

ESSAYS ON SOCIAL INCENTIVES IN ECONOMIC DECISION MAKING

By

Yiqian Wang

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Economics—Doctor of Philosophy

2024

## ABSTRACT

This thesis examines the roles of social incentives in economic decision-making. The first chapter analyzes the impact of (partially) revealing social identity on the behavior of individuals online. The second chapter studies an experiment that targeted leveraging social incentives and social relations to facilitate informal risk-sharing. The last chapter studies the role of social networks in propagating preventive health behaviors, beliefs, and knowledge.

In the first chapter, I study if reducing user anonymity and partially revealing a user's social identity could affect communication on social media. In a motivating signaling framework, I find revealing group-level identity discourages marginalized groups' participation and exacerbates ideological segregation. I test the predictions by leveraging a recent IP location disclosure policy imposed by the Chinese government on mainstream Chinese social media platforms. Empirical findings are consistent with the model predictions using recent text analysis technology advancements. Benchmark analysis through a regression discontinuity in time framework suggests that revealing the user's IP-based geographic location reduced the participation of non-mainland users substantially by nearly 30% one month after the policy implementation. Delving into the mechanism, the policy amplifies geographic-based sorting, possibly leading to ideological segregation. These findings highlight the negative impact of revealing social identity by hurdling intergroup information transmission and intensifying the echo chamber.

The second chapter (co-authored with Prabhat Barnwal and Lex van Geen) studies whether policymakers could engineer the informal risk-sharing between households to mitigate the idiosyncratic risks during shocks. We design a novel experiment to study how *ex-ante commitments* improve risk-sharing in the context of groundwater arsenic poisoning in Bangladesh. In a field experiment conducted across 135 village communities in rural Bangladesh, we evaluate the impact of facilitating *ex-ante commitments* between household pairs to share safe water. On top of the commitment facilitation, we design a two-part randomized trial that first notifies and then implements *peer monitoring* of these commitments before and after testing the arsenic. Well-owners from communities with facilitated commitments reduce arsenic intake by 8.7%, while the effect

spillover to non-well-owners by 16.7%. On the other hand, notifying peer monitoring increases the level of sorting in risk-sharing formation, and implementing peer monitoring eventually hinders risk-sharing facilitated by ex-ante commitment. Our mixed findings imply the complexity of leveraging social incentives and relations to design community-driven programs.

In the last chapter (also co-authored with Prabhat Barnwal and Lex van Geen), we study how societal interactions influence the adoption of preventive health measures. We document three main findings using social networks and geo-location data collected in in-person and phone surveys from 135 villages in rural Bangladesh during the COVID-19 pandemic. First, our results suggest that socially connected and geographically nearby households induce households' adoption of preventive health measures. Second, such peer effects only exist for preventive measures that can be publicly observed. Third, these peer effects tend to disappear when social and geographic distance between households increases. Our findings suggest that the social incentives underlying decisions to adopt preventive health measures are important.

These three chapters highlight my main goal to bridge the recent theoretical studies in the economics of social incentives to novel solutions to real-world problems. I show that social incentives exist, substantially matter for people's decision-making, and, most importantly, could be cleverly engineered to improve the welfare of people in underdeveloped areas. The importance of pursuing further study in this area is threefold. First, social incentives can be correctly engineered to become powerful economic policy tools. Second, while social incentives may matter for all the perceivable economic decisions, our knowledge of predicting to individual and group behavior is still quite limited. Last of all, a deeper study in this area could possibly reshape our understanding of how efficient and economical behavior can be encouraged/distorted by social concerns.

Copyright by  
YIQIAN WANG  
2024

## ACKNOWLEDGEMENTS

I am deeply indebted to my advisors, Chris Ahlin and Prabhat Barnwal, who guided me in understanding economics and encouraged me to pursue my career as a development economist. Chris is an incredibly supportive advisor. His patience with my bad math and often immature theoretical ideas was immensely motivating. He helped me see the bigger picture beyond numbers and equations. Beyond substantial theoretical guidance, Chris dedicated significant time to coaching me on the rigor required to be an applied economist. His mentorship will continue to guide me in the years to come.

Prabhat introduced me to the fantastic world of empirical studies, especially by coaching me to work on RCTs that aim to improve the welfare of poor people. Prabhat taught me the essentials of becoming an economist, including how to think and write like one. He encouraged me to overcome self-doubt and step out of my comfort zone. Most importantly, he recognized and strengthened my ideas, helping me nurture them into several academic outputs. His guidance on proposal writing enabled me to secure the Dissertation Grant from the National Science Foundation, which has been crucial in pursuing my academic career.

I extend my gratitude to Kyoo il Kim and Ken Frank for their thoughtful discussions and guidance, which significantly enriched my understanding of methodologies in social science research. I also wish to express special thanks to Lex van Geen, whose insightful perspectives and occasional critiques of economics frequently remind me to reconsider the role of an economist in a world with a rapidly growing demand for interdisciplinary studies.

I am also thankful to numerous faculty members in the department, including Ben Bushong, Jay Pil Choi, Jon Eguia, Enrique Seira, Hanzhe Zhang, and Oren Ziv. It was a pleasure to engage in thoughtful discussions with them and receive their inspiring suggestions to improve my work.

I am grateful for the exceptional and tireless fieldwork conducted by my local Bangladeshi collaborators from the NGO Forum for Public Health, including Ahasan Habib, Ishrat Jerin, Nishat Juy, and Mir Raihan. Their exemplary work ethic and insightful guidance allowed me to engage in these extraordinarily fascinating and influential projects. I look forward to continuing our

collaboration on future projects, including the one funded by the National Science Foundation.

In addition, I owe an immense debt of gratitude to my friends and family. I thank my dearest friend, Ziang Xiao, for the decade of friendship. Beyond sharing overlapping passions, Ziang, as an exceptional computer scientist and social scientist himself, introduced me to a world of knowledge I could barely encounter by myself. He also provided significant help to my current projects, and I look forward to a long collaboration with him on my growing investigations in social media and digital economics.

Finally, I sincerely thank my parents for their unwavering support in pursuing this somewhat risky career path. While countless times they have charmed me to pick alternative routes, I deeply understand these are their reserved ways of expressing their enduring love and protection. I am glad that I have made them proud. Above all, none of this would be possible without my future wife, Xiangying, who has been my grandest support for almost a decade. Her unwavering faith in me has cured and sustained me through the years of exhausting PhD study and separating from my family and her due to COVID-19. It is my greatest fortune to embark on this new journey with her.

## TABLE OF CONTENTS

CHAPTER 1	TRUST OR STIGMATIZATION? EVIDENCE FROM DE-ANONYMIZING SOCIAL MEDIA IN CHINA . . . . .	1
1.1	Introduction . . . . .	1
1.2	Institutional Background and Context . . . . .	6
1.3	Data . . . . .	9
1.4	Empirical Strategy . . . . .	17
1.5	Results . . . . .	20
1.6	Framework . . . . .	26
1.7	Conclusion . . . . .	31
1.8	Figures . . . . .	33
1.9	Tables . . . . .	44
CHAPTER 2	INFORMAL RISK SHARING TO MITIGATE LOCAL ENVIRONMENTAL RISKS . . . . .	54
2.1	Introduction . . . . .	54
2.2	Research Design . . . . .	59
2.3	Specifications . . . . .	69
2.4	Experiment Findings . . . . .	75
2.5	Conclusion . . . . .	85
2.6	Figures . . . . .	87
2.7	Tables . . . . .	98
CHAPTER 3	PEER EFFECTS IN ADOPTION OF PREVENTIVE MEASURES: EVIDENCE FROM RURAL BANGLADESH . . . . .	112
3.1	Introduction . . . . .	112
3.2	Data . . . . .	117
3.3	Estimation . . . . .	123
3.4	Results . . . . .	128
3.5	Framework . . . . .	132
3.6	Conclusion . . . . .	136
3.7	Figures . . . . .	138
3.8	Tables . . . . .	141
BIBLIOGRAPHY	. . . . .	146
APPENDIX A	CHAPTER 1 . . . . .	154
APPENDIX B	CHAPTER 2 . . . . .	167
APPENDIX C	CHAPTER 3 . . . . .	176

## CHAPTER 1

### TRUST OR STIGMATIZATION? EVIDENCE FROM DE-ANONYMIZING SOCIAL MEDIA IN CHINA

#### 1.1 Introduction

*“The new freedom of expression brought by the Internet goes far beyond politics. People relate to each other in new ways, posing questions about how we should respond to people when all that we know about them is what we have learned through a medium that permits all kinds of anonymity and deception.”*

— Peter Singer, 2005, *Visible Man: Ethics in a World Without Secrets*

Social media platforms provide users with unprecedented levels of anonymity in human interactions. While many public disclosures focus on anonymity-fueled misinformation and toxicity and advocate curtailing, it is also crucial to consider its positive aspects. Anonymity fosters freedom of expression, promotes diversity of opinions, and facilitates intergroup communication. As debates on limiting anonymity intensify, it becomes increasingly critical to understand how anonymity shapes intergroup communication and fosters diversity of opinions on social media.

Our study sheds light on this policy debate by studying the role of identity in the engagement of marginalized and non-marginalized groups on large social media platforms. Specifically, we address two questions. First, does anonymity affect the participation of marginalized groups? Second, does anonymity shape opinion diversity and polarization of viewpoints? We further investigate whether the mechanisms that lead to the outcomes are linked to identity-based discrimination, stigmatization, and segregation that permeate the real world.

We leverage a unique policy shift in China, where the government mandated that all China-based social media platforms publicly disclose the user’s IP-based geographic location. To understand the policy impact, we put together an exclusive dataset of 2.8 million comments posted on 18 thousand popular social media posts from one of the largest Chinese social media platforms. These posts and comments were created a few months before and after the implementation of the policy change. We use the latest text analysis techniques to extract crucial features of these comments to answer



our research questions and explore the underlying mechanisms. These features include the location from where the comments were posted, the level of endorsement received by the comments, the sentiment of the comments, and the political attitude conveyed by the comments.

The primary subject in this study is the comments on the posts. These comments usually reflect individual opinions toward the news or argument contained in the post. Our econometric strategies mainly compare comments made by non-mainland Chinese speakers with those made by mainland Chinese speakers across various posts. In our sample, non-mainland users are considered a cyber minority as they make up only 5% of all comments. Anecdotal evidence further suggests that non-mainland users are marginalized due to their distinct political views, such as pro-Western attitudes, which mainland users may stigmatize. Our text analysis shows a substantial gap of 20% in favor of Western ideology between non-mainland and mainland comments to political posts. This divergence could be due to two factors. First, these two groups live in distinct environments with increasingly polarized political views. Second, these two groups learn news from different media sources, which may slant the narratives differently (Gabore, 2020). In addition, the Chinese government has restricted mainland users' access to global platforms such as Facebook and Twitter for over a decade.

The first strategy exploits the sharp implementation of two phases of policy on the platform. In the first profile de-anonymization phase, the platform reveals users' location publicly on the user's profile page, an information page that is publicly accessible. In the second comment de-anonymization phase, the platform reveals the location in all comments to the public. We identify the policy effects through a regression discontinuity in time design. The policy's local average treatment effect is identified if the content on the social media platform immediately before the time cutoff resembles the content immediately after the cutoff. In the preferred specification, we set the bandwidth to be 30 days. We use a linear function to control for the time trend before and after the cutoff, which is less sensitive to the potential initial impulse response to the policy. We test the identifying assumption using various post topics and poster characteristics. By comparing the features of comments made before and after the cutoff, we identify the policy effects on the

relative presence of non-mainland users and on political attitudes.

The second strategy utilizes the experimental phase of the policy, when the platform made the IP location of comments public only when the posts were related to the Russia-Ukraine conflict. We identify the policy effects through a difference-in-differences design. This method compares the comments made to the affected posts with those made to international political posts unrelated to the conflict. Even though these two types of posts focused on different topics, we can still identify the average effects of the policy on our interested outcomes if these outcomes co-moved in parallel fashion in the absence of the treatment. Further, we investigate the potential selection issue that emerged when posters manipulated the platform's detection algorithm to enroll or evade the revelation.

Our first finding is revelation of location in the user profile substantially discourages the engagement of marginalized groups. The local estimate derived from the RDiT specification suggests the policy reduces the proportion of non-mainland comments by 28.9% with statistical significance at 0.01 level. The reduction is consistently discovered by varying the bandwidth from 20 to 40 days, with the estimated effect ranging from 17.5% to 26.2%.

Surprisingly, we find, on the contrary, the revelation of location in the comment increases the proportion of non-mainlanders' comments by 20.8%. The increment is robust to the choice of bandwidth, and the estimate peaks at 20-day bandwidth with an increment of 30.7%. Meanwhile, we also find that the estimate diminishes with the bandwidth: the estimate falls to 12.2% at 40-day bandwidth. With this observation, together with the large contrast with the estimates using the first policy phase, we hypothesize that the policy effect of comment de-anonymization may have a short-run effect that substantially differed from its long-run effect. Initially, comment de-anonymization might have disproportionately attracted non-mainland comments. Indeed, we find the policy effect turns negative after increasing the bandwidth.

The policy's impact on opinion evaluation may explain the reduction in the first phase. We find the policy differentially affects the endorsements received by comments made by non-mainlanders and mainlanders. Profile de-anonymization increases the standardized number of likes received by

mainland comments by 0.013 standard deviations when commenting on a mainland post<sup>1</sup>. On the other hand, the endorsements received by non-mainland comments was reduced by 0.058 standard deviations when commenting on a mainland post. These effects are economically large, given that the one standard deviation of endorsements equals around 400 likes. These estimates are significant at 0.1 level. At the same time, no statistically significant change happened to the endorsements received by comments on non-mainland posts.

Meanwhile, comment de-anonymization significantly increased the endorsement received by non-mainland comments by 0.055 standard deviations when these comments were made on non-mainland posts. The policy generally makes the non-mainlanders' opinions less favorable. This finding can be explained by identity-based discrimination. The opinion of the marginalized group becomes less favored when the policy exposes their marginalized identity to the dominant group.

Our second finding is both revelation schemes substantially increase anti-Western sentiment detected in the comments of political posts. Respectively, profile de-anonymization increases the sentiment by 12.8%, and comment de-anonymization increases the sentiment by 18.0%. These increments are mostly attributed to the substantial rise in mainland comments made to mainland posts. Anti-Western sentiment increased by 14.2% and 25.5% in these cells for each policy phase. On the other hand, no systematic change was detected in non-mainland comments made to non-mainland posts.

The rise of polarization could be partially explained by the identity-based segregation that the policy exacerbated. We find significant heterogeneity by poster location in the policy's impact on the engagement of non-mainland users. While the policy largely reduced the proportion of non-mainland comments in mainland posts, such a reduction is significantly attenuated in the non-mainland posts. profile de-anonymization reduces the non-mainland comments in mainland posts by 31.1% but has no effect in non-mainland posts. Comment de-anonymization increases the non-mainland comments in mainland posts by 23.2%, and the increment doubles in non-mainland posts. These findings are robust and mostly statistically significant at 0.1 level across different

---

<sup>1</sup>Mainland posts are posts written by a mainlander. Similarly, posts written by a non-mainlander will be shorted as non-mainland posts

bandwidths. Polarization is exacerbated as the users sorted to posts penned by the posters from the same group.

Last of all, we find comment de-anonymization systematically reduces general negativity such as anger, dissatisfaction, and unsupportiveness. In contrast, no similar effect is seen for profile de-anonymization. This suggests that the explicit display of a commenter's geographic identity through their comments may be more effective in regulating emotion on social media. When individuals recognize that their group identity is overtly exposed, they might adjust their behavior due to concerns such as stereotype threat (Steele and Aronson, 1995).

In order to better understand how de-anonymization might change equilibrium behavior, we develop a signaling model based on Bursztyn et al. (2023). This model helps to understand how identity exposure affects marginalized group participation and polarization. We consider an online platform comprised of two groups, e.g., mainland and non-mainland Chinese. The distribution of the individual's benefit from publicly supporting a policy change varies with the group identity. The majority group more prefers the status quo than the minority group. The cost of publicly supporting the policy change comes in the form of social sanctions from people who prefer the status quo. We use this model to examine how the observability of the individual group membership affects equilibrium support.

Through simulation, we derive two important predictions from the model. First, the minority group's willingness to publicly support is reduced by revealing their identity. This theoretical prediction resonates with the reduction of non-mainlanders observed in our data post-policy. Second, the larger the benefit received by the minority group from the policy change relative to its majority counterpart, the larger the social sanction imposed on the group when the identity becomes observable. This finding can explain why de-anonymization increases segregation if posters of different groups intentionally slant the narrative to widen or close the benefit gap.

In general, our study contributes to three bodies of literature. First, our study contributes to the literature on the economics of intergroup interaction. Many influential studies explored the economic return and behavioral change of integrating groups previously segregated by identities

such as wealth level (Rao, 2019) and gender (Dahl, Kotsadam, and Rooth 2021)<sup>2</sup>. We contribute to this literature by looking at the effect of segregating two groups distinct by their location, political attitude, and methods of acquiring information.

Second, our study contributes to the economics of identity and social image. In a related study, Braghieri (2022) experimentally showed that the concern of social image causes the publicly expressed view to deviate from the private view. In another experimental paper, Bursztyn et al. (2020) showed that the payment required for forgoing a socially recognized identity varies with whether the action can be observed. We add to the literature by showing the importance of covering identity to sustain intergroup communication.

Third, our study contributes to the growing literature on the economics of social media. Many recent important studies find social media has real-life effects, such as on mental health and political attitudes or outcomes<sup>3</sup>. Close to our context, Qin, Stromberg, and Wu (2017) and King, Pan, and Roberts (2013) study the role of Chinese social media in facilitating government surveillance. In terms of anonymity, Ederer, Goldsmith-Pinkham, and Jensen (2023) and Wu (2020) study the downside of anonymity on breeding harmful information even from highly-educated groups.

The rest of the paper proceeds as follows. In Section 2, we introduce the institutional background and context in greater detail. In Section 3, we explain our data collection methods and text-analysis procedure. Then, we summarize the data. In Section 4, we describe the empirical strategies. In Section 5, we discuss the empirical findings, and in Section 6, we develop a model to explain the theoretical mechanisms and intuition.

## **1.2 Institutional Background and Context**

### **IP-based Geographic Location Disclosure**

On March 1, 2015, the Cyberspace Administration of China (CAC) issued guidelines to all Chinese social media platforms, known as the Administration of Internet User ID. These guidelines

---

<sup>2</sup>Also other important identities such as ethnicities: Bazzi et al., 2019; race: Schindler and Westcott, 2021; Corno, La Ferrara, Burns, 2022, and social hierarchy: Lowe 2021

<sup>3</sup>For mental health, see: Allcott et al. (2020); Mosquera et al. (2020); Braghieri, Levy, and Makarin, (2022)). For political attitudes or outcomes, see Allcott and Gentzkow (2017); Levy (2021); Bursztyn et al. (2019); Enikolopov, Makarin, and Petrova (2020); Gorodnichenko, Pham, and Talavera (2021)

prohibited user IDs containing keywords linked to sensitive topics, including national security, hate speech against minority groups, pornography, violence, and harassment. Despite granting users considerable freedom regarding profile naming, the new Administration mandated users to verify their real ID before continuing to use the platform.

Over time, the Administration underwent several amendments. Related to this study, in October 2021, the CAC proposed an amendment demanding social media platforms disclose users' IP-based geographic locations. While the primary motive for this change wasn't specified, some alleged that the government believed disclosing IP locations could help curb misinformation fabricated by foreign bots. For instance, credibility would be challenged for someone posting their COVID-19 lockdown experiences in Shanghai when their IP indicated they were based elsewhere.

Sina Weibo, one of China's leading social media platforms, was the first to test the new guideline. From March 4, 2022, the platform started to experiment with the IP revelation function in posts related to Russia-Ukraine. Any post containing the keywords "Russia(俄罗斯)" or "Ukraine(乌克兰)" would have its comments tagged with the commenter's IP-based geographic location. This location is accurate up to the province within China or the country for users outside China.

The primary goal of this action was to reduce misinformation about the Russia-Ukraine War. IP location served as an automatic credibility checker for the unidentified source. However, this strategy had limitations. For instance, a post can evade the keyword filter if it does not include the words "Russia" or "Ukraine." Conversely, users could manipulate the system by adding keywords to unrelated posts and triggering the IP revelation.

*Profile de-anonymization* Despite the controversy raised due to its implementation, Weibo expanded its de-anonymization efforts. Starting March 18, 2022, the platform attaches the user's IP-based geographic location to their profile. Figure A2a shows a standard Weibo profile showcasing self-reported gender, birthday, and location, followed by the system-assigned IP location. This IP location is determined by the latest IP address the user uses. Being inactive for over a month erases the IP location.

*Comment de-anonymization* From midnight on April 28, 2022, the platform started to attach

the IP location underneath all new comments, regardless of whether the post was written before or after April 28. Figure A2b illustrates this policy's sharp implementation. Two comments replied consecutively to a post, with the latter being tagged for replying just half an hour past midnight.

In June 2022, the CAC's amendment was approved, ordering all China-based platforms to show user IP locations from August 1, 2022. While the display method wasn't exactly instructed, most platforms adopted Weibo's approach, tagging comments directly with the user's IP location.

### **Sina Weibo**

Sina Weibo has dominated China's non-communication-focused social media for decades, with a monthly active user count exceeding 300 million. In 2014, Sina Weibo was listed on NASDAQ, the first China-based social media platform to be listed outside China. While international registration is accessible with a verifiable phone number, the platform chiefly serves Chinese speakers residing in China.

Named after the Chinese homonym for "Microblogging," Sina Weibo has an interface and functionality similar to Twitter. Users can craft posts incorporating text, images, and videos and are free to follow, read, and comment on content. An algorithmically-driven timeline showcases posts from followed users, hybridizing time with recommendations. Engaging with content is intuitive, as users can like, share, comment on posts, and interact with other users' comments.

Sina Weibo allows large flexibility in customizing profile pages. Users can share details ranging from their occupations and educational backgrounds to birthdays. A unique feature facilitates experts to authenticate their real-world expertise, granting them an official verification tag atop their profile. Additionally, an IP-based location, presumably derived from the most frequently used IP in the preceding days, is forcibly attached to the user's profile unless the user has been inactive for months.

For content discovery, Weibo offers both keyword and hashtag searches. An advanced search function provides users rich filters, from post types (text, images, videos) to the timeframe within which content was published.

Given its vast user base and rich data, academics, especially social scientists, frequently turn

to Weibo as a prime research subject. Noteworthy studies published in prominent journals<sup>4</sup> have delved into the profound dynamics of media capture, censorship, and the Chinese government's influence over the platform. With the rapid advancements in text analysis technologies, further groundbreaking research is anticipated to spring from this platform.

### **Mainland User vs. Non-Mainland User**

China-based social media platforms are predominantly used by mainland Chinese users while serving a smaller portion of non-mainland Chinese users. Despite the shared language, anecdotal observations reveal a lingering mistrust between these two groups. This division potentially arises from two factors. First, the two groups live in distinct environments with increasingly polarized political views. Most of the non-mainland comments in the data originate from North America and European countries, where the political systems differ largely from mainland China. Second, the groups receive news and information from different media sources. The government has significantly censored the media platforms used by mainland Chinese. The same news could be slanted in different ways. Mainland Chinese users' access to global platforms such as Facebook and Twitter has been restricted for over a decade.

This mistrust manifests most explicitly in discussions of controversial topics. Before revealing the location identity, dissenters were frequently stigmatized as "foreign agents." For instance, frequently observed in COVID-19 posts made in early 2022, commenters who advocated for the termination of the zero-COVID policy were often stigmatized as non-mainlanders. These criticisms usually talk about foreign governments intentionally endangering Chinese public health.

### **1.3 Data**

In this study, we focus on the comments responding to posts, viewing them as representations of individual opinions about the news or daily shared content. Our econometric approach primarily contrasts comments from non-mainland Chinese speakers with those from mainland Chinese speakers across different posts. We developed Python scripts to collect high-frequency text data from social media. Our dataset comprises 17,889 sample posts authored by 509 influential accounts,

---

<sup>4</sup>Such as King, Pan, and Roberts (2017) and Qin, Stromberg, and Wu (2017)



accumulating 2.8 million comments on Sina Weibo. These posts and their respective comments were published between January 1, 2022, and September 1, 2022. This timeframe was chosen to ensure a sufficient number of observations both before and after the policy interventions on March 4th, March 18th, and April 28th, 2022.

### **Topic-oriented process**

Our data collection process highlights a poster-focused approach. This means we initially identify a sample of posters and then gather the posts created by these posters. The more influential the posters, the stronger the policy impact may be detected. This is because the content curated by these posters is usually widely spread and well-commented by a diverse set of commenters.

We face two difficulties in finding influential posters. First, there does not exist a roster or representative survey of influential posters. Second, the impact of posters may change over time. To tackle these challenges, we devise an innovative keyword-oriented procedure to pinpoint a set of influential posters around the policy implementation windows. We start by searching for the most viral posts about an important and well-discussed topic around the policy windows. Then, we identify the posters of these viral posts. Last of all, we collect all viral posts penned by these posters and collect the comments replied to these posts.

One potential challenge is that the type of posters identified through this method may show a strong bias toward our chosen topic. For example, posters identified from viral posts about the Russia-Ukraine conflict might largely specialize in international politics, while those identified from viral COVID-19 posts might predominantly be public health experts. To mitigate this concern, we compile different lists of posters using different topics. We then assess the overlaps between the posters linked to our chosen topic and those associated with other topics. Higher overlap indicates that the listed posters are also likely to be impactful on other topics.

Following a preliminary collection of breaking news during our sample period, it is no surprise that “COVID-19(新冠)” stands out as the preliminary identifier for influential posters. We use Weibo’s search function to look up “COVID-19” and gathered all posts receiving over a hundred likes. From the posters who authored these posts, we exclude accounts that were (1) directly

associated with government entities or news agencies, (2) dedicated celebrity fan pages, or (3) banned. Ultimately, 509 accounts met our criteria. We then collect all posts of these posters curated from January 1 to August 31, 2022<sup>5</sup>.

