes	View and cast close/reopen votes on your own questions

(b) Distance to the Next Threshold

(i) Example 1



(ii) Example 2



3 A Model of Knowledge Contribution

3.1 Overview of the Model

To guide the experiment design, I specify a model that examines how individuals form expectations over time regarding the number of upvotes they will receive for contributing an additional answer on Stack Overflow, the online platform dedicated to technical programming-related questions.

3.2 Model Assumptions

Individual i values cumulative reputation points they already have on the platform at time t, denoted by R_{it}.

In addition, individual i values the number of privileges gained through crossing privilege thresholds, denoted by $q(R_{it})$. Note that $q(.)$ is a deterministic function, in the sense that given R_{it}, the individual i knows exactly how many privileges they have. More specifically, $q(R_{it})$ is a step function that increases by one as R_{it} reaches a certain threshold. Individual i's utility function follows:

$$U_i(R_{it}) = \gamma_i R_{it} + \varphi_i q(R_{it})$$

Intuitively, γ_i represents the social utility of gaining reputation, where it can be due to social status, altruism, warm glow, etc. And φ_i represents the instrumental utility, where additional reputation will help individuals gain additional privileges specified by the platform. Note that $\gamma_i = 0$ is a special case where individual i only values the instrumental utility of gaining additional privileges.

Let s_{it} denote the number of upvotes the individual i may get if they contribute an additional answer at time t. This will give individual i $10s_{it}$ additional reputation points, which means R_{it} evolves according to:

$$R_{it+1} = R_{it} + 10s_{it}$$

The number of upvotes per answer follows a Poisson distribution with parameter λ_{it}, which is also its expected value:

$$s_{it} \sim Poisson\left(\lambda_{it}\right)$$

$$E[s_{it}] = \lambda_{it}$$

Individual i at time period 0 has a prior belief about the distribution parameter λ_{i0} of the number of upvotes s_{i0} they will get if contributing an answer.

$$\lambda_{i0} \sim \Gamma\left(k_{i0}, \theta_{i0}\right)$$

Given the prior, the expected number of upvotes per answer is:

$$E[s_{i0}] = E[\lambda_{i0}] = k_{i0}\theta_{i0}$$

With each answer contribution, i will update their prior belief through Bayesian updating, with posterior:

$$\lambda_{it} \sim \Gamma\left(k_{i0} + \sum_{t=1}^{n} s_{it}, \frac{\theta_{i0}}{n\theta_{i0} + 1}\right)$$

where $\sum_{t=1}^{n} s_{it}$ is the sum of upvotes from historical answers, and n is the number of past answers up till time t.

Given the posterior, the expected number of upvotes per answer is:

$$E[s_{it}] = E[\lambda_{it}] = \left(k_{i0} + \sum_{t=1}^{n} s_{it}\right) \cdot \frac{\theta_{i0}}{n\theta_{i0} + 1}$$

Individual i has an effort cost c_i to contribute answers. ϵ_{it} follows a Type I Extreme Value distribution to account for random variations in the effort cost.

Let $I_{it} = \{\gamma_i, \varphi_i, c_i, R_{it}, k_{it}, \theta_{it}\}$ denotes the information set available to i at time t.

Consider individual i's decision at time $t + 1$.

Individual i will make the contribution decision based on the expected utility to gain from getting the reputation points:

$$\Delta U_{it} = E\left[U_i(R_{it} + 10s_{it}) \mid I_{it}\right] - U_i(R_{it}) - c_i + \epsilon_{it}$$

The individual i will choose to contribute an answer if the expected utility to gain is larger than 0.

$$D_{it} = \begin{cases} 1 & \Delta U_{it} > 0 \\ 0 & \text{otherwise} \end{cases}$$

In the model above, one additional upvote on one of individual i's past answers will shift individual i's posterior belief to:

$$\lambda_{it} \sim \Gamma\left(k_{i0} + \sum_{t=1}^{n} s_{it} + 1, \frac{\theta_{i0}}{n\theta_{i0} + 1}\right)$$

Given the shifted posterior, the expected number of upvotes per answer is:

$$E[s_{it}] = E[\lambda_{it}] = \left(k_{i0} + \sum_{t=1}^{n} s_{it} + 1\right) \cdot \frac{\theta_{i0}}{n\theta_{i0} + 1}$$

In addition, since one additional upvote will give individual i 10 additional reputation points, it will also change individual i's reputation R_{it} and the number of privileges gained through reaching privilege thresholds, $q(R_{it})$. The additional reputation points will also change individual i's distance to the next privilege. D_{it} will therefore be changed.

Below are three model predictions about how an additional upvote will change individual i's decision.

Prediction 1. Given posterior belief, γ_i, φ_i and R_{it}, the lower the n, the increase in individual i's probability of contributing another answer is larger with an additional upvote.

Prediction 2. Given posterior belief, γ_i, φ_i and R_{it}, if the individual i expects to reach the next privilege threshold with one more contribution, then the increase in individual i's probability of contributing another answer is larger with an additional upvote.

Prediction 3. Given posterior belief, γ_i, φ_i and R_{it}, if the individual i expects to reach the next privilege threshold with one more contribution, the higher the φ_i, then, the increase in individual i's probability of contributing another answer is larger with an additional upvote.

With the above predictions in mind, I designed and implemented a field experiment to measure the causal effect of receiving an additional upvote on individuals' online knowledge contribution decisions.

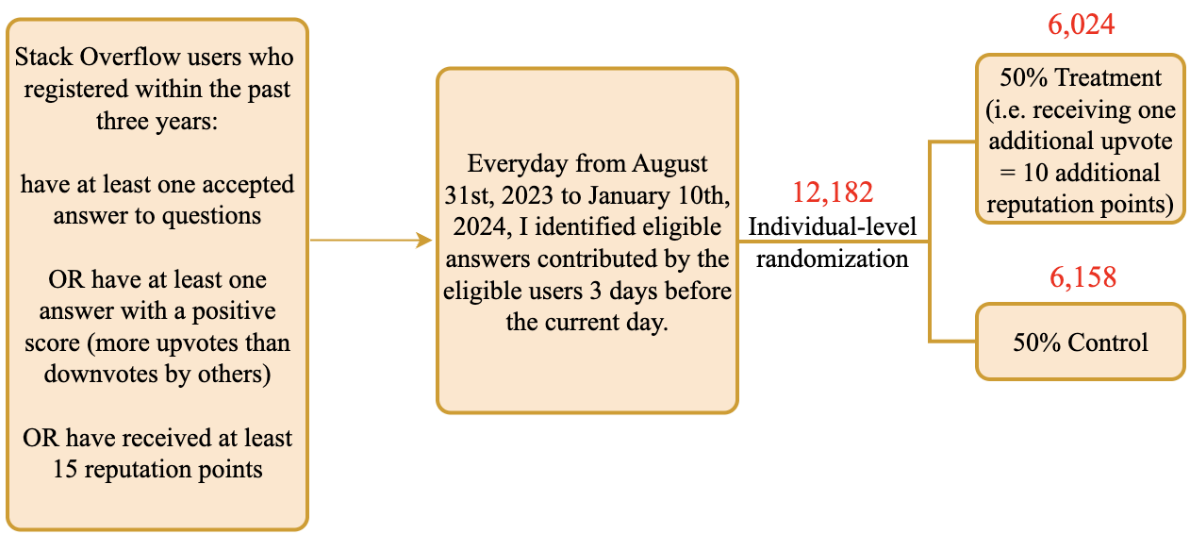
4 Experiment Design

The experiment design is summarized by Figure 3. I first identified the individuals who were eligible to enter my experimental sample. I focused on individuals who registered on Stack Overflow on or before August 31st, 2020, as this subgroup of individuals is relatively newly registered since Stack Overflow's foundation in 2008. I queried Stack Overflow's API to obtain detailed data of all of the 985,841 answers contributed by those 304,617 individuals on or before August 27th, 2023. With the data, I further restricted eligible individuals to those who have at least minimal experience in making meaningful contributions to the platform. More specifically, they had to satisfy at least one of the following criteria:

- Have contributed at least one accepted answer to questions

- Have contributed at least one answer with a positive score (more upvotes than downvotes by other users)
- Have received at least 15 reputation points

Figure 3: Experiment Design



By imposing the minimal criterion above, I was able to include the vast majority of individuals who can make meaningful knowledge contributions to the platform. In this way, I abstracted away from individuals whose contributions receive negative scores (more downvotes than upvotes) from the online community, as those can still be learning about how to use the platform, how to write good answers, how to find good questions to answer, etc.

To implement the experiment, I registered 23 accounts on Stack Overflow. Then, over months, I answered technical programming questions on the platform, thereby gaining the privilege of upvoting others' answers for each account.

From August 31st, 2023 to January 10th, 2024, I collected daily samples of answers contributed three days before the sampling date by the pre-determined set of eligible individuals. Every day at midnight from 0:00 a.m. to 5:00 a.m. Eastern Time, I sent queries to Stack Overflow API to identify answers posted three days before the current date. The three-day lag is to ensure that there is enough time for the answer to be evaluated by other individuals on the platform. During the three-day lag, the individual who contributed the answer can delete the answer, the individual who asked the question can choose to delete the question along with its answers, platform moderators can delete the question or answer if they find them inappropriate, and all individuals on the platform can anonymously upvote the answer if they find it helpful, or anonymously downvote the answer if they find it unhelpful. I restricted the eligible answers to those that are plausibly high-quality (i.e., not being deleted and having no more downvotes than upvotes within three days of posting) by eligible individuals. I used organic decisions by individuals on the platform as a proxy of answer quality. The inclusion of only plausibly high-quality answers ensures there is no deception and no violation of platform policies.

Right after determining the eligible individuals who contributed high-quality answers, I retrieved each individual's detailed full history of past answers, questions, and scores. I conducted randomization on the daily sample at the individual level by assigning 50% of individuals to the control group, and the other 50% to the treatment group. Individuals in the treatment group would receive one manual upvote from one of my accounts for one eligible answer they contributed three days before.

Due to stringent platform moderation policies, I implemented the experiment treatment by myself at midnight for 4-5 hours per day across my 23 accounts for four and a half months.

In the case that an individual in the treatment group contributed multiple plausibly high-quality answers within that day, only the last qualified answer of the day would be treated. Essentially, the randomization is at the individual level, i.e. each treated user would only receive one upvote on one of the answers they posted three days before, and individuals who were included in either the control or the treatment group in previous days would not be sampled again in the future. This is to make sure that the treatment intensity is independent of the individuals' baseline activity level.

To ensure enough newly registered individuals and individuals with relatively low answering experience on the platform are included in the experiment, I expanded the eligible individuals to include those who newly satisfied the criteria during the past week, once per week during the four-month experiment period. This expansion allowed me to include individuals who registered before the experiment began and just satisfied the inclusion criteria in the past week, along with individuals who newly registered and satisfied the inclusion criteria in the past week.

The treatment would give 10 additional reputation points to the individual who contributes the answer and would indicate one additional anonymous individual finds one of their past answers helpful[14].

After implementing the treatment, I tracked each sampled individual's detailed subsequent answers, questions, comments, and profile information at a daily level for several months.

