

# Estimating the Impact of Replies on First-Time Contributors in Online Communities: A Peer Encouragement-Based Approach

*Keywords: Online community, reddit, newcomer, instrumental variable, peer encouragement*

## Extended Abstract

**Introduction.** Many online communities struggle to retain newcomers; positive feedback on early contributions can play a large role in supporting this objective [4]. Positive feedback can serve to reduce uncertainty about the specific norms or potential benefits associated with sharing in the community. First-time contributors in many online communities, however, often fail to receive any feedback at all on their contributions. In Usenet groups, for instance, newcomers are less likely to receive replies to their posts, but receiving replies to contributions increases the likelihood of submitting subsequent posts [1]. On Facebook, newcomers who receive comments on uploaded posts from friends share more posts over the subsequent 12 weeks [2].

Some first-time posters may enter a community with greater domain understanding, increasing the likelihood of both receiving feedback and participating in the future. This type of confounding relationship could introduce bias into studies which rely on observational data, such as those cited above. Experimental approaches enable unbiased causal estimates, but only when received replies are representative of typical responses from peers. ‘Peer encouragement’ experiments, in which a target user’s *peers* are treated with an encouragement to reply to the target, allow an instrumental variable-based approach to estimating effects on the target user [3]. In this work, we estimate the causal impact of community-driven feedback on subsequent newcomer participation within a community, using data from an experiment with a peer encouragement design, conducted across thousands of online communities within Reddit.

**Experiment and Data.** Mobile notifications were sent to Reddit users on four dates in January 2023. On each date, a cohort of ‘eligible posts’ was identified, which had (1) been submitted within the prior 24-72 hours, (2) not been removed, (3) not received any comments, (4) not been marked by the poster as *NSFW*, and (5) been shared in a subreddit created at least one year prior, that had between 5-50000 subscriber contributions over the preceding week.

On each date, these ‘target posts’ were divided evenly, at random, into treatment and control conditions. For ‘treated’ posts, ‘eligible peers’ included all logged-in subscribers to the same subreddit, who had (1) viewed the subreddit on the day the post was shared, (2) had commented on any post in that subreddit within the prior week, and (3) were opted-in to notifications. Eligible peers were filtered at random to ensure that no individual user received more than one notification per day, and that no more than 30 peers were notified for any post.

A total of 37,209 total posts were included in the experiment, shared in a diverse set of 5,369 subreddits spanning 53 topical categories. 18,625 (50.06%) posts were treated, and notifications about these were delivered to 106,025 total recipients, with 24,474 users (23% click-through rate) clicking the notification and 1,628 commenting (6.7% click-to-comment rate) on the target post. 2,389 (12.8%) of the treated posts received 1 or more comments over the following 28 days, compared with 1,041 (5.6%) in the control condition – a significant difference in the fraction of posts receiving replies ( $\chi^2(df = 1, N = 37,209) = 579.43, p < 0.001$ ).

**Statistical Model.** The *instrument* ( $Z$ ) is experiment condition (treatment vs. control), our *exposure* ( $X$ ) is whether the post received 1+ reply (T/F), and our *outcome* ( $Y$ ) is posts by the newcomer in the same subreddit over the subsequent 28 days. As controls ( $L$ ), we include pre-treatment measures describing the newcomer, as outlined in Table X. These are selected to capture individual-level variation in the likelihood of posting to the subreddit in the future.

We fit two models using two-stage least squares (2SLS). Both use logistic regression in Stage 1 ( $X = Z + L$ ), matching a hypothesis that the primary gap in outcomes will be between zero and any replies. Model A uses Poisson regression in Stage 2 to predict the number of future posts in the subreddit ( $Y_{num-posts} = \hat{X} + L$ ); Model B uses logistic regression in Stage 2 to predict the likelihood of future posts ( $Y_{has-posts} = \hat{X} + L$ ). Models are estimated using the `ivtools` package [5] in R, which supports inclusion of a non-linear first stage with a control function that adds estimated residuals from the first stage as a covariate in the second stage.

**Results.** Table 1 summarizes our key results. Results from Model A show that newcomers who receive 1+ reply to their first post in a community go on to submit 3.99 times as many posts to the same community over the subsequent 28 days, compared with newcomers who receive no replies (95% CI: 1.13x - 14.1x,  $p = 0.032$ ). Results from Model B, however find no evidence for a causal relationship between newcomers receiving replies and likelihood of posting again within that community (95% CI Odds Ratio: 0.54x - 2.48x,  $p > 0.05$ ). Repeating the Poisson model (Model A) with the outcome ( $Y$ ) changed to number of posts submitted *outside* the target subreddit, we find no evidence for a similar relationship (95% CI: 0.18x - 1.68x,  $p > 0.05$ ).

**Discussion.** These results suggest that receiving replies on one’s first post in a community has positive, but heterogeneous, effects on subsequent behavior. For those unlikely to post again, replies from the community may offer little encouragement. For those predisposed to post again, receiving replies appears to substantially increase subsequent contribution rates. Our model includes a pre-treatment covariate capturing ‘new’ accounts; we cannot, however, distinguish between users creating accounts to start participating and ‘throwaway accounts’ – one-time accounts created to overcome social or emotional barriers to posting. Future research to classify potential throwaway accounts would likely improve on the estimates presented here.

