Check for updates

Original Article

POMS

Production and Operations Management
1–20
© The Author(s) 2026

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10591478261429201
journals.sagepub.com/home/pao
Sage

Effects of Peer Voting and Followers on User Contribution to Online Knowledge Sharing Platforms: Evidence From a Field Experiment

Mengxia Zhang[1] and Lan Luo[2]

Abstract

Online knowledge-sharing platforms such as Quora and Zhihu (a leading knowledge-sharing platform in China) often use both peer voting and followers to encourage user knowledge contributions. However, these platforms diverge in whether they highlight followers, upvotes, or both as reputation symbols. To better understand the consequences of these platform design choices, we conduct a field experiment with 1,696 focal users on Zhihu where we exogenously increase upvotes or/and followers for treated users over a 53-day intervention period. We monitor focal users' activities for 303 days, covering both pre- and post-intervention periods. We further use a large language model to gauge the quality of 12,998 answers contributed by these users during this time. We find that increasing upvotes significantly boosts users' answer contributions (in volume, total length, and quality), while increasing followers has mostly no overall effect on contributions. Additionally, increasing upvotes can encourage contribution for up to 100 days, particularly among lower-reputation (i.e., with fewer upvotes or followers), less active, female users, and those answering soft topic questions. In contrast, increasing followers only reduces contribution volume (not length or quality) for higher-reputation, more active, and male users. Moreover, while peer voting primarily only affects answer contributions, the follower treatment leads to negative spillovers, reducing users' upvoting on fellow users' answers, following fellow users, or purchasing Zhihu Lives hosted by fellow users. Our research suggests the possibility that users on knowledge-sharing platforms may associate peer voting with contributions, treating it as peer recognition that offers intrinsic motivation. In contrast, users may interpret additional followers as a symbol of status enhancement, which could, in some cases, dampen their motivation to contribute. Overall, our study suggests that platforms aiming to foster active, high-quality contributions should highlight upvotes rather than followers as a reputation symbol. To our knowledge, this study is among the first field experiments to identify and compare the causal effects of peer voting and followers on user contribution to knowledge-sharing platforms.

Keywords

Knowledge Contribution, Peer Voting, Followers, Social Networks, Field Experiment, Platform Design

Date received 15 October 2024; accepted 5 February 2026 after two revisions

Handling Editor: Praveen K. Kopalle

1 Introduction

In today's marketplaces, consumers are not only sharing houses and cars but also sharing invaluable knowledge that significantly impacts our society (Thompson and Hanley, 2018). Unlike houses and cars, shared knowledge online is typically considered a public good (Zhang and Zhu, 2011), often freely accessible to all consumers. However, contributing knowledge online requires effort and is often not monetized. Consequently, online knowledge sharing faces a classic under-provision problem akin to the "private provision of public goods"—where the provision of a good or service often falls short of the optimal amount.

[1]Marketing Group, Ivey Business School, Western University, London, ON, Canada
[2]Marketing Group, Marshall School of Business, University of Southern California, Los Angeles, CA, USA

Corresponding author:
Mengxia Zhang, Marketing Group, Ivey Business School, Western University, 1255 Western Rd, London, ON N6G 0N1, Canada.
Email: mezhang@ivey.ca

Amid this backdrop, numerous knowledge-sharing communities and platforms have emerged on the Internet, including CafeMom, Baidu Knows, Quora, Zhihu, and various Stack Exchange communities (such as Stack Overflow, Mathematics, English Language & Usage, etc.). These platforms are primarily structured around questions and answers, offering a wealth of knowledge covering diverse areas like computer programming, science, languages, home, cooking, parenting, workplace, e-commerce, sports, travel, photography, and more. As such, many of these knowledge-sharing platforms serve millions of people worldwide[1] and have market valuations in the range of billions.[2]

Despite their global success, many knowledge-sharing platforms still struggle to motivate user contribution (Chen et al., 2010). For example, on Stack Overflow, only about 22% of its 16 million registered users had contributed at least one question or answer with any upvotes as of November 16, 2022.[3] The actual rate is likely much lower, as many users browse content without registering. Moreover, engagement commonly declines over a user's lifespan in many online knowledge-sharing communities (Cong et al., 2025; Gallus, 2017; Zhang and Zhu, 2011). The rise of generative AI tools like ChatGPT has further exacerbated the problem, as many now turn to AI for answers (e.g., Burtch et al., 2024; Reuters, 2023; Shorakaei et al., 2025), making it even more critical to sustain user contribution.

To combat these challenges, these platforms often employ peer voting as a potential motivator to incentivize contributions. As Chen et al. (2018) discussed, upvotes given by fellow users can serve as peer recognition, validating the value of answers provided by contributors and thus encouraging their further contribution. In addition, another trendy platform design to encourage user contributions on these platforms is the adoption of followers. Both received upvotes and followers could serve as reputation symbols for users. For example, Zhihu displays both the total upvotes received and the total number of followers on user profiles.

Nevertheless, little is known about the relative impacts of upvotes versus followers on encouraging knowledge contributions, and reputation system designs vary across different platforms. For example, some platforms deliberately avoid incorporating followers on their platforms. Notably, Jeff Atwood, co-founder of Stack Overflow, once expressed his perspective: "We're not building some hyper-viral social networking tool—where we try to game you into hanging around and socializing and building lists of fake friends to get results."[4] As such, Stack Overflow displays reputation solely based on peer voting.[5] In contrast, Quora displays the follower count as a visible reputation, but not the upvote count.

Inspired by such divergence in platform design choices of highlighting upvotes and/or followers as reputation symbols in practice, we aim to investigate the impacts of increasing peer upvotes and increasing follower count on user knowledge contribution on these platforms. To our knowledge, this research is among the first to utilize a field experiment to identify and compare the causal impacts of peer voting and followers on knowledge-sharing platforms. Specifically, we seek to answer the following research questions: Q1. How does increasing the number of upvotes and/or followers affect a focal user's knowledge contributions, and why? Q2. What type of users and topics are more positively or negatively affected by increasing upvotes/followers? Q3. Will the impacts of increasing upvotes/followers spill over to other user behaviors, such as upvoting others' answers, following fellow users, or purchasing live events, which may subsequently affect fellow users' contributions on the platform?

In order to seek answers to our research questions, we carry out a field experiment on 1,696 users of Zhihu, often referred to as China's counterpart of Quora. Specifically, we adopt a 2 (upvote intervention vs. not) by 2 (follower intervention vs. not) between-subjects design, and exogenously send 100 upvotes (followers) to users assigned to the upvote (follower) intervention over a 53-day period using 100 synthetic users. We monitor these users' knowledge contribution and other activities on Zhihu on a daily basis for 303 days (50 days pre-treatment, 53 days during treatment, and 200 days post-treatment). During the 303-day window, a total of 12,998 answers were submitted by these users. We further utilize the OpenAI GPT-4 model to gauge the quality of these answers across a wide range of diverse topics.[6] Our field experiment is inspired by Toubia and Stephen (2013) that investigates whether increasing followers (but not upvotes) via synthetic users increases the number of tweets on Twitter during a less than 2-month post-treatment period.

We face three key challenges in carrying out our field experiment. First, we need to identify a representative sample of users on Zhihu from its massive and hidden sampling space as subjects of our field experiment. Second, the typical random assignment of these users to experimental conditions won't work in our setting because the treatment received by one user in one condition can potentially interfere with the potential outcome of another user in a different condition if the two users are connected through social networks, thereby violating the Stable Unit Treatment Value Assumption (SUTVA) crucial for causal inference. Lastly, in order to implement the upvotes and followers treatments, we need to prepare a group of synthetic users that can be disguised as actual users to the extent possible.

We address the first challenge using the Metropolis-Hastings random walk (MHRW) (Wang et al., 2010), a sampling method from the computer science literature well-known for handling massive and hidden spaces. We further validate that key statistics of the sampled users align with those of the broader Zhihu user base reported by Liu and Cong (2023). We mitigate the second challenge by utilizing a cluster-level treatment assignment method called "graph cluster randomization" (Ugander et al., 2013) to reduce interference between users assigned to different experimental conditions. Last but not least, we hire three research assistants who are active users

of Zhihu.com to meticulously prepare and manage these synthetic users on a daily basis for a period of over 100 days (50 days prior to treatment and 53 days during treatment). Taking advantage of the platform design choice that Zhihu does not display users' tenure or registration dates on its platform, we are able to closely mimic actual user behavior on Zhihu during the 100-day period so that registered Zhihu users cannot distinguish our synthetic users from actual users on the platform.

Our key findings are below. *First*, users who receive the upvote intervention respond positively, being more likely to increase their answer contribution in terms of volume, total length, and quality compared to control users. Notably, the positive treatment effect exists in the first 100 days after the intervention. Nevertheless, the follower intervention has largely no impact on answer contribution. *Second*, the upvote intervention has a stronger positive effect for users with lower reputation (i.e., users with less than median upvotes or followers at the beginning of our observational window), less-active users, female users, or users who provide answers in soft topics. In comparison, higher-reputation, more active, and male users contributed fewer answers when they received more followers. Nevertheless, increasing follower count does not impact answer length or quality. *Lastly*, we find that although the positive impact of peer voting is only applicable to knowledge contribution, increasing followers can induce negative spillovers on other user activities. Specifically, users receiving the follower treatment are less likely to increase upvoting on fellow users' answers, following fellow users, or purchasing Zhihu Lives[7] hosted by fellow users compared with control users, which might subsequently negatively impact the knowledge contribution of these fellow users.

Overall, our findings suggest that peer voting encourages user knowledge contribution, while increases in follower count appear to have mostly no effect. We further supplement our field experiment with a survey of Zhihu users to better understand the underlying mechanisms of the observed effects. In line with theories of intrinsic versus status-related utility (Ryan and Deci, 2000; Toubia and Stephen, 2013; Wasko and Faraj, 2005), our survey suggests that although both upvotes and followers can provide intrinsic and status-related utilities, upvotes may primarily motivate users intrinsically by recognizing specific answer contributions, whereas increases in follower count may function more as status enhancement and thus be perceived as less intrinsically motivating. Our analysis of heterogeneous treatment effects further aligns with this conjecture: users with lower reputations, who may be more responsive to intrinsic utility, respond most positively to peer voting, whereas users with higher reputations, who may derive more utility from status, experience reductions in contribution volume when their follower count increases.

One caveat of our research is that in an organic setting, an increase in the number of followers may lead to an increase in future upvotes from these followers, which may, in turn, increase user contribution. In our field experiment, we intentionally make the increases in upvotes and followers orthogonal to separately identify their effects. As such, our experiment results cannot speak to this potential positive indirect effect of an increase in followers. Nevertheless, we empirically examined this using organic following and upvoting behavior of the control users in our study, and we did not find strong evidence for such an indirect effect.