To assess the potential bias stemming from our keyword-driven approach, we compile another six lists of posters through 31 keywords closely related to major news events in China throughout our sample timeframe (see Table A2). From categorizing all the breaking news listed on Wikipedia of 2022 China News during the sample period, we find all news settled to at least one of the following categories: (1) Technology & Science, (2) Economy, (3) Entertainment & Sports, (4) Social Affairs, (5) Politics, and (6) Education.

We then compute the overlapping coefficient, measured by the proportion of COVID-19 posters appearing in these six categories. Summarized in Table 1.1, the coefficients ranged from 0.225 for Technology & Science to 0.343 for Social Affairs. This suggests roughly one in every five COVID-19 posters also penned at least one viral post about Technology & Science, while one in three did for Social Affairs. These overlap coefficients confirm the comprehensive influence of the posters sourced through searching COVID-19.

For the monthly Top 5 most-liked posts of each poster, we scrape the last 50 pages of comments due to the limitation of the data repository we access. In addition, we scrape the number of likes and the IP location if available.

For the comments that appear after April 28, 2022, the IP location of the commenter is directly scrapable from the location tag attached beneath the comment. However, for comments before April 28, 2022, we instructed the scraper to click into the commenter's user profile and extract the IP location. The IP location data post-treatment was almost perfectly retrieved, while IP location prior-treatment was missed by 36%. This indicates that many users who commented on Weibo before April 28 became inactive for over a month by the time of data scraping.

---

<sup>5</sup>We conduct the scraping from January to June 2023

## **Text Analysis**

### **Text Analysis with Large Language Models**

The recent development of large language models (LLMs) offers a new means for textual analysis. Compared to the traditional machine learning-based approach in text analysis, the LLMs are pre-trained and can often be accessed publicly through API. It does not require a pre-existing annotated dataset for model training while achieving good performance in deductive coding tasks (Ziems et al., 2023; Xiao et al., 2023). In this paper, we adapted a codebook-based approach illustrated in Xiao et al., (2023) for textual analysis, where we combined GPT-4<sup>6</sup> with pre-developed codebooks to assign each post a pre-defined code. Each codebook contains three main elements for each code: name, description, and examples. To leverage LLMs in context learning capability, we included five examples for each code as in context learning examples. The coding process is as follows. First, we randomly sampled 1.3-20% of the data for each coding task and instructed expert coders to assign labels given an existing codebook. Then, we iterate the prompt with the GPT model until it reaches adequate inter-rater agreement with expert coders. We code the rest of the data with the same prompt. To ensure reproducibility, we run the coding process three times, and the final code for each data point is decided based on a majority vote. The model temperature is set to 0<sup>7</sup>.

### **Sentiment**

To investigate the effects of de-anonymization on sentiment, we sampled from the 1.6 million first-stage comments, splitting them into four categories based on whether the users were from the mainland or non-mainland, and whether the comments were made before or after the introduction of comment tagging (April 28, 2022). From each category, we randomly selected 5,000 comments. We then established a codebook differentiating negative, positive, and neutral sentiments. With the assistance of GPT, we identified the overt emotional tone of each comment. We observed that many comments used sarcasm without explicitly displaying strong sentiments. To pinpoint these comments, we expanded our codebook with examples to aid GPT in recognizing sarcasm. For our

---

<sup>6</sup><https://openai.com/research/gpt-4>

<sup>7</sup>The model temperature controls the randomness of the output. For GPT-4, the temperature could be set between 0 and 2. Higher values make the output more random, while lower values make it more deterministic.

data analysis, sarcasm was classified as negative sentiment.

### **Political attitude**

Beyond general sentiment analysis, we further explore a prevalent political attitude on Chinese social media: anti-Western sentiment. This sentiment encapsulates a wide-ranging negative perception towards Western ideologies, governments, individuals, and technological advancements.

To detect them, we first locate around 200 thousand comments on posts related to politics and having at least one foreign country involved. Among these comments, we randomly draw 20,000 mainland comments each before and after the comment revelation (April 28, 2022). We then collect almost all of the non-mainland comments, summing to 4,300. We customize a detailed codebook that encompasses the general aversion towards Western ideology, government, individuals, society, and technology.

### **Data Summary**

In this section, we describe our data in detail. We separately provide details on posts, posters, and comments. Comprehensive summary statistics for each category can be found in Table A1.

#### **Posts**

In total, we amassed 17,889 Weibo posts posted between January 1, 2022, and August 31, 2022. The most popular post received nearly a million likes, over 300 thousand reposts, and over 80 thousand comments. Figure 1.2 illustrates the key attributes of these posts, including where these posts are related, what are the topics of these posts, and the viewing statistics.

To identify the geographic unit involved in the posts, we employed country-specific keywords. For example, a US-related post could have keywords such as "USA," "Biden," "Trump," or "White House." Some posts were linked to multiple countries. The ratios shown in Figure 1.2a are based on considering each country mentioned in a post as a distinct data point. A US-China trade war post would be accounted as one China-specific and one US-specific post. Overall, Mainland China-related topic dominates all other countries and regions, with 65% of the posts being related to issues that happened in China or having China involved. Posts involving countries in North America follow with around 16%. The majority of these posts are US-related. Around 8% of

all posts involve European countries and Russia-Ukraine. Both Hong Kong/Taiwan-related posts and East Asia-related posts occupy around 5%. All the other continents occupy a relatively small portion of posts, and many of these posts also have major countries involved.

To categorize the post topics, we first collect all the breaking news listed on Wiki China News. We categorize all posts in the sample period into seven categories: (1) Technology & Science, (2) Economy, (3) Entertainment & Sports, (4) Social Affairs, (5) Politics, (6) COVID-19, and (7) Other. Unlike how we measured the bias of keyword of choice, we integrate Education-related posts to Social Affairs as they only occupy a small proportion of all posts even though Education-related news is frequently shown. Figure 1.2b illustrates the distribution of post types we categorized by combining keyword matching and GPT coding. Social Affairs-related posts dominate 25% of all posts. Immediately followed the Entertainment-related posts that cover 19% of all posts. COVID-related and Politics-related posts share about 31% of all posts. 11% of posts are categorized as Other, with most of these posts about daily life or emotional expression but not necessarily related to any news topics. In the end, 8% of posts are related to Economy and only 6% are related to Tech&Science.

As depicted in Figure 1.2c, there appears to be a stochastic dominance relationship among the numbers of likes, reposts, and comments. This hierarchy can be attributed to the varying degrees of effort and visibility associated with each action. Pressing Like should be the simplest and least visible form of engagement. Typically, this action remains private and is unobserved by others. Repost should also be effortless, but the action is exposed to at least followers. Comments should be the most costly, and the action is exposed to the wide public when the comment receives a lot of replies or likes.

## **Posters**

Leveraging our keyword-searching approach, we identified 509 posters who are predominantly influential on the platform. On average, these posters own 1.8 million followers, with the most followed poster reaching a staggering 24.8 million followers. Notably, approximately 18% of these posters are linked to non-mainland IP locations—a proportion that's considerably greater than the

ratio of non-mainland comments in our dataset.

As shown in Figure 1.4a, there is no surprise that the average numbers of likes, reposts, and comments of these posters displayed the dominance relationship observed in the post. Figure 1.4b suggests that around 35% of posters only write about China. In addition, the proportion of posters decreases with the proportion of China-related posts, with only one poster's popular posts not at all China-related. On the other hand, as shown in Figure 1.4c, most posters, to some extent, write Entertainment&Sports-related posts and at some point receive a lot of attention. Remarkably, only 10% of the posters in the sample never wrote a popular Entertain-related post. Meanwhile, another 3.5% of posters' popular posts are related to entertainment and sports. This small but distinct set of posters was reached by the keyword COVID-19 because they had viral posts about celebrities contracting the virus.

## **Comments**

We scraped a total of 2.8 million comments made to the almost 18,000 posts. The number of likes received by comments presents a distribution that shows a high skewness to the right, averaging 11.4 likes, with 75% having no likes. Notably, the most liked comment garnered an astonishing 127,000 likes.

We find non-mainland comments only account for 4.22% of all comments in our data. Figure 1.5a shows the geographic distribution of non-mainland comments. The number of comments in each country positively correlates with the number of Chinese immigrants in each country. The US leads the count with over 16,000 comments, surpassing the comment volume from several Chinese provinces. Other countries with significant comment counts also host large Chinese immigrant populations, including Canada, Australia, Japan, various European nations, and countries in South Asia. Notably, while African and South American countries contribute relatively few comments, these comments are distributed widely across both continents.

Figure 1.5b illustrates the distribution of comments from mainland users. As further depicted in Figure A4, we find a noticeable correlation between the number of comments and the population of provinces and municipal cities. However, Beijing and Shanghai stand out as exceptions. Despite

having medium-sized populations compared to other provinces and municipalities, they rank second and third, respectively, in the number of comments. Only Guangdong province, the most populous province in China, has more comments than these two cities.

Figure 1.6 motivates the paper. Each dot represents the daily proportion of non-mainland comments. This figure presents an overall reduction of non-mainland comments, with the decreasing trend starting on March 4, 2022, when Weibo started experimenting with comment revelation in Russia-Ukraine posts. In the following sections, we use different strategies to identify the extent to which the policy causes the reduction.

To evaluate the public's attention to the de-anonymization policy, we calculate the proportion of comments that contain keywords such as "IP" and "IP-location." Figure 1.7 depicts the trend of these comments. Notably, there's a substantial spike in these comments immediately after each policy implementation, suggesting that Weibo users were aware of this change. Second, the surge peaked during the first week of comment tagging, indicating that the comment tagging received the most attention from Weibo users. We do not, however, detect any surges or the gradual uptick in the comments containing IP-related keywords before the policy implementation, suggesting the policy was enforced without prior notifications. This rules out potential anticipation effect that biases the estimation. Also noticeable is a consistent rise in the proportion of IP-related comments after each policy roll-out, suggesting that commenting on one's own or other people's IP location gradually becomes a tradition instead of the initial reaction.

### **Mainland vs. Non-mainland**

Beyond mere representation, significant differences exist between the comments from mainland and non-mainland users. Table ?? reveals that non-mainland users generally display a more subdued sentiment. Their comments are 8% less negative, roughly equally positive, and 10% more neutral compared to those from mainland users. In political posts, non-mainland users also exhibit a lower anti-western sentiment. Only 44% of their responses to international political posts conveyed anti-Western sentiment, in contrast to the 52% observed among mainland comments. This disparity underscores the variations in political attitudes between the two groups.

While the sentiments expressed by the two groups diverge, we do not find a significant difference in the endorsement made towards comments. This may reflect sorting or homophily in that two groups tend to comment on posts where their views are mostly acceptable. In the analysis, we discuss whether sorting happens by poster’s location and its implications.

#### 1.4 Empirical Strategy

The main goal of this paper is to causally identify the impact of the series of de-anonymization policies on the representation of non-mainland users and polarization. In the main results, we estimate the policy impact using a Regression Discontinuity in Time (RDiT) design. The identification assumption is that the content that appeared on the platform shortly before the policy implementation is similar to those that appeared shortly after. So that the only change that could affect the commenting behavior is the commenter’s identity being revealed. We test if there’s a discontinuous change in the proportion of post types at the implementation date. In addition, relying on the uniquely rich high-frequency data we collected, we manage the sample within a relatively narrow time window to rule out potential shocks.

Specifically, we estimate:

$$y_{ijpt} = \beta_1 Treat_t^d + \beta_2 f^a(Time_t^d) + \beta_3 Treat_t^d \times f^b(Time_t^d) + \gamma X_{jpt} + DoW_t + PostType_p + \epsilon_{ijpt}. \quad (1.1)$$

In this equation,  $y_{ijpt}$  is a specific feature of the comment  $i$  made to the poster  $j$ ’s post  $p$  on date  $t$ . In the first main regression,  $y_{ijpt}$  is a non-mainland dummy that equals one if the comment  $i$ ’s IP location is somewhere non-mainland. In the second main regression,  $y_{ijpt}$  is an anti-Western dummy that equals one if the comment  $i$  is detected to contain an anti-Western sentiment. To explore the potential mechanisms, we also let  $y_{ijpt}$  describe the number of likes received by the comment and the negative sentiment detected in the comment.

The treatment variable  $Treat_t^d$  is a set of dummies indicating whether a certain de-anonymization policy has been implemented.  $d \in \{\text{profile, comment}\}$  to distinguish the impact of IP tagging in the profile and the comment separately.  $Treat_t^{profile}$  equals one if the comment  $i$  was made on or after Mar 18, 2022.  $Treat_t^{comment}$  equals one if the comment  $i$  was made on or after April 28, 2022.



The variable  $Time_t^d$  is the date centered around the first day of policy  $d$ 's implementation.  $Time_t^d$  and  $Treat_t^d \times Time_t^d$  separately characterize the time trend before and after the policy implementation. In the preferred specification,  $f^a$  and  $f^b$  are the linear functions, following the recommendation made by Gelman and Imbens (2019). We use a uniform kernel to produce a time trend less sensitive to the impulse policies response to the early policy implementation stage.

In the main specifications, we set the bandwidth to be 30 days. A 40-day gap between the profile and comment revelation naturally caps the bandwidth. We vary the bandwidth from 20 to 40 days to test the robustness of the effect we estimated through Equation 1.1. In addition, we extend the window for estimating the effect of IP revelation in the comment to test the robustness of the result derived from a narrow window.

In a set of covariates  $X_{ijpt}$ , we include the post's number of likes, reposts, and comments to control the post's spread level. In addition, we add a non-entertaining dummy to characterize the degree of seriousness of the post. As the poster may slant the content to cater to the preferences of the potential audiences, we classify the poster's potential anti-western inclination using the ratio of anti-western comments made to them before March 18, 2022. A poster is labeled as "highly anti-western" if over half of the sampled comments made to a political post were detected to contain anti-western sentiment.

We include the Day-of-Week fixed effect in all specifications to rule out cyclical effects within the narrow time window. We further include the fixed effect of post type proxied by the news category the post belongs to.

To test the robustness of the RDiT estimation, we exploit the quasi-randomness in the experimental stage of the policy when the platform revealed the comments' IP location only for Russia-Ukraine-related posts that explicitly included the keyword "Russia" or "Ukraine" in the post content. We can identify the local effects of revealing IP location through a standard difference-in-differences design:

$$y_{ijpt} = \beta_1 Treat_{jpt}^{RU} + \beta_2 Post_t + \beta_3 Treat_{jpt}^{RU} \times Post_t + \gamma X_{jpt} + DoW_t + \epsilon_{ijpt}, \quad (1.2)$$

where  $Treat_{jpt}^{RU}$  indicates whether the post  $j$  includes the triggering keywords and  $Post_t$  is the

conventional status dummy that whether the post  $j$  is written after March 4, 2022, the policy implementation day.

While the ideal comparison should be made between the Russia-Ukraine posts that contained the keywords and those that did not, only around 20% of Russia-Ukraine posts exclude both “Russia” and “Ukraine.” Rather, given all the Russia-Ukraine posts were political during this time frame, we consider all the other international political posts as a fair control. We restrict the sample to all the international political posts written between Jan 1, 2022, and Mar 18, 2022. We include the same set of controls while dropping the post type fixed effect, as all the posts included in the regression are political posts.

One additional source that complicates the interpretation of the average treatment effect estimated from Equation 1.2. First, posters can self-select into or opt out of the treatment by intentionally including or excluding keywords to trigger or evade the function. To assess to what extent the poster takes advantage of this loophole, we predict the likelihood of including the keywords in Russia-Ukraine-related posts using a rich set of poster characteristics. If no set of characteristics strongly predicts the behavior, we can be reassured of the selection concern.

### **Location Heterogeneity**

Equation 1.1 allows a flexible inclusion of interaction terms to study the heterogeneous treatment effects. The primary heterogeneity we examine is whether the treatment effect varies with the user’s location. First, as political preference may be predicted by the poster’s location, with identity being revealed, the group-specific stigmatization may sort users into posters with certain geographic identities. Second, the treatment effects on sentiment and endorsement may vary with the commentators’ geographic location as the group-specific stigmatization varies by group. Hence, we explore how heterogeneity varies across the poster-commentator location pairs. Commenting behavior may display qualitative differences when the comments are on content penned by same-group posters versus different-group posters, leading to heterogeneous treatment effects.

## 1.5 Results

This section discusses the empirical findings. Figure 1.8 illustrates the balance of the daily ratio of posts related to Western news, politics, and COVID-19, residual of the day-of-week fixed effect, and controls. The rest of the balance tests are summarized in Table A3. The continuity of platform content is not perfectly ensured, given that the ratio of political posts significantly declined during profile de-anonymization. Nevertheless, in practice, we control post types using the post type fixed effects.

### 1.5.1 Reduction of Non-mainland users

Table 1.3 reports the local average treatment effect of profile de-anonymization on Non-mainlanders' participation. The first column reports the preferred specification using a bandwidth of 30 days before and after the policy implementation. Columns (2) - (5) report the estimations using bandwidths varying from 20 to 40 days. We find the profile de-anonymization strongly reduces the ratio of non-mainland comments on the platform by 28.9% (-0.017/0.057), statistically significant at 0.01 level. This observed reduction holds consistent when adjusting the bandwidth from 20 to 40 days, with the estimated effect ranging from 17.5% to 26.2%. All but one of the estimates are statistically significant at 0.1 level at least.

A surprisingly drastic contrast emerges when estimating the local effect of comment de-anonymization on the presence of non-mainland comments. Revelation increases the ratio of non-mainland comments by 20.8%, using the 30-day bandwidth. The result is significant at 0.05 level. Meanwhile, the estimated policy effect is sensitive to the choice of bandwidth, with the effect almost halved when extending the bandwidth by ten more days.

The sharp contrast between the effects of policy can be explained by the two policies' different abilities in encouraging comments. Revisiting Figure 1.7, the proportion of comments related specifically to IP location surged around five times higher during the first few days of comment de-anonymization than the first few days of profile de-anonymization. When the IP location is attached to the profile, users can observe the policy change simply by viewing their own profiles. Meanwhile, when the IP location is attached to the comment, users must comment on something to witness the

policy change. The non-mainlanders' positive response to the comment de-anonymization can be explained by their disproportionate interest in testing the policy.

By extending the bandwidth horizon, Figure A5 illustrates the drastic difference when using a narrow bandwidth (30 days) and a wide bandwidth (120 days). While the left panel shows a positive policy impact at the cutoff, the fitted time trend displays a negative slope that consistently evolves, eventually leading to the negative estimate on the right when more weights is concentrated on the downward trend. Table A4 illustrates how the positive effect is gradually displaced by the negative effect, addressing the inconsistent treatment effect issue we encountered when estimating the local policy effect.

Table A5 shows that the structure of post topics and geographic mentions have shifted since the de-anonymization date. If non-mainland users have preferences for commenting in fading sectors, the decline of these sectors' representation could also reduce the number of non-mainland users. For example, political posts were reduced by 16% post-de-anonymization. If non-mainland users prefer to comment on these posts, the reduction of politics-related posts naturally reduces the participation of non-mainland users. Figure 1.9 addresses the concern by showing a heat plot of the aggregate treatment effects estimated from each geographic mention-topic cell, grouping the post with the same geographic mention and topic. More negative estimates are associated with colder colors and vice-versa. Only 10 out of 60 cells obtained a positive estimate, and most of these cells are associated with nonsensitive topics, with three of them coming from Entertainment&Sports mentioning less-discussed geographic locations.

### **Sorting**

Differences in preference seem to affect mainlanders' and non-mainlanders' comments to posts; by the same token, they may also affect the posts written by mainlanders and non-mainlanders. Presumably, a poster is more likely to attract audiences sharing the same geographical identity because of similar interests, experiences, and political attitudes. On the other hand, such identity-based segregation could exaggerate the echo chamber effect. We study whether reducing anonymity furthered segregation.

We saturate Equation 1.1 to explore this prediction by interacting the treatment dummies with a non-mainland post dummy, which indicates whether a non-mainland poster wrote the post being commented upon. Table 1.5 strongly supports the model prediction. First, we find that profile de-anonymization reduces non-mainland comments in mainland posts by 31.1%, but the reduction is insignificant in non-mainland posts. Second, non-mainlanders' impulsive response to the comment de-anonymization was furthered in the sense that the policy increases the non-mainland comments in mainland posts by 23.2%, and the increment doubles in non-mainland posts. These findings are robust and mostly statistically significant at 0.1 level across different bandwidths.

These discernible differences highlight the large heterogeneity of the policy effect based on poster location. The less negative policy effect on the non-mainland poster subset suggests that identity-based segregation is enhanced. We discuss its implications using the following regression results.

### **Endorsement**

We measure the change in endorsement made to the comment by the number of likes standardized within each post. Table 1.6 shows the policy effect on comment endorsement. Columns (1) and (2) in Panel A of Table 1.6 report the direct estimation of Equation 1.1 without and with interacting with the non-mainland post dummy. The non-mainland post dummy indicates whether the post is written by a non-mainland poster, which examines the heterogeneity of policy effect with respect to the poster's identity. These estimates are insignificant and close to zero. However, in column (3), we find the endorsement received by mainland comments increased by 0.013 standard deviations when these comments were appended to mainland posts. At the same time, there is no effect for mainland comments in non-mainland posts. These effects are large, considering that one standard deviation in the control group (post penned by mainland posters) equals over 400 likes. In contrast, column (4) shows that non-mainland comments reduced the endorsement significantly by 0.06 standard deviation, equivalent to around 24 likes. At the same time, the reduction is largely mitigated when commenting on a non-mainland post.

As shown by Panel B of Table 1.6, most estimates become noisier and insignificant from zero

for comment de-anonymization. However, non-mainland comments received a higher endorsement than non-mainland posts.

These results suggest that marginalized non-mainlanders are likely discriminated against in mainland posts, which traditionally attract more mainland audiences. Because of the stronger negative reflections toward their opinions, non-mainlanders naturally reduce their presence on the platform. One caveat is that we cannot qualitatively assess the change in the content of comments. Presumably, while the policy may also systematically change the content of the comments, it is unlikely that the change can take place in a very short time window.

### **1.5.2 Anti-Western sentiment**

The reduction of opinions by non-mainland users would amplify the voice of mainland users, who are, on average, more inclined to an anti-western narrative when discussing political issues. The strengthened segregation could further polarization.

To empirically test the policy impact on polarization, we estimate Equation 1.1 with the anti-western dummy as the outcome. We use the same specification for comment endorsement, except that the post-type fixed effect is mechanically omitted, given that the outcome variable is only measured in political posts. Panel A of Table 1.7 shows that profile de-anonymization increases the overall anti-Western sentiment by 12.8%, with statistical significance at 0.01 level. Column (2) shows that the increment happens in both types of posts. Column (3) further shows that mainland comments are the main driver for increasing anti-Western sentiment. On the other hand, column (4) suggests the policy reduces the anti-Western sentiment expressed among non-mainland comments. However, the results are not statistically significant, perhaps due to the small number of observations.

More robust results are found in Panel B of Table 1.7. Revealing in comments significantly increased anti-Western sentiment overall by 18.1% (0.086/0.476). Consistent with previous results, the increase happens in both types of posts. Interestingly, as shown by column (3), while the mainland comments show a similar increase in sentiment in mainland posts, this increase is smaller in non-mainland posts. The estimated reduction of 10.9% (-0.054/0.495) is both economically

meaningful and statistically significant at the 0.05 level.

Combining these two findings, we show that reducing anonymity causes stronger polarization, measured by the anti-Western sentiment. Specifically, anti-Western users sort into mainland posts that presumably lean towards an anti-Western narrative, and pro-Western users sort into non-mainland posts that presumably lean towards a pro-Western narrative. The immediate and substantial policy effect is potentially consistent with a selection effect in which the policy incentivizes the commenters to be more gathered to their favored posts instead of the real-time attitudinal change.

### **1.5.3 Sentiment**

Last of all, we study the policy impact on general sentiment, measured by the negativity in the comments detected by our text analyzer. Negativity, such as anger, dissatisfaction, and unsupportiveness, reflects explicit attitudes people express toward others. Having the identity revealed by the policy may change people's explicitness of expression.

Panel A of Table 1.8 shows no significant change in the sentiment of comment due to the profile de-anonymization. On the other hand, comment de-anonymization significantly reduced the negative tone in comments in all columns except column (3). While noisily estimated, column (3) still presents estimates that are signed in the same direction as the other three columns, suggesting that, in general, the stronger revelation is making people comment more mildly. Such changes are consistent with the theory when individuals recognize that their group identity is overtly exposed, they might adjust their behavior due to concerns such as stereotype threat (Steele and Aronson, 1995).

### **1.5.4 Robustness**

Before the universal implementation of IP-location revelation, the website briefly experimented with comment de-anonymization with only posts that included keywords "Russia" or "Ukraine." The experiment started on March 4, 2022, two weeks before the profile de-anonymization. We test whether the findings from the previous sections are replicated in this short-period experiment through a difference-in-differences design. Given that almost all Russia-Ukraine posts included

these two keywords and that all of the Russia-Ukraine posts that appeared during this period are political, we use all the other international political posts as the control group. Figure 1.11 shows the pretrends test for the four interested outcomes. We do not find notable pretrends that could confound the analysis.

Table 1.9 shows the difference-in-differences estimates of impacts of comment de-anonymization on four outcomes among the posts that included keywords “Russia” and “Ukraine.” The variable  $RU\_KW \times Post\ RU$  reports the DID estimate. We find the estimated impacts on non-mainland and negative comments are similar in magnitude to the estimates reported in Table 1.3 and Table 1.8. Last, we do not observe that the policy increased the anti-Western sentiment in these posts.

A major challenge of the identification of using DID in our context is that the poster can intentionally select into or opt out of the treatment by including or excluding the keywords “Russia” and “Ukraine.” To address this concern, we predict the likelihood of using the keywords in Russia-Ukraine-related posts using the poster’s baseline (before March 4, 2022) characteristics. Figure 1.10 presents the coefficients of the estimate of regressing whether the poster included two keywords in the Russia-Ukraine posts after the policy on a set of poster’s baseline characteristics. We find no strong evidence that a particular type of poster is more likely to include the RU keywords other than the ones that already using these words in the baseline.