Table 1: Balance Table of Variables for the Treatment and Control Groups

Variable	**Control Mean**	**Treatment Mean**	**t-Statistic**	**p-Value**
Profile View Count	40.875	42.368	-0.186	0.852
Reputation	325.568	282.192	0.997	0.319
Reputation Change Last Week	9.865	10.043	-0.395	0.693
Reputation Change Last Month	17.034	16.262	0.643	0.520
Reputation Change Last Quarter	33.760	31.602	0.790	0.430
Reputation Change Last Year	115.916	102.839	1.010	0.313
Number of Past Answers	19.187	17.088	0.931	0.352
Number of Past Questions	3.008	3.045	-0.263	0.792
Profile Information Filled	0.551	0.542	1.051	0.293
Location Information Filled	0.335	0.334	0.195	0.845
Number of Users	6158	6024		

Notes: The t-stats reported follow the Welch's two-sample t-test.

In total, I sampled 12,182 individuals during the experiment period of over four months.

[14]The experiment carefully followed Stack Overflow's terms of usage and received IRB approval from the University of California, Berkeley before implementation. The IRB approval number is 2023-04-16245.

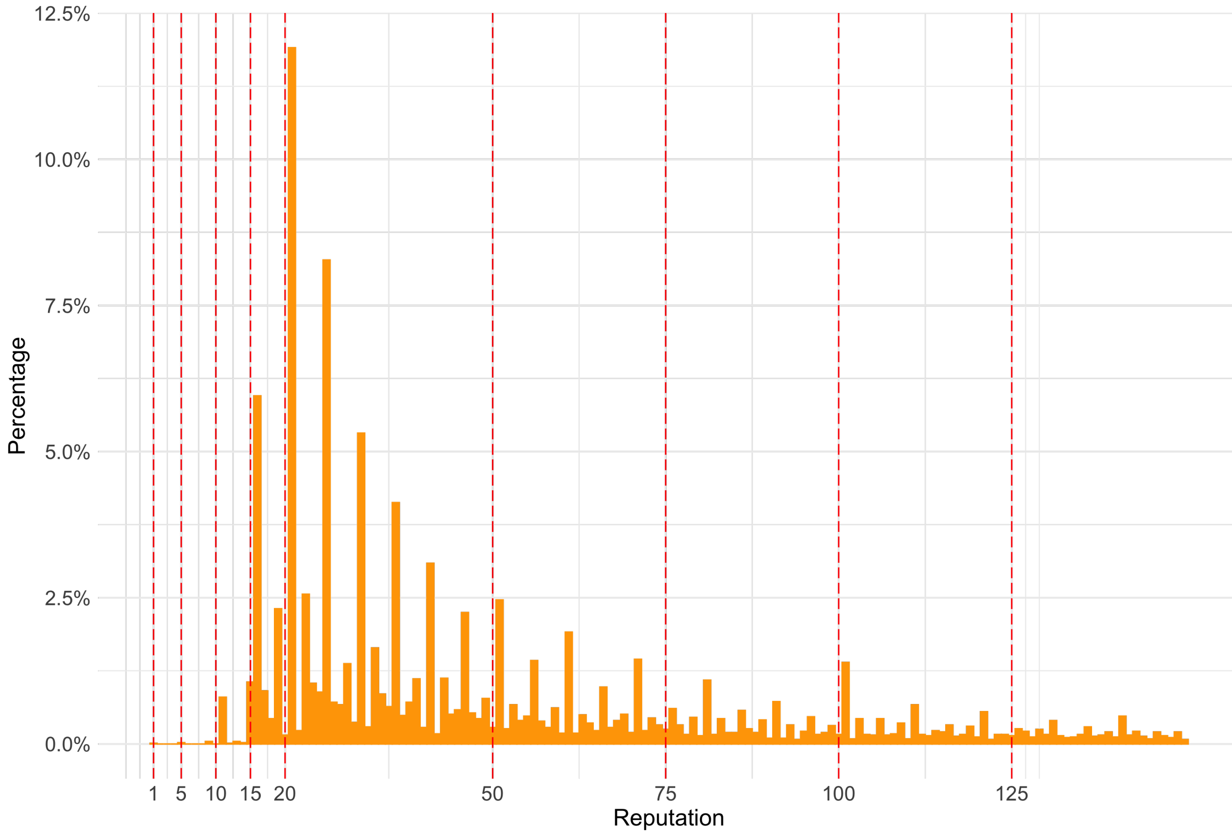
Among the sampled individuals, 6,024 are in the treatment group, and 6,158 are in the control group. I implemented the experiment for 132 consecutive days, and on average, I sampled 92 individuals per day. Table 1 shows the control group and treatment group are balanced on all key variables: profile view count by other users, reputation points, recent reputation changes, number of past answers, number of past questions, and availability of profile and location information.

Figure 4b shows the distribution of the total number of past answers by each individual at the time of being sampled. The distribution is right-skewed, with a majority of individuals having a low level of past answering experience, and a few individuals having a large number of past answers. The median individual in the sample contributed 4 answers in the past, and the average of the number of past answers is 18.15. Figure 4a shows the distribution of reputation for each individual at the time of being sampled. The red dashed vertical lines correspond to the privilege thresholds in Figure 2a. There is a mass of individuals with reputation points right above the thresholds. For example, there is a high proportion of individuals with reputation points right above 15, 20, and 50. However, there is also a substantial number of individuals who are less than 10 reputation points below the next privilege threshold, which means that one upvote from the experiment will make them cross the next privilege threshold and face another threshold - 3,044 individuals fall into this category. Overall, the distribution of reputation is right-skewed, with a median of 51 and a mean of 304.1. The additional 10 reputation points by the experiment treatment will move up the median individual's reputation by around 19.61%.

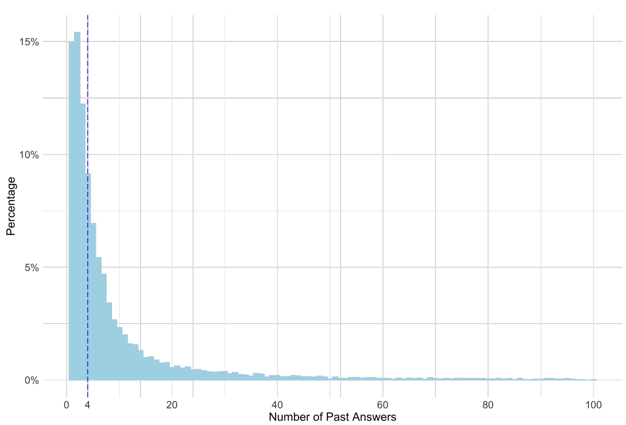
Figure 4c shows the distribution of the past average answer score (the number of upvotes minus the number of downvotes) for each individual at the time of being sampled. The distribution is also right-skewed, with a median of 0.50 and a mean of 0.78.

Figure 4: Distributions of Reputation Points, Number of Past Answers, and Average Score for the Experimental Sample

(a) The Percentage Distribution of Reputation Points at the Time of Being Sampled



(b) The Percentage Distribution of Number of Past Answers at the Time of Being Sampled



(c) The Percentage Distribution of Average Past Score at the Time of Being Sampled

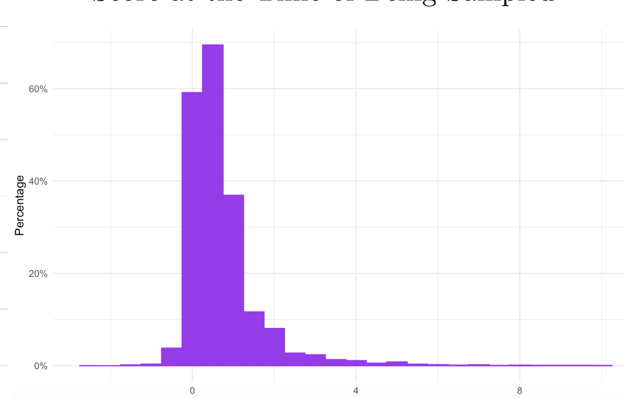
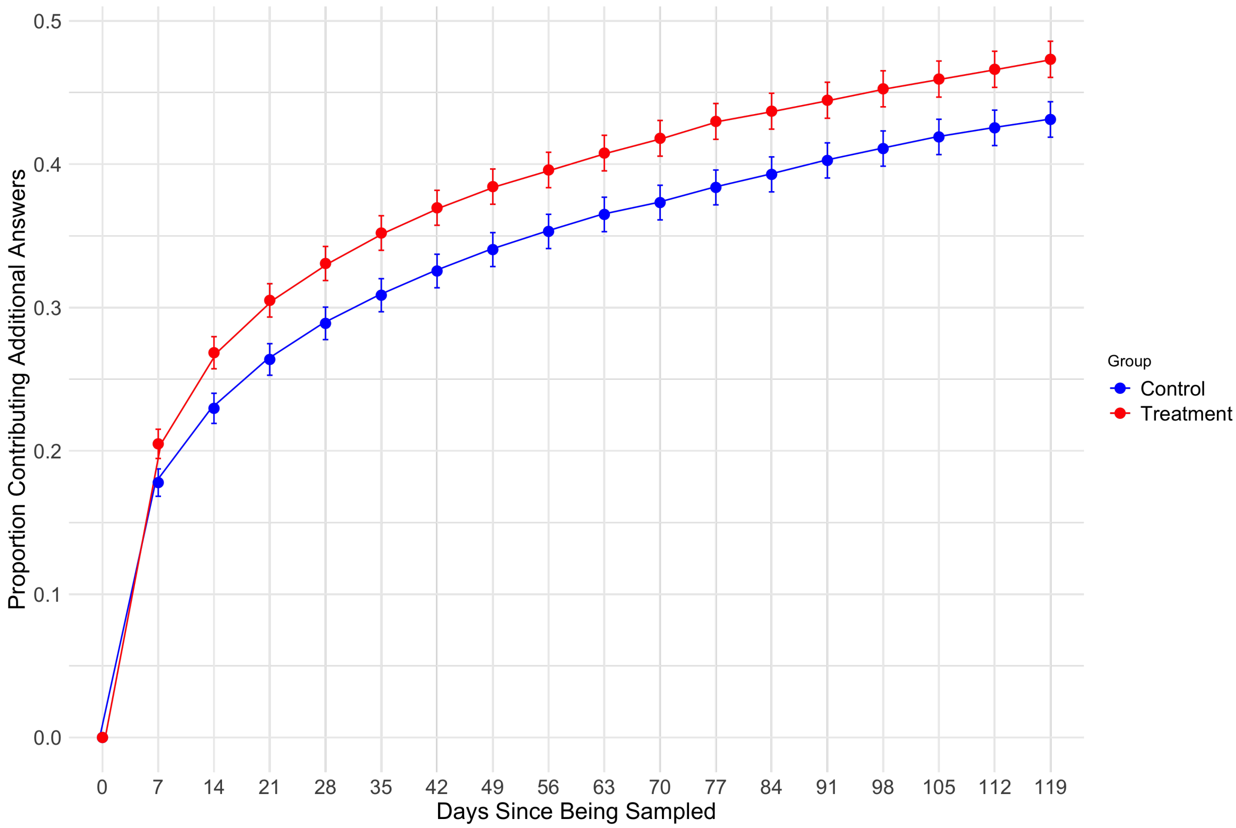


Figure 5: The Proportion of Individuals Contributing Additional Answers in Treatment and Control Groups Over Time



5 Results

5.1 Main Results - Extensive Margin of Knowledge Contribution

Receiving one additional anonymous upvote significantly increases the probability of contributing another answer. Figure 5 shows the proportion of individuals contributing at least one additional answer in treatment and control groups after being sampled, measured over different time horizons. The proportion in both the control and treatment grows over time. In the control group, there are 17.8% individuals who contributed additional answers within 7 days, and the number rises to 26.1% within 21 days, and 35.3% within 56 days. In the treatment group, there are 20.5% individuals who contributed additional answers within 7 days, and the number rises to 30.5% within 21 days, and 39.6% within 56 days. The difference in the proportion between the treatment and control groups is statistically significant. It gradually increases from 2.7 percentage points within the first 7 days, to 4.1 percentage points within the first 21 days and remains stable at around 4.3 percentage points later, up to at least four months after being sampled. This suggests that the treatment effect is not driven by intertemporal substitution over time (i.e. individuals decide to make a sooner contribution with the treatment, but they would have contributed later had they not received the treatment).