Our research offers valuable actionable insights for designing platform reputation systems. Because peer voting could be more closely tied to intrinsic utility than status-related utility, displaying upvote counts as a reputation symbol appears to be a viable strategy. In contrast, prominently displaying follower counts could highlight user status, which does not encourage contribution. Yet the null effect of our follower intervention should be interpreted with caution. In our field experiment, synthetic followers may not fully engage with the focal users in the way real followers could have in real life. As a result, the effects associated with follower increases should be viewed conservatively. More broadly, our findings do not imply that knowledge-sharing platforms should avoid integrating social networks. Prior research shows that connecting users to external social networks can reduce cold-start problems by attracting new users (Chen, 2021). Moreover, users may gain both monetary (e.g., Sun et al., 2017) and non-monetary (e.g., Shriver et al., 2013; Toubia and Stephen, 2013) rewards from social engagement. For example, contributions from followees can motivate users to contribute more (Jin et al., 2022), and nudges from followers can stimulate content creation and foster appreciation for other creators (Zeng et al., 2023). Our findings speak only to follower count as a reputation symbol, not to the broader functionality of social networks. Our result suggests that rather than using follower count to signal user reputation, knowledge-sharing platforms should consider leveraging alternative levers of social networks (e.g., social nudges) to encourage followers to interact with the focal users.

The remainder of the article is organized as follows. Section 2 discusses how our paper adds to the extent literature. Section 3 describes the data and field experiment design. Section 4 presents the average treatment effects of peer voting and followers on user knowledge contribution. Section 5 examines heterogeneous treatment effects and spillover effects on other user behaviors. Section 6 concludes the article.

2 Contributions to Extant Literature

Our research contributes to the following three streams of literature. The first stream is prior work in operations management (OM) that investigates reputation systems as platform design choices (Brosig-Koch and Heinrich, 2014; Tang and Whinston, 2020). Reputation system design aims to reward user behavior by providing signals about individuals' past performance and trustworthiness (Resnick and Zeckhauser, 2002). Recent research in OM has explored different types of reputation signals, such as seller transaction history

(Deng et al., 2023), the number of followers or likes (Cheng et al., 2025), online search volume (Cheng et al., 2025), and product reviews or buyer feedback (Kokkodis et al., 2022) within the empirical context of different digital platforms. Additionally, researchers have also studied the effects of reputation on sales outcomes (Cheng et al., 2025), the interplay between different forms of reputation (Cheng et al., 2025), or strategies to improve reputation systems (Deng et al., 2023; Kokkodis et al., 2022). We contribute to this literature by examining two widely used reputation symbols—peer voting and follower count—as platform design choices.

The second stream focuses on how peer voting and follower count as reputation systems relate to digital content contribution. As shown in Table 1, prior studies have typically examined peer voting (e.g., Chen et al., 2018; Deolankar et al., 2024) and follower count (Moqri et al., 2018; Toubia and Stephen, 2013) in isolation, with limited research comparing or disentangling their respective effects. One such exception is Jin et al. (2022) that applies a hidden Markov model on observational data from Zhihu to characterize contribution dynamics in response to changes in received upvotes and followers over time. We differ from Jin et al. (2022) by employing a field experiment to identify the causal impacts of peer voting and increasing follower count on knowledge contribution (volume, effort, and quality), the evolution of these impacts, as well as spillover effects on other behaviors that potentially affect fellow users' contributions. Utilizing a field experiment offers two benefits. First, by making the increases in upvotes and followers exogenous, we can avoid potential confounders of treatment effects, such as user-level time-variant unobservables (e.g., users enhance domain-specific knowledge over time) that jointly influence upvotes received/follower requests as well as user contributions. Second, by orthogonalizing the typically co-determined variables of new upvotes and new followers in observational data, we are able to clearly identify and distinguish the effect of peer voting from that of increased follower count.

We also add to a third research stream that focuses on knowledge contribution on digital platforms. Within this literature, researchers have studied the impact of either non-reputation-related or reputation-related platform design choices on user contribution. For example, in the former group, researchers have examined the extent to which different non-reputation-related design features such as content monetization (Cong et al., 2025), live events (Cong et al., 2023), monetary rewards (Garnefeld et al., 2012), best practices (Li and Sandino, 2025), and recommendation algorithms (Liu and Cong, 2023) influence user contributions on knowledge-sharing platforms. Our research particularly contributes to the latter group. In addition to examining the relative roles of peer voting and follower count as reputation symbols, we also contribute to the third stream of literature by being the first to leverage modern LLMs to evaluate answer quality. This stream of research has typically relied on answer acceptance or upvotes as proxies for answer quality (e.g., Jiang,

Table 1. Relations of reputation systems (peer voting and follower count) to content contribution.

Paper	Data source	Data type	Key independent variables		Key dependent variables	Effect evolution	Effects on other behaviors
			Peer voting	Follower count			
Chen et al. (2018)	Super user	Observational data	✓		Volume		
Deolankar et al. (2024)	Reddit	Observational data	✓		Volume and intensity		
Zhu et al. (2024)	Anonymous	Observational data	✓		Volume and quality		
Shen et al. (2025)	Amazon	Observational data	✓		Volume and quality	✓	
Jiang (2024)	Stack Overflow	Field experiment	✓		Volume, quality, and length	✓	✓ vote, login, and comment
Moqri et al. (2018)	Github	Observational data		✓	Volume	✓	
Toubia and Stephen (2013)	Twitter	Field experiment		✓	Volume		
Jin et al. (2022)	Zhihu	Observational data	✓	✓	Volume and quality		
This study	**Zhihu**	**Field experiment**	✓	✓	**Volume, quality, and length**	✓	**✓ upvote, follow, and purchase**

2024; Jin et al., 2022; Safadi et al., 2021). These measures are useful but imperfect, as they depend heavily on answer exposure, which can be influenced by topic popularity, the author's social connectedness, and a range of other factors. Hiring domain-specific experts to rate quality across the wide range of topics on these platforms is also practically infeasible. Leveraging recent advances in large language models, we use GPT-4 to provide quality assessments for each answer. GPT-4 has demonstrated strong performance across diverse knowledge domains (OpenAI, 2023). Our use of GPT-4 offers a pragmatic and scalable proxy for answer quality, illustrating the potential for LLMs to support research on knowledge-sharing platforms. Nonetheless, LLM-based quality assessments should be interpreted with caution, as they remain subject to inherent measurement error.

3 Data and Experiment Design

We conduct our field experiment on Zhihu.com, the counterpart of Quora in China. Zhihu.com was launched in 2011 and has become China's most prominent question-and-answer style knowledge-sharing platform since then,[8] with 200 million registered users in 2018[9] and a 3.5 billion USD valuation in 2019. Similar to Quora, users on Zhihu can ask questions, answer questions, vote on answers, and follow other users. Upvote and follower counts are prominently displayed as reputation symbols.

3.1 Sampling Users for Our Field Experiment

Given that there are 200 million registered users on Zhihu, it is impossible to carry out our field experiment for the entire set of users. Instead, we aim to sample a representative subset of users from Zhihu. In such a case, the most convenient way is to sample nodes uniformly at random from the entire space of users on the platform. Nevertheless, the space of Zhihu users is massive and hidden, and the Zhihu API does not offer any avenue to randomly sample users. However, we could find a subsample of users on the platform (e.g., by checking users following a range of topics on the platform), and the API enables us to explore the neighbors of these users.

In such situations, two types of sampling methods are available (Hasan et al., 2013): (1) graph traversal techniques, which sample users without replacement (e.g., forest fire, snowball sampling, and depth-first search), and (2) random walk techniques, which sample users with replacement (e.g., classic random walk, random walk with jumps, and MHRW). Chen et al. (2013), Ebbes et al. (2016), and Aral (2016) compare these sampling methods. Their consensus is that the sampling strategy depends on the goal of the experiment. We aim to generalize our findings on the effects of peer voting and follower count on user knowledge contribution on Zhihu and similar online knowledge-sharing platforms. We therefore seek to obtain a representative sample of the Zhihu population. The MHRW sampling method mimics uniform sampling, which helps us sample representative users of the platform (Gjoka et al., 2010, 2011; Hasan et al., 2013; Wang et al., 2010). In contrast, other graph traversal and random walk techniques tend to oversample users with more links (i.e., followers or followees), because there are more paths from the seeding users to the more-connected users (Hasan et al., 2013).

Specifically, we first randomly pick one seeding user from the followers of each of the ten most popular topics on Zhihu.com, including education, economics, investment, law, healthcare, science & technology, culture, game, art, and anime. For each seeding user, we run an MHRW process for 11,000 iterations, resulting in 10 parallel MHRW tracks. Following Wang et al. (2010), we define the degree of a user u, denoted d_u, as the number of neighbors (i.e., total of followers and followees). In each iteration, starting from user u, we randomly select a neighbor v, and move to v with probability $\min\{1, \frac{d_u}{d_v}\}$. This walk favors transitions toward lower-degree users (when $\frac{d_u}{d_v} > 1$), and reduces transitions toward high-degree users (when $\frac{d_u}{d_v} < 1$), thus removing the bias toward highly connected users. It has been shown that under MHRW, in expectation, every user is sampled with equal probability $\frac{1}{n}$, where n is the total number of users in the network (Hasan, 2016; Wang et al., 2010).

Next, we conduct the Geweke test (Geweke, 1991) and the Gelman-Rubin test (Gelman and Rubin, 1992) to examine convergence of each MHRW track (i.e., whether the degrees of sampled users have reached a stable distribution). The Geweke test shows that all ten tracks converged after 800 iterations, and the Gelman-Rubin test confirms convergence across tracks after 200 iterations. To be conservative, we discard the first 1,000 iterations per track as burn-in. To reduce autocorrelation, we apply a thinning procedure and retain one sample every five iterations. Note that since MHRW samples with replacement, some users appear multiple times. After this procedure, we obtain 3,918 unique users. Because inactive users may not respond to recognition or rewards, we oversample active users.[10] We define a user as "active" if they contributed at least one answer or question in the two months before our data collection began. This yields 1,190 active users. We then randomly select 810 inactive users from the remainder, resulting in 2,000 users. Finally, we remove users whose pages are unavailable, users banned or flagged by the platform, and users with extremely high follower counts (i.e., $> 20{,}000$).[11] This leaves us with 1,696 focal users: 966 active and 730 inactive.