In parallel to the robustness test, the already-treated Russia-Ukraine Posts also present a credible placebo test. As the comments made to these posts are already tagged with commentators’ IP locations, the later universal policy implementations should not have any impact other than spillovers. In Table 1.10, we show that for all of the outcomes we studied, comment-level de-anonymization only significantly increased the non-mainland comments in these posts. The drastic increase, instead of being caused by the policy, may be explained by the spillover of the temporary increase in non-mainland comments throughout the whole platform.

In general, we find no strong evidence that the effects on non-mainlanders’ presence, comment endorsement, anti-Western sentiment, and negativity could have resulted from other shocks.



## 1.6 Framework

To motivate the empirical results, we study a stylized signaling model of individuals that differ by both their private and group-specific benefits from supporting a policy change. We focus the impact of making group membership observable on an individual's decision to express support for the policy.

### Signaling framework

Expressing support for the policy triggers social sanctions from the opposing party. In our empirical context, when an individual expresses unfavorable opinions on social media, the opposing parties criticize the individual's motivation. Further, when the individual's group membership is observable, opposing parties could amplify the criticism by stigmatizing the group the individual belongs to.

Burnsztyn et al. (2023) provide a simple signaling framework to describe the decision to publicly support the policy under the threat of social sanctions. We extend the framework to consider a commonly encountered scenario where individuals differ by group memberships. Individual preference is partially determined by group membership. For example, younger voters, on average, lean towards the progressive political agenda in US, while senior voters are more likely to support the conservative agenda. In our context, non-mainland Chinese may prefer the termination of the zero-COVID policy more than mainland Chinese as they may benefit from the positive economic impact and be less influenced by the negative health impacts. The analysis becomes more complicated when group preferences overlap, and our proposed model helps explore the decision to support publicly when group membership is revealed and when it is kept anonymous and develops equilibrium implications of de-anonymization.

Consider a society that comprises a continuum of individuals from two groups,  $G = \{L, F\}$ , with measures  $\mu$  and  $1 - \mu$ , respectively. Group  $F$  is also considered the minority with  $\mu \geq 0.5$ . In our context,  $L$  would be the local mainland group, and  $F$  would be the foreign or non-mainland group. The only difference distinguishing the two groups is the fixed benefit,  $\omega^G$ , from a policy change. The benefit could be understood as the expected benefit from policy change or the potential

gain of successful persuasion. The policy is more favored by  $F$ :  $\omega^F > \omega^L$ .

In addition to the fixed benefit, the individual obtains a private benefit,  $t \sim U[-T, T]$ , from supporting the policy change. The private benefit can be regarded as the political preference of the individual, with higher  $t$  associated with larger benefit from supporting.<sup>8</sup> We assume that the private benefit is distributed uniformly and symmetrically around zero to make the model more tractable. Overall, any individual  $i$ 's benefit from publicly supporting the policy is

$$b_i^G = \omega^G + t_i,$$

where  $b_i^G \sim U[-T + \omega^G, T + \omega^G]$ .

We assume that the benefit of publicly supporting the policy change is proportional to the benefit of policy change. In addition, we normalize the proportion to be equal to one for simplicity.

Neither group has unanimous support or opposition towards the policy change, reflected by  $\omega^L + T > 0$  and  $\omega^F - T < 0$ . It would be improbable, for instance, to find a society without even one young individual backing a conservative agenda or to find no member of the diaspora supporting the political party of the home country. These restrictions reflected the previous examples that while the majority of young voters prefer progressive social change, a small portion of them may prefer a conservative agenda, and while the majority of non-mainland Chinese prefer the termination of the zero-COVID policy, some insist on the continuation.

The supporting decision is binary. An individual can either support the policy with  $d_i = 1$  or remain status quo with  $d_i = 0$ <sup>9</sup>.

Expressing support for the policy change triggers social sanctions from the party that prefers the status quo. In our context, making dissent views on social media incurs criticisms from people who hold mainstream values. We model the sanction imposed by each sanctioner to a supporter as the distance between her benefit and the supporter's expected benefit from the policy change. Suppose there are two individuals,  $i$  and  $j$ .  $j$  supports the status quo, then the social sanction imposed on  $i$

---

<sup>8</sup>The magnitude of private benefit could be plausibly inferred through repeated real-life interactions. However, the interaction is much less frequent in cyberspace. In this case, people do not have a good prior on others' private benefits but rely on visible identities and opinions.

<sup>9</sup>Here, we simplify complicated social responses to debatable issues. Specifically, we group the silent individuals with those who support the status quo.

is zero if  $i$  also supports the status quo. However, if  $i$  supports the change, then  $j$  impose a social sanction that is equal to:

$$|E(b_i|d_i = 1, *) - b_j|,$$

where  $E(b_i|d_i = 1, *)$  is the expected benefit of  $i$ , signaled by  $i$ 's supporting decision  $d_i = 1$ .

Essentially, this absolute difference models a realistic scenario in the sanction imposed to a progressive supporter of policy change is higher from a far-right conservative with a very low  $b$  but lower from a moderate conservative with a relatively high  $b$ . The total sanction,  $S_i$ , a supporter  $i$  received sums all the individual sanctions:

$$S_i = \underbrace{\mu \int_{b:d_L(b)=0} |E(b_i|d_i = 1, *) - b| dF_{bL}}_{\text{from status quo L}} + \underbrace{(1 - \mu) \int_{b:d_F(b)=0} |E(b_i|d_i = 1, *) - b| dF_{bF}}_{\text{from status quo F}},$$

where  $b : d_G(b) = 0$  is the set of total benefits  $b$  that induces opposition for members in group  $G$ . This sanction formula models two important aspects of the sanction that happened in real life. First, the strength of the sanction varies with the sanctioner's group representation. In our setup,  $L$  and  $F$ 's sanction ability is proportional to their representations,  $\mu$  and  $1 - \mu$ . Second, the extent to which the sanction is imposed is positively associated with the magnitude of the discrepancy. When dissenters are viewed to differ more significantly from the values of the sanctioner, the sanction is heavier.

Individual  $i$  from group  $G$  chooses to support if the total benefit outweighs the social sanction:

$$b_i^G - \kappa S_i \geq 0.$$

The parameter  $\kappa$  is a fixed parameter that ranges  $(0, 1)$ . It assumes that the strength of social sanction is linear in  $S_i$  and bounded by  $(0, 1)$ .

### **Anonymity**

Inference about a supporter's type depends on whether group membership of supporters is revealed. In our context,  $G$  is only observable after the platform tags the user's comment with an IP-based geographic location. Before de-anonymization, inference relies only on the expected total

benefit signaled by the supporting decision. Thus, the sanction is determined by

$$|E(b_i|d_i = 1) - b|.$$

When group identity is de-anonymized, this information further facilitates the inference, and the sanction is determined by

$$|E(b_i|d_i = 1, G) - b|.$$

## Equilibrium

In both scenarios, the equilibrium determines who will support and how much social sanction will be imposed. Specifically, the equilibrium is characterized by partitioning the support of  $b$ , with one set preferring the status quo and imposing sanctions on the other supporting the policy change.

**Definition 1** *The equilibrium for group  $G$  under observability  $A$  is defined by a mapping  $d_G^A : b \rightarrow \{0, 1\}$  such that*

$$d_G^A(b_i) = \begin{cases} 1 & \text{if } b_i - \kappa S^* \geq 0, \\ 0 & \text{if } b_i - \kappa S^* < 0, \end{cases} \quad (1.3)$$

where

$$S^* = \mu \int_{b:d_L^S(b)=0} |E(b_i|d_i = 1, *) - b| dF_{b^L} + (1 - \mu) \int_{b:d_F^S(b)=0} |E(b_i|d_i = 1, *) - b| dF_{b^F}.$$

Note that in equilibrium, the total sanction  $S^*$  that a supporter receives does not depend on his or her own type but depends on the perceived social average. Therefore, in any equilibrium, if  $b_i$  induces support and  $b_j$  induces status-quo preference, it must be that  $b_i > b_j$ . In other words, we can characterize the equilibrium through thresholds. In addition, when group membership is unobserved, the supporting decision only depends on one common threshold for both  $L$  and  $F$  as they are indistinguishable by the sanctioners.

**Lemma 1** *When the identity is unobservable, there is a common threshold  $\tilde{b}$  such that,*

$$d_G^U(b_i) = 1, \text{ iff } b_i \geq \tilde{b}.$$

When identity is observable, there is a set of thresholds  $\{\tilde{b}_L, \tilde{b}_F\}$  such that,

$$d_G^O(b_i) = 1 \quad \text{if } b_i \in G \quad \text{and} \quad b_i \geq \tilde{b}_G. \quad (1.4)$$

With Lemma 1, we can immediately show:

**Lemma 2** *In every case,*

$$E(b_i | d_i = 1, *) \geq b \quad \forall b \quad \text{s.t.} \quad d(b) = 0.$$

The proof is straightforward through contradiction. In either case, if the expected benefit of the supporter is lower than an opposer, then there must exist at least one supporter whose benefit is lower than that opposer. However, he or she should not be a supporter. This proposition provides a convenient condition to solve the equilibrium as the differences inside of absolute values are always positive.

**Proposition 1** *There is a unique equilibrium when group membership is unobservable.*

The proof of uniqueness is provided in the Appendix. The idea is to observe that the equilibrium punishment can be represented as the difference between the mean benefit of the supporter and the mean benefit of the population and then use the fact that the first derivative of the conditional mean of a uniform distribution is  $\frac{1}{2}$  to obtain the uniqueness.

**Conjecture 1** *There is a unique equilibrium when group membership is observable.*

Given Proposition 1, we also conjecture that the equilibrium of the observable case is also unique. We provide support for the conjecture by simulating  $\{t_L, t_F\}$  on the grid of  $\{\mu, \omega^G, T, \kappa\}$  within an economically reasonable range<sup>10</sup>. Figure A6 demonstrates a selection of parameters and shows the uniqueness of  $\{t_L, t_F\}$  when the individual's group identity is observable.

## Predictions

We characterize the model through simulation, which is made possible by Proposition 1 and Conjecture 1. Figure A6 depicts the comparative statics of induced private benefit thresholds,  $t^G$ , by varying the sanction discount factor  $\kappa$ . We allow  $T$  from 1 to 50 to vary the noise level in

<sup>10</sup>We simulate the solution using Python package SciPy. The numerical solver of this package finds all the solutions if there are many. In our case, it produces a single solution for all the set of parameters we entered.

inference. For each  $T$ , we pick  $\omega^L$  and  $\omega^F$  arbitrarily up to the overlapping condition. We start  $\omega^L$  from a negative value to zero to positive to model situations in which policy change benefits  $F$  at the expense of  $L$ , does not influence  $L$ , and potentially benefits  $L$ . A greater threshold value  $t^G$  indicates a higher private benefit is required to sustain public support. A greater  $\kappa$  is associated with people being more sensitive to social sanction.

**Prediction 1** *Minority group's willingness to support declines when the group membership is revealed.* This prediction is obtained by observing that the curve  $t_{obs}^F$  is continuously higher than  $t_{unobs}^F$ , meaning that the private benefit required to sustain support is higher when group membership becomes publicly observable. This prediction matches the empirical result that non-mainlanders post less following de-anonymization.

## 1.7 Conclusion

This article examines the effects of reducing the anonymity of a social media platform. It identifies the causal effects by leveraging three unique opportunities. First, the sharp and unexpected policy implementation creates a quasi-experiment. Second, the access to high-frequency text data around the policy's implementation date. Third, groundbreaking advancements in natural language processing enable swift content and sentiment recognition.

We find that reducing anonymity reduces the participation of non-mainland Chinese users on one of the largest Chinese social media platforms. Besides the impact of de-anonymization on the comments' endorsement and sentiments, the effects also vary with the commentator's geographic identity. Eventually, we show that the identity-based discrimination and segregation ignited by de-anonymization could explain these two stylized findings.

The findings have two implications for understanding social media ecology and policy design. First, a moderate level of anonymity reduction can affect the demographic structure of the social media platform. Precise identification is nearly impossible, given that Chinese users are identified at the provincial level and overseas users at the country level. Therefore, systematic behavioral change is likely triggered by preferential treatment or discrimination against specific geographic groups. While our study is China-specific, the result may be generalizable, given that many global

platforms have implemented verification programs to inhibit the transmission of misinformation and harmful action and may consider universally reducing anonymity within a legal scope.

Second, as one type of public opinion, social media opinion could have been leveraged for policy design, especially in places with a high system barrier to aggregate public opinion. In this paper, we show that regulation could amplify one opinion favored by the domestic group through two channels. First, it reduces out-group participation, who favor the dissent opinion. Second, it further increases the representation of one opinion within the domestic group. Both channels crowd out dissents and could have normalized the extreme by exacerbating the echo chamber, causing the social media opinion to deviate from public opinion and distort the policy design based on social media data.

Last of all, we believe several extensions could promisingly increase the significance of this literature. First and foremost, in this article, we only study the potential sorting happening within the platform. It is valuable to study the potential inter-platform migration resulting from the policy. Specifically, while our data repository, Sina Weibo, pioneered the de-anonymization in early March, many other major Chinese social platforms did not follow the lead until the end of the year, when the Administration issued the mandatory order. Anecdotal evidence also suggests that different social media platforms differentiate with different norms or ideologies.

Second, more efforts could be exerted to characterize the poster types. Influential social media posters are now regarded as important news outlets, while their individuality permits stronger political bias and slant. Therefore, a more detailed characterization of posters should provide new perspectives on how social media policies vary with the poster identities.

## 1.8 Figures

Figure 1.1 The prompt structure incorporates meta-prompt tuning and in-context learning. Content in [] was replaced by codebook and data when performing the deductive coding task. When analyzing Chinese data, we translated the prompt into Chinese

### Text Analysis Prompt Structure

Imagine you are an expert in qualitative analysis. Your task is to code the text into one code. Please make sure you read and understand these instructions carefully. Please keep this document open while reviewing, and refer to it as needed.

Coding Steps:

1. Read the text thoroughly to understand its content and context.
2. Review the codebook to understand the codes and their definitions. The codebook serves as a guide for coding the text consistently and systematically.
3. Assign codes from the codebook to input text. This process involves matching the content of the text to the codes and their definitions in the codebook. It's essential to be consistent and systematic in this step to ensure the reliability of the analysis.
4. Assign the final label from one of the [CODE-1, CODE-2, ..., CODE-n]
5. Review the coded text and refine the coding if necessary.

Classify the [text] into one of the [CODE-1, CODE-2, ..., CODE-n] Based on the following codebook that includes the description of each code and a few examples. Note that the code is only for the [text], not the context, but the context should be considered when making the decision.

Codebook: CODE: [CODE-1]; DESCRIPTION: [CODE-1-DESCRIPTION]; EXAMPLES: [CODE-1-EXAMPLES].  
CODE: [CODE-2]; DESCRIPTION: [CODE-2-DESCRIPTION]; EXAMPLES: [CODE-2-EXAMPLES]. ... CODE: [CODE-n];  
DESCRIPTION: [CODE-n-DESCRIPTION]; EXAMPLES: [CODE-n-EXAMPLES].

Context: [CONTEXT]

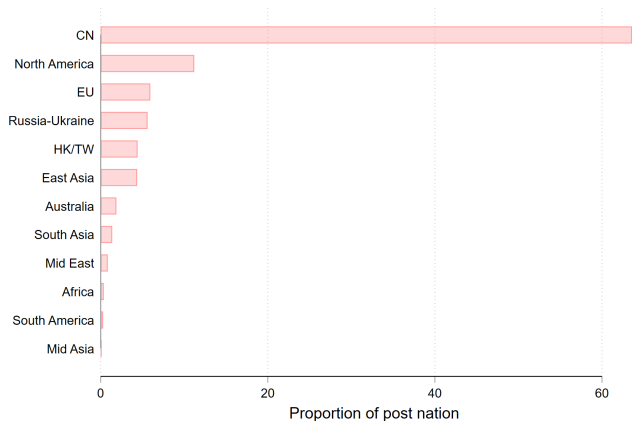
Text: [TEXT]

Choose from the following candidates: [CODE-1, CODE-2, ..., CODE-n]

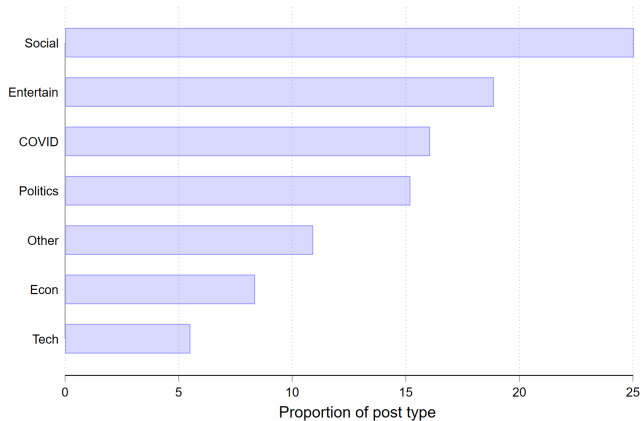


Figure 1.2 Post Statistics

(a) The Distribution of Geographical Identifiers



(b) The Distribution of Post Topics



(c) The Distribution of # of Likes, Reposts, and Comments

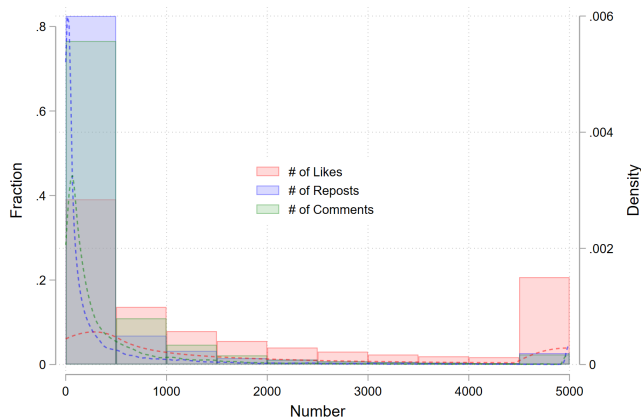
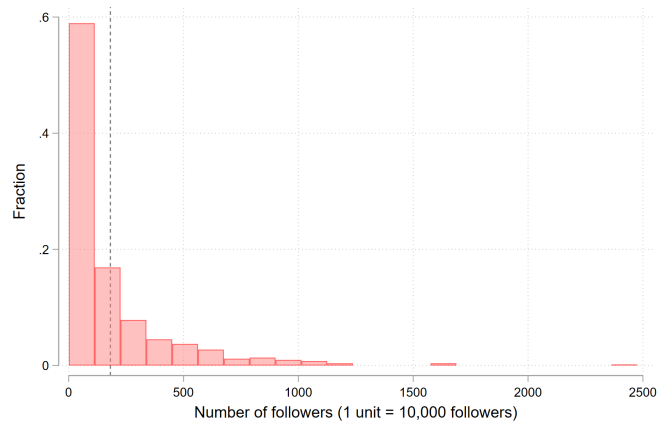


Figure 1.3 Poster's Basic Statistics

(a) The Distribution of # of Followers



(b) The Poster's IP-Location

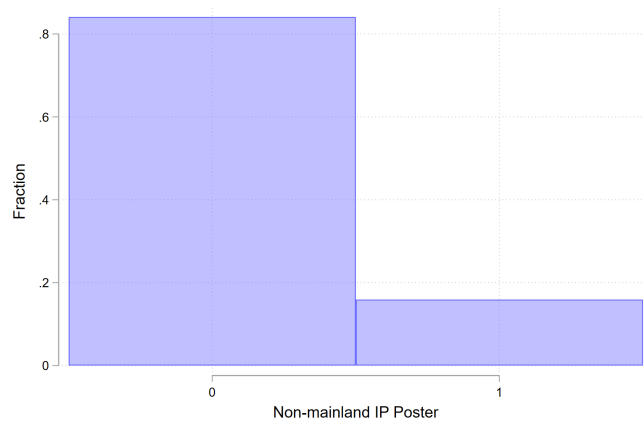
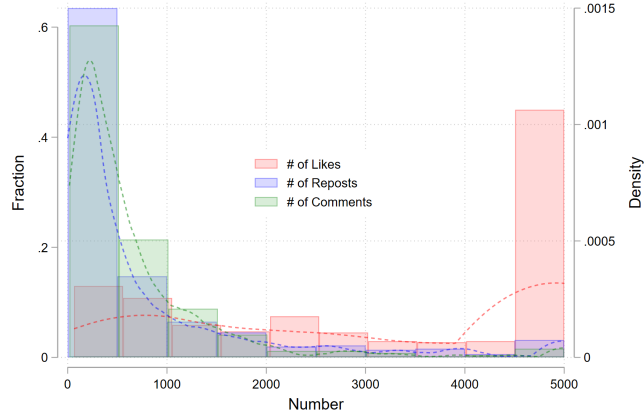
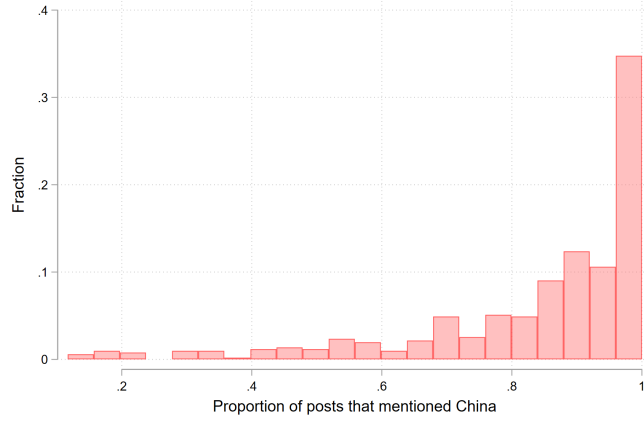


Figure 1.4 Poster's Post Statistics

(a) The Distribution of # of Likes, Reposts, and Comments



(b) The Distribution of Poster's China-related Post



(c) The Distribution of Poster's Non-Entertain Posts

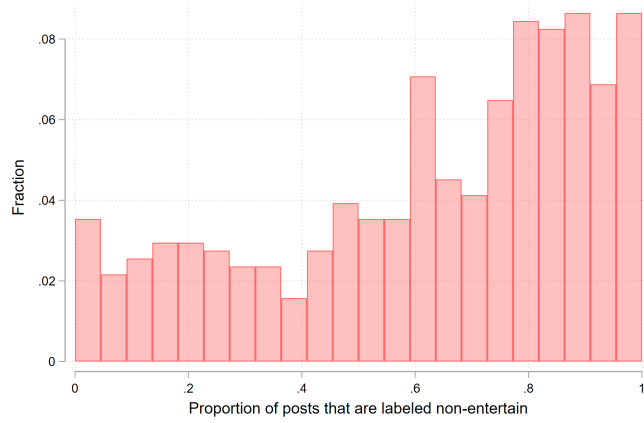
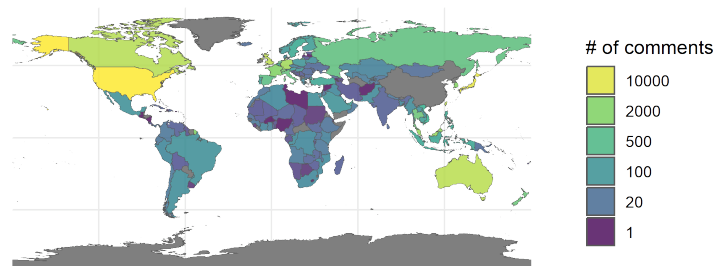
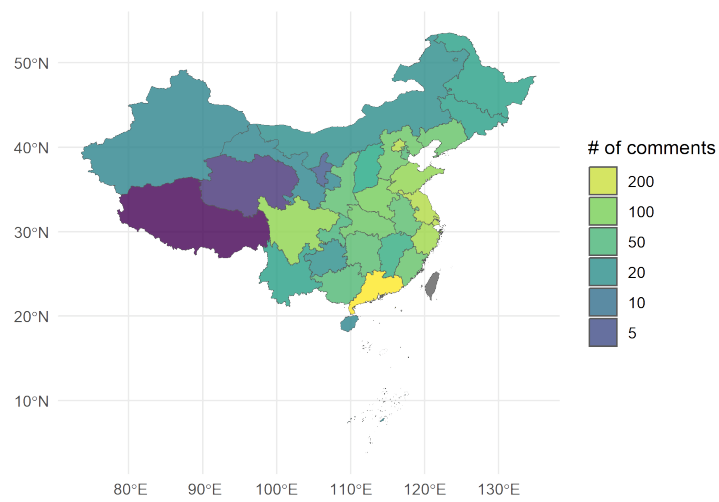


Figure 1.5 The Number of Comments Made According to the IP Location

(a) Non-mainland



(b) Mainland



Note: One unit of comment equals to 10,000 of comments.