Table 2: Contributing Additional Answers Over Time (Extensive Margin)

	Dependent variable: I(Additional Answers)			
	Within 7 Days	Within 14 Days	Within 21 Days	Within 120 Days
Treatment	0.027***	0.040***	0.042***	0.042***
	(0.007)	(0.008)	(0.008)	(0.009)
Observations	12,182	12,182	12,182	12,182
R^2	0.061	0.065	0.061	0.038
Adjusted R^2	0.047	0.051	0.048	0.024
Dep. Var. Mean	0.178	0.229	0.261	0.432
Percentage of Baseline	15.169	17.467	16.092	9.722
Sample Day FE	*Yes*	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*	*Yes*

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

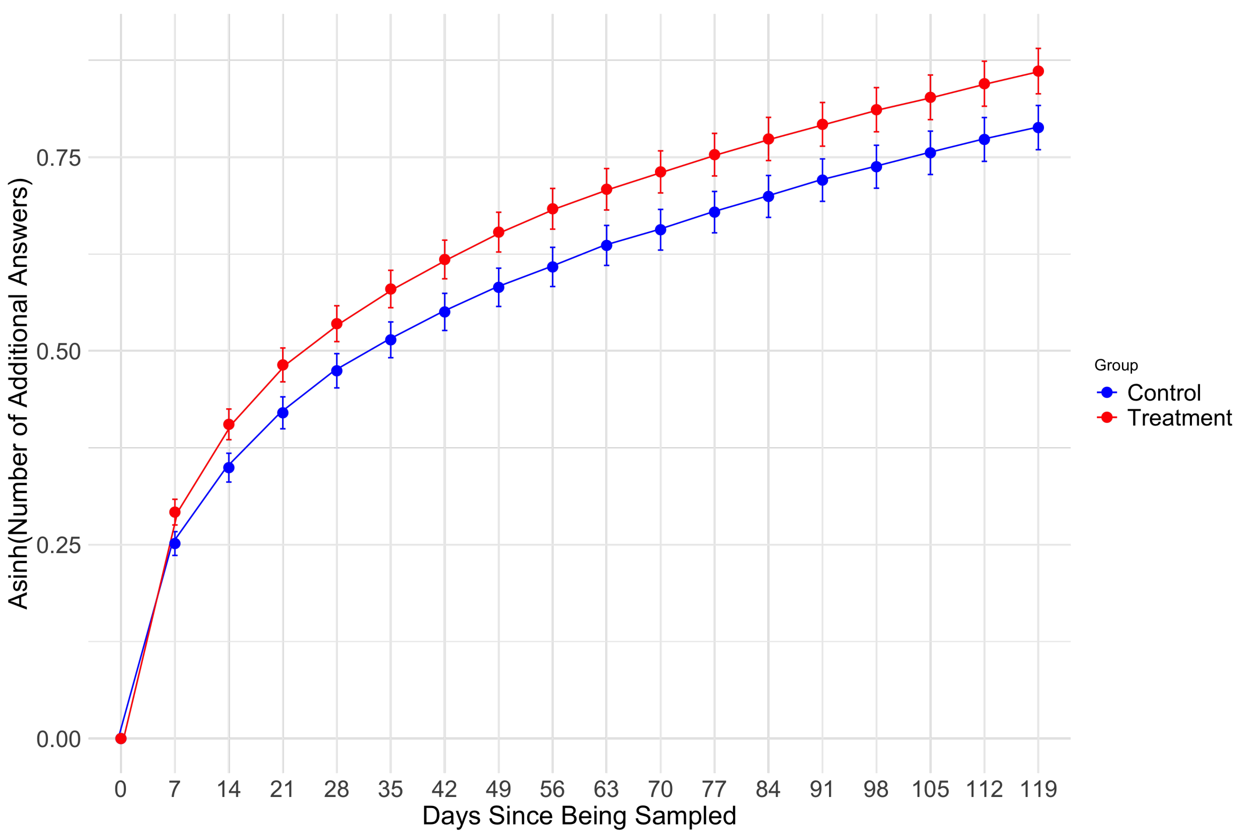
Note: The outcome variables are whether the individual contributed at least one additional answer within different numbers of days of being sampled. The standard error is clustered at the sample day level.

Table 2 shows linear probability regressions with the dependent variable being whether the individual contributed at least one additional answer within 7 days, 14 days, 21 days, and 120 days since being sampled. The regressions included sample-day fixed effects and registration-month fixed effects, with standard errors clustered at the sample-day level. Each of the 12,182 observations in the regressions corresponds to an individual from either the treatment or control group. The treatment of an additional upvote increases the probability of contributing answers within 7 days by 2.7 percentage points, within 14 days by 4.0

percentage points, within 21 days by 4.2 percentage points, and within 120 days by 4.2 percentage points. These treatment effects measured with different time windows correspond to 15.2%, 17.5%, 16.1%, and 12.2% of the proportion who contributed answers in the control group within the corresponding time window. Consistent with the pattern in Figure 5, the treatment effect gradually increases over time and stabilizes at around 21 days after being sampled, as it may take time for treated individuals to notice the additional upvote and find appropriate questions to answer. The subsequent analysis will use 21 days as the time period for outcome measurement.

5.2 Main Results - Intensive Margin of Knowledge Contribution

Figure 6: The Number of Additional Answers Per Individual in Treatment and Control Groups Over Time



Receiving one additional anonymous upvote significantly increases the number of answers

Table 3: The Number of Additional Answers Contributed within 21 Days (Intensive Margin)

	Dependent variable: Number of Additional Answers			
	Asinh(Number)	Poisson(Number)	Number	Number(Winsorized at 5)
Treatment	0.063***	0.141***	0.162**	0.093***
	(0.016)	(0.017)	(0.082)	(0.025)
Observations	12,182	12,182	12,182	12,182
R^2	0.070		0.094	0.087
Adjusted R^2	0.057		0.081	0.074
AIC		53,469.810		
Dep. Var. Mean	0.420		1.033	0.639
Sample Day FE	*Yes*	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*	*Yes*

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

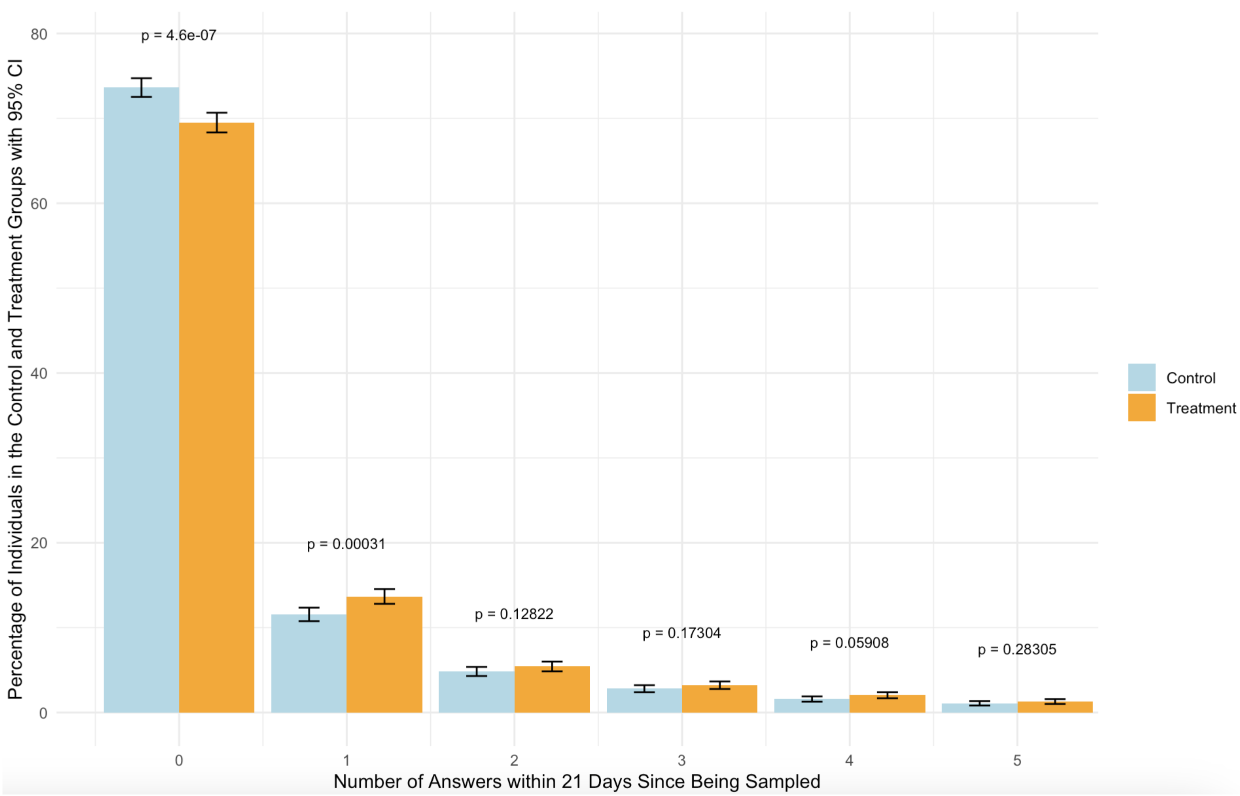
Note: The outcome variables are the number of answers, the number of answers winsorized at 5, and the arcsinh of the number of answers contributed by the individual within 21 days of being sampled. The standard error is clustered at the sample day level for the first, third, and fourth columns.

contributed. Figure 6 shows the average asinh of the number of additional answers per individual in treatment and control groups over time. The gap between the treatment and control groups gradually widens over time within the first 21 days since being sampled, and then stays stable over time for at least four months. Table 3 presents several different specifications with outcomes being the number of answers contributed within 21 days of being sampled. Overall, receiving one additional upvote increases the number of answers per individual contributed within 21 days by 0.171, corresponding to around 16.554% of the number for the control group.