We record the user ID, number of followers, number of received upvotes, contributed answers, users' self-reported gender, and other behaviors not regarded as knowledge contribution but that may influence fellow users' contributions—including the number of posted questions, number of upvotes to fellow users' knowledge contributions, number of followed users, and number of purchased Zhihu Lives.[12] Table 2 presents the summary statistics of these user characteristics on Day 1 of our observation window.[13] Figure 1 shows the histograms and log-log plots for users' total number of

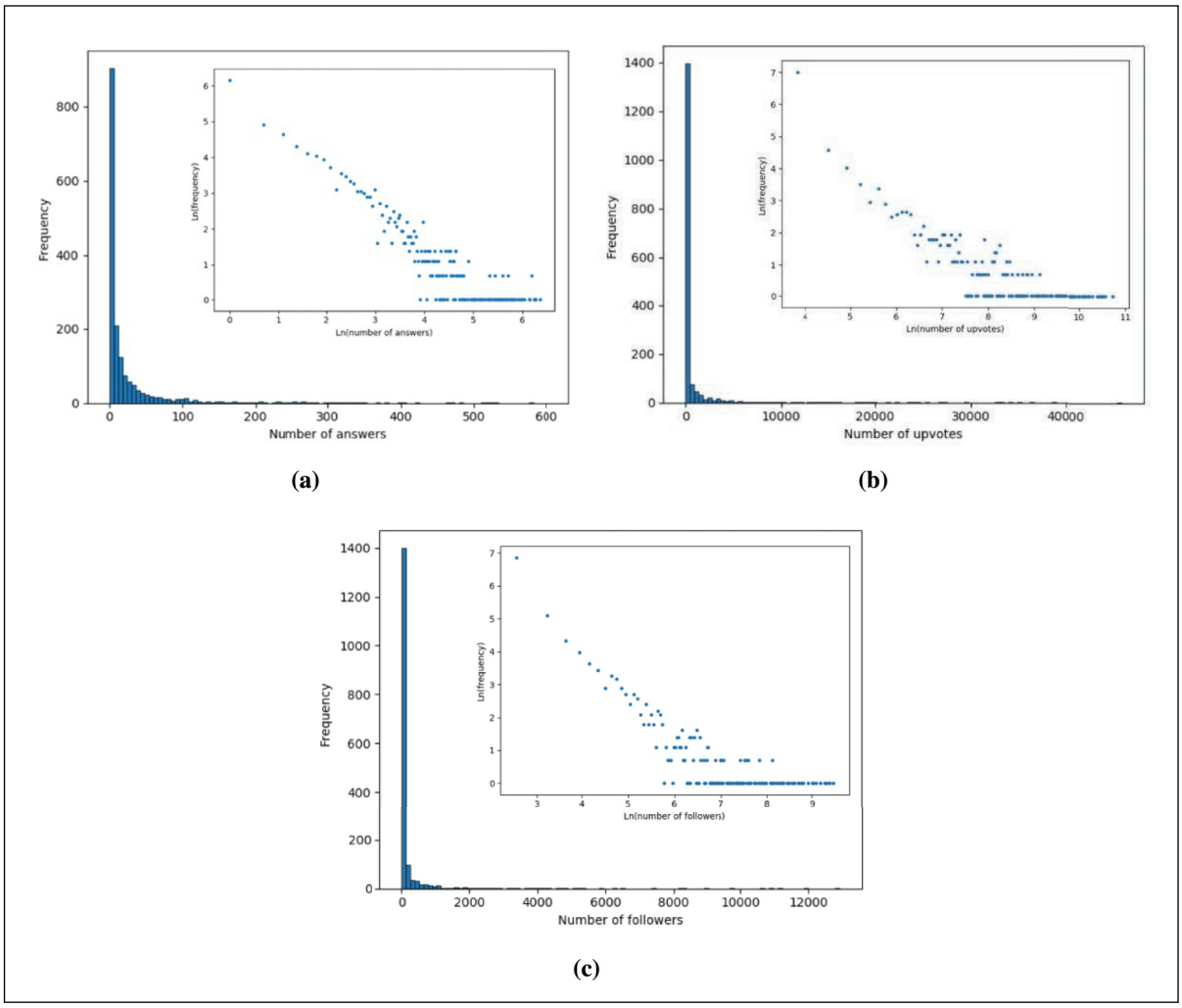

Figure 1. Histograms and Ln–Ln plots of total number of answers, received upvotes, and followers on day 1 of the observation window across all users.

Table 2. Summary statistics of user characteristics on day 1 of the observation window.

	N	Mean	Median	Standard deviation	Min	Max
Total number of upvotes	1,696	930.474	7	3626.526	0	45,734
Total number of followers	1,696	240.019	8	1018.905	0	12,820
Total number of answers	1,696	25.871	5	62.015	0	576
Total number of questions	1,696	2.068	0	8.335	0	212
Total number of followees	1,696	93.513	25	292.287	0	5,000
Total number of purchased Zhihu Lives	1,696	1.035	0	5.553	0	112
Gender (female = 1)	1,228	0.471	0	0.499	0	1

Notes: Gender row has fewer observations because 27.6% of users do not report their genders.

answers, received upvotes, and followers on Day 1. We observe that the distributions of all three variables are right-skewed, which is common in online platforms (Stephen and Toubia, 2009). To validate the representativeness of the sampled users with respect to the broader Zhihu user base, we compare their statistics with those reported by Liu and Cong (2023) (see Web Appendix A). This comparison demonstrates close alignment across key characteristics.

3.2 Experiment Conditions and Assignment

We design a randomized between-subject field experiment with 2 (upvote intervention vs. not) by 2 (follower intervention

Table 3. Randomization check.

	Control		Upvote		Follower		Upvote+Follower		*p*-value
N	1,390		106		102		98		
Total number of upvotes	900.96	(95.44)	973.73	(380.23)	1144.70	(411.40)	1079.37	(375.97)	0.89
Total number of followers	221.31	(25.56)	366.75	(134.49)	339.46	(118.40)	265.96	(127.28)	0.37
Total number of answers	25.95	(1.67)	24.91	(5.57)	29.14	(7.55)	22.35	(4.89)	0.89
Total number of questions	1.99	(0.18)	3.95	(2.22)	1.51	(0.27)	1.76	(0.31)	0.11
Total number of followees	91.70	(8.07)	121.86	(26.54)	100.62	(30.33)	81.13	(14.07)	0.73
Total number of purchased Zhihu Lives	1.09	(0.16)	1.01	(0.41)	0.35	(0.15)	0.95	(0.31)	0.63
Gender (female = 1)	0.46	(0.02)	0.52	(0.06)	0.51	(0.06)	0.57	(0.06)	0.21

Notes: Standard errors in parentheses.

vs. not) conditions. We name the four conditions *control*, *upvote*, *follower*, and *upvote+follower* conditions. One challenge faced by experiments conducted on social networks is that the treatment received by one user in one condition can interfere with the potential outcome of another connected user in a different condition (Aral, 2016). This violates the Stable Unit Treatment Value Assumption (SUTVA, Rubin, 1986) necessary for causal inference, which requires that the potential outcome of one user should be unaffected by the particular assignment of treatments to other users.

To minimize possible interference between experimental conditions, we use *graph cluster randomization* (Aral, 2016; Eckles et al., 2017; Ugander et al., 2013) to assign users to the four conditions. The intuition is to randomly assign treatment at the cluster rather than the individual level so that users will be maximally surrounded by others assigned to the same treatment. Following Ugander et al. (2013), we apply the ϵ-net graph clustering algorithm with $\epsilon = 1$ to partition the network of 1,696 focal users. This algorithm groups users into clusters (or "balls") of radius 1, each consisting of a center user and their directly connected neighbors, on the social network formed by our focal users. The 1-net clustering approach is well-suited for cluster-level randomization, as it yields relatively small and similarly sized clusters, facilitating balanced distributions of user characteristics across experimental conditions. Using this method, we obtain 1,088 clusters from the 1,696 users.

We then randomly assign clusters to experimental conditions. For implementation feasibility (see Section 3.3), we want each treatment condition to have about 100 users. In the end, 106 users are assigned to the upvote condition, 102 users to the follower condition, and 98 users to the upvote+follower condition. The remaining 1,390 users are assigned to the control condition. To verify that users are similar across experimental conditions, in Table 3, we compare user characteristics on Day 1 of the observation window across experimental conditions. The *p*-values from one-way ANOVA tests show that the sample means of the different groups are not significantly different, indicating users are similar across experimental conditions.[14]

3.3 Experiment Timeline and Intervention Implementation

We observe the 1,696 users for 303 days and collect data on a daily basis during the observation window. Figure 2 illustrates the timeline of our field experiment. Specifically, there are 50 days before the experimental intervention. We then implement the intervention for 53 days. Following that, we compare users' contributions in the 50 days after the intervention to the 50 days before the intervention to gauge the causal impacts of increasing upvotes and/or followers on user contributions. Lastly, we continue to monitor the users for another 150 days to examine how the treatment effects evolve over time.

To implement the experimental interventions, inspired by prior research (Bohren et al., 2019; Qiu and Kumar, 2017; Toubia and Stephen, 2013) we aim to give users receiving the upvote (follower) intervention 100 extra upvotes (followers), using 100 synthetic users. We choose 100 as the target quantity for both interventions to ensure comparability and practical feasibility. We have also examined the relative salience of the interventions based on percentage increase (see Section 4.1). We hire three research assistants who are active users of Zhihu.com to manage these synthetic users.

The research assistants made the following efforts over the period of 50 days before the experiment interventions to prepare these synthetic users so that they were as realistic as possible. First, we curated each synthetic user's name, gender, portrait, headline, and industry to make them look coherent for each user and diversified across users. Second, we let the synthetic accounts follow each other so that each synthetic user has some followers and followees. Additionally, we let the synthetic users follow some topics, questions, columns, collections, and some popular users on Zhihu.com. Lastly, each synthetic user upvoted some top answers[15] within the topics followed by the synthetic user. In Web Appendix B, we present an example of a synthetic Zhihu user and report results from a survey of 118 Zhihu users assessing the perceived realism of our synthetic users. The results show that Zhihu users cannot distinguish synthetic users from users in our control group and their followers. Another related concern is that followers of a typical Zhihu user might be friends of the focal

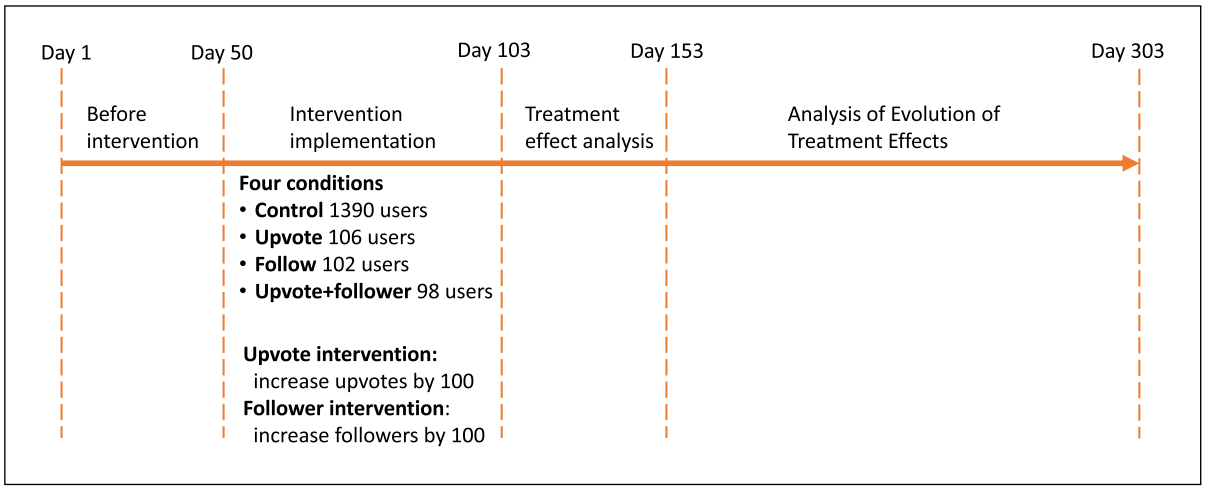

Figure 2. Experiment design and timeline.

user, so treated users might react to received upvotes or followers from our synthetic users differently than they would normally. To examine this possibility, we asked the same 118 participants to indicate the approximate percentage of their followers who are friends on Zhihu. The average percentage is 27.03% (s.e. = 2.13%), indicating that most followers of a typical Zhihu user are strangers.