Figure 1.6 The Daily Proportion of Non-mainland Comments

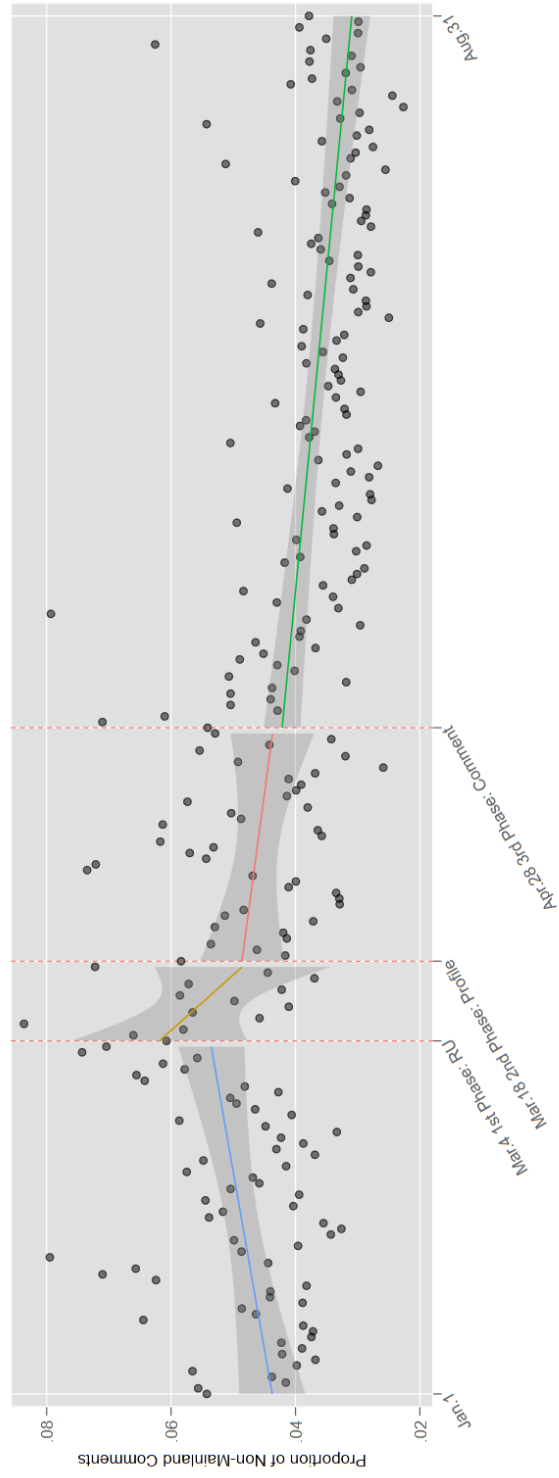


Figure 1.7 The Daily Proportion of IP-related Comments

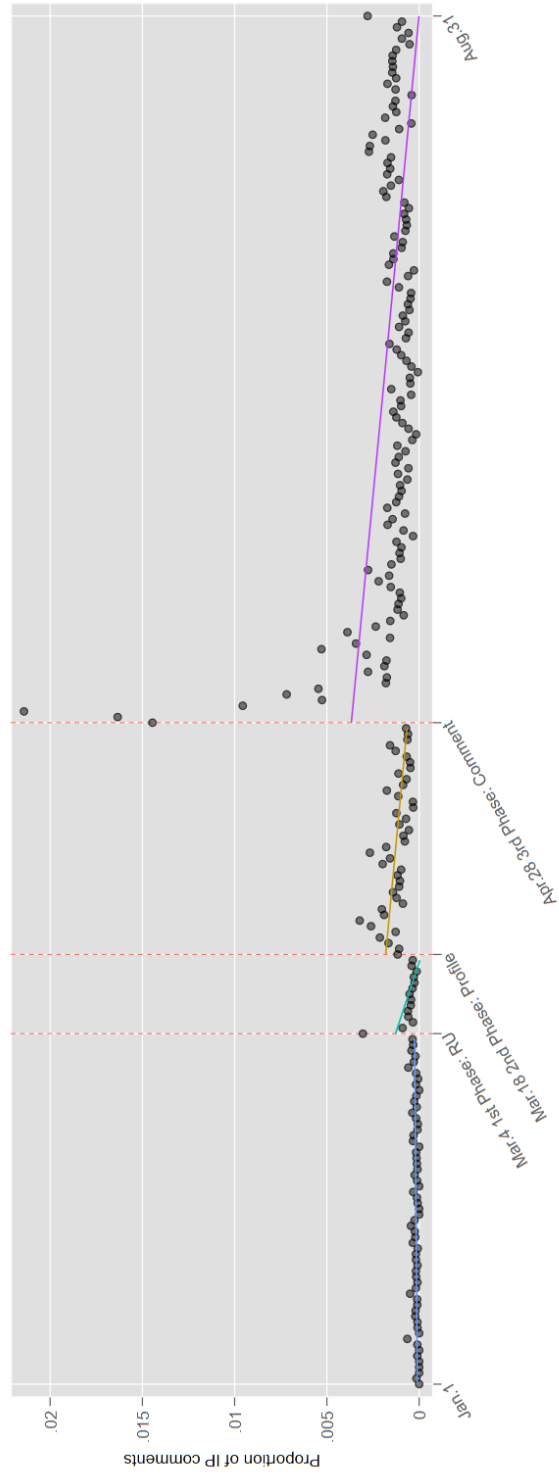
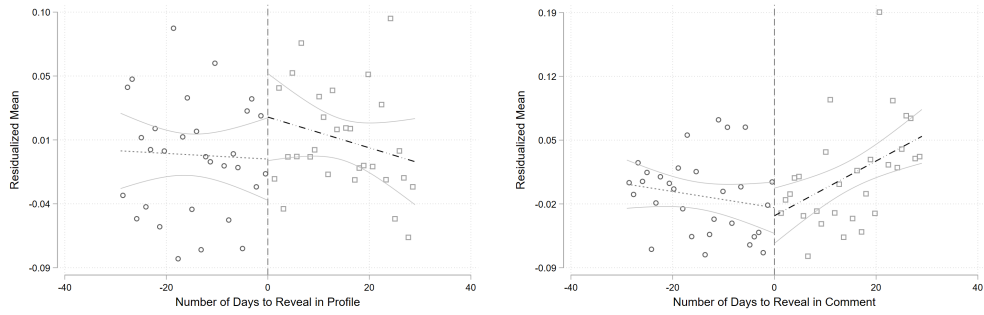
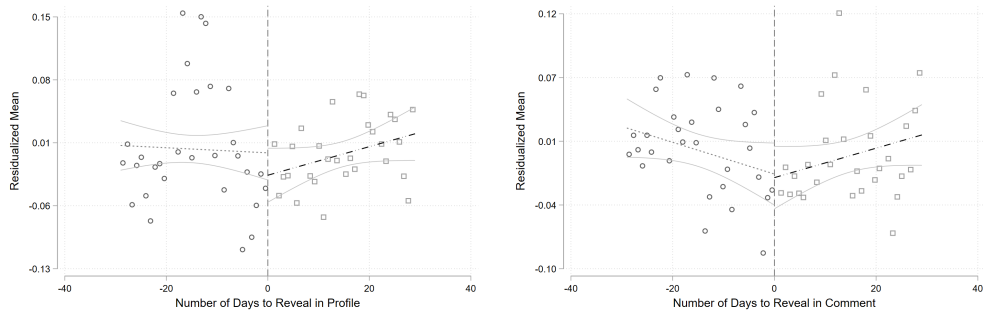


Figure 1.8 RD Balance

(a) Western-related Posts



(b) Politics Posts



(c) COVID Posts

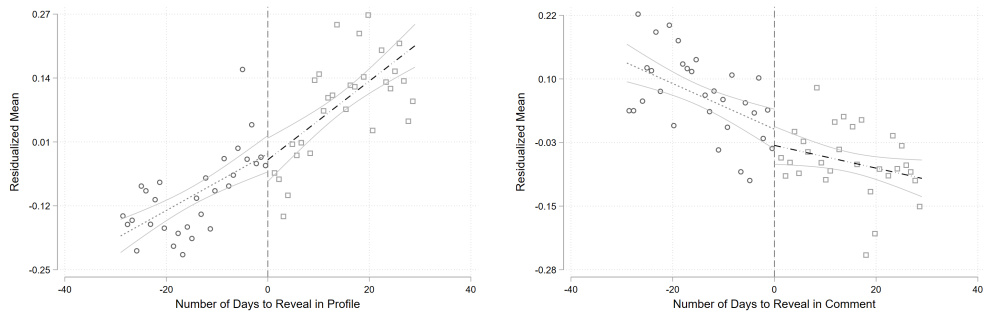


Figure 1.9 Heatplot of De-anonymization Effect on Non-mainland Presence in Each Sector

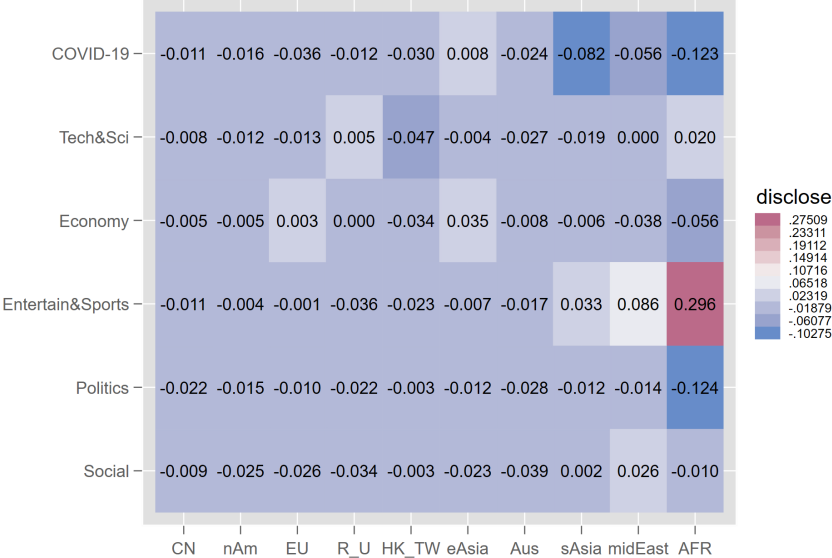




Figure 1.10 Predicting the Poster's Likelihood to Include Keywords

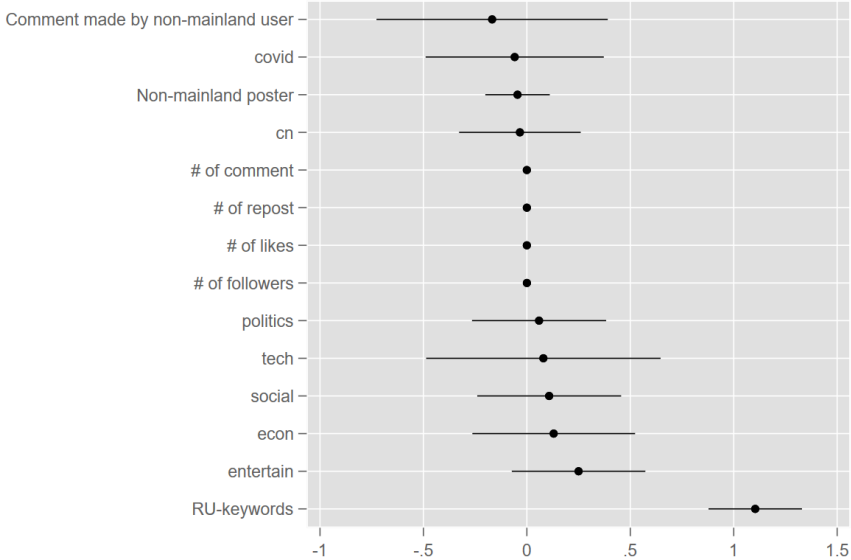
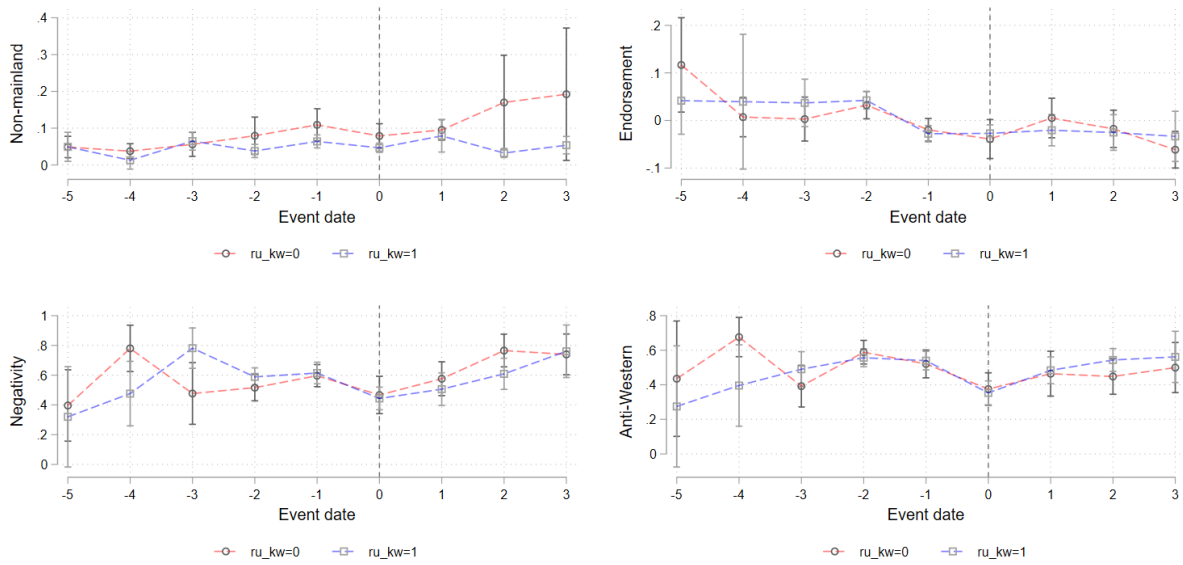


Figure 1.11 Parallel Pretrend Test



## 1.9 Tables

Table 1.1 Overlapping Coefficients of COVID-Influential Posters and Other Influential Posters

C=	Tech	Econ	Entertain	Social	Politics	Edu
	.225	.275	.258	.343	.312	.368

*Notes:* Overlapping coefficients of COVID-19 posters and posters compiled by other seven categories are calculated by equation  $\frac{|COVID \cap C|}{|COVID|}$ , where  $C$  is the set of posters in category  $C$ .

Table 1.2 Mainland vs. Non-mainland

	Mainland			Non-mainland			diff
	N	mean	sd	N	mean	sd	
<b>General Characteristics</b>							
# of likes	2,183,756	11.97	356.03	96241	11.76	316.39	-0.210
Comment length	2,183,756	22.473	25.71	96241	24.918	28.06	2.445***
<b>Sentiments</b>							
Negative	13,209	0.40	0.49	12,505	0.36	0.48	-0.037***
Neutral	13,209	0.39	0.49	12,505	0.44	0.50	0.046***
Positive	13,209	0.21	0.40	12,505	0.20	0.40	-0.009*
Anti-western	40,000	0.52	0.50	4,300	0.44	0.50	-0.075***

Notes: The comment length is calculated by the STATA *ustrlen* command, which assigns a length of 1 to each Chinese/English character and punctuation. The length of English comments is inflated due to this nature but non-Chinese comments are extremely rare in our data

The difference is calculated by regressing the characteristic on a non-mainland dummy without adjusting the standard error.

Table 1.3 Impact of Reveal IP Location in Profile on Non-mainland Participation

Bandwidth Days =	Non-mainland Comment				
	30	20	25	35	40
IP in Profile $t \geq Mar18$	<b>-0.0165***</b> (0.00627)	-0.0106 (0.00854)	-0.0127* (0.00702)	-0.0146** (0.00574)	-0.0114** (0.00544)
Observations	<b>436,879</b>	307,171	375,402	492,426	552,794
Control Mean	<b>0.0570</b>	0.0606	0.0590	0.0557	0.0553

Notes: This table presents the main results from estimating the effect of profile de-anonymization on the presence of non-mainland comments. It reports the estimation results from the main regression discontinuity model using the non-mainland comment dummy as the outcome. The control mean is the proportion of non-mainland comments within the bandwidth days before the policy implementation. The regression includes the set of controls defined in the previous section, as well as day-of-week and post-type fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 1.4 Impact of Reveal IP Location in Comment on Non-mainland Participation

Bandwidth Days =	Non-mainland Comment				
	30	20	25	35	40
IP in Comment $t \geq \text{Apr.28}$	<b>0.0101**</b> <b>(0.00478)</b>	0.0142** (0.00623)	0.0155*** (0.00573)	0.00823* (0.00469)	0.00590 (0.00446)
Observations	<b>464,000</b>	281,924	358,752	561,447	642,200
Control Mean	<b>0.0486</b>	0.0462	0.0508	0.0483	0.0482

Notes: This table presents the main results from estimating the effect of comment de-anonymization on the presence of non-mainland comments. It reports the estimation results from the main regression discontinuity model using the non-mainland comment dummy as the outcome. The control mean is the proportion of non-mainland comments within the bandwidth days before the policy implementation. The regression includes the set of controls defined in the previous section, as well as day-of-week and post-type fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 1.5 Policy Impacts on Non-mainland Participation with Poster Location Heterogeneity

Bandwidth (Days) =	Non-mainland Comment				
	30	20	25	35	40
<i>Panel A. Profile Revelation</i>					
Revelation	-0.0142** (0.00635)	-0.00909 (0.00857)	-0.00945 (0.00709)	-0.0115** (0.00583)	-0.00778 (0.00549)
Revelation X NM post	0.00541 (0.00834)	0.00813 (0.0102)	0.00191 (0.00906)	0.00412 (0.00777)	-0.00145 (0.00715)
Observations	436,879	307,171	375,402	492,426	552,794
Control Mean	0.0457	0.0492	0.0476	0.0450	0.0441
<i>Panel B. Comment Revelation</i>					
Revelation	0.00902* (0.00478)	0.0113* (0.00630)	0.0116** (0.00586)	0.00719 (0.00466)	0.00458 (0.00439)
Revelation X NM poste	0.0177* (0.00960)	0.0264** (0.0107)	0.0287** (0.0115)	0.0148* (0.00878)	0.0157* (0.00802)
Observations	464,000	281,924	358,752	561,447	642,200
Control Mean	0.0388	0.0376	0.0408	0.0386	0.0387

Notes: This table presents the effect of policy on the presence of non-mainland users. Panels A and B investigate the effect of profile and comment de-anonymization separately. They report the estimation results from the main regression discontinuity model using the non-mainland comment dummy. Each column reports an estimate using the bandwidth shown on the top row. Revelation is the dummy of the policy implementation. NM post indicates if a non-mainland poster writes the post. The Control Mean is the proportion of non-mainland comments within the bandwidth days before the policy implementation. The regression includes the set of controls defined in the previous section, as well as day-of-week and post-type fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 1.6 Policy Impacts on Comment Endorsement

	Comment Endorsements			
	(1)	(2)	(3)	(4)
<i>Panel A. Profile Revelation</i>				
Revelation	0.00738 (0.00643)	0.00820 (0.00651)	0.0128* (0.00662)	-0.0584* (0.0324)
Revelation X NM post		-0.00507 (0.00589)	-0.0103* (0.00612)	0.0217 (0.0227)
Observations	431,765	431,765	408,565	23,200
Control SD	436.7	465.1	438.5	405.8
Comment	All	All	Mainland	Non-mainland
<i>Panel B. Comment Revelation</i>				
Revelation	-0.000243 (0.00479)	-0.00329 (0.00600)	-0.00269 (0.00613)	-0.0219 (0.0296)
Revelation X NM post		0.000462 (0.00538)	-0.00850 (0.00556)	0.0547** (0.0218)
Observations	775,854	467,073	445,689	21,384
Control SD	363.9	305.7	338.5	63.2
Comment	All	All	Mainland	Non-mainland

Notes: This table presents the effect of policy on comment endorsement. The comment endorsement is measured by the number of likes received by the comment, standardized within the post the comment replied to. Panels A and B investigate the effect of profile and comment de-anonymization separately. They report the estimation results from the main regression discontinuity model using the comment endorsement as the outcome and a 30-day bandwidth. Revelation is the dummy of the policy implementation. NM post indicates if a non-mainland poster writes the post. The Comment row shows the comment sample used in the estimation. All stands for all the comments. Mainland (Non-mainland) is the sample that only includes mainland (Non-mainland) comments. The control SD is one standard deviation of the number of likes received by the comments that appeared in some posts within the bandwidth days before the policy implementation. The regression includes the set of controls defined in the previous section, as well as day-of-week and post-type fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.



Table 1.7 Policy Impacts on Political Attitude

	Anti-Western Sentiment			
	(1)	(2)	(3)	(4)
<i>Panel A. Profile Revelation</i>				
Revelation	0.0626** (0.0253)	0.0645** (0.0261)	0.0704*** (0.0272)	-0.0308 (0.100)
Revelation X NM post		-0.00571 (0.0215)	-0.00364 (0.0226)	-0.0407 (0.0749)
Observations	10,555	10,555	9,930	625
Control Mean	0.4887	0.4911	0.4974	0.3333
Comment	All	All	Mainland	Non-mainland
<i>Panel B. Comment Revelation</i>				
Revelation	0.0855*** (0.0286)	0.0957*** (0.0293)	0.126*** (0.0313)	-0.124 (0.0862)
Revelation X NM post		-0.0364 (0.0243)	-0.0541** (0.0268)	0.0461 (0.0589)
Observations	4,401	4,401	3,814	587
Control Mean	0.4758	0.4875	0.4950	0.4077
Comment	All	All	Mainland	Non-mainland

Notes: This table presents the effect of policy on political attitude, measured by the anti-Western sentiment detected in the comment. The anti-Western sentiment is a dummy variable that indicates the inclination of anti-Western sentiment classified by our text-analysis algorithm. Panels A and B investigate the effect of profile and comment de-anonymization separately. They report the estimation results from the main regression discontinuity model using the anti-Western dummy and a 30-day bandwidth. Revelation is the dummy of the policy implementation. NM post indicates if a non-mainland poster writes the post. The Comment row shows the comment sample used in the estimation. All stands for all the comments. Mainland (Non-mainland) is the sample that only includes mainland (Non-mainland) comments. The Control Mean is the proportion of anti-Western comments that appeared in the international political posts within the bandwidth days before the policy implementation. The regression includes the set of controls defined in the previous section, as well as day-of-week and post-type fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 1.8 Policy Impacts on General Sentiment

	(1)	Negative Sentiment		(4)
		(2)	(3)	
<i>Panel A. Profile Revelation</i>				
Revelation	-0.0243 (0.0199)	-0.0171 (0.0205)	-0.0438 (0.0277)	0.00703 (0.0308)
Revelation X NM post		-0.0285 (0.0197)	-0.0107 (0.0297)	-0.0343 (0.0271)
Observations	9,966	9,966	5,225	4,741
Control Mean	0.4796	0.5167	0.5029	0.4546
Comment	All	All	Mainland	Non-mainland
<i>Panel B. Comment Revelation</i>				
Revelation	-0.0708*** (0.0287)	-0.0508* (0.0292)	-0.0604 (0.0414)	-0.0342 (0.0417)
Revelation X NM post		-0.0745*** (0.0243)	-0.0166 (0.0472)	-0.0917*** (0.0301)
Observations	4,361	4,361	2,121	2,240
Control Mean	0.4297	0.4444	0.4536	0.4330
Comment	All	All	Mainland	Non-mainland

Notes: This table presents the effect of policy on general sentiment, measured by the negativity detected in the comment. The observations are taken from a sample of 40,000 comments equally comprised by randomly drawn comments from mainlanders and non-mainlanders. The negativity sentiment is a dummy variable that indicates the presence of negativity, such as anger, dissatisfaction, and mistrust, that is classified by our text-analysis algorithm. Panels A and B investigate the effect of profile and comment de-anonymization separately. They report the estimation results from the main regression discontinuity model using the negativity dummy and a 30-day bandwidth. Revelation is the dummy of the policy implementation. NM post indicates if a non-mainland poster writes the post. The Comment row shows the comment sample used in the estimation. All stands for all the comments. Mainland (Non-mainland) is the sample that only includes mainland (Non-mainland) comments. The Control Mean is the proportion of negative comments that appeared in the posts within the bandwidth days before the policy implementation. The regression includes the set of controls defined in the previous section, as well as day-of-week and post-type fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table 1.9 Robustness Check Using Treated Russia-Ukraine Posts

	Non-mainland	Endorsement	Anti-Western	Negativity
$Treat^{RU}$	-0.0241***	0.00249	0.0229	0.0408
[=1 if keywords]	(0.00829)	(0.00720)	(0.0224)	(0.0283)
Post	0.0499*	-0.0338***	-0.0848**	0.0785
[=1 if t>Mar04]	(0.0257)	(0.0119)	(0.0337)	(0.0555)
$Treat^{RU} \times Post$	<b>-0.0455*</b>	<b>-0.00494</b>	<b>-0.0437</b>	<b>-0.170***</b>
	(0.0263)	(0.0144)	(0.0448)	(0.0658)
Observations	116,830	194,350	15,483	4,348
Control Mean	0.0710	-	0.5216	0.5696
Control SD	-	1.0	-	-

Notes: This table presents the difference-in-differences estimate of comment de-anonymization during Weibo experimented with the function only on Russia-Ukraine Posts. The outcomes are the four comment features used in the RD regressions. The estimation sample comprises all the Russia-Ukraine posts and international politics posts that appeared in our data between January 1 and March 4, 2022.  $Treat^{RU}$  indicates if the post includes the keyword "Russia" or "Ukraine." Post indicates if the post was written on or after March 4, 2022.  $Treat^{RU} \times Post$  captures the DID estimate. The regression includes the set of controls as well as the day-of-week fixed effect. Heteroskedasticity robust standard errors are reported below point estimates.

Table 1.10 Placebo Test Using Treated Russia-Ukraine Posts

	(1) Non-mainland		(2) Endorsement		(3) Anti-Western		(4) Negativity	
	profile	comment	profile	comment	profile	comment	profile	comment
Reveal	0.0147 (0.00957)	0.0213** (0.00996)	0.0250 (0.0382)	0.00156 (0.0379)	-0.0293 (0.0723)	0.0387 (0.0665)	-0.0213 (0.0811)	-0.0158 (0.175)
Observations	27,669	20,044	27,667	20,042	2,346	1,631	1,443	287
Control Mean	0.0466	0.0385	-	-	0.0443	0.0925	0.3063	0.2900
Control SD	-	-	184.9	49.7	-	-	-	-

Notes: This table presents the effects of policy on all four key outcomes. The estimation sample is the Russia-Ukraine posts with the keyword "Russia" or "Ukraine" that triggered the comment de-anonymization on March 4, 2022. The table shows the estimation results from the main regression discontinuity model using the 30-day bandwidth. The Control Mean shows the average of the outcome variable within the bandwidth days before the policy implementation. The regression includes the set of controls as well as the day-of-week fixed effect. Heteroskedasticity robust standard errors are reported below point estimates.

## CHAPTER 2

### INFORMAL RISK SHARING TO MITIGATE LOCAL ENVIRONMENTAL RISKS

#### 2.1 Introduction

Informal risk-sharing within social networks is observed in low-income countries where public goods, institutional protection, and formal insurance markets are often inadequate. Households secure consumption by helping each other during income, health, or environmental shocks. Classic models predict, for instance, that pooling risks by a whole community would lead to a more efficient level of risk-sharing than pooling by small subgroups. However, the degree to which the risk is being shared and mitigated informally in practice is usually lower than the most efficient level of risk-sharing (Townsend, 1994; Udry, 1994; Fafchamps and Lund, 2003). This observation motivates questions on how informal risk-sharing evolves and whether it can be expanded through simple interventions.<sup>1</sup>

Our primary motivation is understanding how informal risk sharing bonds form and are sustained in a community. First, we study the role of commitments in increasing informal risk sharing and eventual risk mitigation. Commitments here refer to explicit agreements (commitments) households make to share resources before realization of an idiosyncratic environmental risk. On the one hand, commitments may increase the likelihood of cooperating to reduce risks. On the other hand, people may be reluctant to make explicit commitments when they are concerned that explicit commitments may increase reprisals for deviations. Second, we explore how such commitments are sustained in communities. Specifically, we evaluate the role of peer monitoring within communities in following up on ex-ante commitments and mitigating risk.