To further understand how the treatment shifts the distribution of the number of answers within 21 days of being sampled, I generate Figure 7 for the percentage distribution of the number of answers by individuals in the control and treatment groups. The x-axis is truncated at the 95th percentile, 5, focusing on the most common values. Overall, in the control group, 73.6% individuals do not contribute any answer, 11.6% individuals contribute

1 additional answer, and the remaining 14.8% contribute more than 1 additional answer. In the treatment group, 69.5% individuals do not contribute any answer, 13.7% individuals contribute 1 additional answer, and the remaining 16.8% contribute more than 1 additional answer. Comparing the distribution between the treatment and control groups, the treatment group has a significantly smaller percentage of individuals who did not contribute any additional answer, and a significantly higher percentage of individuals contributing one additional answer, compared with the control group. For those who contributed more than one additional answer, the distributions in the treatment and control groups appear very similar, and the treatment group has slightly more individuals contributing more than one answer.

Figure 7: The Percentage Distribution of Number of Answers within 21 Days Since Being Sampled



5.3 Main Results - Other Outcomes

Receiving one additional anonymous upvote has no significant effect on whether the individual decides to login at least once within 21 days of being sampled (Table 4a). This is as expected - there is no email or other external notification being sent to individuals when they receive an upvote, so the treatment effect on the primary outcome, answer contribution, is not through a pure notification channel where individuals are notified to login to the platform. However, the treatment does significantly increase the number of days individuals login to the platform with 21 days of being sampled. The magnitude is small - the number of days is increased by 0.255, corresponding to around 3.9% of the baseline.

Receiving one additional anonymous upvote does not significantly impact the probability of posting additional questions or the number of questions being posted within 21 days (Table 4b). If individuals merely enjoy posting content with more reputation or expect to use reputation points as a signal to others to attract more attention to the questions they post, they would have also increased the probability of posting questions or the number of questions. The null effect suggests this is not the case. The fact that receiving an additional upvote on a past answer does not impact an individual's decision to post a question also suggests that individuals do not update their belief about reputation to gain from questions based on reputation gained from their answers. However, the treatment significantly increases the probability of posting comments by 3.3 percentage points, and the number of comments within 21 days by around 5.8% of the control group (Table 4b). The reason is that to answer a question, individuals may need to comment on the question to ask for clarifying details, and upon answering, individuals may need to comment to reply to others who comment on their answers.

Receiving one additional anonymous upvote also significantly impacts the voting behavior within 21 days (Table 5). For upvoting, the treatment significantly increases the probability of making at least one upvote by 2.7 percentage points, corresponding to 7.30% of the

Table 4: Summary of Login Activities and Additional Contributions

(a) Panel A: Login Activities

	Dependent variable:	
	I(Login)	Asinh(Num. Login Days)
Treatment	0.001	0.039**
	(0.004)	(0.018)
Observations	12,176	12,176
R^2	0.023	0.037
Adjusted R^2	0.009	0.023
Dep. Var. Mean	0.929	2.559

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

(b) Panel B: Contributing Additional Questions and Comments

	Dependent variable:			
	I(Questions)	Asinh(Num. Questions)	I(Comments)	Asinh(Num. Comments)
Treatment	0.008	0.007	0.033***	0.058***
	(0.005)	(0.006)	(0.008)	(0.020)
Observations	12,182	12,182	12,182	12,182
R^2	0.018	0.018	0.004	0.004
Adjusted R^2	0.049	0.055	0.028	0.057
Dep. Var. Mean	0.090	0.100	0.254	0.485

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Note: In panel A, the outcome variables are whether the individual login at least once and the arcsinh of the number of days the individual login within 21 days of being sampled. The standard error is clustered at the sample day level. In panel B, the outcome variables are whether the individual contributed at least one additional question, arcsihn of the number of additional questions contributed, whether the individual contributed at least one additional comment, and arcsinh of the number of additional comments contributed within 21 days of being sampled.

baseline. The number of upvotes also increase by around 5% of the baseline. For downvoting, the treatment significantly increases the probability of making at least one downvote by 1.1 percentage points. Although the magnitude of the effect on downvoting is smaller than the magnitude of the effect on upvoting, this is mostly due to a very low baseline - only 5.5% of individuals in the control group make at least one downvote within 21 days, and the treatment increase the probability of having at least one downvote by around 20% of the baseline. The treatment also increases the number of downvotes by 2.1%.

Table 5: Upvoting and Downvoting

	Dependent variable:			
	I(Upvotes)	Asinh(Num. Upvotes)	I(Downvotes)	Asinh(Num. Downvotes)
Treatment	0.027***	0.050**	0.011***	0.021**
	(0.008)	(0.020)	(0.004)	(0.009)
Observations	12,050	12,050	12,050	12,050
R^2	0.021	0.022	0.035	0.037
Adjusted R^2	0.007	0.008	0.021	0.023
Dep. Var. Mean	0.370	0.662	0.055	0.086
Sample Day FE	*Yes*	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*	*Yes*

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Note: The outcome variables are whether the individual did at least one upvote, arcsihn of the number of additional upvotes, whether the individual did at least one downvote, and arcsihn of the number of additional downvotes within 21 days of being sampled.

5.4 Results - Comparative Statics, Answer Experience

Receiving one additional anonymous upvote has a heterogeneous effect on individuals with different levels of answer experience. Table 6a presents the results splitting the sampled individuals by quartiles of the number of past answers they have contributed at the time of being sampled. Individuals with 1 - 2 past answers have an insignificant treatment effect of 2.36 percentage points, around 16.28% of the control group. Individuals with 3 - 4 past answers experience a significant treatment effect of 4.61 percentage points, around 19.79% of the control group. Individuals with 5 - 10 past answers have the largest treatment effect of 6.30 percentage points, corresponding to around 23.16% of the control group. Individuals who are more experienced with more than 10 past answers at the time of being sampled have a slightly lower treatment effect of 3.77 percentage points, around 9.15% of the control group.

5.5 Results - Comparative Statics, Reputation

Receiving one additional anonymous upvote has a heterogeneous effect on individuals with different reputation points. Table 6b presents the results splitting the sampled individuals by quartiles of reputation points at the time of being sampled. Individuals with 1 - 26 reputations have a significant treatment effect of 3.32 percentage points, around 15.16% of the control group. Individuals with 27 - 51 reputation points have a larger and more significant treatment effect of 5.69 percentage points, around 24.85% of the control group. Individuals with 52 - 142 reputation points have the largest treatment effect of 6.84 percentage points, corresponding to around 28.62% of the control group. Individuals who have most reputation points belonging to the fourth quartile with more than 142 reputation points at the time of being sampled have a smaller and insignificant treatment effect of 3.27 percentage points, around 9.19% of the control group.

5.6 Results - Comparative Statics, Distance to Privilege Thresholds

Receiving one additional anonymous upvote has a heterogeneous effect on individuals with different distances to the next privilege threshold. Table 7 presents results where I split individuals based on their distance to the next privilege threshold at the time of being sampled, and how the experimental upvotes will change their motivations.

Table 6: Heterogeneity of Treatment Effect on Answer Contributions

(a) Panel A: Contributions by Previous Experience

	Dependent variable: I(Additional Answers)			
	1-2 Past Answers	3-4 Past Answers	5-10 Past Answers	Above 10 Past Answers
Treatment	0.025**	0.044***	0.063***	0.038**
	(0.012)	(0.015)	(0.015)	(0.018)
Observations	3,596	2,529	3,017	3,040
R^2	0.064	0.125	0.165	0.163
Adjusted R^2	0.017	0.061	0.115	0.113
Dep. Var. Mean	0.145	0.233	0.272	0.412
Percentage of Baseline	16.276	19.785	23.162	9.150
Sample Day FE	*Yes*	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*	*Yes*

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

(b) Panel B: Contributions by Reputation

	1-26 Reputation	27-51 Reputation	52-142 Reputation Reputation	Above 142 Reputation
Treatment	0.0332*	0.0569***	0.0684***	0.0327
	(0.0149)	(0.0166)	(0.0166)	(0.0172)
R^2	0.0724	0.1093	0.1122	0.1573
Adj. R^2	0.0215	0.0510	0.0552	0.1059
Num. obs.	3312	2799	2850	2993
Dep. Var. Mean	0.219	0.229	0.239	0.356
Percentage of Baseline	15.160	24.847	28.619	9.185
Sample Day FE	*Yes*	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*	*Yes*

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: In panel A, the outcome variables are whether the individual contributed at least one additional answer within 21 days of being sampled, split by the number of past answers the individual has contributed before being sampled. In panel B, the outcome variables are whether the individual contributed at least one additional answer within 21 days of being sampled, split by the number of reputation points the individual has before being sampled.

Figure 8: The Change of Social and Instrumental Motivations with the Treatment

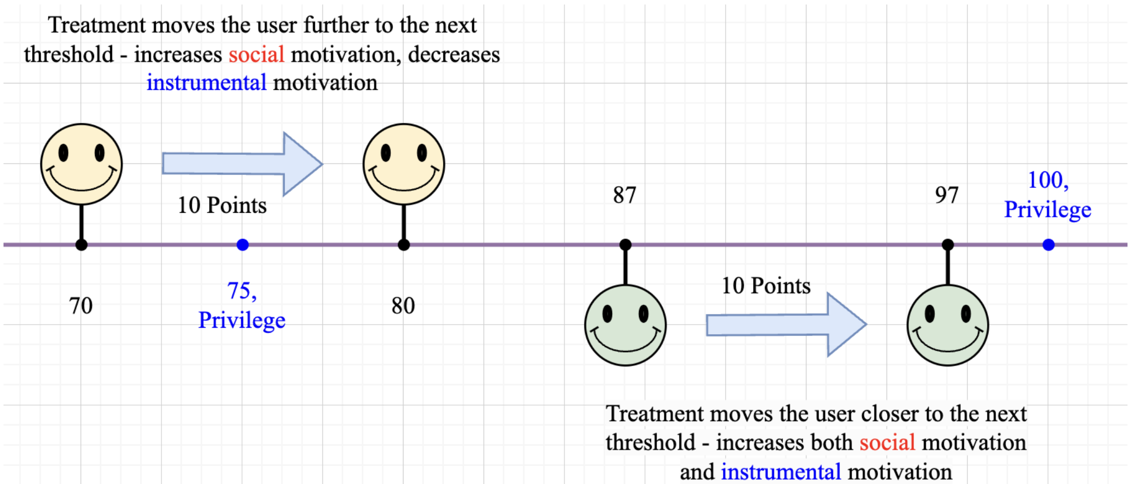


Table 7: Heterogeneity of Treatment Effect on Answer Contributions by Distance to the Next Privilege Threshold Before Being Treated

	Dependent variable: I(Additional Answers)		
	1-10 Distance	11-20 Distance	More than 20 Distance
Treatment	0.040**	0.055***	0.044***
	(0.020)	(0.017)	(0.011)
Observations	1,969	2,796	6,334
R^2	0.152	0.107	0.086
Adjusted R^2	0.070	0.049	0.060
Dep. Var. Mean	0.241	0.257	0.278
Percentage of Baseline	16.598	21.401	15.827
Sample Day FE	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Note: The outcome variables are whether the individual contributed at least one additional answer within 21 days of being sampled, split by distance to the next threshold, before being treated/or not. Since the privilege thresholds are very concentrated below 20, the above results are for individuals with more than 20 reputation.