We implement the upvote and follower interventions from Day 51 to Day 103 of the observation window. Following Toubia and Stephen (2013), to enhance the perceived authenticity of these interventions, we vary the number of upvotes and follow requests from synthetic users assigned to each treated user per day, ranging from 2 to 6, with some days receiving none. In addition, for each treated user, each upvote or/and follow request is given by a different synthetic user over the course of the intervention. All treatment conditions are implemented concurrently each day by the three research assistants. For the upvote intervention, research assistants are instructed to upvote the most recent answer by a treated user that appears to be of acceptable quality at the time of upvoting. As a result, a treated user may receive upvotes on one or more answers as they would organically.

3.4 Extracting Answer Quality Using LLMs

The focal users contributed 12,998 answers in our 303-day observation window, covering diverse topics. Hence, it is challenging to hire experts in each field to rate each answer's quality. To overcome this challenge, we use GPT-4, an LLM developed by OpenAI, to evaluate answer quality.

We measure each answer's overall quality and five quality dimensions—readability, relevancy, completeness, credibility, and objectivity—on 7-point Likert scales (with 1 being the lowest and 7 being the highest). Following prior literature (Goes et al., 2014; Jin et al., 2022; Khern-am-nuai et al., 2018; Singh et al., 2014), we define overall quality as the overall effectiveness and usefulness of the answer in addressing the question; readability as how easily the answer can be read and understood by the audience; objectivity as the impartiality and lack of bias in the answer; completeness as whether the answer covers all essential aspects of the question comprehensively; credibility as the reliability and trustworthiness of the information provided in the answer; and relevancy as how closely the content of the answer aligns with the question asked.

We validate the answer quality ratings provided by GPT-4 by comparing them with ratings from a human expert and alternative LLMs in Web Appendix C. We find that GPT-4 demonstrates a high degree of alignment with human expert ratings on the subsample of human-labeled answers. Furthermore, the ratings of leading LLMs (e.g., GPT-4 and Gemini 2.0) show strong consistency with each other. We choose to use GPT-4 for our study as it exhibits the highest agreement with human expert ratings among the LLMs in comparison.

4 Average Treatment Effects of Peer Voting and Follower Count

We measure user contribution using three metrics: answer volume, length, and quality. The volume captures the frequency of contribution, the length captures users' effort spent in writing the answers, and the quality reflects the effectiveness and usefulness of the answer in addressing the question.

4.1 Average Treatment Effects: 50 Days After Versus 50 Days Before the Intervention

To evaluate the impacts of our experimental interventions (i.e., upvote intervention and follower intervention), we compare contribution volume, length, and quality over the 50 days following the experimental intervention to the 50 days preceding it. Following Toubia and Stephen (2013), we focus on comparing content contribution before versus after the intervention period, because user behavior during the intervention might change as users gradually accumulate more upvotes/followers

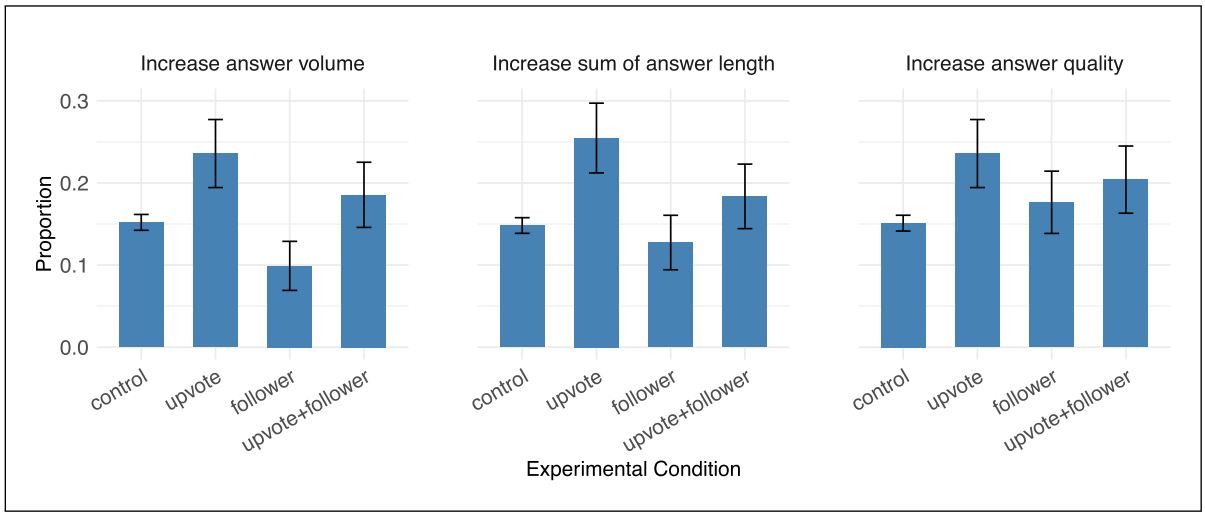

Figure 3. Changes in knowledge contribution by experimental conditions. Notes: The error bars represent ± standard error of the mean for each group.

during the treatment period. For example, users receiving the upvote intervention may initially refrain from answering more questions as they monitor engagement (i.e., upvotes).[16]

In line with Toubia and Stephen (2013), we examine changes in user contributions by calculating the proportion of users whose contributions increase in the post-intervention period relative to the pre-intervention period. Our modeling choice is motivated by well-known platform priorities such as maximizing the proportion of active users (Chen et al., 2024). This operationalization also reduces sensitivity to outliers with extreme changes in contributions, which is important given the highly skewed distribution of user activity typical of online platforms (Liu and Cong, 2023; Toubia and Stephen, 2013).[17] Additionally, we use total answer length as a proxy for a user's total effort, a meaningful outcome for knowledge sharing platforms (Wasko and Faraj, 2005).[18]

Figure 3 presents the model-free evidence from the four experimental conditions, showing that a greater proportion of users in the upvote condition increase their answer volume, length, and quality compared to those in the control condition. Conversely, users in the follower condition exhibit a lower or similar proportion of increase in knowledge contributions relative to those in the control group. Users in the upvote+follower condition generally fall between those in the upvote condition and those in the follower condition, indicating a mixed impact when both interventions are applied together.

We then formally test the results by estimating the following model:

$$\begin{aligned}\text{ansInc}_i = \beta_0 + \beta_1 \cdot \text{upvoteInt}_i + \beta_2 \cdot \text{followInt}_i + \beta_3 \\ \cdot \text{upvoteInt}_i \cdot \text{followInt}_i + \varepsilon_i. \end{aligned} \quad (1)$$

Let i denote the user. $\text{ansInc}_i = 1$ if the user increases the number of answers per day (or the sum of answer lengths per day, or the average answer quality) in the 50-day post-intervention period compared to the 50-day pre-intervention period; $\text{ansInc}_i = 0$ otherwise.[19] upvoteInt_i is the dummy for the upvote intervention, and followInt_i is the dummy for the follower intervention. β_0 captures the baseline, namely, the percentage of users in the control condition who increased their contribution organically after the intervention period. β_1 captures the average treatment effect of the upvote intervention, and β_2 captures that of the follower intervention. β_3 represents the additional effect of receiving both interventions compared to receiving only one. ε_i is the error term. We address heteroskedasticity of the linear probability model by employing robust standard errors.

Column (1) of Table 4 shows the treatment effects on answer volume. We observe that users who receive the upvote intervention are 8% more likely to increase answer volume after the intervention compared to the control group. Users receiving the follower intervention are marginally less likely to increase answer volume. The muted effect of follower count on answer volume is consistent with findings by Toubia and Stephen (2013). Moreover, the coefficient for the interaction term is not significantly different from zero, indicating that the two interventions are largely independent in their effects. In other words, one intervention neither enhances nor diminishes the effect of the other, and users receiving both interventions experience a simple sum of their respective effects.

Column (2) of Table 4 shows the treatment effect on answer length. We find that users receiving the upvote intervention are 11% more likely to increase their contribution effort than those in the control condition. Users receiving the follower intervention are not significantly different from users in the control condition. We also observe that the coefficient of the interaction term is not significant, indicating again that the two interventions are in general independent in their effects on the overall contribution effort.

Table 4. Average effects of increasing upvotes or/and followers on answer contribution.

	(1) Increase answer volume	(2) Increase sum of answer length	(3) Increase answer quality
(Intercept)	0.152***	0.148***	0.151***
	(0.00964)	(0.00953)	(0.00961)
Upvote intervention	0.0838*	0.107*	0.0848*
	(0.0423)	(0.0434)	(0.0423)
Follower intervention	−0.053·	−0.0208	0.0254
	(0.0312)	(0.0344)	(0.0389)
Upvote intervention * Follower intervention	0.00273	−0.0503	−0.0572
	(0.0651)	(0.0671)	(0.0698)
N	1,696	1,696	1,696
Log likelihood	−681.475	−680.825	−705.296

Notes: Robust standard errors in parentheses. $\cdot p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Column (3) of Table 4 shows the treatment effect on answer quality. We see that users receiving the upvote intervention are 8% more likely to increase answer quality than the control condition, but the follower intervention does not have a significant effect on answer quality on average. We also test the effects of the two interventions on specific answer quality dimensions in the Web Appendix C Table A8, which shows that the upvote intervention has a positive impact on answer readability, relevancy, and completeness, while the follower intervention does not have a significant impact on any quality dimensions on average.

So far our results suggest that increasing upvotes stimulates user contributions, while an increase in followers does not significantly impact or may even marginally reduce contributions. An alternative explanation is that the upvote intervention has a more positive impact because this intervention might be more salient to treated users than the follower intervention. To examine this possibility, we first investigate how Zhihu informs users when they receive new upvotes and followers. On Zhihu, users who received a new upvote are only notified when they log into their Zhihu account. In contrast, users who received a new follower are notified not only on the Zhihu app but also by email, making follower notifications presumably more salient than upvote notifications. Second, we compare the intensity of the intervention by examining the percentage (rather than absolute) increase for each intervention. Based on the data in Table 3, an average user in the upvote condition has 974 upvotes, so receiving 100 upvotes represents a 10% increase.[20] In contrast, an average user in the follower condition has 339 followers, so receiving 100 followers represents a 29% increase, larger than the percentage increase from the upvote intervention. Last but not least, we compare how Zhihu displays total upvote and follower counts to the focal user. Both reputation symbols are prominently displayed on user's profile page. Overall, we do not find strong evidence for the alternative explanation.