In our setting of rural Bangladesh, households consuming groundwater face the idiosyncratic risk of arsenic poisoning by drinking water from their household well. A considerable proportion of these wells are contaminated with naturally occurring arsenic (Figure. 2.6). Chronic exposure

---

<sup>1</sup>Limited commitment and hidden income are often cited in the economics theory literature as critical factors that restrain the informal risk-sharing (Coate and Ravallion, 1993; Kocherlakota, 1996; Ligon, 1998; Ligon, Thomas, and Worrall, 2002)

to arsenic by drinking from contaminated well water damages human health early (spontaneous abortions, still-births, increased infant mortality, diminished intellectual function) and later in life (cardiovascular disease, cancers of the lung and liver), reduces labor productivity, and impedes the accumulation of human capital (Carson et al., 2011; Abdul et al., 2015; Pitt et al., 2021). However, given that arsenic is a colorless and odorless contaminant, a vast majority of households in our sample did not know their well's arsenic status before our intervention. Therefore, testing a well for arsenic and sharing the result with users of the well naturally creates consumption shocks in these households.

One way for households to reduce arsenic exposure and mitigate health risks is to 'switch' to nearby low-arsenic wells (van Geen et al., 2002). Arsenic contamination in groundwater aquifers varies at short distances even within a village. This means an affected household can potentially use safe water from nearby wells. Prior studies indeed show that households switch to nearby wells in response to arsenic testing of their wells. However, the switching level varies widely from village to village – from 20% to 70% (Madajewicz et al., 2007; Barnwal et al., 2017), Tarozzi et al., 2021, and apparently not only because of differences in the distance to a safe well (Pfaff et al., 2017). This suggests other barriers to reducing arsenic exposure and therefore insufficient risk-sharing.

We implement two interventions through a randomized controlled trial in 135 rural communities in Bangladesh (Figure 2.2). By 'community' we refer here either to a whole village or to a village subdivision known as a para when the village is too large. In the control group of 36 communities, we sent generic voice call and text messages regarding the arsenic contamination, test wells for arsenic and inform households of the result. We expect some informal risk-sharing to emerge naturally from implicit commitments through existing social ties. Specifically, some safe-well-owning households may voluntarily share their tubewells with neighbors who own poisonous wells. The sharing and switching level in the control group communities provides us with the business-as-usual counterfactual scenario.

In contrast to implicit commitments, our first intervention is to facilitate the formation of ex-ante mutual commitments between households. This intervention is *ex-ante* because the participating

households commit to sharing risk *before* arsenic tests are conducted for their wells. We implement the intervention using a water sharing commitment contract that we call Water Sharing Coupons<sup>2</sup>. The pre-printed coupon states that when two well-owning households exchange coupons with each other before testing, and one of them turns out to have a high arsenic well, the household owning the safe well will allow the household with the unsafe well to collect water from the safe well for drinking and cooking. We distribute these coupons among well-owning households in 99 communities and ask households to exchange them with neighbors of their choice.

In a subset of 33 village communities (out of 99 communities with coupons), we implement our second intervention by enhancing peer monitoring. This entails making coupon exchange information public within the community, but in a limited way only. We ask for households' consent for this future revelation in advance. To further isolate the effect of selection into commitment contracts due to the consent for peer monitoring, we ask for consent in an additional 33 communities (out of 99 communities with coupons) but do not implement peer monitoring in these communities. Figure. 2.3 shows the experiment design.

In 99 villages where we intervened in the ex-ante commitment through distributing coupons, households exchanged 5.7 coupons with their neighbors. A set of dyad analyses reveals the households' strategies for forming risk-sharing. We find socioeconomic status negatively correlated with the number of coupons exchanged by the household. Previous social interaction, close geographic distance, and similarity in socioeconomic status and social preferences predict household coupon exchange. On the other hand, the experimentally manipulated enforcement environment, namely peer monitoring, marginally reduces the household's willingness to exchange coupons by 5.5%. Specifically, the notification of peer monitoring exacerbates the positive assortative matching in geographic distance and socioeconomic status.

We timed our intervention to take place before the government's own on-going blanket testing of 8 million wells for arsenic across the country. This Arsenic Risk Reduction Project (ARRP) is implemented by the Department of Public Health Engineering of Bangladesh. The arsenic in

---

<sup>2</sup>Will be short-termed as "coupons" in the rest of the paper.

South Asian groundwater is of natural origin (Fendorf et al., 2010). Although concentrations of arsenic are generally stable over time, concentrations can vary even at short neighbor-to-neighbor distances (van Geen et al. 2002). These tests are not regularly available to households. The last blanket testing of wells for arsenic was conducted by the government about two decades ago. Due to replacement and new installations, almost all of the wells in rural Bangladesh were untested before ARRP. Therefore, the testing program naturally created a welfare shock to households by revealing the tubewell's arsenic status, which they typically did not know. During the testing, a team of NGO workers tested each well using a field kit for arsenic in water (Pfaff et al., 2017). These tests provide information on arsenic in three ways – verbal communication to the well owner, painting the well spout red (more than  $50\mu\text{g}/\text{L}$  arsenic) or green (less than or equal to  $50\mu\text{g}/\text{L}$  – the Bangladesh national standard for arsenic in drinking water), and a printed result card (Figure 2.6).

We collect data in multiple baseline surveys. We first map all wells in the study community to the household that owns them at baseline. We also collect social network information from all households in our sample. In our study communities, there are about 120 households per community on average. Out of 16,054 households, 11,975 reported complete or partial ownership of at least one tubewell. We collect households responses to questions measuring altruism, reciprocity, and several other critical behavioral traits using the toolkit introduced by Falk et al. (2016, 2018). We also collect data on norms about well-sharing in the community. In the endline survey, we collect data on the household's current water source to construct a binary variable indicating whether the household is using a high or low arsenic well. In a sample of households, we also conduct audit testing by asking for a glass of water and testing that water.

We evaluate our intervention on several outcomes. First, we evaluate the reduction in exposure from switching and sharing wells in the different treatment arms. Second, we evaluate the formation of mutual risk-sharing agreements and we measure the impact of these agreements on well switching. Specifically, we estimate the impact of notification for peer-monitoring on the likelihood of formation of agreements between two households in the same community.



Through a set of RCT regressions that we pre-registered at the Journal of Development Economics, we find coupon exchange reduces the well-owning households' Endline consumption of arsenic by 8.7%. Unexpectedly, the mitigations in arsenic consumption among non-well-owners are compellingly two times larger. On the other hand, we find implementing peer monitoring vastly attenuates the positive effects of coupon-facilitated commitment. We hypothesize that the stronger monitoring practice crowds out safe well-owners' intentions to share.

In general, our study provides experimental evidence on a relatively less-studied topic – how to encourage informal risk-sharing among households facing idiosyncratic risks.<sup>3</sup> Our contribution here is to evaluate whether facilitation of the ex-ante commitments can lead to more informal risk-sharing. Specifically, we contribute to three strands of literature.

First, we contribute to understanding limited commitment in informal risk-sharing within communities. In previous studies, commitment devices and contracts help improve savings (Ashraf et al., 2006), quit addictive products (Gine et al., 2010), and increase labor productivity (Kaur et al., 2015). At the same time, other papers (e.g., Kinnan 2022) reject the role of limited commitment for low risk-sharing between households. We extend the knowledge about whether simple commitment contracts can overcome the commitment problems that hinder informal risk-sharing.

Second, we provide empirical evidence on how peer monitoring shapes risk-sharing between individuals in the absence of institutions. Classic studies emphasized peer monitoring's crucial role in the (micro) credit market (Stiglitz, 1990; Hermes et al., 2005). More recently, the impact of peer monitoring on individual financial decisions has been studied (Breza and Chandrasekhar, 2019). We contribute to this literature by testing the effectiveness of peer monitoring on informal risk-sharing and studying the channels through which peer monitoring operates.

Third, we contribute to the risk-sharing literature in two separate ways. One, our dyad-level data allows us to experimentally trace how risk-sharing networks formed under a manipulated contract enforcing environment. The two most related papers provide empirical evidence from

---

<sup>3</sup>Feigenberg et al. (2013) provide experimental evidence on economic returns of increasing social interaction. Meghir et al. (2022) study the effect of migration subsidies on informal risk-sharing within villages using a field experiment.

framed lab-in-field experiments (Barr and Genicot, 2008; Attanasio et al., 2012). Two, we add to the literature on assessing the role of social networks and preferences in forming mutual insurance (Fafchamps and Gubert, 2007; Attanasio et al., 2012).

## **2.2 Research Design**

We implement an intervention to answer the proposed research questions and evaluate them using a clustered-randomization trial in rural Bangladesh. Our first treatment is on increasing commitment before wells are tested using coupons. A successive second treatment aims to enhance peer monitoring by making commitment information public in a limited way.

### **2.2.1 Sample selection**

The experiment takes place in a central region of Bangladesh affected by arsenic. We selected a subset of villages with an intermediate proportion of wells contaminated by arsenic relative to the national standard of  $50\mu\text{g}/\text{L}$  (or parts per billion - ppb) based on blanket testing under Bangladesh Arsenic Mitigation and Water Supply Program (BAMWSP) conducted in 2000-05 (Fig. 2.6). Most households no longer knew the status of their well by the time of the baseline survey conducted in January 2020 as 5-10% of these wells are replaced each year (van Geen et al., 2014). Whereas BAMWSP testing and various interventions reduced arsenic exposure throughout the country, there are reasons to believe that level of exposure to arsenic was still significant in the study region (Jamil et al., 2019).

We selected five contiguous Upazilas (sub-districts) based on BAMWSP data and geological similarity to our previous studies conducted in Arai hazar Upazila. We first shortlisted villages by excluding villages with less than 20% of wells elevated in arsenic. We then collected central GPS coordinates for these villages to determine in Google Earth which villages were not too large or too small for blanket testing. We eventually narrowed down the sample to 135 villages with 25-95% of wells high in arsenic according to BAMWSP (Fig.2.2). Some of the villages in this sample were still too large to survey entirely, and we conducted additional field work to delineate individual paras, which are informally demarcated geographical partitions of a village. We then randomly chose one para per village in case of these large villages. Our final sample of 135 village

communities includes 103 full villages and 32 village paras.

The first survey entailed a complete census of 16,054 households, defined as immediate relatives sharing the kitchen, and mapping and tagging a total of 11,154 wells with a unique ID embossed on a stainless steel tag (Fig. 2.4). Overall, we identified a subset of 11,975 households who fully or partly owned at least one well in the survey. While most households (55%) own exactly one well, about 4% households, own more than one. Table 2.7 and Table 2.8 summarize the household demographics, primary well characteristics, and household asset.

Well ownership can be complicated sometimes. When multiple families claim one specific well as their own (e.g., people who contributed towards digging the well, or, siblings who inherited the well), we determine the main ownership after discussing it with all stakeholders.

A significant number of households use more than one well. In the census, we asked households to identify the well they primarily use for drinking water. We call them households' "primary well". We also asked about other wells households use for drinking, and call them secondary wells.

A well is defined safe if the test kit shows an arsenic concentration of  $50\mu\text{g}/\text{L}$  or less. For all higher values, the well is defined as unsafe.

## **2.2.2 Interventions**

### **Commitments to share water**

During our first intervention, we asked households to exchange coupons as a mutual commitment contract. The exchange of coupon between two households indicates the agreement that households will share safe wells water, regardless of the outcome of the arsenic test. In most cases, we distributed 10 coupons to 6,979 well-owning households and asked them to exchange them with each other forming a mutual insurance pair.

A coupon is illustrated in Fig. 2.5. It states that when two well-owners exchange their coupons, the safe well owner is expected to share the well water with the unsafe well owner after the testing for arsenic. Following the statement are four lines: the first prints the coupon owner's well tag number, the second the coupon owner's name, the third, has the name of risk-sharing partners chosen by mutual agreement printed, and the indicates the date of the coupon exchange.

We provided 10 coupons to each household owning a single well. Ideally, households should be allowed to exchange as many coupons as they like, but this would have become logistically difficult. Our pilot work and calculations considering the arsenic risk indicated that ten coupons per well would adequately insure households in most areas against arsenic exposure. In practice, since several types of well-ownership were identified in the baseline survey, we slightly modified the number of coupons distributed according to self-reported well ownership. For those wells that claimed sole ownership, the owners received ten coupons for them to share. For the jointly-owned wells, each owner received five coupons. The number of coupons each well-owned household receives is based on the number of wells claimed by the household. For example, if a household claimed primary ownership of one well and joint ownership of another, then the household would receive 15 coupons.

The field team gave households three days to exchange coupons with other households in their villages. On the first day, the field agents distributed coupons and encouraged well-owning households to exchange them with other well-owning households. Three days later, field agents returned to the village and recorded the tubewell and household IDs from the coupons the households received. Field agents checked for consistency by determining if the number of received and remaining coupon of the household still summed to ten for each well (or five for joint ownership). Households kept all exchanged or not-exchanged coupons as a record of the agreements that were made.

### **Peer-Monitoring**

The peer-monitoring intervention sends information about coupon exchanges for a given household to at most two other households (‘monitors’) in the same village. We hypothesize that these monitors may help increase the extent to which previous commitments are respected. The monitors are randomly chosen from the first- and second-order neighbors in the coupon network of a given household.<sup>4</sup>

---

<sup>4</sup>coupon networks are networks generated by exchanging coupons by households. A given household A’s Nth-order-(coupon)-neighbors are all the others in the coupon network connected to household A with a length-N shortest path. For example, the first-order neighbors are the households that exchanged coupons with household A. The second-order neighbors are the households that exchanged coupons with household A’s first-order neighbors but not with

Households in the 66 study villages that treated with peer-monitoring or peer-monitoring notification were invited for participation before the coupon distribution. Table 2.3 describes the contents of consent received by households lived in different treatment arms.

The monitoring information was sent by text messages. Specifically, all well-owning households in the 135 study villages received a generic automated voice phone message describing the impact of arsenic on human health. In 33 villages assigned to the peer-monitoring treatment group, households received two additional types of customized text messages: The first type of message we call a “monitor message” was sent to randomly selected monitors of a named household. The message includes the total number of coupon exchanges a named household head made along with the names of at most two other household heads who exchanged coupons with the named household.<sup>5</sup>

The second type of message we called a “receipt message” was sent to the named household itself. The receipt message informed these named households of who their monitors were and what information was sent to these monitors.

Table 2.4 describes the text messages and the voice call we sent to households. Whereas each household only received one receipt message, it could simultaneously be a monitor to several other households. For example, if household A is the center node of a star-style coupon network, where all the other households in the network only exchange with household A, household A becomes the monitor of all these other households.

### **2.2.3 Experimental Design**

Our interventions follow a clustered-randomization design (Figure. 2.3 and Table 2.2). We randomly assigned 135 villages to the following three treatment arms and one control arm. In all 135 villages, the study team informed people about the adverse health impact of arsenic and recommended households switch to low arsenic wells nearby if the well test shows high arsenic in

---

household A itself. If household A exchanged at least one coupons and had more than one second-order neighbor, one of the first-order neighbors and one of the second-order neighbors were randomly selected as household A’s monitor. If household A has at least one first-order neighbor and no second-order neighbor, one of its first-order neighbors was randomly selected as the monitor.

<sup>5</sup>The message would include two names if the household exchanged coupons with at least two other households. If the named household only exchanged a coupon with one other household, the message includes one name only.

their primary well. All household wells in 135 villages were tested for arsenic.

Control (C): 36 villages were randomly assigned to the control group, where we did not implement any intervention. We will record well switching and sharing at the same time as in treatment villages to measure the magnitude of business-as-usual risk-sharing.

coupons (T): 99 village villages were randomly selected to the coupon group. We distributed coupons to well-owning households and asked them to exchange them with other households as a form of commitment to share wells.

From these 99 villages, we further assigned villages to the second set of treatment groups.

coupons only (T1): In 33 coupon villages, no other treatment was provided.

coupons + Peer Monitoring Notification (T2): In this group of 33 coupon villages, we notified well-owning households that a peer-monitoring program may be implemented in the future that would make their coupons exchanges public in a limited way. This group only received the notification, but not the actual peer monitoring treatment.

coupons + Peer Monitoring Notification+Peer Monitoring (T3): In the remaining 33 coupon villages, we first notified well-owning households about the peer-monitoring program and then implemented it once wells were tested.

## **2.2.4 Survey**

We collected data in multiple rounds of surveys in 2020-22.

### **Household census and well listing**

Between January 15 and February 3, 2020, enumerators were able to reach 16,054 households in 135 villages in a door-to-door campaign, accounting for 92% of the total of 17,538 households identified in these villages. Following consent, the name, age, gender, and relationship of each household member, as well as the GPS coordinates of each house, were recorded electronically. Up to two mobile phone numbers from all consenting households were also recorded. Most households were subsequently recontacted using one of these numbers, or in some cases, alternative numbers provided by their neighbors.

During the same household and well survey, enumerators attached one metal well tag to each

well owned or partially by that household. Each stainless steel well tag was embossed with a unique number, and that number was recorded by the enumerator, along with a photo of the well and another of the mounted tag. After mounting a tag, the enumerator asked questions regarding well ownership. For the 10,098 wells privately owned by a single household, the well tag number was linked to that household. For the 1,056 wells jointly owned by multiple households, the well tag number was linked to these households. A subset of 4,077 household did not report owning a well and used the well owned by a neighbor. We collected data for a total of 11,459 wells and, among these, track the subset of 9,771 wells identified as primary wells throughout our experiment <sup>6</sup>.

### **Baseline Surveys**

We conducted two subsequent household surveys over the phone to comply with the government lockdown order and out of concern for the safety of enumerators and participants. We recorded a detailed set of household-level demographics, including deaths, illness, migration, health, asset ownership, well information, risk preferences, and social contacts within the community. As summarized in Table 2.6, we also mapped the social networks of each community based on the social contacts we collected from each of the households. These questions are slightly modified from the social network questionnaire used by Banerjee et al. (2013), from which we record that on average, each household has 4.1 close social contacts. Finally, we addressed a series of COVID-19-related questions including local COVID exposure, knowledge of infection, adoption of preventive measures, and economic impact to 20% of sampled households in the first round and 28% of sampled households in the second round.

During the first phone survey conducted from May 8 to June 7, 2020, a total of 14,551 (91%) of the households surveyed in person in January 2020 could be contacted and consented to respond. During the second phone survey conducted between October 27 and December 14, 2020, 11,933 (74%) households responded and consented. The repeated household census and COVID-related questions led us to show that there was no detectable increase in COVID-related mortality in our

---

<sup>6</sup>The survey response from households suggests 1,383 wells were a secondary well, and 305 wells were installed by the government or NGOs in the study villages.

135 study communities in 2020, although there was a very significant economic impact (Barnwal et al., 2021).

### **Coupons distribution and information recording**

The first round of coupon distributions and recording took place in 44 villages between March 1 and April 3, 2020, the second round in another 52 village communities between May 25 and June 26, 2021. The third round in the last 3 village communities took place in between August 21 and August 28, 2021. Enumerators entered the targeted villages with well maps drawn using the GPS data collected from the listing survey. Upon reaching a well, the enumerator found the owner(s) of the well and invited them to the coupon exchange program. Three days after the enumerator's visit for coupon distribution, the enumerator returned to these villages to record the coupon exchange information and ask a few more questions regarding coupon sharing. In 94 out of 135 villages, water samples were collected from a random subset of 400 wells for laboratory testing and comparison with the kit results after recording the coupon exchange information.

In the coupon distribution, we first recorded how participants perceived the extent of arsenic contamination within the neighborhood. Second, we asked households for information about their well including its respect to arsenic, well depth, age, and location (locked inside the living place or exposed to the public). Third, we elicited households' social preferences including altruism, trust, reciprocity, and risk preference using the survey toolkit developed by Falk et al. (2016, 2018). We then elicited households' willingness to switch to and share water sources through a set of willingness-to-pay and privacy concern questions. In the end, we elicited households' perceptions of social norms regarding sharing and compliance to publicly accepted actions.

In the coupon information recording survey, we first recorded each participating household's coupon exchange information. Photos of exchanged coupons were taken for the record and back-checking if needed. We then asked participants to list the primary reason for each coupon exchange, the chance of the exchanging well containing arsenic, and to what extent the household knows its risk-sharing partner's exchange. Further, we asked norm questions regarding complying or reneging contracts. We successfully collected the coupon exchange information from 7,216 wells in the 99



treated villages. On average, each household exchanged 5.56 coupons, with 9.9% of households did not exchange while 17.3% of households exchanged all 10 coupons.

New wells continue to be installed in these villages for greater convenience or to replace old wells that no longer function (van Geen et al., 2014). Enumerators also encountered undocumented wells that were missed from the listing survey. We managed both these situations by attaching new well tags to the handpumps. We repeated the asset questions to a small but random set of households to check the quality of the phone call surveys.

### **Well testing**

All functioning wells in the 135 study village communities were tested with a colorimetric field kit for arsenic. The result was reported orally to the household on the spot and by leaving a card with the household showing both the categorical (safe/unsafe) outcome and the arsenic concentration that was measured (Fig. 2.6).

Due to long delays in government testing in a subset of our study villages, our implementing partner NGO Forum carried out the well test. In the remaining villages, our enumerators shadowed the government well testing team to collect the data. This also means that two different types of kits were used for testing. In 95 villages, the established ITS EconoQuick kit (George et al., 2012) was used by field staff trained and hired by NGO Forum. In the remaining 40 villages, government testers were shadowed and the results were recorded electronically by NGO Forum staff in order to be able to access the data as in the other villages. In these 40 villages, the testing was conducted with the Macherey-Nagel QuantoFix kit, which is more recent but relies on the same Gutzeit reaction, by staff hired by the local government. The EconoQuick kit testing comes with a visual calibration scale discrete bins at 0, 10, 25, 50, 100, 200, 300, 500, and 1000  $\mu\text{g}/\text{L}$ . The calibration scale of QuantoFix kit is expressed in milligram per liter (or parts per million - ppm) in bins similar to those of the EconoQuick kit but with one additional bin at the low end of the spectrum: 0.000, 0.005, 0.010, 0.025, 0.050, 0.100, 0.250, 0.500  $\text{mg}/\text{L}$ .

Well testing was conducted in 108 villages from December 18, 2021 to February 15, 2022. In the remaining 27 villages, testing was delayed and extended fitfully between May 22, 2022 and July

3, 2022 because of a shortage in the government's supply of kits.

In all study villages, our well testing team wrote the the kit reading on the results card along with a reminder that the national standard is  $50 \mu\text{g}/\text{L}$  (Fig. 2.6). In addition, the spout of each tested hand pump was painted green or red, depending on whether the test result was  $\leq 50$  or  $\geq 100\mu\text{g}/\text{L}$ , respectively. We expect in a subset of 95 villages, our testing campaign will be followed by the government's testing and potential repainting of the spout of the well according to their new test result.

The number tested was somewhat lower than the original number wells surveyed because of disrepair or because wells could not be found again. As summarized in Table 2.5, we find that about 40% of the 9,839 tested wells contains less than  $50\mu\text{g}/\text{L}$  arsenic, which is the Bangladesh national standard for drinking water. For consistency with government policy, we use this definition for safe household wells even though the World Health Organization guideline for arsenic is  $10\mu\text{g}/\text{L}$ .

### **Endline Survey**

After the well-testing and text message interventions are completed, the experiment has been concluded with the Endline survey conducted during September to December of 2022. During the survey, we elicit the switching and sharing status for each household, well-owning households and households that do not own a well. In addition to questions about well switching and sharing, we ask each household the reasons for switching or sharing and the frequency of switching or sharing. We also elicit each household's perceptions of the local severity of the arsenic problem. These perceptions include some direct questions such as the average arsenic status in the community and the arsenic status of their coupon cosigners, and some indirect questions such as their willingness to pay for new wells and the local depth below which a new well would have to be installed to likely be low in arsenic <sup>7</sup>.

Following the switching and sharing questions, we will repeat the social preferences and norms elicitation conducted during the coupon distribution and information collection phase. Further, we

---

<sup>7</sup>Due to geological factors, there is within given villages often a well-defined depth below which groundwater is likely to be low in arsenic. This depth can vary considerably between neighboring villages, however (Gelman et al., 2004)

will test each participating household's knowledge of the extent to which other households switch or share wells with increasing network distance to document information dissemination within different treatment groups.

### **2.2.5 Randomization and power calculation**

#### **Randomization**

The treatment arms are randomized through a Stata 16 built-in command, *splitsample*, which randomly allocate a treatment to each community. We stratified villages within upazillas (sub-districts).

In Table 2.9 and Table 2.10, we summarize the balance across different treatment arms. In the first 12 columns, we show the number of observations, mean, and standard deviation of relevant household characteristics of four treatment arms. Then, we compare the balances across different treatment arms. These comparisons include the balance between each treated villages with the control, the balance between treated villages, and the balance between any treated villages with the control. The vast majority of the comparisons show small differences, suggesting that the interventions are randomly distributed in the sample.

#### **Power calculation**

The ratio of unsafe owners eventually switching to safe wells, or the switching rate, is the outcome of interest. This ratio not only depends on the actual ratio of safe wells in the community but also on the number of coupons the villagers exchange with each other. We predict the switching rate based on a stylized model by fixing the number of coupons villagers on average exchange and how likely they eventually switch when there are any available safe wells. We show that the switching rate will be high enough for our sample to achieve a conventional level of power ( $\beta = 0.8$ ).

We assume that every community has the same average number of 80 well-owning households. Every household has one well and exchanges the same number of coupons. The probability of having an unsafe well is  $p$ , which is commonly known by the households. However, households do not have private information about the arsenic status of their own well or that of other wells. We set the baseline switching rate to 0.28 based on Barnwal et al. (2017). The sharing made

by exchanging coupons is fully enforced so that everyone follows the coupon exchange result and testing outcome.