The first group consists of individuals who are 1-10 reputation points below the next privilege threshold, and if they are upvoted, they will reach the next privilege threshold and face another threshold that is further away. The second group consists of individuals who are 11-20 reputation points below the next privilege threshold, and if they are upvoted, they will become closer to the next privilege threshold, with less than 10 points away. The third group consists of individuals who are more than 20 reputation points below the next privilege threshold, and if they are upvoted, they will become closer to the next privilege threshold, but still more than 10 points away.

The experimental upvotes can affect an individual's future contribution decision at the intensive margin through two main channels - by shifting belief that relates to social motivation, also by giving additional reputation points, and thus changing an individual's distance to the next privilege threshold, that is, the instrumental part of the motivation. To assess the importance of the latter part, I estimate the treatment effect at the extensive margin separately for three groups of individuals: the first group, who move across the nearest reputation threshold due to the experimental upvote and are thus further from the next threshold than they were before. If instrumental motivation is critical, this could decrease future contributions. The second and third groups, who get closer to the threshold due to the experimental upvote, and thus should increase future contributions with increased instrumental incentives.

Figure 8 presents a graphical illustration. The yellow user has a reputation of 45; upon receiving the experimental treatment, he will have a reputation of 55. In this case, the additional upvote helps him gain the privilege of commenting everywhere at 50, but he is further away from the next privilege threshold, at 70. If instrumental motivation is critical, this could decrease future contributions. The yellow user is an example of those within 1-10 reputation points below the next privilege threshold. The blue user with a reputation of 57 is different, and with an experimental upvote, he will have a reputation of 67. In this case, the additional upvote does not help him gain the privilege but moves him closer to the next privilege threshold at 70, and thus should increase future contributions with increased instrumental incentives. The blue user is an example of those who are more than 10 reputation points below the next privilege threshold.

I find that the treatment effect on the extensive margin is positive and significant in the three groups. Moreover, statistical tests cannot reject the hypothesis that the effects are equivalent across the groups, although the effect appears marginally larger for the second and third groups that are moved closer to the next privilege threshold. This suggests that the instrumental motivation to reach the next threshold is only part of the motivation to contribute. I return to this in the structural estimation later.

5.7 Results - Overall Future Answer Quality and Efforts

Conditional on contributing at least one answer within 21 days since being sampled, I calculate the average answer quality and efforts for each individual. As for answer quality, I track whether an answer was accepted by the individual who posted the corresponding question, and the scores of the answer rated by other individuals (total number of upvotes minus the total number of downvotes) on the platform. As for answer efforts, I calculate the number of words included in the answer, and the number of code examples included in the answer. As shown in Table 8, there is no significant difference between the treatment group and the control group. This suggests that the treatment does not lead to different average quality or efforts from individuals, once they decide to contribute additional answers.

However, since this analysis is conditional on contributing additional answers, the effect could arise from two sources. First, the treatment may induce different individuals to contribute, and second, within an individual, the treatment may induce the individual to contribute higher or lower quality answers or make more or less effort in answering, compared to their own past answers. I try to decompose the effect by exploring the two sources in subsection 5.8 and subsection 5.9.

Table 8: Quality and Efforts of Additional Answers (Conditional on Contributing)

	Accepted	Scores	Body Words	Codes
Treatment	-0.003	0.012	-4.658	0.030
	(0.009)	(0.019)	(4.825)	(0.115)
R^2	0.063	0.046	0.058	0.057
Adj. R^2	0.012	-0.006	0.007	0.006
Num. obs.	3327	3327	3327	3327
Dep. Var. Mean	0.118	0.150	164.063	2.683
Sample Day FE	Yes	Yes	Yes	Yes
Reg. Month FE	Yes	Yes	Yes	Yes

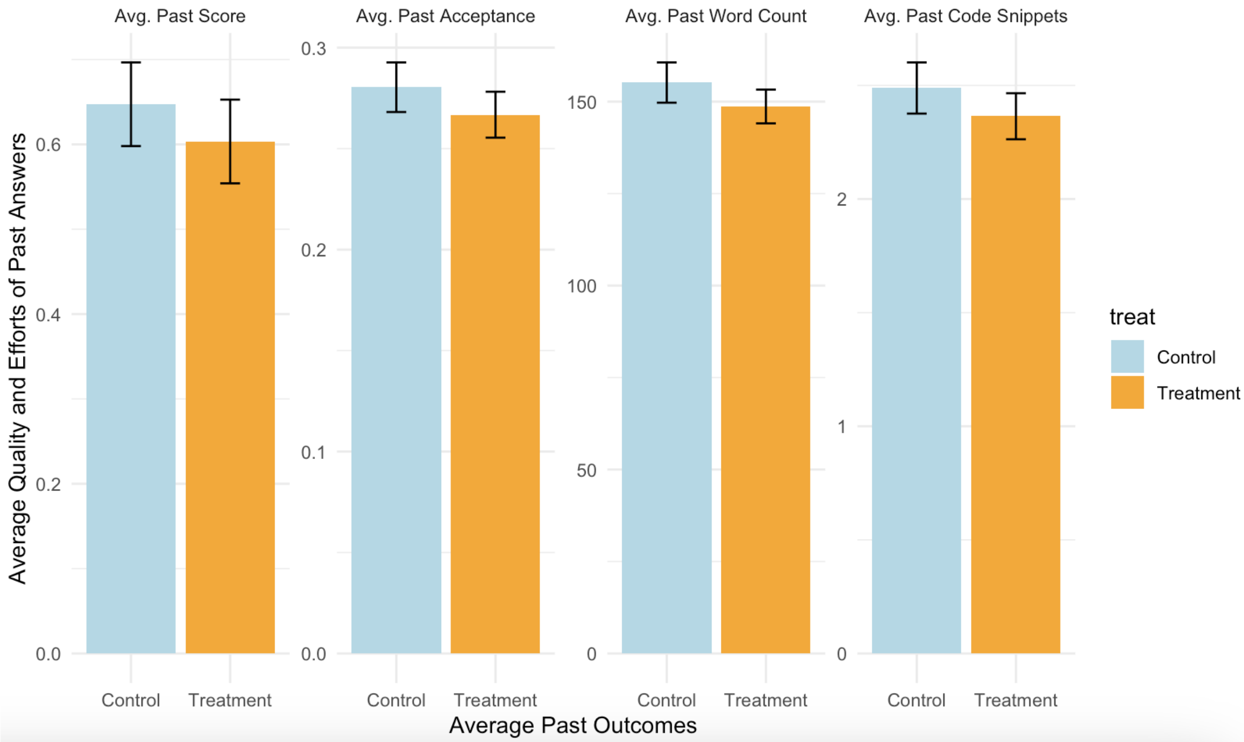
$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: The outcome variables are measured within 21 days of being sampled. Accepted is defined as the proportion of accepted answers, scores is defined as the average scores of answers, body words is defined as the average number of words in answers, and codes is defined as the average number of code pieces included in answers, for each individual who contribute at least one answer within 21 days of being sampled.

5.8 Results - Who is Motivated to Contribute by the Treatment?

To decompose the treatment effect on overall future answer quality and efforts, I track comprehensive information of all 221,093 past answers at the time the individuals are sampled. I calculate the average past answer quality and efforts for each individual, conditional on them contributing at least one answer within 21 days since being sampled. Figure 9 shows the average past answer score, the average proportion of answers being accepted, the average word count of past answers, and the average number of code examples included in the answers, averaging across individuals in the control and the treatment group. The individuals in the treatment group, on average, have lower quality and lower efforts in their past answers than those of the control group. Although the difference is not statistically significant, this can be interpreted as suggestive evidence that the treatment of one additional upvote (raising the answer score by one) induces marginal individuals with lower past answer scores to contribute additional answers, consistent with the model.

Figure 9: Individuals in the Control and Treatment Groups, Conditional on Contributing within 21 Days of Being Sampled



5.9 Results - Does the Treatment Lead to Lower Answer Quality and Efforts?

To further decompose the treatment effect on overall future answer quality and efforts, I explore whether within an individual the treatment leads to different levels of answer quality and efforts. I calculate the change in average answer quality and efforts for each individual, conditional on contributing at least one answer within 21 days since being sampled. The change in average answer quality and efforts is calculated by subtracting each individual's past average answer quality and efforts from the individual's average answer quality and efforts within 21 days of being sampled. Table 9 shows that conditional on contributing, the individuals in the treatment group increased their answer quality and efforts relative to their past answers more than those of the control group. However, the difference is not statistically significant.

Table 9: Change of Quality and Efforts Compare to Average Past Answers (Conditional on Contributing)

	ΔAccepted	ΔScores	ΔBody Words	ΔCodes
Treatment	0.014	0.063	1.390	0.115
	(0.011)	(0.039)	(4.698)	(0.104)
R^2	0.061	0.093	0.054	0.047
Adj. R^2	0.008	0.042	0.001	−0.006
Num. obs.	3327	3327	3304	3304
Dep. Var. Mean	−0.164	−0.493	8.593	0.187
Sample Day FE	*Yes*	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*	*Yes*

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: The outcome variables are measured within 21 days of being sampled. Accepted is defined as the proportion of accepted answers, scores is defined as the average scores of answers, body words is defined as the average number of words in answers, and codes is defined as the average number of code pieces included in answers, for each individual who contribute at least one answer within 21 days of being sampled.

6 Identification and Estimation

To quantify the relative importance of social motivation of gaining reputation (due to social status, altruism, warm glow, etc.) and instrumental motivation given by the platform (gaining more privileges with more reputation) in contribution decisions, I structurally estimate the model of individual decisions with heterogeneous utility parameters to match with moments from the experimental data.

The separate identification of utility parameters, γ_i and φ_i, is derived from comparing the contributing proportion of individuals within 10 reputation points of the next privilege threshold in the treatment and control groups. While the expected additional reputation points are independent of the distance to the next threshold, the expected additional privilege is not. The treatment group, which receives the experimental upvote, reaches the next privilege threshold and then faces a new, distant threshold. Thus, their short-term contribution decision is driven by the belief in gaining additional reputation rather than reaching the next threshold. In contrast, the control group, with less than 10 reputation points to the next threshold, is motivated by both gaining additional reputation and reaching the next privilege threshold.

The identification of the standard deviation of γ_i and φ_i arises from heterogeneous treatment effects across different subgroups of individuals. According to the individual learning model, holding γ_i and φ_i constant, if an individual does not expect to reach the next privilege threshold with their next contribution, their response to the treatment should decrease as the number of past contributions increases. The discrepancies between model predictions assuming homogeneous parameters and actual treatment effects indicate heterogeneous utility parameters, helping to identify the standard deviation of γ_i. Similarly, heterogeneous treatment effects among individuals within 10 reputation points of the next privilege threshold before the treatment help separately identify the standard deviation of φ_i.