4.2 Why Do Received Upvotes Encourage Knowledge Contribution, But Followers Do Not?

Prior literature suggests that, besides economic incentives, two types of utilities may motivate platform users to contribute content: *intrinsic* utility and *status-related* utility. Intrinsic utility stems from the act of contributing for its inherent satisfaction (e.g., self-efficacy and the joy in helping others) rather than for some separate consequence such as rewards (e.g., Kankanhalli et al., 2005; Ryan and Deci, 2000). Status-related utility, as one type of extrinsic utility, reflects the benefits that users may derive from reputation or prestige (e.g., Fershtman and Gandal, 2007; Wasko and Faraj, 2005). Based on this literature, we hypothesize that users on knowledge-sharing platforms also receive intrinsic utility and status-related utility.

The key difference between intrinsic vs. status-related utilities, as suggested by Toubia and Stephen (2013), is that whereas intrinsic utility is derived from *posting* content viewed by many users, status-related utility is derived from *having* a high reputation (e.g., a larger follower count or upvote count). Hence, if intrinsic utility dominates, users will contribute more after the intervention because now they enjoy more intrinsic utilities (e.g., more confidence after receiving upvotes, a larger audience after receiving followers) each time they contribute. In contrast, if the status-related utility dominates, users may contribute less after the intervention because now they have a higher status and may become less eager to continue contributing to earn more status. The diminishing marginal return of status is consistent with the experimental evidence that satiating the need for social acceptance leads to a reduction in the drive to satisfy that need (DeWall et al., 2008).

Based on the observed average treatment effects, we conjecture that, in the upvotes intervention, intrinsic utility may outweigh status utility, whereas in the follower intervention, status utility may play a comparatively larger role. This conjecture rests on the idea that upvotes are tied to specific answers, potentially functioning as direct signals of peer recognition that could foster intrinsic motivation. In contrast, followers are awarded to users, which may elevate perceived status but may

not offer the same immediate intrinsic reinforcement for contributions. Indeed, follower count is widely seen as a marker of status, frequently used in prior research to quantify influence and popularity (Cha et al., 2010; Kwak et al., 2010) and commonly perceived as such by the public (Beck, 2009; Leonhardt, 2011).

We carry out two investigations to test our conjecture. First, we invite Zhihu users to answer a survey to better understand their thought process upon receiving upvotes vs. followers on Zhihu. Second, we seek empirical support for the underlying mechanisms using heterogeneous treatment effects along user reputation and the spillover effects to other activities on the platform. We report findings from the survey below and results from heterogeneous treatment effects and spillover effects in Section 5.

Specifically, we conduct a survey of 100 Zhihu users recruited via InsightWorks (a Chinese research company). See Web Appendix E for the survey we use. These users are first asked about how they associate upvotes and followers with answers or personal characteristics, respectively. When asked about "Zhihu users often receive upvotes and gain followers. Which of the two do you think is more a result of their answers?," 78% of participants chose "upvotes." In contrast, 68% of participants chose "followers," when asked "which of the two do you think is more a result of their personal characteristics?."

We further present participants with a list of questions that ask to what extent they attribute a set of intrinsic and status utilities to receiving upvotes vs. followers on Zhihu. We follow Kankanhalli et al. (2005) to develop the battery of statements associated with intrinsic versus status utilities. We then provide each participant with a slider scale anchored at -5 and 5, with -5 (5) representing the participant attributing the statement towards receiving followers (upvotes). For example, one question asks the extent to which they associate receiving upvotes or followers with the statement: "Makes me feel that my contributions are recognized by peer users on Zhihu." We choose a slider scale rather than a single choice question under the consideration that receiving upvotes (or followers) can have both intrinsic and status utilities, and the key is whether participants attribute the two types of utilities more towards receiving upvotes or followers. In line with our conjecture above, we observe that users attribute intrinsic utilities more towards receiving upvotes and status utilities more towards getting followers ($\text{mean}_{\text{intrinsic}} = 0.42$, $\text{mean}_{\text{status}} = -0.8$, p-value = 0.000).

Our findings offer tentative implications for the design of reputation systems on knowledge-sharing platforms. If follower count functions as a status signal that may have mostly no effects or reduce some users' motivation to contribute, platforms might consider experimenting with ways to downplay follower count as a prominent reputation symbol. Conversely, platforms might consider placing greater emphasis on peer voting, as users could be more likely to associate upvotes with recognition for their specific contributions, which may help reinforce intrinsic motivation to contribute. However, the implications should be viewed conservatively, as it is also possible that we do not detect a positive effect of adding followers because our synthetic followers did not fully engage with users as real followers might have.

4.3 Evolution of the Average Treatment Effects Over Time

Prior research has shown that the effect of rewards can change over time (Bénabou and Tirole, 2006). Specifically, rewards may stimulate performance in the beginning, but impair performance in the long run (Kruglanski, 2015) or when the rewards stop (He et al., 2021). For example, in a survey of a variety of programs for getting people to lose weight, stop smoking, or wear seat belts, Kohn (1993) find that people who received rewards showed better compliance at the beginning, but worse compliance in the long run than people who did not receive the rewards. One potential reason is that extrinsic rewards may crowd out people's intrinsic motivation (Bénabou and Tirole, 2006). Hence, in our context, it is important to examine whether the effects of the upvote or follower intervention change over time, particularly whether they may have a negative impact once the intervention ends.

To this aim, we track all users for another 150 days, divide the 150 days into three 50-day intervals, and compare each interval with the 50 days before the experiment. For each comparison, we estimate the same specification as in equation (1), except that $\text{ansInc}_i = 1$ if the user increases the number of answers per day (sum of answer lengths per day, or average answer quality) in the focal interval compared to the 50-day pre-intervention period; $\text{ansInc}_i = 0$ otherwise.

Figure 4 presents the treatment effects across four intervals: days 1–50, 51–100, 101–150, and 151–200 following the intervention. The treatment effects in the first 50 days after intervention are the same as those in Table 4; we include them here for comparison. We observe that: (1) the positive effects of increasing upvotes on answer volume and quality are significant for the first 100 days after the intervention and become insignificant after that; and (2) the positive effect of increasing upvotes on answer length is only significant for the first 50 days after the intervention.

In comparison, we find that the effects of increasing followers on answer volume, length, and quality are largely insignificant or marginally negative, and eventually fade out. The results suggest that the upvote intervention can stimulate answer contribution for around 50–100 days and does not have a negative impact in the long run. Meanwhile, the follower intervention has a more muted effect on answer contribution, and this effect does not reverse in the long run.

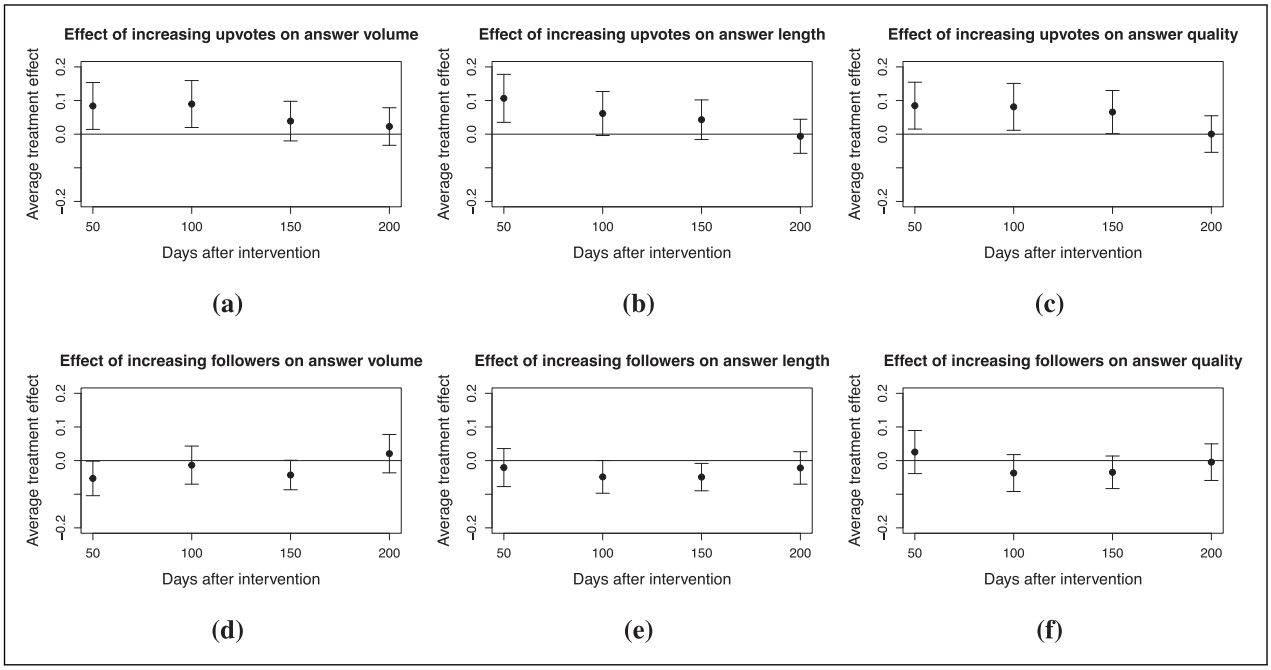

Figure 4. Evolution of the effects of increasing upvotes/followers on knowledge contribution. Notes: The error bars represent the 90% confidence intervals.

4.4 Investigating Potential Effect of an Increase in Followers on Upvotes Received

So far our results show that increasing follower count does not encourage contribution, but it is possible that an increase in followers might lead to more future engagement (e.g., viewing and upvoting) from these followers, which in turn might increase user contribution. Because we do not observe whether the followers have viewed the content created by the focal users, we look into whether the focal users receive more upvotes after a fellow user becomes a follower in an organic setting.

To examine this possible indirect effect, we first look at model-free evidence by directly examining whether and when organic followers give upvotes to followed users using data from the 1,390 control users over a 303-day observation window. Because we track the followers each focal user has on a daily basis, we are able to identify new organic followers by extracting differences in the followers from two consecutive days. Similarly, to identify new upvotes, we compare upvotes from two consecutive days and record the differences. We then match new upvotes with new followers by user IDs.

Next, we compare the number of upvotes given before, during, and after an organic follower follows the focal user. We find that most upvotes from new followers are given on the day they start following the focal user: on average, a new organic follower gives 0.032 upvotes to the focal user that day. Very few upvotes occur before following: on average, 0.003 upvotes per new organic follower in the 50 days before following. More importantly, interactions remain rare after following: in the 50 days post-following, a new organic follower gives an average of 0.003 upvotes to the focal user. This indicates that, even among organic followers, interactions in the form of upvotes are infrequent after the initial following event on Zhihu.