For example, suppose that the targeting community has 80 well-owning households, and the probability, unobserved to the villager, of having a safe well is 0.3. If we assume each household exchanges 5 coupons, the probability after testing of accessing no safe well is  $0.7^5 \approx 0.17$ . Therefore, 83% of households will be able to access at least one safe well. Conservatively, we assume half of the unsafe well households will not be able to switch because the agreement breaks down for some reasons. The switching rate will then be around 40%, a 12 percentage point increase from baseline.

*Minimal number of villages* We calculate the minimal number of treatment villages to guarantee enough power ( $\beta = 0.8$ ) to reject the null at the significance level  $\alpha = 0.05$ . To reflect our design, we set the number of control clusters equal to number of treatment clusters with each cluster contains 80 participants.

Conservatively assuming that half of the risk-sharing agreement made by exchanging coupons will be executed, we calculate minimal number of treatment to detect the predicted treatment effects calculated from different combinations safe well probability, number of coupons exchanged, and intracluster correlations. In Fig. A7 to Fig. A10, we show the minimal number of treatment clusters with the probability of safe well ranging from 0.2 to 0.6, number of coupons exchanged ranging from 2 to 10, and the ICC ranging from 0.05 to 0.5. These figures show that the current number of treatment (99 villages) and control (36 villages), should guarantee enough power to detect the predicted treatment effects in majority of cases <sup>8</sup>.

### **2.3 Specifications**

In this section, we map each core research question to specifications and data obtained from the survey. We analyze our core research questions with the stages of the experiment. Therefore, we first study the formation of risk-sharing networks under the manipulated experimental environment. Second, we study whether the experiment affects the household's awareness and knowledge about

---

<sup>8</sup>Given ICC, the minimal number of the clusters needed is decreasing with the proportion of safe wells and the number of coupons exchanged.

arsenic contamination. Last, we study whether our interventions to the ex-ante commitment and peer monitoring mitigate the arsenic.

### **Peer-monitoring affects risk-sharing networks formation**

Does the formation of risk-sharing agreement depend on social enforcement? When a household knows that their compliance with ex-ante commitments will eventually be monitored by their peers, they may alter the strategy in deciding which households to choose for ex-ante commitments. Thus, the peer-monitoring treatment may alter the formation of risk-sharing networks. We hypothesize that this information – that commitments are going to be made partially public – may have a significant effect on a household’s willingness to commit and on decisions about whom to commit with.

We informed households in 66 villages about peer monitoring and obtained their consent right before coupon distribution (treatment groups T2 and T3), while 33 villages were provided coupons only (treatment T1). Thus, we can test our hypothesis by comparing coupon exchanges in peer-monitoring notification villages (66 villages in T2+T3) with the same in non-peer-monitoring-notification villages (33 villages in T1).

We present two specifications. In the first specification (Equation 2.1), we test whether the aggregate level of coupons exchanged in the communities is affected by peer-monitoring commitment. In the second specification (Equation 2.2), we estimate a richer household dyad-level model that allows us to control for household pair-specific factors.

On the aggregate level:

$$y_{iv} = \beta_0 + \beta_1 * \mathbb{1}[T2_v = 1 \text{ or } T3_v = 1] + X_{iv}\gamma + \xi_v + \epsilon_{iv}. \quad (2.1)$$

Outcome variable  $y_{iv}$  is the number of coupons exchanged by household  $i$  in village  $v$ .  $X_{iv}$  controls for household-level characteristics as in Equation 2.5. We include upazila-level FE and cluster standard errors at the village level. The variable  $\mathbb{1}[T2_v = 1 \text{ or } T3_v = 1]$  indicates whether the village receives peer-monitoring notification only (treatment T2) or notification as well as the peer-monitoring treatment (treatment T3).  $\beta_1$  captures the treatment effect of peer-monitoring

notification on several risk-sharing partners the household seeks. Hence  $\beta_1$  captures a aggregate-level change in the network.

*Sample* We use data from 99 coupon villages to estimate Eq. 2.1. By design, we do not have data on coupons from the 36 villages in the control group.

On the Household-dyad level:

$$y_{ijv} = \beta_0 + \beta_1 * \mathbb{1}[T2_v = 1 \text{ or } T3_v = 1] + \beta_2 G_{ijv} + |X_{iv} - X_{jv}| \gamma_1 + |X_{iv} + X_{jv}| \gamma_2 + \xi_v + \epsilon_{ijv}. \quad (2.2)$$

Outcome variable  $y_{ijv}$  in this model indicates whether two households,  $i$  and  $j$ , in village  $v$ , exchanged coupons. The components  $|X_{iv} - X_{jv}|$  and  $|X_{iv} + X_{jv}|$  captures the “difference” and “average characteristics” of household  $i$  and  $j$ . For example, if  $X$  is the wealth level, then the first component measures the wealth difference, and the second component measures the average wealth level. The corresponding  $\gamma$ s thus capture the magnitude of assortative matching and the level effect. Our coefficient of interest is  $\beta_1$ .

Extending this empirical specification, we will include an interaction of the difference component with the treatment indicator  $\mathbb{1}[T2_v = 1 \text{ or } T3_v = 1]$ . The corresponding estimated coefficients would indicate in which direction and through which channel peer-monitoring notification affects the formation of risk-sharing networks. If the estimated coefficients on the interaction terms are negative, it would suggest that when households expect stronger monitoring in the future, they become more homophilic when selecting their risk-sharing partners.

*Sample* We use all household pairs (dyads) from 99 coupon villages to estimate Equation 2.2.

### **Peer-monitoring increases awareness and knowledge**

We test whether peer monitoring increases the arsenic awareness and knowledge of local arsenic contamination.

In the Endline survey, we asked all households if they had discussed well-sharing with their neighbors before well-testing. If our intervention successfully increases the awareness of cooperation against arsenic, we expect a significant increase in confirmation from treated villages. We estimate the treatment effects using:

$$y_{iv} = \beta_0 + \beta_1 T1_v + \beta_2 T2_v + \beta_3 T3_v + X_{iv}\gamma + \eta_u + \epsilon_{iv} \quad (2.3)$$

In Equation 2.3,  $y_{iv}$  indicates whether the household self-reports the discussion, and  $\beta_1$  to  $\beta_3$  captures the treatment effects.

In the 99 coupon villages, the Endline survey elicits each participant's knowledge of some of their first-, second-, and third-degree coupon neighbors' (a) well arsenic and (b) switching/sharing. To the extent that peer monitoring facilitates information transmission, households in T3 should be better informed of neighbors' well status. The omitted group here is the village communities that were only treated with coupons (treatment groups T1).

We estimate:

$$y_{iv} = \beta_0 + \beta_1 T2_v + \beta_2 T3_v + X_{iv}\gamma + \eta_u + \epsilon_{iv} \quad (2.4)$$

Outcome variable  $y_{iv}$  now measures the household's knowledge about its neighbors' arsenic status. We construct a score for each household indicating whether they could correctly report the well status of their first-, second-, and third-order neighbors in the coupon networks. The parameter of interest  $\beta_2$  reflects the level of well and switching-related information disseminated in the local community in peer-monitoring villages (T3) when compared with non-peer-monitoring villages (T1). Should peer-monitoring indeed facilitate the transmission of information,  $\beta_2$  signs positive.

*Sample* We estimate this equation for all households exchanging coupons in the 99 coupon villages.

### **Ex-ante commitment and risk mitigation**

In the experiment, two households make an ex-ante commitment to share safe wells to reduce their arsenic risk by exchanging coupons. We postulate that coupons lead to more sharing of safe wells. On the one hand, committing to share risk explicitly before revealing the status of wells could possibly increase the cost of deviation afterwards. This partially resolves the limited commitment problem. On the other hand, coupons may not increase risk sharing if households only exchange them with other households with whom they would share even without the coupons. Hence, we

test the hypothesis with the specification:

$$y_{iv} = \beta_0 + \beta T_v + X_{iv}\gamma + \eta_u + \epsilon_{iv}. \quad (2.5)$$

Outcome variable  $y_{iv}$  indicates the consumption of safe water by household  $i$  in village  $v$ , as collected in the Endline survey. We define this variable in four different ways –

- (1) the arsenic content (continuous variable) of the primary well households use for drinking water
- (2) the status relative to the  $50\mu g/L$  standard of the primary well households use for drinking water
- (3) whether household  $i$  owning an unsafe well switched to a safe well (sample restricted to households owning high-arsenic wells)
- (4) the number of neighbors household  $i$  owning a safe well shared it with (sample - restricted to households owning low-arsenic wells).

We keep this distinction between sharing and switching (in #3 and #4 above) because one safe well owner can potentially share water with multiple unsafe well owners.

$T_v$  indicates the coupon treatment assignment i.e., whether village  $v$  receives coupon intervention. Hence  $\beta$  is the estimated treatment effect that compares outcome in treatment (99 communities) with the same in control (36 communities).  $\eta_u$  is the Upazila-level FE. We cluster standard errors at the village-level. *Covariates.* We include an array of household-level baseline covariates, denoted by  $X_{iv}$  above:

- (1) Wealth proxy: asset index
- (2) Demographic variables: Household members' primary education completion rate, household size, male ratio (number of males over size of household), child ratio (number of children over size of household)
- (3) Health risk preference (Falk et al., 2016, 2018)

*Sample.* We use data from all well-owning households in the whole 135 village communities to estimate Equation 2.5. The reason we restrict to well-owning households identified in the baseline is because only these households were provided with coupons to exchange. When the outcome variable  $y_{iv}$  represents switching by households owning an unsafe well, our sample includes only



households owning an unsafe well. When the outcome variable  $y_{iv}$  represents sharing by households owning a safe well, our sample includes only households owning a safe well.

### **Peer-monitoring facilitates risk-sharing**

We estimate the impact of induced peer-monitoring on the formation of risk-sharing networks as well as the eventual switching/sharing (treatment T3). Peer-monitoring may potentially strengthen cooperation by increasing the expected cost to a deviating household (i.e., the household that promised to share water ex-ante, but didn't do so ex-post). The monitoring households ('monitor') were randomly selected conditional on their connectedness to the monitored household in the coupon networks. Given that the peer-monitoring was only implemented in half of the 66 villages that were notified about the peer monitoring program, there are three treatment arms considered in this specification: 33 villages with only coupons (treatment T1), 33 villages with coupons and peer monitoring notification (treatment T2), and 33 villages with coupons and peer monitoring (treatment T3). Control villages are in the omitted group. We estimate the treatment effects of peer-monitoring using the following household-level regression specification (Equation 2.6):

$$y_{iv} = \beta_0 + \beta_1 T1_v + \beta_2 T2_v + \beta_3 T3_v + X_{iv}\gamma + \eta_u + \epsilon_{iv} \quad (2.6)$$

Similar to Equation 2.5, outcome variable  $y_{iv}$  is defined in four different ways,  $T1_v$  is the binary variable that indicates whether village  $v$  received the coupon intervention only,  $T2_v$  indicates that households in village  $v$  received the coupon intervention along with only the notification for peer-monitoring,  $T3_v$  indicates that households in the village  $v$  received coupons, notification and the peer-monitoring treatment.

$X_{iv}$  contains the household-level characteristics. The  $\beta$ s capture the treatment effect of each intervention.  $\beta_1$  would show the direct effect of exchange coupons on risk-mitigation.  $\beta_2$  would indicate the mean water sharing/switching in the coupons+Notification group.  $\beta_3$  captures the treatment effect of providing peer-monitoring treatment, notification, and coupons.

The primary insight of this hypothesis is that we use a two-stage design to separately identify the effect of anticipated monitoring and actual monitoring on risk mitigation. Here,  $\beta_2$  captures the effect of any change in strategy during coupon exchange, in response to anticipated monitoring.

Anticipating peer-monitoring in the future, households in T2 villages may seek a different set of neighbors when making water-sharing commitments. Since peer monitoring was eventually not implemented in T2 villages, any change in risk mitigation can be attributed to the effect of anticipated monitoring through altered risk-sharing network structure (also see Hypotheses H6). We test this hypothesis by comparing corresponding coefficient on T2 with T1, i.e.,  $\beta_2 - \beta_1 = 0$ .

Further, to test whether peer monitoring alone (i.e., net of any anticipation effect at coupon exchange stage) has a significant impact on water sharing or switching, we test whether  $\beta_3 - \beta_2 = 0$ . The net effect of peer monitoring, however, would include both anticipated and actual peer monitoring. To that end, we test whether  $\beta_3 - \beta_1 = 0$ .

Similar to the first hypothesis, we include Upazila FE and cluster standard errors at the village level. We will test for heterogeneous treatment effect in the same way as we specified for Equation 2.5.

*Sample* The sample we used to estimate Equation 2.6 is the same as the sample we used in Equation 2.5. When  $y_{iv}$  represents switching by a household owning an unsafe well, our sample includes only households owning an unsafe well. When  $y_{iv}$  represents sharing by households owning a safe well, our sample includes only households owning a safe well.

## 2.4 Experiment Findings

This section presents the main findings based on the specifications outlined previously. We will proceed through the stages of the experiment. First, we discuss the formation of ex-ante commitment, measured by coupon exchange behavior. Our findings show that the number of coupons exchanged is influenced by pre-existing household characteristics and the experimentally manipulated environment. Next, we demonstrate how assortative matching, measured by differences in dimensions such as household characteristics, affects coupon exchanges between households. Additionally, we provide evidence that the experimentally varied enforcement environment intensifies positive assortative matching.

Second, we examine how our experiment impacts participants' awareness and knowledge. We evaluate participants' awareness through a cooperation question in the endline survey. The results

show that the treatments significantly increased the inclination for well-sharing cooperation before arsenic testing. Beyond self-reported data, we present findings from a set of novel questions designed to assess households' knowledge about the arsenic condition of their coupon-networked neighbors' wells. Our results indicate that peer monitoring via SMS campaigns enhances knowledge of local arsenic contamination, thereby increasing overall awareness.

Finally, we present the major findings on arsenic mitigation. Our first set of analyses compares the arsenic consumption in the endline. We find that ex-ante commitment moderately reduced overall arsenic consumption. The effects varied with the addition of peer monitoring practices. Interestingly, the households that benefited most from the experiment were those without ownership of the primarily used well at baseline. Although these households were initially ineligible for the experiment due to their lack of ownership, they benefited substantially from the intervention, indicating a significant spillover effect.

Our secondary set of analyses compares the measured switching based on the change of primarily used wells. Among the four types of switching previously specified, we found that the treatments significantly reduced the switching to higher arsenic wells, suggesting that households are more likely to seek safer wells when they switch.

**Coupon number** The number of coupons households exchange provides an "intensive margin" understanding of our interventions. We first hypothesize that the characteristics of the households affect the coupon exchange. Figure 2.9 shows the association between household characteristics and the number of coupons being exchanged by regressing the coupon numbers on a detailed set of household characteristics obtained from the Baseline survey. On average, the participating households exchanged 5.7 coupons. We find that multiple factors negatively correlate with the number of coupons exchanged. Specifically, a 10% increase in the primary education ratio is associated with a reduction of 0.07 coupons exchanged. Additionally, a one standard deviation increase in wealth level, measured by the asset index, is associated with a reduction of 0.02 coupons exchanged. These negative correlations suggest that individuals of higher socioeconomic status may rely on other methods to mitigate risk, or they possess private information about the quality of their

wells that may inhibit them from sharing in the future. Conversely, self-arsenic belief, measured by whether the household believes their well contains poisonous levels of arsenic, positively correlates with the number of coupons exchanged. This suggests that people perceive coupon exchange as a means to mitigate arsenic risks.

Second, we test if the variation in the enforcement environment, manipulated by our intervention, can change the willingness to exchange coupons. We pool households in T2 and T3 together and compare their coupon exchange against households in T1. T2 and T3 households are pooled together in the analysis as they are unaware of the actual implementation of peer monitoring through the SMS campaign when they exchange coupons. Therefore, the difference in coupon exchanges can be attributed to the anticipation of higher enforcement due to peer monitoring. Figure 2.7 summarizes this comparison. On average, households in T1 villages exchanged 5.913 coupons, while households in T2+T3 villages exchanged 5.589 coupons, indicating that the anticipation of higher enforcement reduced coupon exchange by around 5.5%.

Figure 2.8 characterizes the probability distribution of the coupon exchange in these two groups. Panel A shows the proportion of coupons exchanged between zero and ten across two groups. We find households have a strong motivation to participate, with the distributions of both groups shifting to the right. More than 16% households from both groups exchanged all ten coupons, and only less than 9% of households from both groups quit by exchanging zero coupons. The red and blue dash lines correspond to the mean of coupon exchange in T1 and T2+T3 villages. The distribution of coupon numbers exchanged in T1 dominates T2 on the higher end. Panel B plots the cumulative function of coupon numbers from two groups and shows a near stochastic dominance of coupon numbers exchanged in T1.

**Pairwise coupon exchange** Previous comparisons suggest that households in T1 villages are more willing to exchange coupons than their counterparts from T2 and T3 villages. We explain this phenomenon by arguing that people anticipating stronger enforcement become more selective about whom they contract for risk-sharing today. To verify this hypothesis, we use the dyad regression specified in the previous section to conduct two analyses.

In the first analysis, we estimate a benchmark model by regressing the dummy variables for coupon exchange status between every pair of households from the same village on a rich set of dyad variables. These variables include dummies for social interaction, geographic distance, and the absolute differences and sums of multiple household characteristics. The coefficients of the latter two sets of variables measure the sorting and level effects.

In the second analysis, we estimate benchmark specifications with the interaction of T2+T3 treatment and household characteristics to test if a stronger enforcement environment affects sorting behavior. This analysis provides insight into the "extensive margin" of our intervention's impact on the risk-sharing formation.

Figure 2.10 shows the results from the benchmark dyad regressions which pooled all the sorting and level variables together. Among the 171,666 dyads included in the analysis, 7.54% exchanged coupons. The coefficients show the direction and magnitude of assortative matching. A negative coefficient determines a positive assortative matching with respect to a characteristic in that two households are more likely to exchange coupons when they present a similarity in that characteristic.

Panel A provides three important findings. First, the intention to exchange coupons or form risk-sharing decreases with distance; a 50-meter increase in the distance reduces the intention to exchange coupons by 2.7 percentage points. Second, consistent with previous comparisons, living in a T2+T3 village reduces this propensity by 1.5 percentage points. However, prior interactions significantly predict coupon exchange, as two households that reported prior interaction with each other had a 26.3 percentage point increase in the propensity to exchange coupons.

Panel B shows that matching in socioeconomic status continues to predict coupon exchange. Larger absolute differences in education levels and wealth reduce the propensity to exchange coupons. Specifically, a ten percentage point difference in education levels and a one standard deviation difference in wealth reduce the propensity by 0.245 percentage points and 0.266 percentage points, respectively. However, these associations are substantially weaker than the covariates presented in Panel A.

Interestingly, as shown in Panel C, we find that matching social preferences also predicts coupon

exchange, with magnitudes similar to those of socioeconomic status. These social preferences are measured using qualitative questions recommended by Falk et al. (2018), and we standardize the summation of responses to create the index. A one standard deviation difference in altruism, trust, and positive and negative reciprocity is associated with a 0.1 percentage point lower likelihood of exchanging coupons. One explanation for this could be that demographic characteristics and socioeconomic status are highly correlated with the social preferences of a household, leading to matches based on these implicit preference factors. Another explanation could be that households interact with like-minded and similarly valued neighbors.

Table 2.11 and Table A7 show the regression results of Equation 2.2 fully saturated with the peer monitoring notification treatment dummy. These tables report the estimates of differences and their interactions with the notification treatment dummy. These interaction terms measure changes in the matching pattern in notification villages T2 and T3 when compared to T1 villages, where we do not send the notification. Column (1) shows that living in T2+T3 villages substantially enhanced the geographic distance sorting by around 36.9% (0.00827/0.0224). On the other hand, living in these 66 villages does not affect the coupon exchange between households who have already interacted with each other. This may suggest that existing social relations are less influenced by the externally manipulated enforcement environment. While we do not obtain significant findings for most household characteristics dyad variables, we find stronger positive assortative matching in asset level happens in the notification villages when compared to the non-notification villages.

Column (2) reports the same estimation with additional two-way-individual fixed effects  $\xi_i$  and  $\xi_j$ . The addition of these sets of fixed effects eliminates the unobserved heterogeneity at the individual levels. We find the interactions of geographical distance and wealth level with the treatment dummy are also significant and similar to the estimates from Column (1). This result provides the robustness to the estimates obtained in Column (1).

What are the implications of stronger positive assortative matching in our context? First, stronger sorting in geographic distance reduces the effectiveness or risk-sharing when the arsenic contamination is spatially correlated. Risk-sharing with physically close neighbors becomes in-

effective when the naturally occurring contamination affects all the water sources in the small neighborhood. Second, stronger sorting exacerbates inequality. Households with higher socioeconomic status may be able to afford and adopt more advanced technology to mitigate arsenic risks, such as deep wells and filtering systems. However, these resources will only be shared among higher socioeconomic groups when people of similar socioeconomic status sort together.

### **Awareness and knowledge**

In this section, we discuss the experimental findings regarding the households' awareness and knowledge of arsenic contamination (in the neighborhood). The analyses in this section serve two purposes.

First, we show that our interventions affect households' awareness of arsenic. The increase in awareness is first reflected by the significant self-reported willingness to cooperate against arsenic contamination. We further show that the intervention increases awareness by showing households from T3 villages. In these villages, we implemented peer monitoring through the SMS campaign, and households scored higher in correctness when guessing the direct neighbors' well status.

Second, the positive findings in awareness and knowledge verify that we have correctly administered the intervention as initially designed. Therefore, any unexpected result should not be attributed to the mis-administration of the experiment.

Column (1) of Table 2.12 shows that while all three treatments substantially increased the well-owning household's self-reported discussion of well-sharing with neighbors before arsenic testing, the magnitudes of the increase do not vary across different treatments. 29% of households in the 36 control villages discussed well-sharing with their neighbors. T1, which solely forms the ex-ante commitments through coupon exchange, increases the discussion propensity by 49.7%. Respectively, by notifying and implementing peer monitoring, T2 and T3 achieved increases of 55.5% and 53.4%. However, these three magnitudes are not statistically different from each other.

Interestingly, as shown in column (2), we further discovered the treatment effect spillover to the non-well-owning households. Households categorized as non-well-owning households did not report at least partial ownership of the primary well used in the baseline. Therefore, we did

not distribute coupons to these households as they should not be granted the responsibility to "share" the well with neighbors. Meanwhile, we find these households in T1 villages reported a 21.6% (0.0604/0.28) higher propensity to discuss well-sharing than their counterparts from the control villages. Even though insignificant, the estimates from T2 and T3 are both positive and economically large, 16.8% (0.0470/0.28) and 12.4% (0.0346/0.28) respectively. The attenuation of the spillover effect in T2 and T3 may have resulted from peer monitoring further inhibiting the "leakage" of the program towards the non-well-owners, as they are not eligible for the program and may be less willing to expose themselves through the spread of the text message.

Our second test examines the households' knowledge regarding their coupon-networked neighbors. In the Endline survey, we asked randomly selected households from 99 villages treated with ex-ante commitment the well arsenic status of random first-, second-, and third-order coupon networked neighbors. Households answered either safe or poisonous, and we compared the household's answer with our testing data. We find households in T3 villages, where we spread the coupon exchange information through text messages, scores significantly higher for guessing the first-order neighbor's well.

Figure 2.11 shows our findings. Compared to the scores achieved by the households from T1 villages, where we only distributed the coupons, the red coefficients with confidence intervals connected by the red dash lines represent the additional scores achieved by the households in T2 villages, where we notified the peer monitoring but did not implement. Households from T2 did not score better in any of the three questions. However, households from T3, represented by the blue set of lines, showed a significantly higher knowledge about their first-order neighbor's well arsenic status than their counterparts from T1 and T2 villages. Specifically, 50.5% households from T1 correctly guessed the first-order neighbor's well status. The proportion is close to random guessing. Households in T2 perform worse but insignificantly by 3.1% (-0.0159/0.505). On the other hand, households in T3 perform significantly better by 12.3% (0.0619/0.505). We also observe a plausible reduction in knowledge of higher-order neighbor well status among T3 households.

This comparison implies that T3 households acquired a better knowledge of the local arsenic



status than T1 and T2 households due to the text message. While the text message did not contain any information about their neighbors' well status, it may incentivize T3 households to acquire such information by themselves, as they may have more complied with the intervention, which asked them to switch if their wells are unsafe and to share if their coupon-neighbors' wells are unsafe.

### **Main outcome 1: Arsenic mitigation**

The main policy interest of our intervention is to facilitate the ex-ante commitment to water-sharing before the arsenic testing to mitigate the arsenic consumption after the test. Unlucky households now get help from lucky neighbors who exchange coupons with them. Therefore, hypothetically, coupon exchange facilitates the ex-ante commitment, so the households' primary well used post-intervention should contain lower arsenic.

Table 2.13 summarizes the major findings <sup>9</sup>. Panel A shows the pooled reduction from coupon-facilitated ex-ante commitment. Columns (1) and (2) show a moderate reduction of arsenic consumption among the targetted households, who reported the ownership of the primary well in the Baseline survey. Column (1) shows that coupon exchange reduces the well-owning households' endline consumption of arsenic by 8.7%(-18.45/210.88), regardless of the arsenic status of the primary well they used in the Baseline. However, if we only consider the well-owning households whose baseline well contains poisonous levels of arsenic, the reduction is even less salient. We only find a reduction of 4.9% among these households, which is neither statistically significant nor economically meaningful.

We unexpectedly find the mitigations in arsenic consumption are compellingly larger among non-well-owners. Columns (3) and (4) of Panel A show that non-well owners switched to wells that are much safer than the previous wells. Column (3) shows that non-well-owners have their arsenic consumption reduced by 16.7%, and reduced by 11.5% had they used a high arsenic well in the baseline. On the other hand, as shown in columns (5) and (6), we fail to observe any reduction in arsenic among households who had already used a low arsenic well from the baseline. In general, the substantial reduction among non-well-owners suggests the treatment not only spillover

---

<sup>9</sup>We dropped observations without the full set of controls. For estimation without the controls, please refer to Table A9

to awareness but also the eventual reduction.