I simplify the effort cost, c_i, to be heterogeneous across individuals but constant with respect to the length of answers or the number of code examples included. This assumption is based on the finding that the estimated treatment effect on answer efforts is not statistically significant. It suggests that the primary effort cost of contributing an answer is the cognitive cost of understanding the question and devising a solution. Once an individual decides to contribute, the additional effort of typing more words or including code examples is relatively minor compared to the cognitive cost. Furthermore, I assume individuals have a rational prior when posting their first answer by calibrating k and θ based on the complete historical scores of all past answers by sampled individuals.

For estimation, I categorize individuals based on their reputation levels, proximity to the next privilege threshold (within 10 reputation points or not), and whether they are in the control or treatment group. I then compute the proportion of individuals in each subgroup who contributed within 21 days of sampling. Table 10 provides a complete list of the empirical moments used for the simulated method of moments, along with the corresponding number of observations for each moment. For each individual, I simulate 1,500 heterogeneous utility parameters, γ_i and φ_i, drawing from log-normal distributions. I predict individuals' decisions to contribute additional answers based on their history of answer scores and reputation at the time of sampling. Then, I predict their decisions to contribute after being treated with one additional upvote. I calculate the proportion of individuals contributing with and without the treatment. Finally, I compare the model-predicted proportions of contributors with and without treatment to the actual contributing proportions estimated from the treatment and control groups.

Table 10: Empirical Moments for the Simulated Method of Moments

	Proportion	Number Obs.
Reputation Below 19 - Control Group	0.231	532
Reputation Below 19 - Treatment Group	0.265	543
Reputation Between 20 and 39 - Control Group	0.229	2039
Reputation Between 20 and 39 - Treatment Group	0.266	1973
Reputation Between 40 and 49 - Control Group	0.267	450
Reputation Between 40 and 49 - Treatment Group	0.265	452
Reputation Between 50 and 64 - Control Group	0.240	504
Reputation Between 50 and 64 - Treatment Group	0.334	458
Reputation Between 65 and 74 - Control Group	0.219	215
Reputation Between 65 and 74 - Treatment Group	0.332	250
Reputation Between 75 and 89 - Control Group	0.221	253
Reputation Between 75 and 89 - Treatment Group	0.343	251
Reputation Between 90 and 99 - Control Group	0.282	110
Reputation Between 90 and 99 - Treatment Group	0.307	137
Reputation Between 100 and 114 - Control Group	0.270	222
Reputation Between 100 and 114 - Treatment Group	0.262	221
Reputation Between 115 and 124 - Control Group	0.274	95
Reputation Between 115 and 124 - Treatment Group	0.176	108
Reputation Between 125 and 189 - Control Group	0.231	459
Reputation Between 125 and 189 - Treatment Group	0.298	473
Reputation Between 190 and 199 - Control Group	0.211	38
Reputation Between 190 and 199 - Treatment Group	0.224	49
Reputation Between 200 and 239 - Control Group	0.257	74
Reputation Between 200 and 239 - Treatment Group	0.324	74
Reputation Between 240 and 249 - Control Group	0.235	17
Reputation Between 240 and 249 - Treatment Group	0.300	10

7 Structural Estimates

Table 11: Structural Estimates

	Estimate	Standard Error
$ln(\gamma)$ mean	-6.702	1.160
$ln(\gamma)$ sd	7.111	0.803
$ln(\varphi)$ mean	-19.640	9.962
$ln(\varphi)$ sd	14.895	7.423
c	2.840	0.733

Table 11 shows structural estimates. γ_i is assumed to follow a log-normal distribution, with an estimated log-mean of -6.702 and an estimated log-standard of deviation 7.111. φ_i is assumed to follow a log-normal distribution, with an estimated log-mean of -19.640 and an estimated log-standard deviation of 14.895. The effort cost parameter c is estimated to be 2.840. All of the five parameters are precisely estimated. The estimates suggest that the instrumental preference parameter, φ_i, follows a more dispersed distribution compared to the distribution of the social preference parameter, γ_i. Figure A1 shows model fit.

8 Counterfactuals

8.1 Measuring the Relative Importance of Social Motivation and Instrumental Motivation

To understand the relative importance of social motivation and instrumental motivation, I consider the counterfactual scenarios of removing social motivation and removing instrumental motivation respectively. First, I simulate the scenario where social motivation is absent to. The results indicate that the proportion of individuals willing to contribute drops to 26% of the baseline control group. Next, I simulate the removal of instrumental motivation. In this scenario, the proportion of contributing individuals remains at 84% of the baseline control group. The substantially larger reduction observed with the removal of social motivation suggests that social motivation plays a more critical role in influencing contribution decisions.

In subsection 8.2 and subsection 8.3, I consider two platform strategies designed to enhance social motivation: amplifying the number of upvotes for answers that already have at least one upvote and giving an additional upvote for high-quality answers.

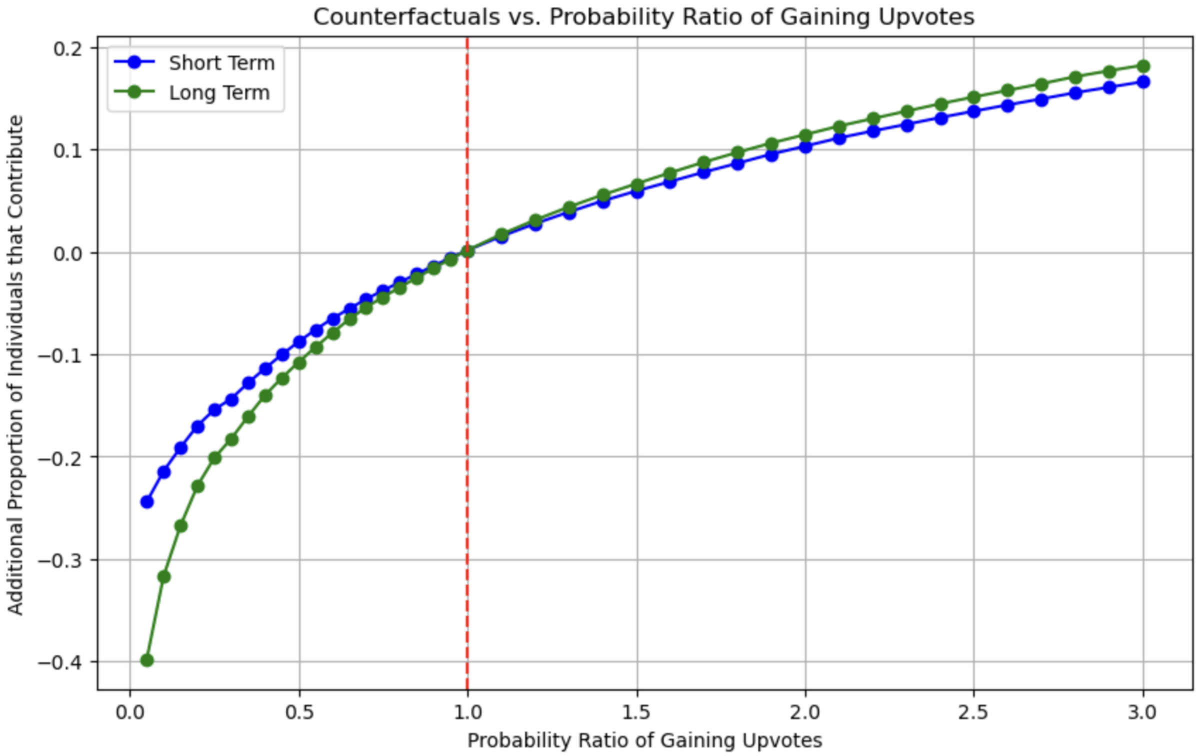
8.2 Changing the Probability of Gaining Upvotes by Highlighting Certain Answers

How will individuals' contribution decisions change if the probability of gaining upvotes changes? Practically, platforms change the probability of gaining upvotes by choosing to give certain answers more or less exposure than others or display notifications to nudge or discourage individuals from upvoting when they view certain answers. In this counterfactual exercise, I consider changing the probability of gaining upvotes for all past answers with a positive number of upvotes by individuals in my sample.

Figure 10 shows the change in the proportion of individuals that contribute, with counterfactual levels of probability of gaining upvotes relative to the baseline. The probability levels to the right of the vertical red correspond to increasing the probability of gaining upvotes, while the probability levels to the left of the vertical red line correspond to decreasing the probability. The blue line corresponds to the short-run, when individuals' prior has not adjusted, while the green line denotes the long-run, where individuals' prior has already adjusted according to the changed upvoting probability.

In the short run, increasing the probability of gaining upvotes on all past answers by 50% will induce an additional 5% of individuals to contribute. A 100% increase in this probability will result in about 9% more individuals contributing, and a 150% increase will lead to approximately 12% additional contributors. The efficacy of increasing the probability of upvotes shows diminishing returns. In the long run, as the prior adjusts, the proportion

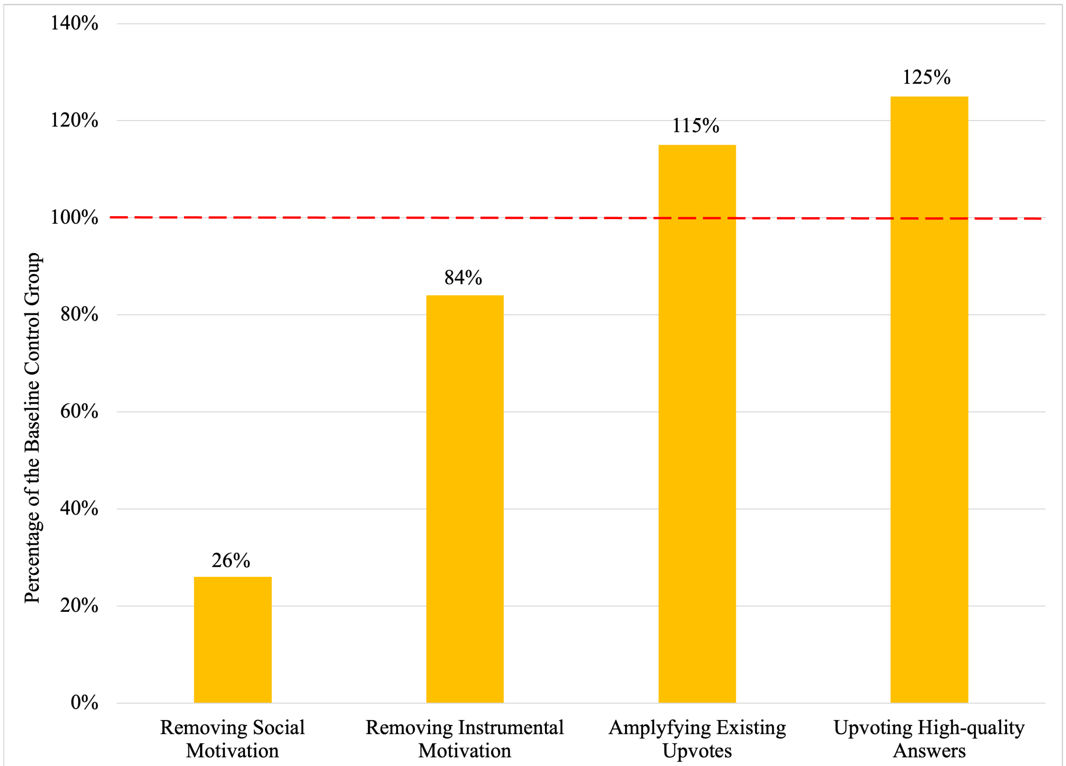
Figure 10: Counterfactuals



of contributing individuals increases slightly more.