In addition, we formally test the relationship between an increase in followers on upvotes received in the future using the data from all control users during the 303-day observation window. Given that not many users on Zhihu are active on a daily basis, we analyze the data at the user-week level. We first calculate the correlation between the received followers and received upvotes in the subsequent week. The correlation is 0.078 (p-value $< 2.2 \times 10^{-16}$), suggesting that, without considering other factors, they are positively correlated, albeit that the correlation is small. We then regress the number of upvotes received in a given week on the number of new followers gained last week, while controlling for factors such as lagged upvotes, lagged contribution, and user and week fixed effects. As reported in Web Appendix D, we find that after accounting for these factors, newly acquired followers do not lead to an increase in future upvotes.

5 Understanding Effect Heterogeneity and Spillover Effects to Other Activities

We further probe into the heterogeneous treatment effects on knowledge contribution and spillover effects to other behaviors. This investigation has two goals. First, it enables the platform to better understand the heterogeneous effects of increasing upvotes and followers across different users and

topics. Second, it helps to unveil the potential mechanisms underlying the average treatment effect.

5.1 Heterogeneous Treatment Effects by User Reputation

We group users into low- and high-reputation groups using a median split based on the total number of upvotes received on Day 1 of our observation window, resulting in 846 low-reputation users and 850 high-reputation users.[21] Following Cong et al. (2025), we modify equation (1) to examine the heterogeneous effect of upvote and follower interventions by user reputation. Let us denote G_{il} and G_{ih} as whether user i belongs to the low or high group. We then replace the intercept β_0 with group-specific intercepts β_{0l} and β_{0h} to capture potentially different baseline contributions. We also interact the term upvoteInt_i with G_{il} and G_{ih} (i.e., $\text{upvoteInt}_i \cdot G_{il}$ and $\text{upvoteInt}_i \cdot G_{ih}$) to allow for heterogeneous treatment effects of the upvote intervention, and make similar modifications for the term followInt_i and the interaction $\text{upvoteInt}_i \cdot \text{followInt}_i$.

The top panel in Table 5 shows the results. We observe that for answer volume, the upvote intervention has a significant positive effect on users with lower reputation. Conversely, the follower intervention has a significant negative effect on users with higher reputation, and the effect is robust after we remove users with the top 1%, 5%, and 10% of the total number of upvotes (or followers) (Web Appendix F Table A11). Regarding answer length and quality, the upvote intervention also has a significant positive effect on users with lower reputation, but the follower intervention does not have a significant effect on either group.

The above results are consistent with our conjecture discussed in Section 4.2. We find that the positive average treatment effects of the upvote intervention on answer contributions are primarily driven by low-reputation users, who may be more strongly motivated by intrinsic utility. In contrast, the marginally negative effect of the follower intervention on answer volume appears to be driven largely by high-reputation users, who likely derive greater utility from status and may become less motivated to contribute once their status is further enhanced by additional followers. This heterogeneous effect aligns with (Toubia and Stephen, 2013), which suggests that status-related utility could become increasingly salient as users accumulate more followers. Taken together, these results tentatively suggest that displaying upvote count as a visible reputation symbol may encourage contributions from low-reputation users, whereas displaying follower count may be demotivating, specifically with respect to contribution volume, for high-reputation users.

5.2 Heterogeneous Treatment Effects Across Less and More Active Users

Motivating inactive users to contribute is crucial for the vibrancy of a knowledge-sharing platform (Chen et al., 2018; Jin et al., 2022). We classify users into less versus more active users using a median split of the total number of answers contributed up to Day 1 of our observation window, yielding 844 less active and 852 more active users.[22] We modify equation (1) in the same way as in Section 5.1 to examine the heterogeneous effects by user activity level, except that G_{il} and G_{ih} now represent less active and more active users, respectively.

The mid panel of Table 5 provides the results. We observe that increasing upvotes has a positive effect on answer length and quality for less active users. In contrast, increasing followers has a marginally negative effect on answer volume for more active users. These results suggest that using peer voting as a reputation symbol can motivate less active users to contribute, whereas relying on follower count as a reputation symbol may discourage more active users.

5.3 Heterogeneous Treatment Effects by Gender

Prior research has found that women tend to be underrepresented, less active, and less rewarded per answer in online knowledge-sharing platforms (Bohren et al., 2019; Gallus and Bhatia, 2020; May et al., 2019; Vasilescu et al., 2014). These disparities highlight the necessity of understanding gender-specific responses to platform features for creating a more inclusive online environment. We group users by their self-reported gender on Zhihu into three groups: female ($n = 579$), male ($n = 649$), and missing self-reported gender ($n = 468$). We modify equation (1) in the same way as in Section 5.1 to examine heterogeneous effects by gender, except that we now define three user groups: $G_{i,\text{female}}$, $G_{i,\text{male}}$, and $G_{i,\text{missing}}$.

The bottom panel in Table 5 summarizes the results. We observe that increasing upvotes has a significantly positive effect on females' answer volume and length. In contrast, increasing followers has a significant negative effect on answer volume for male users. The results suggest that, compared to men, women may respond more positively to peer voting as a reputation symbol, deriving greater intrinsic utility from it, while placing less value on follower count as a status-related reputation.

5.4 Heterogeneous Treatment Effects by Topic Type

So far, we have focused on how the treatment effects vary by user type. Meanwhile, the treatment effects of increasing upvotes and followers could also vary by topic type. Prior research has shown that soft topics (e.g., sports, entertainment, and travel) have a lower entry barrier than hard topics (e.g., science, technology, and law) (Bakshy et al., 2015), and user behaviors across soft and hard topics can be differently responsive to changes in platform design or experimental manipulations, such as the launch of Livestreaming (Cong et al., 2023) and initial ratings of a post (Muchnik et al., 2013).

We categorize topics of answers based on the topics of the parent questions.[23] Similar to Cong et al. (2025) and Bakshy et al. (2015), we consider 10 topic categories[24] as hard topics and the remaining 15 categories[25] as soft topics. During our

Table 5. Heterogeneous effects of increasing upvotes or/and followers on answer contribution.

	Increase answer volume			Increase sum of answer length			Increase answer quality		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Upvote intervention * Low reputation	0.103*			0.153**			0.156**		
	(0.0483)			(0.0539)			(0.0538)		
Upvote intervention * High reputation	0.0826			0.0704			0.0201		
	(0.07)			(0.0688)			(0.0662)		
Follow intervention * Low reputation	0.00665			0.00422			0.0263		
	(0.034)			(0.0334)			(0.0379)		
Follow intervention * High reputation	−0.111*			−0.0421			0.0293		
	(0.0517)			(0.0588)			(0.0656)		
Upvote intervention * Less active		0.0378			0.0817 .			0.0832 .	
		(0.0415)			(0.0493)			(0.0493)	
Upvote intervention * More active		0.0983			0.106			0.0648	
		(0.0645)			(0.0645)			(0.0631)	
Follow intervention * Less active		−0.00888			−0.00745			−0.00601	
		(0.0279)			(0.0278)			(0.0278)	
Follow intervention * More active		−0.0943 .			−0.0255			0.0489	
		(0.0554)			(0.0618)			(0.0679)	
Upvote intervention * Female			0.157 .			0.161*			0.128
			(0.0803)			(0.0802)			(0.0782)
Upvote intervention * Male			0.122			0.125			0.102
			(0.0836)			(0.0836)			(0.0811)
Follow intervention * Female			−0.00348			0.000829			0.0478
			(0.0602)			(0.0601)			(0.0668)
Follow intervention * Male			−0.111*			−0.0564			0.00538
			(0.0469)			(0.0573)			(0.065)
Group baseline controlled	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Three-way interactions controlled	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1,696	1,696	1,696	1,696	1,696	1,696	1,696	1,696	1,696
Log likelihood	−623.359	−612.624	−665.81	−628.707	−619.252	−668.775	−628.707	−619.252	−668.775

Notes: Robust standard errors in parentheses. Group baselines and three-way interactions are omitted for simplicity. In general, the high group and male group have higher baselines, and the three-way interaction terms are not significant. $.p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 6. Heterogeneous effects of increasing upvotes or/and followers by topic type.

	Increase answer volume		Increase sum of answer length		Increase answer quality	
	Soft topics (1)	Hard topics (2)	Soft topics (3)	Hard topics (4)	Soft topics (5)	Hard topics (6)
(Intercept)	0.122***	0.0993***	0.137***	0.108***	0.137***	0.104***
	(0.00879)	(0.00802)	(0.00921)	(0.00832)	(0.00921)	(0.0082)
Upvote intervention	0.0758.	0.0234	0.0803.	0.0336	0.0709.	0.0372
	(0.0397)	(0.0329)	(0.0411)	(0.0349)	(0.0405)	(0.0348)
Follower intervention	−0.00466	−0.0307	0.0104	−0.0393	0.0202	−0.0259
	(0.0331)	(0.0263)	(0.0363)	(0.0264)	(0.0372)	(0.0279)
Upvote intervention * Follower intervention	−0.0404	0.0101	−0.0335	0.0304	−0.0542	0.0476
	(0.0626)	(0.0514)	(0.0672)	(0.0549)	(0.0663)	(0.0576)
N	1,696	1,696	1,696	1,696	1,696	1,696
Log likelihood	−547.296	−356.567	−635.715	−427.961	−631.084	−422.927

Notes: Robust standard errors in parentheses. .$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 7. Spillover effects of increasing upvotes or/and followers.

	Increase question volume (1)	Upvote more users (2)	Follow more users (3)	Purchase more Zhihu lives (4)
(Intercept)	0.0303***	0.247***	0.231***	0.0259***
	(0.0046)	(0.0116)	(0.0113)	(0.00426)
Upvote intervention	−0.00196	−0.0305	0.0619	0.0024
	(0.0168)	(0.0417)	(0.0456)	(0.0167)
Follower intervention	0.00934	−0.0808*	−0.0919*	−0.0259***
	(0.0199)	(0.0387)	(0.0362)	(0.00426)
Upvote intervention * Follower intervention	−0.00672	0.119.	0.0469	0.0078
	(0.0311)	(0.071)	(0.072)	(0.0195)
N	1,696	1,696	1,696	1,696
Log likelihood	571.843	−964.464	−932.498	793.129

Notes: Robust standard errors in parentheses. The reason for the positive log-likelihoods in Columns (1) and (4) is that the outcomes are mostly zero, and the residuals are highly concentrated. For a normal distribution with a small variance, its density can be > 1 and thus its log-likelihood can be > 0. .$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

observation window, 44% of answers are from hard topics and 75% are from soft topics.[26]

We define two outcome variables: whether users increase their contributions to soft topics and whether they increase contributions to hard topics. We estimate equation (1) separately for each outcome variable. As users can contribute to both types of topics, all users are included in both estimations. Table 6 shows that increasing upvotes has a significantly positive effect on the volume, length, and quality of answers within soft topics, but increasing followers does not have a significant effect on answer contributions within either type of topic. These results suggest that (1) contributions within soft topics are more responsive to peer voting than those within hard topics; and (2) using peer voting as a reputation symbol can encourage user contributions in soft-topic areas, whereas increasing followers does not.