We further find that these benefits may be localized to the communities in which replies are received; there is not a clear generalized effect on posting rates across the service. This is consistent with our assumption that community feedback would increase subsequent contribution rates by reducing uncertainty about the norms and benefits of participating in that specific community. We note that due to the structure of this experiment, our findings apply only to communities in which members are influenced by an encouragement to reply.

## References

- [1] ARGUELLO, J., BUTLER, B. S., JOYCE, E., KRAUT, R., LING, K. S., ROSÉ, C., AND WANG, X. Talk to me: Foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2006), pp. 959–968.
- [2] BURKE, M., MARLOW, C., AND LENTO, T. Feed me: Motivating newcomer contribution in social network sites. *CHI 2009*.
- [3] ECKLES, D., KIZILCEC, R. F., AND BAKSHY, E. Estimating peer effects in networks with peer encouragement designs. *Proceedings of the National Academy of Sciences* 113, 27 (2016), 7316–7322.
- [4] KRAUT, R., BURKE, M., RIEDL, J., AND RESNICK, P. Dealing with newcomers. *Evidencebased Social Design Mining the Social Sciences to Build Online Communities 1* (2010), 42.
- [5] SJOLANDER, A., AND MARTINUSSEN, T. Instrumental variable estimation with the r package `ivtools`. *Epidemiologic Methods* 8, 1 (2019).

Stage 1 (Exposure) Model [Logistic]				
	$\beta$	SE	Odds Ratio [95% CI]	$p$
<i>Intercept</i>	-1.385	0.076	0.25 [0.22:0.29]	<0.001
<i>(Z) Assignment Treatment</i>	0.921	0.039	2.51 [2.33:2.71]	<0.001
<i>(L) Account &lt;= 1 Day Old</i>	-0.036	0.096	0.96 [0.80:1.16]	0.710
<i>(L) Prior 28 Days Visited Anywhere</i>	0.012	0.002	1.01 [1.01:1.02]	<0.001
<i>(L) Prior 28 Days Has Posts Anywhere</i>	-0.129	0.039	0.88 [0.81:0.95]	<0.005
<i>(L) Prior 28 Days Has Actions Target Sub</i>	0.139	0.049	1.15 [1.04:1.26]	<0.005
<i>(L) Target Sub Topic "Mature"</i>	0.099	0.081	1.11 [0.94:1.29]	0.222
<i>(L) Target Sub Weekly Contributors (Log)</i>	-0.564	0.023	0.57 [0.54:0.60]	<0.001
Stage 2 (Outcome) Model A [Poisson]				
	$\beta$	SE	Rate Ratio [95% CI]	$p$
<i>Intercept</i>	-1.125	0.179	0.32 [0.20:0.41]	<0.001
<i>(X) Received Comments on Post</i>	1.383	0.644	3.99 [1.13:14.10]	0.032
<i>(L) Account &lt;= 1 Day Old</i>	0.621	0.117	1.86 [1.48:2.34]	<0.001
<i>(L) Prior 28 Days Visited Anywhere</i>	-0.030	0.003	0.97 [0.96:0.98]	<0.001
<i>(L) Prior 28 Days Has Posts Anywhere</i>	0.597	0.057	1.82 [1.63:2.03]	<0.001
<i>(L) Prior 28 Days Has Actions Target Sub</i>	0.528	0.060	1.67 [1.51:1.91]	<0.001
<i>(L) Target Sub Topic "Mature"</i>	0.892	0.102	2.44 [2.00:2.98]	<0.001
<i>(L) Target Sub Weekly Contributors (Log)</i>	-0.003	0.047	1.00 [0.91:1.09]	0.956
<i>R [Control Function]</i>	0.020	0.097	1.02 [0.84:1.23]	0.843
Stage 2 (Outcome) Model B [Logistic]				
	$\beta$	SE	Odds Ratio [95% CI]	$p$
<i>Intercept</i>	-1.875	0.107	0.16 [0.12:0.19]	<0.001
<i>(X) Received Comments on Post</i>	0.148	0.388	1.16 [0.54:2.48]	0.704
<i>(L) Account &lt;= 1 Day Old</i>	0.312	0.071	1.37 [1.19:1.57]	<0.001
<i>(L) Prior 28 Days Visited Anywhere</i>	-0.023	0.002	0.98 [0.97:0.98]	<0.001
<i>(L) Prior 28 Days Has Posts Anywhere</i>	0.509	0.034	1.66 [1.56:1.78]	<0.001
<i>(L) Prior 28 Days Has Actions Target Sub</i>	0.751	0.038	2.12 [1.97:2.28]	<0.001
<i>(L) Target Sub Topic "Mature"</i>	0.456	0.062	1.58 [1.40:1.78]	<0.001
<i>(L) Target Sub Weekly Contributors (Log)</i>	-0.031	0.027	0.97 [0.92:1.02]	0.250
<i>R [Control Function]</i>	0.098	0.052	1.10 [1.00:1.22]	0.060

Table 1: Summarized results from two 2SLS models estimating the causal relationship between receiving replies on a first post and subsequent posting behavior, for newcomers to a community. Both models share a first stage, which estimates the effect of the experiment condition (Z) on whether newcomers received one or more replies (X). Stage 2A estimates the effect of receiving 1+ reply on the number of future posts; Stage 2B estimates the effect on the likelihood of future posts. Both Stage 2 models include a control function (estimated residuals from the first stage). Several variations on these models produced similar results.