However, decreasing the probability of gaining upvotes can have an increasingly negative marginal effect. As shown by the counterfactual probability levels to the left of the vertical line, a 10% decrease in the probability of gaining upvotes will reduce the proportion of contributing individuals by 2% in the short run, and a 50% decrease will reduce it by 8%. The slope becomes steeper as the probability further decreases. Additionally, in the long run, as the prior adjusts to the lower probability of gaining upvotes, the magnitude of the negative effect increases. The gap between the short-run (blue line) and long-run (green line) effects to the left of the vertical line widens as the probability further decreases.

Figure 11: Counterfactual Proportion in Percentage of Baseline



8.3 Upvoting High-quality Answers by Introducing Expert Evaluations

When individuals decide which question to contribute an answer to, they can choose from a wide range of questions - every day, there are thousands of new questions posted, along with millions of existing questions. If an individual chooses to answer a less popular question or post an answer at a time when the visits to the site are low, their answers may be less likely to be noticed and upvoted by others, even though the quality of the answers may be high.

In this counterfactual, I consider giving one additional upvote to answers of high quality, defined by answers with at least one code example. This definition is consistent with existing research using data from Stack Overflow, such as Bregolin (2022). I show that giving one additional upvote to those answers will increase the proportion of contributing individuals by 25% of the baseline.

In the future, I plan to assess the quality of approximately 220,000 past answers contributed by individuals in my sample using advanced language models, such as ChatGPT. Subsequently, I will simulate a counterfactual scenario of upvoting answers deemed high-quality by the language model but which had not received any upvotes from other users at the time of sampling. This counterfactual analysis will provide insights into how the lack of recognition impacts contributions to online knowledge platforms.

In summary, Figure 11 illustrates the counterfactual proportion as a percentage of the baseline for four scenarios: removing social motivation, removing instrumental motivation, tripling the number of existing upvotes, and giving high-quality answers one additional upvote.

9 Conclusions

In this paper, I design and conduct a large-scale field experiment involving 12,182 individuals on one of the largest online question-and-answer platforms by collecting and treating daily samples for four and a half months. The treatment gives one additional upvote to eligible answers. I then track comprehensive data on individuals' subsequent daily posting behavior on the platform for four months.

I find that receiving one more upvote on an answer substantially impacts individuals' subsequent behavior. At the extensive margin, the probability of individuals making additional contributions increases by around 15% of the baseline. At the intensive margin, receiving one more upvote on an answer can increase the number of answers contributed within three weeks by around 6%. Both of the effects are sustainable over two months and, thus, are not driven by intertemporal substitutions.

The treatment effect is heterogeneous. It is larger on individuals with the number of past answers and reputation in the low-to-middle range and is slightly larger on individuals close to the next reputation threshold after being treated.

In terms of other outcomes, receiving one more upvote on an answer does not significantly impact whether individuals log in to the platform within 21 days of being sampled, but it slightly increases the number of days they log in by 0.55 days, around 3.9% of the baseline. The treatment has no effect on question contributions. However, it significantly increases both the probability of posting comments and the number of comments posted, with the effect magnitude on posting comments comparable to that on contributing answers. Additionally, the treatment significantly impacts individuals' voting behavior within 21 days. For upvoting, the treatment increases the probability by 7.30% of the baseline and the number of upvotes by around 5% of the baseline. For downvoting, the treatment increases the probability of making at least one downvote by around 20% of the baseline and the number of downvotes by 2.1%. The treatment does not seem to induce significantly less effort or worse answers. Conditional on individuals deciding to contribute, there is suggestive evidence that individuals increase efforts.

To quantify how both motivators affect the primary outcome of interest - the extensive margin of whether or not to contribute additional online knowledge, I structurally estimate the model of individual decisions with heterogeneous utility parameters to explain my experimental findings. The estimates suggest that the extent to which individuals value social motivation and instrumental platform privileges varies widely across individuals.

Using estimates from the structural model, I explore a set of counterfactuals. I simulate scenarios where social motivation and instrumental motivation are removed, respectively. When social motivation is absent, the proportion of contributors drops to 26% of the baseline, whereas removing instrumental motivation retains 84% of the baseline. This larger reduction with the absence of social motivation suggests its more critical role in influencing contributions.

Then, I explore two platform strategies designed to enhance social motivation: amplifying the number of upvotes for answers that already have at least one upvote by highlighting those answers, and giving an additional upvote to high-quality answers by introducing expert evaluations. My findings indicate that tripling the number of upvotes for answers with at least one upvote can increase the proportion of contributing individuals by 15% compared to the baseline. Additionally, awarding one extra upvote to high-quality answers results in a 25% increase in contributions relative to the baseline.

Overall, the study suggests the important role of social motivation in motivating online content contributions and highlights the role of platform intervention.

References

Ahn DY, Duan JA, Mela CF (2016) Managing user-generated content: A dynamic rational expectations equilibrium approach. *Marketing science* 35(2):284–303.

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

Athey S, Ellison G (2014) Dynamics of open source movements. *Journal of Economics & Management Strategy* 23(2):294–316.

Bregolin J (2022) Communication effort and the cost of language: Evidence from stack overflow. *Working Paper* .

Burtch G, He Q, Hong Y, Lee D (2022) How do peer awards motivate creative content? experimental evidence from reddit. *Management Science* 68(5):3488–3506.

Burtch G, Hong Y, Bapna R, Griskevicius V (2018) Stimulating online reviews by combining financial incentives and social norms. *Management Science* 64(5):2065–2082.

Burtch G, Lee D, Chen Z (2023) The consequences of generative ai for ugc and online community engagement. *Available at SSRN 4521754* .

Cabral L, Li L (2015) A dollar for your thoughts: Feedback-conditional rebates on ebay. *Management Science* 61(9):2052–2063.

Chen Y (forthcoming) *Digital Public Goods: Crowdsourcing, Crowdfunding, and User-Generated Content.*

Chen Y, Farzan R, Kraut R, YeckehZaare I, Zhang AF (2023) Motivating experts to contribute to digital public goods: A personalized field experiment on wikipedia. *Management Science* .

Chen Y, Harper FM, Konstan J, Li SX (2010) Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review* 100(4):1358–1398.

Conti A, Gupta V, Guzman J, Roche MP (2023) Incentivizing innovation in open source: Evidence from the github sponsors program. *Working Paper* .

Deolankar V, Fong J, Sriram S (2023) Content generation on social media: The role of negative peer feedback. *Available at SSRN 4522092* .

Eckles D, Kizilcec RF, Bakshy E (2016) Estimating peer effects in networks with peer encouragement designs. *Proceedings of the National Academy of Sciences* 113(27):7316–7322.

Gallus J (2017) Fostering public good contributions with symbolic awards: A large-scale natural field experiment at wikipedia. *Management Science* 63(12):3999–4015.

Goes PB, Guo C, Lin M (2016) Do incentive hierarchies induce user effort? evidence from an online knowledge exchange. *Information Systems Research* 27(3):497–516.

Guo M, Ni J, Shen Q, Xu Y (2023) Quantity vs diversity in online content production: Evidence from a knowledge sharing platform. *Johns Hopkins Carey Business School Research Paper* (22-12).

Huang JT, Kaul R, Narayanan S (2022) The causal effect of attention and recognition on the nature of user-generated content: Experimental results from an image-sharing social network. *Stanford University Graduate School of Business Research Paper* .

Jiménez Durán R (2021) The economics of content moderation: Theory and experimental evidence from hate speech on twitter. *Available at SSRN 4044098* .

Jin J, Li Y, Zhong X, Zhai L (2015) Why users contribute knowledge to online communities: An empirical study of an online social q&a community. *Information & management* 52(7):840–849.

Lacetera N, Macis M (2010) Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. *Journal of Economic Behavior & Organization* 76(2):225–237.

Lerner J, Tirole J (2002) Some simple economics of open source. *The journal of industrial economics* 50(2):197–234.

Mummalaneni S, Yoganarasimhan H, Pathak V (2023) How do content producers respond to engagement on social media platforms? *Available at SSRN 4173537* .

Paridar M, Ameri M, Honka E (2024) More, faster, and better? effects of rewards on incentivizing the creation of user-generated content. *Available at SSRN 4580732* .

Srinivasan K (2023) Paying attention. *Working Paper* .

Sun Y, Dong X, McIntyre S (2017) Motivation of user-generated content: Social connectedness moderates the effects of monetary rewards. *Marketing Science* 36(3):329–337.

Toubia O, Stephen AT (2013) Intrinsic vs. image-related utility in social media: Why do people contribute content to twitter? *Marketing Science* 32(3):368–392.

Wang Y, Goes P, Wei Z, Zeng D (2019) Production of online word-of-mouth: Peer effects and the moderation of user characteristics. *Production and Operations Management* 28(7):1621–1640.

Xu L, Nian T, Cabral L (2020) What makes geeks tick? a study of stack overflow careers. *Management Science* 66(2):587–604.

Zeng Z, Dai H, Zhang DJ, Zhang H, Zhang R, Xu Z, Shen ZJM (2023) The impact of social nudges on user-generated content for social network platforms. *Management Science* 69(9):5189–5208.

Zhang M, Luo L (2024) Effects of peer voting and followers on user contribution to online knowledge sharing platforms: Evidence from a field experiment. *Available at SSRN* .

Zhang X, Zhu F (2011) Group size and incentives to contribute: A natural experiment at chinese wikipedia. *American Economic Review* 101(4):1601–1615.

Appendices

Table A1: Heterogeneity of Treatment Effect on Answer Contributions by Distance to the Next Privilege Threshold Before Being Treated

	Dependent variable: I(Additional Answers)		
	1-10 Distance	11-20 Distance	More than 20 Distance
Treatment	0.042***	0.055***	0.044***
	(0.014)	(0.017)	(0.011)
Observations	3,044	2,796	6,334
R^2	0.092	0.107	0.086
Adjusted R^2	0.038	0.049	0.060
Dep. Var. Mean	0.238	0.257	0.278
Percentage of Baseline	17.647	21.401	15.827
Sample Day FE	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Note: The outcome variables are whether the individual contributed at least one additional answer within 21 days of being sampled, split by distance to the next threshold, before being treated/or not. This table includes all individuals, as a complement to Table 7.

Table A2: Average Quality and Efforts of Past Answers (Conditional on Contributing)

	Accepted	Scores	Body Words	Codes
Treatment	-0.018^{*}	-0.044	-5.801	-0.087
	(0.009)	(0.037)	(3.788)	(0.080)
R^2	0.069	0.085	0.065	0.069
Adj. R^2	0.018	0.035	0.014	0.019
Num. obs.	3327	3327	3304	3304
Dep. Var. Mean	0.282	0.643	155.556	2.495
Sample Day FE	*Yes*	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*	*Yes*

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: The outcome variables are measured within 21 days of being sampled. Accepted is defined as the proportion of accepted answers, scores is defined as the average scores of answers, body words is defined as the average number of words in answers, and codes is defined as the average number of code pieces included in answers, for each individual who contribute at least one answer within 21 days of being sampled.