5.5 Spillover Effects of Peer Voting and Follower Interventions

We investigate whether the impacts of upvote and follower interventions spill over to other user activities, including posting questions, upvoting answers from fellow users, following fellow users, and purchasing fellow-user-hosted Zhihu Lives. To do so, we estimate equation (1) with different outcome variables. For posting questions, we compare each focal user's average daily question volume in the 50 days after the intervention to the 50 days before it. Similarly, for upvoting answers and following, we assess whether a focal user upvotes more answers from fellow users or follows more users per day after the intervention. For Zhihu Lives purchases, we examine whether the focal user purchases more Zhihu Lives per day after the intervention.

Table 7 shows that the upvote intervention does not significantly affect question posting, upvoting, following, or Zhihu

Live purchases of treated users. This pattern suggests that peer voting can encourage answer contributions without reducing engagement in other platform activities, consistent with our conjecture that upvotes provide contribution-specific recognition and may reinforce intrinsic motivation. In contrast, the follower intervention decreases users' upvoting, following, and Zhihu Live purchases. The overall null effect of increased followers on answer contributions, coupled with reduced engagement elsewhere, may reflect the stronger status-related utility associated with follower count. As users accumulate followers, they may feel more satisfied with their existing status and thus less motivated to pursue certain reputation-enhancing behaviors, such as answering questions, following or upvoting others for potential reciprocation, or participating in Zhihu Lives to acquire knowledge for future contributions. Taken together, these findings suggest that not all reputation symbols on knowledge-sharing platforms are equally beneficial: reputation signaled through upvotes may support continued engagement, whereas reputation signaled through follower count may not.

6 Conclusions

We investigate and compare the roles of peer voting and followers, two common reputation symbols, on users' knowledge contribution. Utilizing a randomized field experiment on a major knowledge-sharing platform, we find that peer voting boosts user knowledge contribution (in volume, length, and quality) for up to 100 days, whereas increasing follower count has no overall effect on contributions. Peer voting is especially effective for lower-reputation, less-active, and female users, and in soft topic areas; in contrast, increases in follower count appear to reduce contribution volume (not length and quality) for higher-reputation, more-active, and male users. We further show that peer voting primarily encourages answer contributions, whereas additional followers generate negative spillovers by reducing users' engagement with others. Overall, our findings suggest that upvotes may activate stronger intrinsic utility by providing recognition for specific contributions, whereas additional followers may enhance status-related utility, potentially making high-type users feel more content with their status and thus less motivated to contribute volume. To our knowledge, our research is among the first to employ a field experiment to disentangle these effects. We also contribute to the literature by leveraging an LLM to assess answer quality at scale.

Consequently, our research offers an operational perspective on how knowledge-sharing platforms might design reputation systems to better motivate user contributions. Because peer voting is tied to specific answers, it is more likely to signal peer recognition and provide intrinsic motivation, rather than function as a broad status cue. Platforms seeking to encourage contributions may therefore consider prioritizing or designing upvotes as a reputation symbol that emphasizes contribution-level feedback. In contrast, although social networks can help platforms attract users and facilitate interactions, platforms should be cautious about over-relying on follower count as a reputation symbol. While the magnitudes of the estimated effects may vary across platforms, the underlying insights may generalize to similar Q&A communities such as Quora or Reddit. Notably, these platforms differ in what they display as reputation metrics: Quora features follower count prominently, whereas Reddit emphasizes cumulative peer voting. Our findings suggest that the former approach may unintentionally strengthen the association between followers and status, while the latter approach may be more effective for sustaining user contributions.

Lastly, our research is subject to limitations and offers several promising avenues for future research. First, our conjecture regarding why followers may have no or muted effects on answer contribution remains speculative. The followers we added were synthetic and may not have fully engaged with our focal users as real followers might, which may partially explain why follower count exhibits limited effects. Second, the answer quality variable is constructed using an LLM and may be subject to measurement error. Third, our field experiment is carried out on Zhihu, which may be subject to some unique characteristics. For example, while Quora is known for its diverse user base, Zhihu's users are often better educated, and the discussion topics on Zhihu tend to gear more towards professional knowledge (Zhu, 2024). Thus, future research could examine the generalizability of our findings in other Q&A communities. More importantly, future work may experimentally vary platform designs to explore how the effects of peer voting and followers are moderated by other design elements, such as less salient follower displays, more prominent upvote visibility, or reduced reliance on followers for user or post exposure. Additionally, future studies could also explore other forms of follower engagement, such as content viewing, to gain a more comprehensive understanding of the role followers play in influencing focal users' knowledge contributions. Addressing these questions can help platforms better encourage knowledge contribution while preserving social connectivity.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

ORCID iDs

Mengxia Zhang https://orcid.org/0000-0002-4901-0635
Lan Luo https://orcid.org/0000-0003-1002-8237

Supplemental Material

Supplemental material for this article is available online (doi: 10.1177/10591478261429201).

Notes

1. For example, StackExchange had 5.1 billion visits and 9.6 billion pageviews in 2018 (https://stackexchange.com/about), and Stack Overflow alone had 27 million answers as of June 27, 2019 (https://stackexchange.com/sites). Quora had 300 million unique users every month as of September 2018 (https://en.wikipedia.org/wiki/Quora) and 89 million questions as of March 17, 2021 (https://www.quora.com/How-many-questions-have-been-asked-on-Quora-1). Zhihu had 300 million active users every month as of March 2017 (http://tech.sina.com.cn/roll/2017-05-20/doc-ifyfkkmc9878547.shtml).
2. For example, in 2019, Quora was valued at $2 billion (https://en.wikipedia.org/wiki/Quora), and Zhihu (the Chinese counterpart of Quora) was valued at $3.5 billon (https://www.caixinglobal.com/2019-08-13/chinese-quora-zhihu-completes-434-million-fundraising-101450117.html). In 2021, StackOverflow was sold for $1.8 billion (https://en.wikipedia.org/wiki/Stack_Overflow).
3. Based on the Stack Overflow user page (https://stackoverflow.com/users), about 22% of users had at least 10 reputation points as of November 16, 2022. According to Stack Overflow's reputation rules (https://stackoverflow.com/help/whats-reputation), each upvote gives ten points.
4. https://stackoverflow.blog/2008/12/01/why-cant-i-accept-my-own-answer/
5. https://stackoverflow.com/help/whats-reputation
6. OpenAI GPT has been known for working well across multiple languages (Goli and Singh, 2024). Our prompt is in English, except for the answer and question text which are in Chinese. The GPT-4 is able to understand the prompt. We further validate GPT-4-generated answer quality ratings by comparing them with ratings from a human expert and alternative large language models (LLMs).
7. Zhihu Lives are interactive live streaming sessions where speakers share knowledge of their areas of expertise. Zhihu Live participants can interact with the speaker in real time. Although Zhihu users post and read questions and answers for free, users often need to pay a small fee to join Zhihu Lives.
8. https://www.techinasia.com/zhihu-chinas-answer-quora
9. https://en.wikipedia.org/wiki/Zhihu
10. The oversampling of active users does not change our main findings. Please refer to Web Appendix A.
11. Only seven users have >20k followers at the time of our user sampling, so we cannot randomly assign them to the four experimental conditions to achieve balanced user characteristics. Also, given that we found increasing followers' effect is the most salient for high-reputation users, removing the seven users makes our results more conservative.
12. Very few of the sampled users have ever hosted livestreaming events themselves, so we focus on their purchased Zhihu Lives hosted by other users.
13. Note that the platform does not summarize the total number of upvotes to fellow users, so we count the number of outgoing upvotes by checking the user's activity page daily. Hence, the total number of outgoing upvotes on Day 1 was zero for all focal users and omitted in Table 2.
14. We also conducted pair-wise t-tests and tests using log-transformed variables, and the conclusion is the same.
15. Top answers are answers ranked high on a topic page, often answers with many upvotes and belong to questions with many views.
16. See more details in Web Appendix H where we examine the effects of upvote and follower interventions on user contribution during the intervention period.
17. If we use the raw change in the number of answers (post- vs. pre-treatment) as the outcome variable, we still find that the upvote intervention has a positive average treatment effect on answer volume, whereas the follower intervention has no effect.
18. Because this measure reflects the combined impact on both raw answer volume and effort per answer, we also report results for mean and median answer length in Web Appendix G. The results for the follower intervention on mean and median answer length, for both average and heterogeneous effects, are consistent with our main model specification. The average treatment effect of the upvote intervention is directionally consistent but weaker than in our main results, and its heterogeneous treatment effects are in line with those from our main model specification.
19. In cases where a user does not contribute any answers within a given period, we assign an answer quality score of 0, assuming that any contribution is better than no contribution. This assumption is grounded in prior research (Chiu et al., 2006; Wasko and Faraj, 2005), which suggests that user-generated content on knowledge-sharing platforms carries two main types of value: information value and social value. Even when an answer has low informational quality, it can still provide social value by fostering engagement and interaction. Meanwhile, we also examine answer quality conditional on contribution, where the analysis focuses only on answers actually posted by users. We find that answers in the upvote condition show a more positive change in quality relative to the control condition ($M_{upvote} = 0.291$, $M_{control} = -0.044$, $p = 0.074$), where M denotes the change in mean answer quality (after–before). In contrast, answers in the follower condition do not exhibit a significant improvement compared to the control ($M_{follower} = -0.303$, $M_{control} = -0.044$, $p = 0.269$). This pattern is consistent with the results reported in Table 4.
20. We calculate the percentage increase at the user level for both upvotes and followers, because the total number of upvotes and the total number of followers are prominently displayed on a Zhihu user's profile page on both the Zhihu website and mobile app. In contrast, the average number of upvotes per answer is not visible to users.
21. We also group users by follower count, and the results (Web Appendix F, Table A10) are consistent with our findings here.
22. We also categorize less versus more active users based on whether they contributed any answers or questions in the two months prior to the observation window. The results, presented in Web Appendix F Table A10, are consistent with results reported here. Re-estimating the models on these subsamples separately yields essentially the same results.
23. We categorize a question's topics based on its topic tags. When users post a question on Zhihu, the platform automatically attaches related topic tags, and users can choose additional tags. We observe 15,933 unique topic tags in our dataset. We classify these topic tags into 25 topic categories following Cong et al. (2025), and use GPT-4 to assign each topic tag to one of the 25 categories.