Panel B dissects the effects of each of the treatments. Columns (1) and (2) shows that both the effects of T1, only distributing the coupons, and T2, only notifying the peer monitoring, are weakly larger than the pooled effects. T1 and T2 reduced the overall arsenic consumption of well-owners by 11.4% (-23.98/210.88) and 13.7% (-28.94/210.88), respectively. They reduced the arsenic consumption of well-owners whose baseline wells were high in arsenic by 6.5% (-21.07/321.97), respectively.

Most surprisingly, T3, where we implement peer monitoring, vastly attenuates the pooled effects. We find that well-owning households from T3 villages consume the same level of arsenic-contaminated water as they did in the baseline. Such a finding contradicts the previous results of increasing awareness and knowledge in T3 villages, suggesting a strong negative force counteracts people's incentive to switch to safer wells, even being better aware of the arsenic contamination and local arsenic status.

Meanwhile, the negative force in T3 should not have affected the non-well-owning households. We find substantial reductions in arsenic consumption for non-well-owning households in all treatment arms. As shown in Column (3) of Panel B, T1 reduces the non-well owners' consumption of arsenic by 20.2% (-43.68/216.68). T2 and T3 reduce their consumption of arsenic by, a similar magnitude, 14.5% (-31.31/216.68) and 14.2% (-30.78/216.68). Column (4) shows a weakly smaller reduction in all three groups, with the reduction of T2 and T3 insignificant due to the lack of statistical power. None of the treatment effects have spilled over to households that used safe wells in the baseline.

Therefore, what drives the negative effect of T3 on well-owning households? We hypothesized that the stronger monitoring practice reduced the well-owners' intention to switch and share, while non-well-owners were not affected as they were excluded from the program. While we cannot find similar evidence in the health risk mitigation literature, there is a rich literature on regulation and monitoring that crowds out the socially beneficial behaviors as soon as these behaviors signal intrinsic motivations. For example, Cardenas, Stranlund, and Willis (2000) find environmental

regulation crowds out other-regarding behaviors. Dickinson and Villeval (2008) shows monitoring crowds out labor effort in the laboratory. When well sharing is perceived as other-regarding behaviors, safe well owners may become reluctant to share the wells in the village where the well sharing has been peer-monitored. Consequently, unsafe well owners may be discouraged from seeking safe wells in equilibrium.

This hypothesis is partly supported by columns (1) and (2) of Table A8, in which we only estimate the treatment effects among the switchers. While switchers may be endogenously motivated by the interventions, Columns (1) and (2) show that T3 substantially reduced the well-owning households' endline arsenic consumption had they switched. Overall, existing evidence suggests that the stronger enforcement environment unexpectedly crowds out the well switching and sharing, leading the peer monitoring to eliminate the positive treatment effect from coupon-facilitated ex-ante commitment fully.

## **Main outcome 2: Switching**

We approach arsenic mitigation through different dimensions in the second set of the main analyses. Instead of comparing the arsenic concentration of the well used by households in the endline, we directly investigate the well switching.

A household is considered to have the well switched if the primary well reported in the Endline survey differs from the one they used in the Baseline survey. We further categorize well-switching into four different categories. They are switching (1) from an unsafe well to a safe well, (2) to a well with lower arsenic, (3) from a safe well to an unsafe well, and (4) to a well with higher arsenic. Effective intervention will increase the first two types of switching but reduce the latter two types of switching. The first type of switching only considers the households that used unsafe wells in the baseline, and the third type only considers the households that used safe wells in the baseline.

Table 2.14 summarizes the findings <sup>10</sup>. Most of the specifications suffered from the power shortage as households that did not switch at all comprised the largest proportion of all four categories. We do find some evidence suggest that our interventions moderately improve the

---

<sup>10</sup>We dropped observations without the full set of controls. For estimation without the controls, please refer to Table A11

well-switching. Column (7) shows the treatment effects of the interventions on the households' propensity to switch to wells contaminated with higher arsenic. 2% of the households in the control group switched to wells with arsenic concentration higher than the wells they used primarily in the baseline. According to Panel A, coupon-facilitated ex-ante commitment overall reduces this harmful direction of switching by 52% (-0.0104/0.02).

Panel B shows the effects separately. All three interventions exhibit similar impacts on switching to higher arsenic wells. T1, T2, and T3 reduced such a switching by 56.5% (-0.0113/0.02), 44.2% (-0.0089/0.02), and 54.5% (-0.0109/0.02), respectively.

The low switching rate explains the moderate arsenic reduction discussed in the previous subsection. Columns (1) and (2) show that T3 negatively affects the switching from unsafe wells to safe wells and the switching to safer wells. These negative impacts are too small to be precisely estimated. On the other hand, they are economically meaningful, given the exceptionally low switching rate among the 36 control villages.

Overall, the measured switching provides us a different angle to validate the results that we established using the endline arsenic consumption.

## **2.5 Conclusion**

Arsenic contamination was first discovered in Bangladesh in 1993. Soon after its discovery, the government and global health organizations spent innumerable human efforts and funding to discover plausible solutions to mitigate arsenic risks. However, thirty years later, it is still estimated that over 50 million Bangladeshis drink water contaminated with a poisonous level of arsenic. This paper explores the effectiveness of a novel community-driven approach. This approach theoretically facilitates informal water risk-sharing through village networks by nudging villagers to form ex-ante commitments on well-sharing before the well testing. In addition to facilitating commitment, we further implement a peer monitoring program, as the traditional agency theory suggests that stronger monitoring leads to higher compliance with commitment.

Through a large-scale clustered randomized controlled trial, we discovered mixed evidence about facilitating ex-ante commitment and peer monitoring in water risk-sharing. While facilitating

ex-ante commitment moderately improves the risk-sharing, peer monitoring sabotage the gain through two channels. In the first channel, anticipating peer monitoring reduces people's willingness to share risks, possibly due to the concern of higher punishment for renegeing the commitment in the future. Such a concern exacerbates the positive assortative matching in the risk-sharing formation, which increases the vulnerability of risk-sharing against natural shocks. In the second channel, peer monitoring may reduce the other-regarding behavior that sustains the risk-sharing networks. The classic crowd-out theory explains that stricter regulations crowd out the individual's intrinsic motivation to benefit the neighborhood, reducing the critical risk-sharing tool in our context.

Our novel findings have important policy implications that explore the community-driven approach to mitigate health risks. In addition to encouraging other-regarding behaviors, regulations, such as monitoring, may disincentivize individuals to share risks and resources due to selection and motivation crowd out. The effectiveness of enforcement, such as peer monitoring, needs a more thorough examination, such as policy experiments. Overall, our findings highlight the complexity of facilitating risk-sharing, even in small villages, when confronting norms, informal institutions, and anticipated enforcement.

## 2.6 Figures

Figure 2.1 The testing result by Bangladesh Arsenic Mitigation and Water Supply Program (BAMWSP)

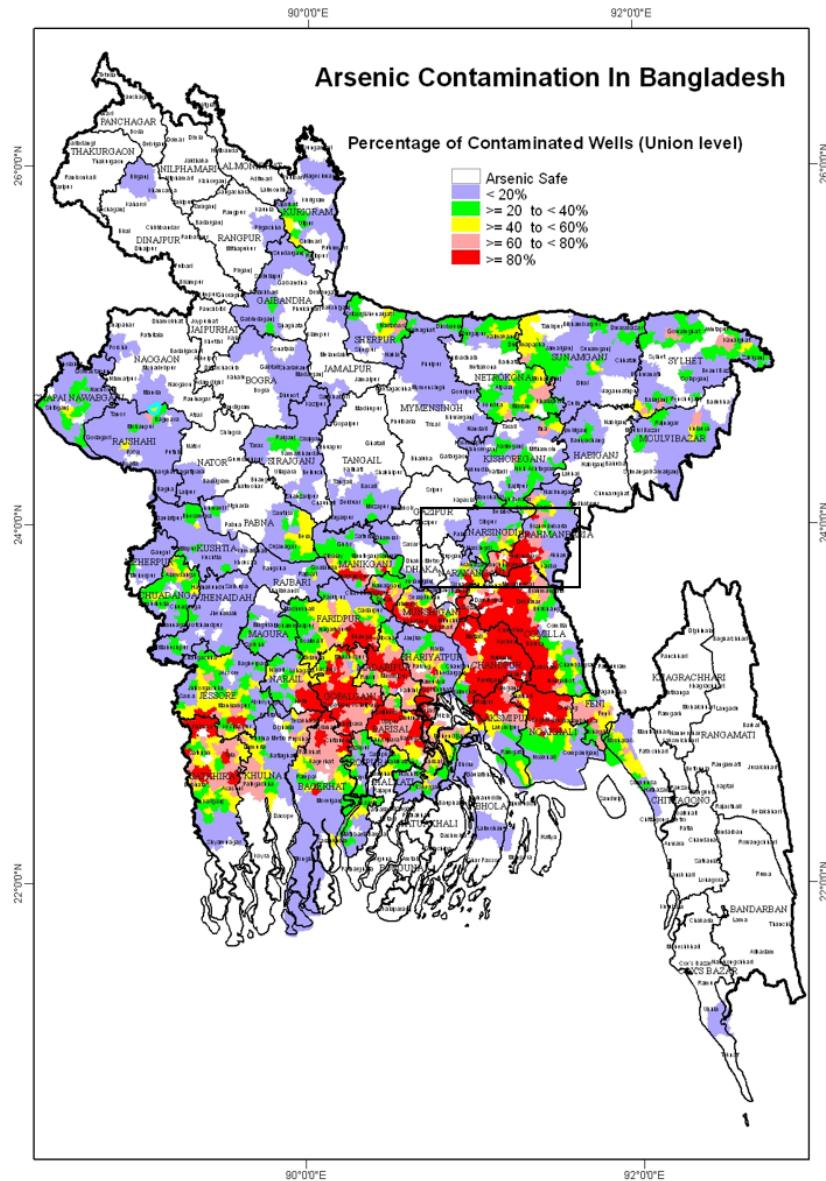


Figure 2.2 The map of 135 village communities

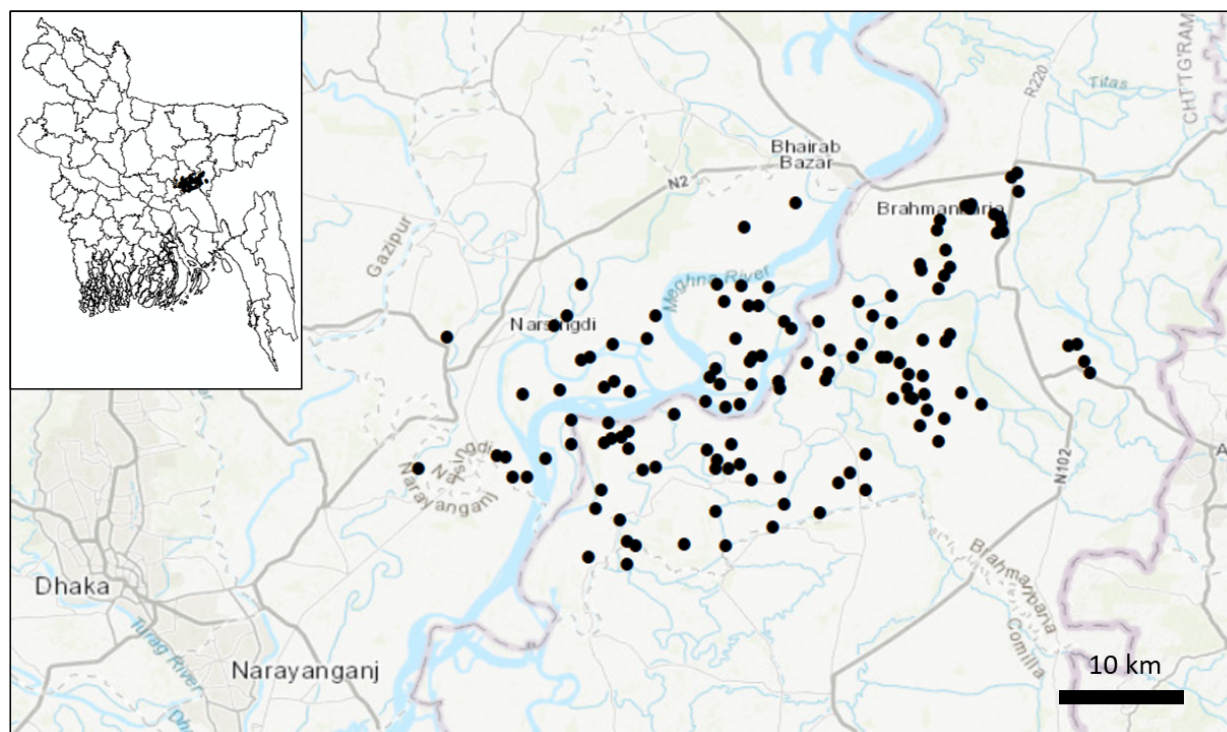


Figure 2.3 Treatment Arms

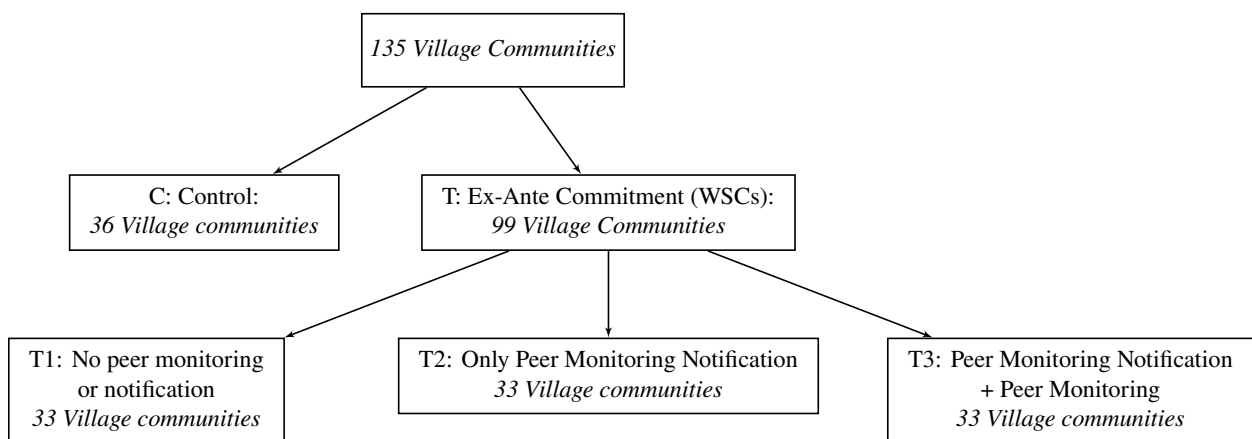


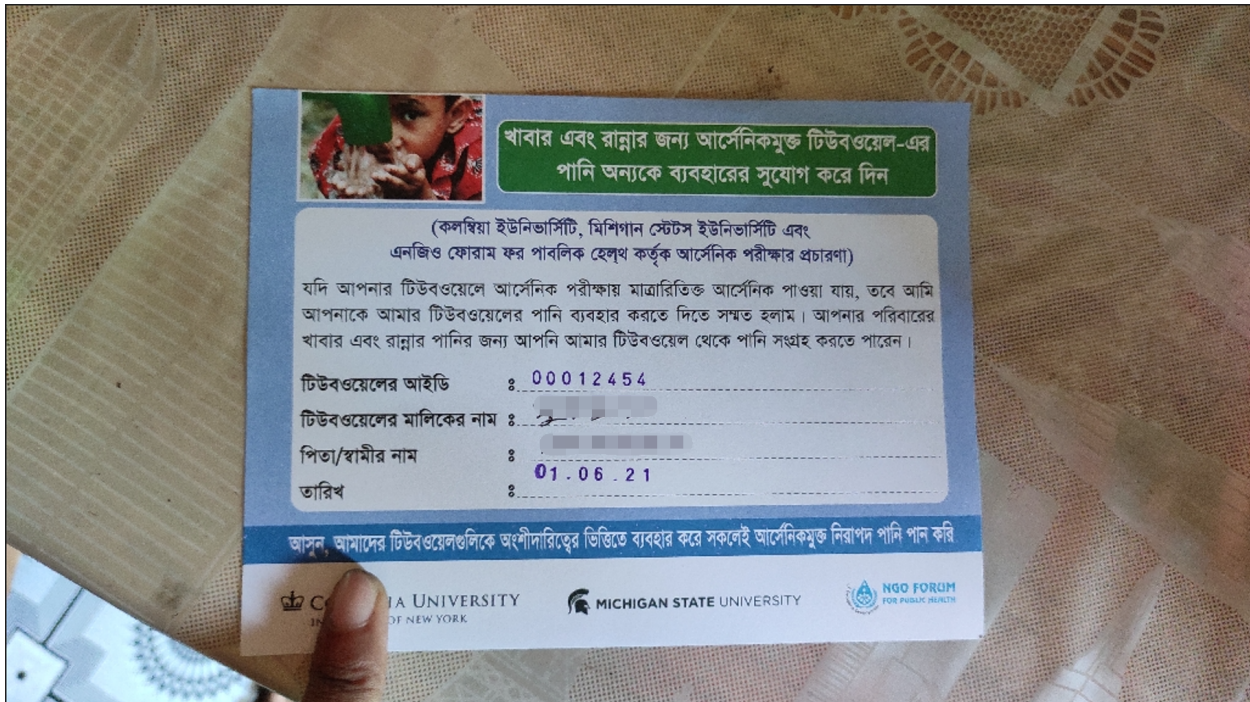


Figure 2.4 A well tag



Notes: A well tag is a metal tag installed on a well with a number engraved on the tag. Each tag is engraved with a unique number. Therefore by recording the well tag we identify the well. During each visit, enumerators only attach a new tag to any well without a tag.

Figure 2.5 Water sharing coupon




Notes: We distribute ten coupons to each household who claimed at least partial ownership to the well that they used primarily in the baseline. On top of the coupon printed a statement saying that whoever receive the coupon has the right to use the household's well, has the well tested to be safe in arsenic. Underneath the statement prints the household's well tag ID, household head's name, and the date.

Figure 2.6 Test result card

## টিউবওয়েলের পানির আর্সেনিক পরীক্ষার ফলাফল


কেয়ারটেকারের নাম: .....

টিউবওয়েল আইডি:




০ - ১০ পিপিবি  
আর্সেনিক  
নিরাপদ পানি

মি.গ্রা./ লিটার




> ১০-৫০ পিপিবি  
আর্সেনিক  
নিরাপদ পানি

মি.গ্রা./ লিটার




> ৫০ পিপিবি  
আর্সেনিক  
দূষিত পানি


মি.গ্রা./ লিটার



COLUMBIA UNIVERSITY  
IN THE CITY OF NEW YORK



MICHIGAN STATE UNIVERSITY



NGO FORUM  
FOR PUBLIC HEALTH

### পানির উৎসের তথ্য

প্রযুক্তির ধরন	: নলকূপ
স্থাপনের সাল	: ..... গভীরতা: ..... (ফুট)
গ্রাম	: .....
ইউনিয়ন	: .....
উপজেলা	: রায়পুরা/নবীনগর/বাঞ্ছারামপুর/ব্রাহ্মণবাড়ীয়া
জেলা	: নরসিংদী/ব্রাহ্মণবাড়ীয়া

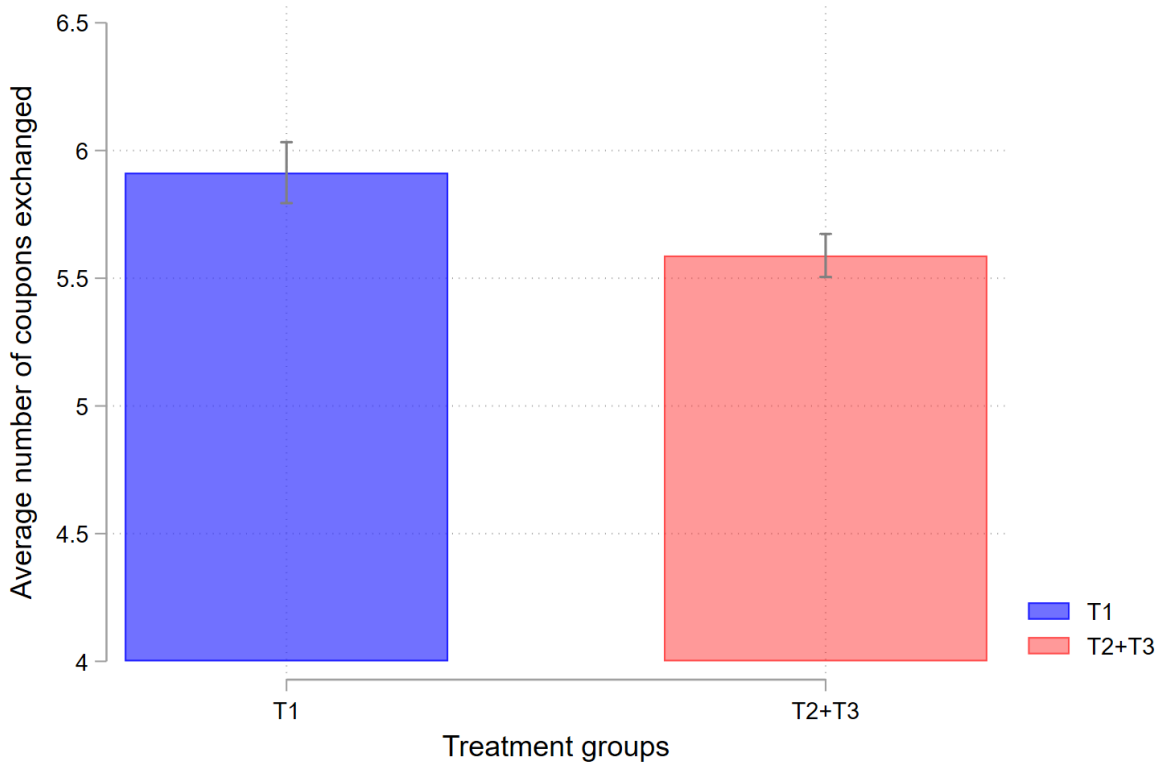
### পানি পরীক্ষার ফলাফল

পরীক্ষার তারিখ	পরীক্ষার তথ্য			সহনীয়মাত্রা	মন্তব্য
	কিটের নাম	পরীক্ষার নাম	ফলাফল		
	ITS Econo-Quick Kit	আর্সেনিক	পিপিবি	৫০	

### আর্সেনিকমুক্ত পানি পান করুন ও সুস্থ থাকুন

- বাংলাদেশ সরকার কর্তৃক প্রণীত ইনভায়রনমেন্টাল কনজারভেশন রুল ১৯৯৭ অনুযায়ী আর্সেনিকের গ্রহণযোগ্য মাত্রা ০.০৫ মিলিগ্রাম/লিটার বা ৫০ পিপিবি।
- লালমুখো/লাল প্র্যাকার্ডযুক্ত টিউবওয়েলের পানি আর্সেনিকমুক্ত। তাই লালমুখো/লাল প্র্যাকার্ডযুক্ত টিউবওয়েলের পানি পান ও রান্নার কাজে ব্যবহার করবেন না।
- সবুজমুখো/সবুজ বা নীল প্র্যাকার্ডযুক্ত টিউবওয়েলের পানি ব্যবহার করুন ও সুস্থ থাকুন।
- শরীরের কোন জায়গায় আর্সেনিকোসিস রোগের লক্ষণ (হাত ও পায়ে বৃষ্টির ফেটার মত কাল দাগ, হাত-পায়ের তালুর চামড়া শক্ত হয়ে যাওয়া) দেখা দিলে সাথে সাথে ডাক্তারের সাথে যোগাযোগ করবেন।
- আর্সেনিকোসিস রোগের কোন সুনির্দিষ্ট ঔষধ নাই। নিরাপদ পানি, পুষ্টির ও ভিটামিনযুক্ত খাবার এই রোগ থেকে আরোগ্য লাভে সহায়তা করে।
- আর্সেনিকোসিস রোগ কোনভাবেই বংশগত, ছোঁয়াচে বা সৃষ্টিকর্তার অভিধাণ নয়। আর্সেনিকোসিস রোগীর সাথে খাওয়া, মেলামেশা ও একসাথে বসবাস করলে এই রোগ ছড়ায় না।

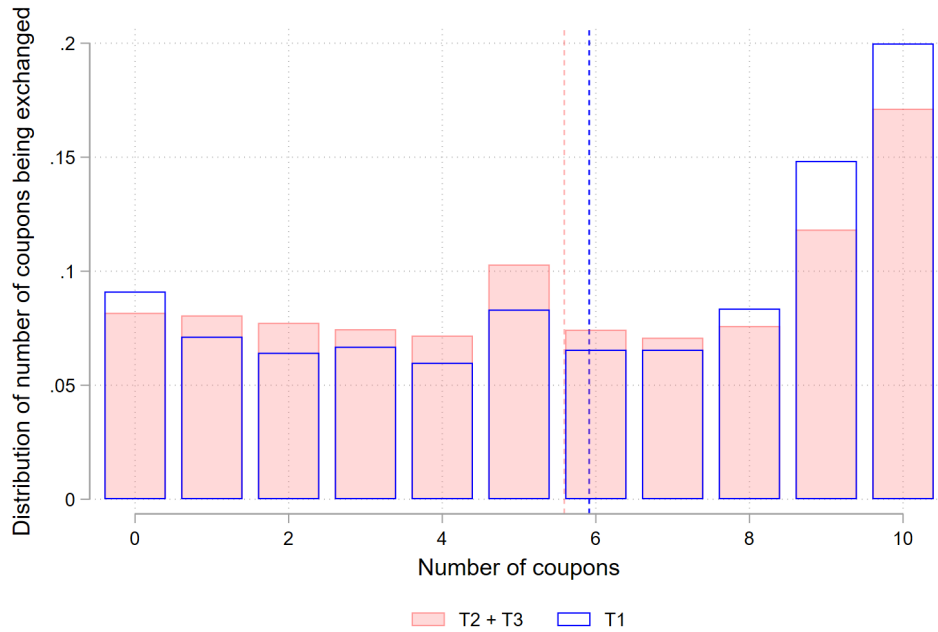
Figure 2.7 Number of coupons being exchanged



Notes: The comparison comes from 6,574 households that exchanged at least one coupon in the 99 villages where we facilitate the ex-ante commitment through the coupon exchange. On average, households in T1 village exchanged 5.913 coupons with their neighbors, and households in T2+T3 village exchanged 5.589 coupons with their neighbors. The confidence interval, constructed by adjusting the standard deviation with the sample size, indicates a significant difference across the groups.