Table A3: Ruling out an Alternative Mechanism - Treatment Effect on Answer Contributions by Reputation and Proportion of Reputation Gained from Answers

	Dependent variable: I(Additional Answers)		
	Group 1	Group 2	Combined
Treatment	0.067***	0.052***	0.050***
	(0.024)	(0.012)	(0.012)
High Reputation with Low Proportion			-0.030
			(0.020)
Treatment:High Reputation with Low Proportion			0.003
			(0.026)
Observations	1,185	4,625	5,810
R^2	0.172	0.064	0.063
Adjusted R^2	0.031	0.027	0.034
Dep. Var. Mean	0.179	0.250	0.235
Percentage of Baseline	37.430	20.800	21.277
Sample Day FE	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Note: The outcome variables are whether the individual contributed at least one additional answer within 21 days of being sampled. The first column includes individuals who have an above-median reputation and with less than 50% of their reputation gained from answers, before being treated/or not. The second column includes individuals who have a below-median reputation and with more than 50% of their reputation gained from answers. If the heterogeneous treatment effect by answer experience/reputation in Table 6 is driven by satiation of reputation, the first column would have a smaller treatment effect. This is ruled out by the test in the third column, showing that the difference between the treatment effects of the two groups is statistically insignificant.

Table A4: Ruling out an Alternative Mechanism - Treatment Effect on Individuals Near Privilege Thresholds Unrelated to Contributing Answers

	Dependent variable: I(Additional Answers)		
	Within 10 Below Irrelevant Privileges	More than 10 Below Irrelevant Privileges	Combined
Treatment	0.098^{**}	0.071^{***}	0.075^{***}
	(0.049)	(0.019)	(0.019)
Near			-0.041
			(0.026)
Treatment:Near			0.011
			(0.047)
Observations	518	2,074	2,592
R^2	0.312	0.143	0.132
Adjusted R^2	-0.010	0.066	0.070
Dep. Var. Mean	0.230	0.265	0.258
Percentage of Baseline	42.609	26.792	29.070
Sample Day FE	Yes	Yes	Yes
Reg. Month FE	Yes	Yes	Yes

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Note: The outcome variables are whether the individual contributed at least one additional answer within 21 days of being sampled. I focus on three privilege thresholds that are unrelated to contributing additional answers: setting bounties at 75 (offering some of reputation as bounty on a question), viewing close votes at 250 (viewing and casting close/reopen on ones own questions), and accessing review queues at 500 (accessing the first posts and late answers review queues). The first column includes individuals who have a reputation within 10 points below the three privilege thresholds, before being treated/or not. The second column includes individuals who have a reputation above the previous threshold, but more than 10 points below the three privilege thresholds. If the treatment effect in the first column of Table 7 is completely driven by additional privilege making it easier for people to contribute answers, then the first column in this table would have an insignificant treatment effect, when the additional privileges are unrelated to contributing answers. This is ruled out by the statistically significant positive treatment effect in the first column, and the test in the third column, showing that the difference between the treatment effects of the two groups is statistically insignificant.

Table A5: Ruling out an Alternative Mechanism - Treatment Effect on Individuals Contributing No More than One Additional Answer

	Dependent variable: I(Additional Answers)		
	1-10 Distance	11-20 Distance	More than 20 Distance
Treatment	0.026^{**}	0.041^{**}	0.029^{***}
	(0.013)	(0.016)	(0.009)
Observations	2,628	2,344	5,284
R^2	0.070	0.088	0.040
Adjusted R^2	0.005	0.016	0.008
Dep. Var. Mean	0.126	0.125	0.139
Percentage of Baseline	20.635	32.800	20.863
Sample Day FE	*Yes*	*Yes*	*Yes*
Reg. Month FE	*Yes*	*Yes*	*Yes*

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Note: The outcome variables are whether the individual contributed at least one additional answer within 21 days of being sampled, split by distance to the next threshold, before being treated/or not. This table includes individuals who contributed no more than one additional answer within 21 days of being treated. This corresponds to 84.198% of the full sample.

Figure A1: Model Fit

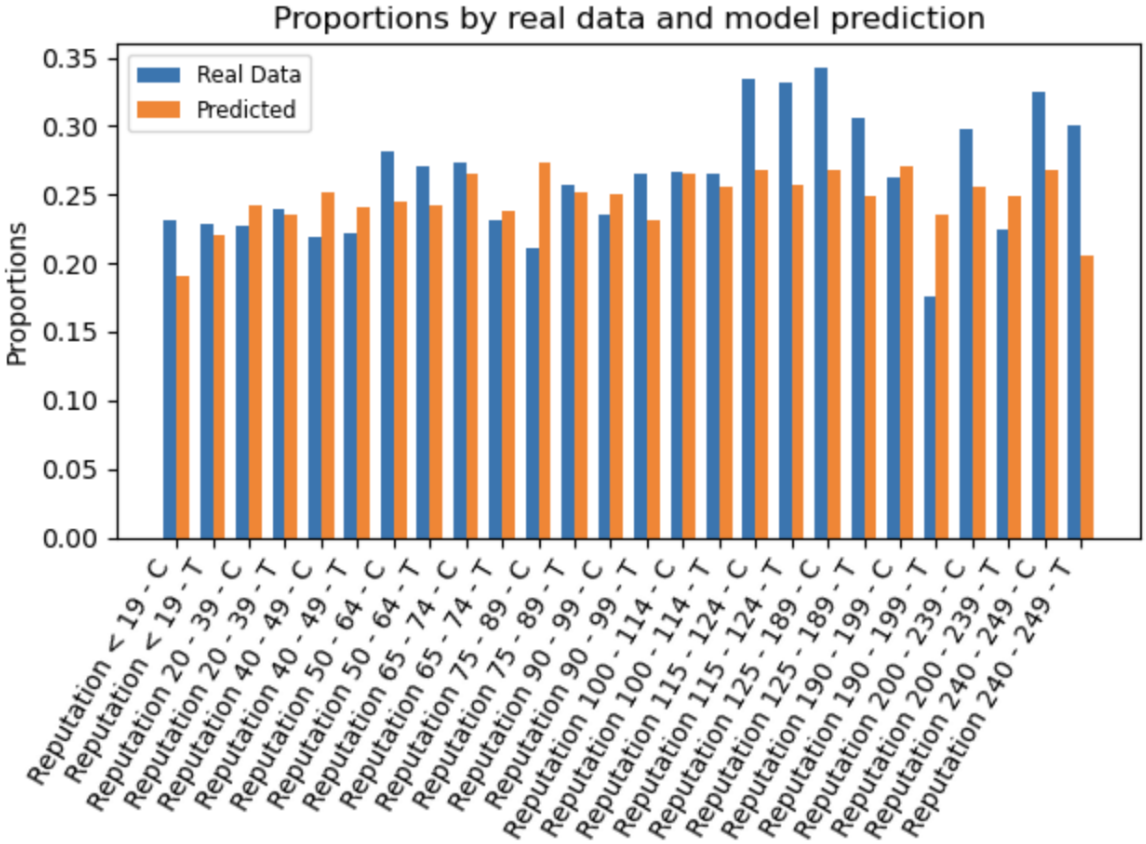
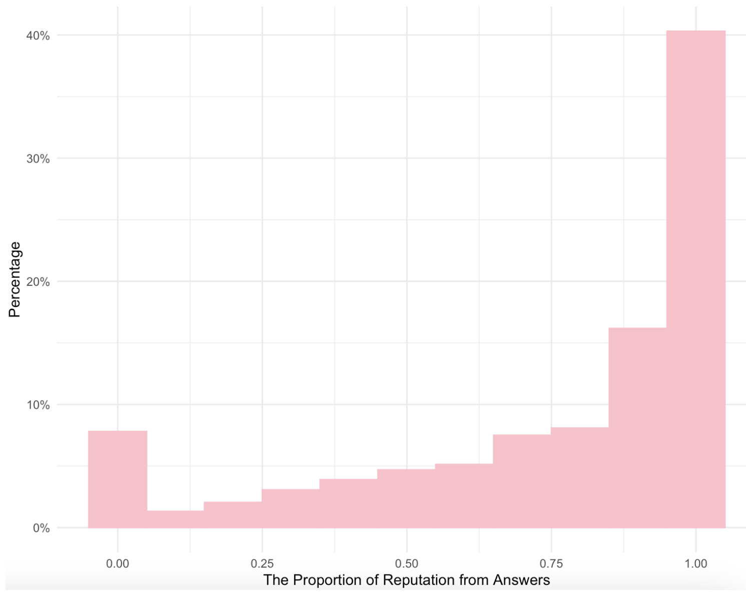
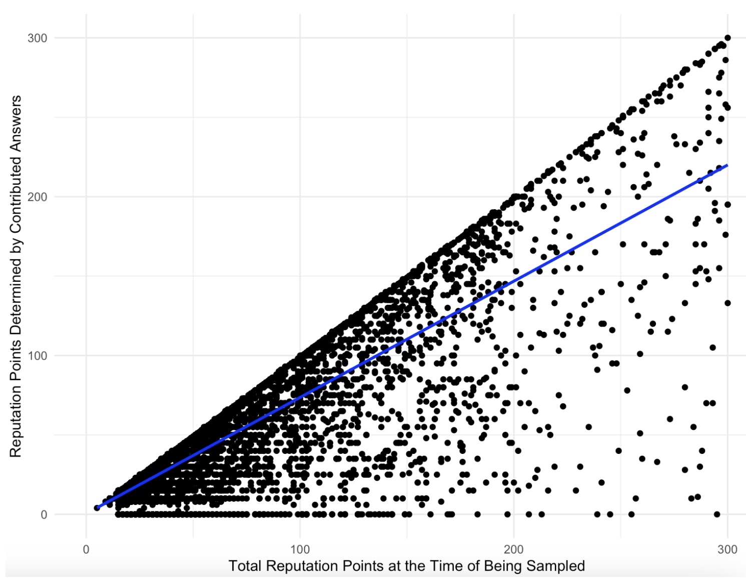


Figure A2: Reputation from Answers and Total Reputation at the Time of Being Sampled

(a) The Percentage Distribution of the Proportion of Reputation from Answers



(b) Reputation from Answers vs. Total Reputation



Note: Panel (a) of the figure shows the percentage distribution of the proportion of reputation from contributed answers for each individual at the time of being sampled. In addition to answering questions, individuals can gain reputation points by posting questions and suggesting edits, and they can lose reputation by downvoting others. Overall, 25% of individuals gain less than 59.52% of their reputation points through contributing answers. The median individual in the sample gains 90.91% of reputation through contributing answers. The mean proportion is 74.25%. Panel (b) of the figure shows a scatter plot of reputation points from answers versus total reputation points at the time of being sampled. Each point represents one individual.