24. Law, Business, Economics & Finance, Healthcare, Architecture & Interior Design, Art Appreciation & History, Science & Technology, Sports, Investment, and Literature.
25. Internet, Psychology, Education, Travel, Lifestyle, Food & Cuisine, Career, Reading & Writing, Entertainment, Animal & Nature, Beauty & Fashion, Culture & Society, Politics, Religion, and others.
26. The two percentages do not sum to 100% because questions can span both soft and hard topics. For example, "What amazing programs or games have been developed using Python?" has five topic tags: game, software, Python, coding, and program. The first maps to *Entertainment* (a soft topic), while the rest map to Science & Technology (a hard topic).

References

Aral S (2016) Networked experiments. In: Bramoullé Y, Galeotti A and Rogers B (eds.) *The Oxford Handbook of the Economics of Networks*, pp.376–411. Oxford University Press.

Bakshy E, Messing S and Adamic LA (2015) Exposure to ideologically diverse news and opinion on facebook. *Science (New York, N.Y.)* 348(6239): 1130–1132.

Beck H (2009) New way to gauge popularity. *The New York Times*. https://www.nytimes.com/2009/10/21/sports/basketball/21nba.html. Published October 21.

Bénabou R and Tirole J (2006) Incentives and prosocial behavior. *American Economic Review* 96(5): 1652–1678.

Bohren JA, Imas A and Rosenberg M (2019) The dynamics of discrimination: Theory and evidence. *American Economic Review* 109(10): 3395–3436.

Brosig-Koch J and Heinrich T (2014) Reputation and mechanism choice in procurement auctions: An experiment. *Production and Operations Management* 23(2): 210–220.

Burtch G, Lee D and Chen Z (2024) The consequences of generative AI for online knowledge communities. *Scientific Reports* 14(1): 10413.

Cha M, Haddadi H, Benevenuto F, et al. (2010) Measuring user influence in Twitter: The million follower fallacy. In: *Proceedings of the international AAAI conference on web and social media*, Vol. 4. pp.10–17.

Chen A (2021) *The Cold Start Problem: How to Start and Scale Network Effects*. New York, NY: Harper Business.

Chen G, Chan T, Zhang D, et al. (2024) The impact of a more diversified recommender system on digital platforms: Evidence from a large-scale field experiment. *Available at SSRN 4365121*.

Chen W, Wei X and Zhu K (2018) Engaging voluntary contributions in online communities: A hidden Markov model. *MIS Quarterly* 42(1): 83–A8.

Chen X, Chen Y and Xiao P (2013) The impact of sampling and network topology on the estimation of social intercorrelations. *Journal of Marketing Research* 50(1): 95–110.

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

Cheng X, Jin Y, Xu Y, et al. (2025) Express: Revisiting seller reputation: Impact of internal and external reputation on sales in online knowledge payment markets. *Production and Operations Management* 34(8): 2180–2198.

Chiu CM, Hsu MH and Wang ET (2006) Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems* 42(3): 1872–1888.

Cong Z, Liu J and Manchanda P (2023) The role of 'live' in livestreaming markets: Evidence using orthogonal random forest. Working Paper.

Cong Z, Zhao Y and Zhang Z (2025) Understanding users' content contribution behavior when knowledge can be priced. *Marketing Science* 44(3): 655–670.

Deng J, Ghasemkhani H, Tan Y, et al. (2023) Actions speak louder than words: Imputing users' reputation from transaction history. *Production and Operations Management* 32(4): 1096–1111.

Deolankar V, Fong J and Sriram S (2024) The effect of downvotes on content creation: Evidence from social media. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4522092.

DeWall CN, Baumeister RF and Vohs KD (2008) Satiated with belongingness? Effects of acceptance, rejection, and task framing on self-regulatory performance. *Journal of Personality and Social Psychology* 95(6): 1367.

Ebbes P, Huang Z and Rangaswamy A (2016) Sampling designs for recovering local and global characteristics of social networks. *International Journal of Research in Marketing* 33(3): 578–599.

Eckles D, Karrer B and Ugander J (2017) Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference* 5(1): 20150021.

Fershtman C and Gandal N (2007) Open source software: Motivation and restrictive licensing. *International Economics and Economic Policy* 4: 209–225.

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

Gallus J and Bhatia S (2020) Gender, power and emotions in the collaborative production of knowledge: A large-scale analysis of wikipedia editor conversations. *Organizational Behavior and Human Decision Processes* 160: 115–130.

Garnefeld I, Iseke A and Krebs A (2012) Explicit incentives in online communities: Boon or bane? *International Journal of Electronic Commerce* 17(1): 11–38.

Gelman A and Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4): 457–472.

Geweke J (1991) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical Report Vol. 196, Federal Reserve Bank of Minneapolis, Research Department, Minneapolis, MN, USA.

Gjoka M, Kurant M, Butts CT, et al. (2010) Walking in facebook: A case study of unbiased sampling of osns. In: *Proceedings IEEE Infocom*. pp.1–9.

Gjoka M, Kurant M, Butts CT, et al. (2011) Practical recommendations on crawling online social networks. *IEEE Journal on Selected Areas in Communications* 29(9): 1872–1892.

Goes PB, Lin M and man Au Yeung C (2014) "Popularity Effect" in user-generated content: Evidence from online product reviews. *Information Systems Research* 25(2): 222–238.

Goli A and Singh A (2024) Frontiers: Can large language models capture human preferences? *Marketing Science* 43(4): 709–722.

Hasan MA (2016) Methods and applications of network sampling. In: *Optimization challenges in complex, networked and risky systems*, pp.115–139. INFORMS.

Hasan MA, Ahmed N and Neville J (2013) Network sampling: Methods and applications. In: *Tutorial of the 19th ACM SIGKDD*

conference on knowledge discovery and data mining. Chicago, IL, USA.

He C, Ozturk OC, Gu C, et al. (2021) The end of the express road for hybrid vehicles: Can governments' green product incentives backfire? *Marketing Science* 40(1): 80–100.

Jiang Y (2024) Dynamics in online knowledge contributions – Evidence from a field experiment. Working Paper.

Jin Y, Tan Y and Huang J (2022) Managing contributor performance in knowledge-sharing communities: A dynamic perspective. *Production and Operations Management* 31(11): 3945–3962.

Kankanhalli A, Tan BCY and Wei KK (2005) Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly* 29(1): 113–144.

Khern-am-nuai W, Kannan K and Ghasemkhani H (2018) Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research* 29(4): 871–892.

Kohn A (1993) *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes*. Boston, MA: Houghton Mifflin.

Kokkodis M, Lappas T and Kane GC (2022) Optional purchase verification in e-commerce platforms: More representative product ratings and higher quality reviews. *Production and Operations Management* 31(7): 2943–2961.

Kruglanski AW (2015) Issues in cognitive social psychology. In: *The hidden costs of reward*, pp.19–29. Psychology Press.

Kwak H, Lee C, Park H, et al. (2010) What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on World wide web*, pp.591–600.

Leonhardt D (2011) A better way to measure twitter influence. *The 6th Floor (blog)*. http://6thfloor.blogs.nytimes.com/2011/03/24/a-better-way-to-measure-twitter-influence. March 24.

Li SX and Sandino T (2025) The effect of a system for sharing best practices within pre-existing peer networks. *Management Science*. https://doi.org/10.1287/mnsc.2023.00990.

Liu J and Cong Z (2023) The daily me versus the daily others: How do recommendation algorithms change user interests? Evidence from a knowledge-sharing platform. *Journal of Marketing Research* 60(4): 767–791.

May A, Wachs J and Hannák A (2019) Gender differences in participation and reward on stack overflow. *Empirical Software Engineering* 24: 1997–2019.

Moqri M, Mei X, Qiu L, et al. (2018) Effect of "following" on contributions to open source communities. *Journal of Management Information Systems* 35(4): 1188–1217.

Muchnik L, Aral S and Taylor SJ (2013) Social influence bias: A randomized experiment. *Science (New York, N.Y.)* 341(6146): 647–651.

OpenAI (2023) Gpt-4 technical report. *arXiv preprint*.

Qiu L and Kumar S (2017) Understanding voluntary knowledge provision and content contribution through a social-media-based prediction market: A field experiment. *Information Systems Research* 28(3): 529–546.

Resnick P and Zeckhauser R (2002) Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system. In: *The Economics of the internet and e-commerce*, Vol. 11, pp.127–157. Emerald Group Publishing Limited.

Reuters (2023) ChatGPT sets record for fastest-growing user base – analyst note. Available at: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01 (accessed: 24 November 2024).

Rubin DB (1986) Comment: Which ifs have causal answers. *Journal of the American Statistical Association* 81(396): 961–962.

Ryan RM and Deci EL (2000) Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology* 25(1): 54–67.

Safadi H, Johnson SL and Faraj S (2021) Who contributes knowledge? Core-periphery tension in online innovation communities. *Organization Science* 32(3): 752–775.

Shen W, Liu Y and Wang Y (2025) Penalty or reward? The effect of social disincentives on online users' contributions. *Management Science* 71(8): 6770–6792.

Shorakaei H, Zhang M, Bayley T, et al. (2025) Empowering or eroding contributions? How generative ai impacts user-generated content across diverse communities. *How Generative AI Impacts User-Generated Content Across Diverse Communities (March 24, 2025)*.

Shriver SK, Nair HS and Hofstetter R (2013) Social ties and user-generated content: Evidence from an online social network. *Management Science* 59(6): 1425–1443.

Singh PV, Sahoo N and Mukhopadhyay T (2014) How to attract and retain readers in enterprise blogging? *Information Systems Research* 25(1): 35–52.

Stephen AT and Toubia O (2009) Explaining the power-law degree distribution in a social commerce network. *Social Networks* 31(4): 262–270.

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

Tang Q and Whinston AB (2020) Do reputational sanctions deter negligence in information security management? A field quasi-experiment. *Production and Operations Management* 29(2): 410–427.

Thompson N and Hanley D (2018) Science is shaped by wikipedia: Evidence from a randomized control trial. Working Paper.

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

Ugander J, Karrer B, Backstrom L, et al. (2013) Graph cluster randomization: Network exposure to multiple universes. In: *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining*, pp.329–337.

Vasilescu B, Capiluppi A and Serebrenik A (2014) Gender, representation and online participation: A quantitative study of stackoverflow. *Interacting with Computers* 26(5): 488–511.

Wang T, Chen Y, Zhang Z, et al. (2010) Unbiased sampling in directed social graph. In: *ACM SIGCOMM computer communication review*, Vol. 40. pp.401–402.

Wasko MM and Faraj S (2005) Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly* 29(1): 35–57.

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

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

Zhu K, Khern-am nuai W and Yu Y (2024) Negative peer feedback and user content generation: Evidence from a restaurant review platform. *Production and Operations Management* 10591478231224941.

Zhu M (2024) What is Zhihu? Our guide to China's Q&A platform. https://www.nativex.com/en/blog/what-is-zhihu-our-guide-to-chinas-qa-platform. Nativex, Blog Post.

How to cite this article

Zhang M and Luo L (2026) Effects of Peer Voting and Followers on User Contribution to Online Knowledge Sharing Platforms: Evidence From a Field Experiment. *Production and Operations Management* XX(XX): 1–20.