Figure 2.8 The distribution of the number of coupons being exchanged in T1 vs T2+T3 villages

(a) Panel A. Distribution of Number of Coupons Exchanged in T1 vs T2+T3



(b) Panel B. CDF Comparison

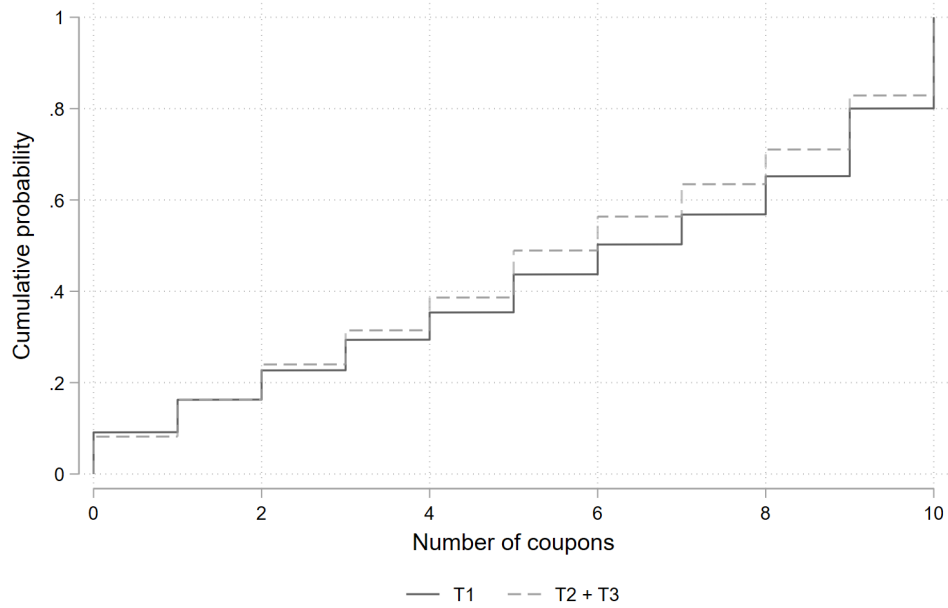
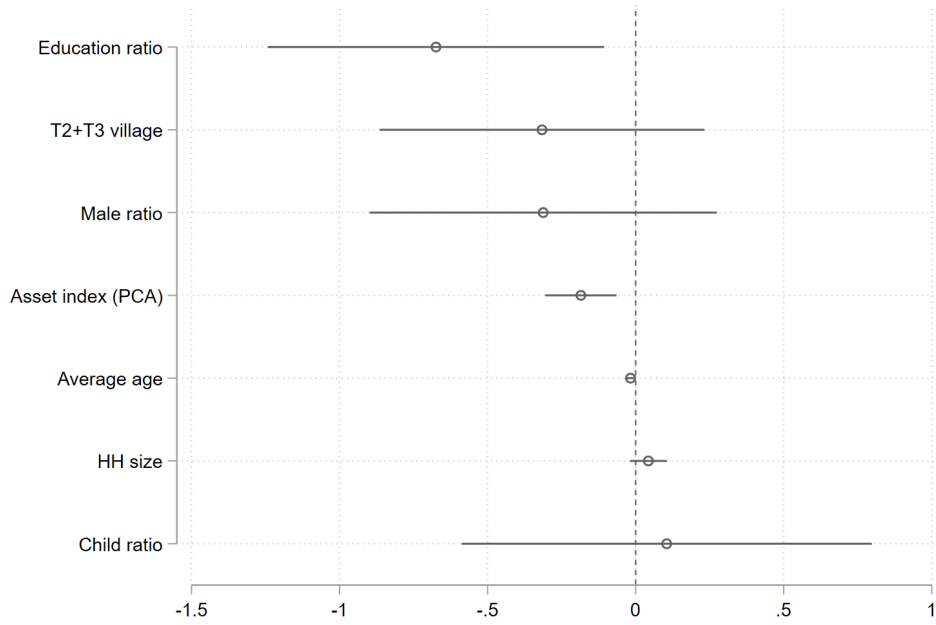


Figure 2.9 Associations of treatment status, demographic characteristics, and social preferences on the coupon exchange

(a) Panel A. The demographic characteristics and treatment status



(b) Panel B. The social preferences

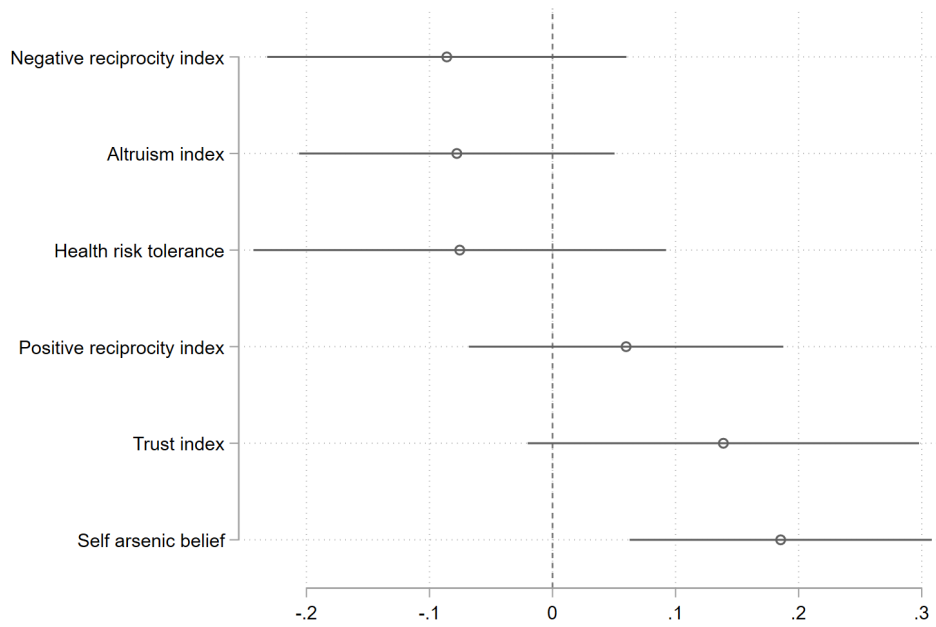
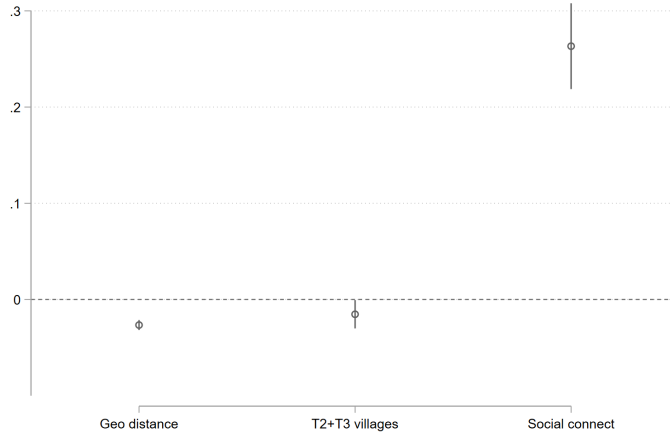
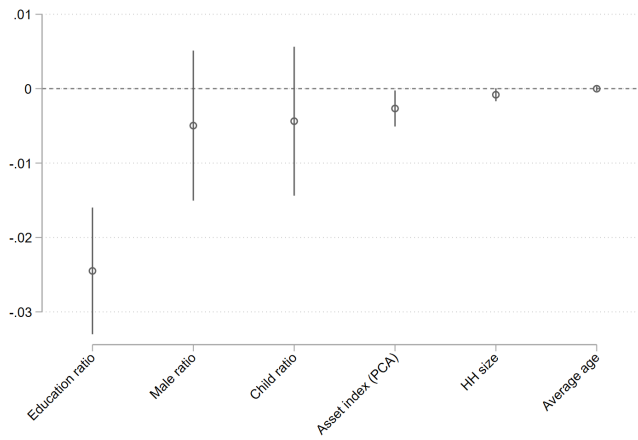


Figure 2.10 The impacts of social connection, distance, and differences in demographics and social preferences on the propensity to exchange coupons

(a) Panel A. Social connection, geographic distance, and treatment group



(b) Panel B. Differences in demographic characteristics



(c) Panel C. Differences in social preferences

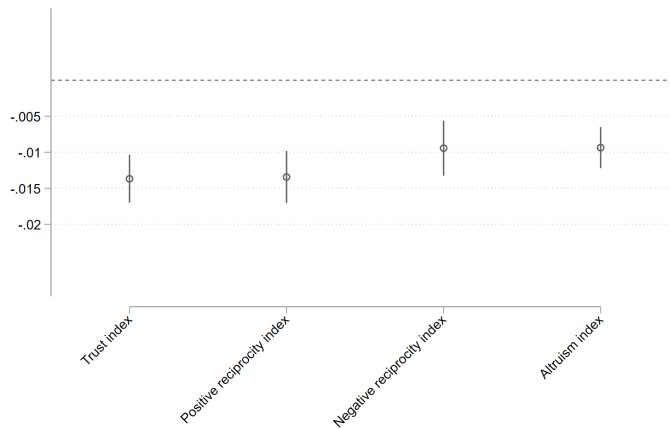
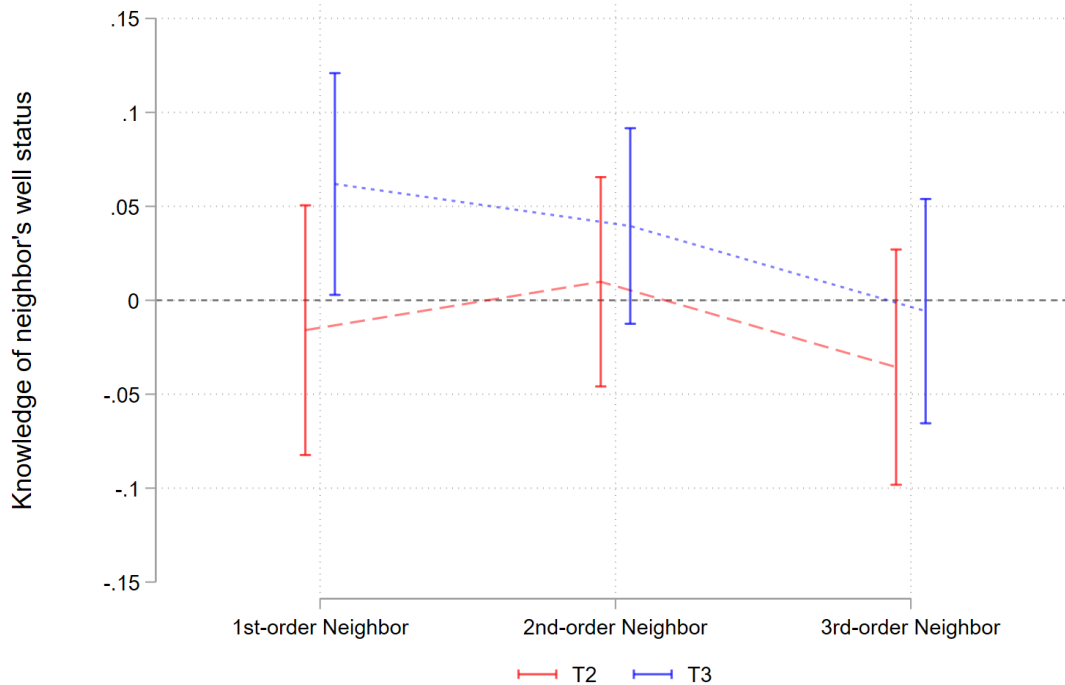


Figure 2.11 Participants' knowledge about the neighbor's well's arsenic status



Notes: This figure shows the impacts of the peer monitoring through the SMS campaign on the household's local knowledge of arsenic contamination. We randomly asked one-third of households who exchanged at least one coupon in the 99 villages where we facilitate the ex-ante commitments. The outcome variable is a score calculated based on the correctness of the household, answering whether or not its 1st-order, 2nd-order, and 3rd-order neighbor's well is safe. The neighbor's order is defined by the distance in the network constructed by the coupon exchange. Households that correctly answer the question receive 1 point but 0 otherwise. The full score is 3. The coefficients that plotted in the figure are obtained from regressing the household's score on treatment dummies, T2 and T3, pre-specified controls, and the Upazila FEs. The standard errors are clustered at the village level. The zero line can be interpreted as the baseline score participants from T1 villages achieved.



## 2.7 Tables

Table 2.1 Study Timeline

<b>Phases</b>	<b>Expected Completion</b>	<b>Status</b>
Household listing	January 2020	Completed
Baseline surveys		
First-round	May 2020 - June 2020	Completed
Second-round	November 2020 - December 2020	Completed
Intervention: WSC exchange	March 2021 - August 2021	Completed
Well testing	December 2021 - July 2022	Completed
Intervention: text message	March 2022 - July 2022	Completed
Endline survey	September - December 2022	Completed

Table 2.2 Treatment arms and number of households

Treatment Arms	Number of Villages	Number of Households
Control (C)	36	4,363
Ex-ante Commitment (T)	99	11,691
No notification or peer-monitoring (T1)	(33)	(4,154)
Only notification, but no peer monitoring (T2)	(33)	(3,689)
Notification and Peer-monitoring (T3)	(33)	(3,848)
Total:	135	16,054

Table 2.3 ‘Notification for Peer-monitoring’ treatment in T2+T3 communities

Treatment Arms	Content
135 villages (T+C)	<p>Hello, my name is XXX and I work for NGO Forum for Public Health, Bangladesh, a non-profit organization and collaborating with researchers from Columbia University and Michigan State University.</p> <p>The goal of our project is to reduce exposure to arsenic by drinking contaminated well-water. We expect that the government’s Department of Public Health and Engineering will test your tubewell for arsenic in the coming weeks. Members of our team visited you for a household survey a first time in January 2020 and followed up with two rounds of phone calls.</p>
99 villages (T)	<p>This time we would like to invite you to take part in an activity designed to encourage you and your neighbors to share those wells that turn out to be low in arsenic after DPHE testing.</p> <p>If you agree to participate, we will give you 10 water sharing coupons that we encourage you to exchange with any other well-owner who would be willing to let you use his well and with whom you would be willing to share your own well. You will have 3 days for exchanging coupons, after which we will return to record the coupons you have exchanged as well as the coupons you have not exchanged. We will also record your response to a number of questions about your household and water usage using tablets or smartphones. We will recheck your name and phone number(s) for future reference as well. We may occasionally send you text (SMS) messages to remind you of the health risks from drinking tubewell water with toxic levels of arsenic and to share information about this study.</p>
66 villages (T2 + T3)	<p>We may also send you a text (SMS) message to a maximum of 3 of your neighbors listing the names of households you agreed to share your well with using the coupons. Conversely, you may receive a text (SMS) message listing the names of households up to 3 of your neighbors agreed to share their well with using the coupons.</p>
135 villages (T+C)	<p>If you have no objection, please allow NGO Forum for Public Health to take “GPS” coordinates of where we are conducting the interview.</p> <p>This means we will use our phone to mark our interview location on a map. Once the survey work is completed your identifying information will be masked before data is analyzed, so no one will be able to identify you from your answers. The results of this study may be published or presented at professional meetings, but the identities of all research participants will remain anonymous in any research presentation or publication.</p> <p>The following entities will have access to the data: Researchers and Research Staff, Institutional Review Board (IRB). Implementing agencies- NGO Forum for Public Health, Bangladesh and Innovations for Poverty Action (IPA)</p>

Table 2.4 Text message and voice call

Treatment	Coverage	Content
First General Text Message	135 villages (T+C)	Hello! We, NGO Forum, recently visited your para for arsenic testing of wells. Here is some information regarding arsenic. Arsenic is toxic for you and your children's health and well-being. Drink low arsenic water for the sake of your well-being!
Second Text Message	66 villages (T1+T2)	Recall that, before well testing, you and your peers in this para agreed to share water by exchanging water-sharing coupons.
Third Text Message	33 villages (T3)	Recall that, before well testing, you and your peers in this para agreed to share water by exchanging water-sharing coupons. We will share the names of people, who exchanged coupons in your para through SMS.
Customized monitor message	33 villages (T3)	Hello! Mr/Ms AAA, XXX in your para exchanged coupons with YYY, ZZZ, and N other households.
Customized receipt message	33 villages (T3)	Hello! Mr/Ms XXX, We have shared names of people, who you exchanged coupons with, to ABC and DEF.
Voice message	135 villages (T + C)	Hello! We, NGO Forum, recently visited your para for arsenic testing of wells. Here is some information regarding arsenic. Arsenic is toxic for you and your children's health and well-being. Drink low arsenic water for the sake of your well-being!
Second voice message	66 villages (T1 + T2)	Recall that, before well testing, you and your peers in this para agreed to share water by exchanging water-sharing coupons
Third voice message	33 villages (T3)	Recall that, before well testing, you and your peers in this para agreed to share water by exchanging water-sharing coupons. We will share the names of people, who exchange coupons in your para through SMS.

Table 2.5 Well arsenic test results

Arsenic concentration ( $\mu\text{g}/\text{L}$ )	Count	Percentage	Cumulative
0	1234	12.54%	12.54%
0-10	1130	11.48%	24.03%
10-50	1467	14.91%	38.94%
50-300	4028	40.94%	79.88%
300-	1980	20.12%	100.00%
Total	9839	100%	

Note: We used two testing kits in the field with different scales. Therefore this table consolidates the results from two kits.

Table 2.6 Social Networks

	N	mean	sd
Socialization	11,933	3.017	2.495
Discuss farming issues	11,933	0.788	1.433
Discuss health issues	11,933	0.879	1.433
Discuss financial issues	11,933	0.84	1.387
Borrow or lend daily necessities	11,933	1.336	1.706
Borrow or lend money	11,933	1.162	1.532
<i>Total Degree</i>	11,933	4.086	3.970

Note: We asked households who they interact with in respect to each category. For example, *Discuss health issues* means that respondents were asked to elicit the households they often discussed health issues with.

Table 2.7 Summary Statistics: Household and well characteristics

	count	mean	sd	min	max
Household size	16054	5.10	2.05	1	21
Average age	16054	27.07	10.74	6	100
Male ratio	16054	0.48	.18	0	1
Child ratio	16054	0.39	0.21	0	1
Primary edu ratio	16054	0.29	0.26	0	1
Risk tolerance	14039	1.86	1.16	1	5
Asset PCA Index	13294	0.00	1.00	-2.64	9.99
Number of wells	16054	0.80	0.52	0	4
Well depth	7716	131.43	91.99	1	1000
Well age	9732	9.90	7.41	1	81
Well tested for arsenic	10032	0.07	0.263	0	1

Notes: A household is defined as a group of relatives that share the same kitchen. In a few cases where households owned multiple wells, we used the well that was reported as the primary well to measure the depth, age, and whether it tested for arsenic.

Table 2.8 Summary Statistics: Assets

	count	mean	sd	min	max
Rooms	13741	2.70	1.30	0	10
Electricity <sup>1</sup>	13823	.98	.15	0	1
Fans	13778	2.33	1.23	0	10
Mobilephone	13733	1.89	1.12	0	10
Smartphone	13720	0.87	0.96	0	10
Cycle or rickshaw	13709	0.11	0.39	0	10
Motorcycle	13709	0.06	0.27	0	8
TV	13759	0.47	0.53	0	8
Refrigerator	13759	0.52	0.53	0	8



Table 2.9 Balance across treatment arms: Household and well characteristics

	C	T1	T2	T3	Differences p-value					
					T1-C	T2-C	T3-C	T3-T2	T2+T3-T1	T1+T2+T3-C
Household size <sup>1</sup>	5.1 (2.06) [4363]	5.06 (2.09) [4154]	5.11 (2.04) [3689]	5.07 (1.99) [3848]	-0.039	0.008	-0.035	-0.043	0.025	-0.023
Average age	27.1 (10.76) [4363]	27.29 (10.52) [4154]	26.95 (10.84) [3689]	27.24 (10.96) [3848]	0.191	-0.151	0.14	0.291	-0.193	0.066
Male ratio	0.42 (0.2) [4363]	0.41 (0.2) [4154]	0.41 (0.2) [3689]	0.42 (0.2) [3848]	-0.01	-0.003	0.007	0.01	0.012**	-0.002
Child ratio	0.4 (0.21) [4363]	0.39 (0.22) [4154]	0.4 (0.21) [3689]	0.4 (0.21) [3848]	-0.009	0.006	0	-0.005	0.011	-0.001
Primary edu ratio <sup>2</sup>	0.28 (0.25) [4363]	0.3 (0.27) [4154]	0.28 (0.25) [3689]	0.29 (0.26) [3848]	0.022	-0.003	0.009	0.011	-0.019	0.01
Risk tolerance <sup>3</sup>	1.86 (1.15) [3822]	1.85 (1.17) [3616]	1.87 (1.17) [3273]	1.86 (1.14) [3328]	-0.007	0.012	-0.001	-0.013	0.013	0.001
Asset PCA Index	0.00 (1.03) [3547]	0.04 (0.99) [3510]	0.00 (1.00) [3074]	-0.04 (0.97) [3163]	0.04	0.001	-0.035	-0.036	-0.057	0.003
Number of wells	0.81 (0.52) [4363]	0.79 (0.54) [4154]	0.79 (0.53) [3689]	0.81 (0.49) [3848]	-0.019	-0.014	-0.002	0.012	0.011	-0.012
Well depth <sup>4</sup>	124.64 (116.36) [1981]	126.38 (100.28) [1739]	134.64 (103.35) [1785]	123.64 (85.46) [1834]	1.74	9.997	-1.003	-11	2.682	3.552
Well age	10.19 (7.67) [2669]	10.08 (7.55) [2330]	9.85 (7.42) [2249]	10.21 (7.40) [2395]	-0.114	-0.345	0.018	0.363	-0.044	-0.143
Well tested for arsenic	0.07 (0.25) [2768]	0.09 (0.29) [2509]	0.07 (0.26) [2302]	0.06 (0.24) [2453]	0.027*	0.007	-0.003	-0.01	-0.025	0.011

Table 2.10 Balance across treatment arms: Assets

	C	T1	T2	T3	Differences p-value					
					T1-C	T2-C	T3-C	T3-T2	T2+T3-T1	T1+T2+T3-C
Number of wells	0.81 (0.52) [4363]	0.79 (0.54) [4154]	0.79 (0.53) [3689]	0.81 (0.49) [3848]	-0.019	-0.014	-0.002	0.012	0.011	-0.012
Rooms	2.71 (1.31) [3698]	2.73 (1.34) [3607]	2.68 (1.31) [3146]	2.66 (1.22) [3290]	0.024	-0.033	-0.053	-0.02	-0.068	-0.019
Electricity <sup>1</sup>	0.98 (0.14) [3718]	0.98 (0.15) [3624]	0.98 (0.15) [3165]	0.98 (0.15) [3316]	-0.001	-0.004	-0.002	0.002	-0.002	-0.002
Fans	2.34 (1.23) [3707]	2.35 (1.25) [3613]	2.34 (1.24) [3156]	2.29 (1.20) [3302]	0.008	0.001	-0.05	-0.051	-0.033	-0.013
Mobile phone	1.92 (1.19) [3694]	1.9 (1.07) [3606]	1.87 (1.13) [3145]	1.86 (1.09) [3288]	-0.019	-0.043	-0.053	-0.01	-0.029	-0.038
Smartphone	0.86 (1.00) [3689]	0.86 (0.91) [3600]	0.9 (0.98) [3144]	0.86 (0.92) [3287]	0.003	0.039	-0.003	-0.042	0.015	0.012
Cycle or rickshaws	0.1 (0.38) [3681]	0.12 (0.37) [3602]	0.1 (0.4) [3141]	0.13 (0.39) [3285]	0.011	-0.01	0.022	0.032	-0.004	0.008
Motorcycle	0.06 (0.28) [3681]	0.08 (0.32) [3602]	0.05 (0.22) [3141]	0.06 (0.24) [3285]	0.02	-0.014	-0.005	0.009	-0.029*	0.001
Television	0.46 (0.53) [3694]	0.51 (0.51) [3613]	0.49 (0.56) [3152]	0.44 (0.54) [3300]	0.048	0.03	-0.013	-0.042	-0.041	0.022
Refrigerator	0.49 (0.53) [3694]	0.54 (0.52) [3613]	0.53 (0.52) [3152]	0.5 (0.55) [3300]	0.051	0.043	0.014	-0.029	-0.023	0.036

Table 2.11 Assort to coupon groups under peer-monitoring pressure

	Whether households $i$ and $j$ exchanged coupons	
	(1)	(2)
Social connect	0.272*** (0.0508)	0.266*** (0.0423)
Social connect $\times$ Notification	-0.0152 (0.0549)	-0.0310 (0.0463)
Geo distance	-0.0224*** (0.00344)	-0.0314*** (0.00650)
Geo distance $\times$ Notification	-0.00827* (0.00456)	-0.0152* (0.00782)
Asset diff	0.00115 (0.00204)	-0.000446 (0.00187)
Asset diff $\times$ Notification	-0.00593** (0.00250)	-0.00626** (0.00246)
Observations	171,666	171,666
R-squared	0.130	0.246
FEs	Village	Village+Individual

Notes: This table shows the coefficients from a dyad regression in which the outcome variable indicates whether two households from the same village exchanged coupons. The table reports the coefficients of the social connections, geographic distance, and difference in asset level, and their interactions with a treatment dummy of the peer monitoring notification (T2+T3 villages), **Notification**. Standard errors are clustered at the village community level.

Column (1) reports the regression coefficients of social connection, geographic distance, asset level, and their interactions with the Notification dummy. The regression includes four sets of variables: (1) social connection, geographic distance, asset level, and their interactions with the Notification dummy; (2) absolute differences of household characteristics and their interactions with the Notification dummy; (3) sums of household characteristics and their interactions with the Notification dummy; and (4) village dummies.

Column (2) reports the same regression coefficients with Village FEs and two-way-individual FEs,  $\xi_i$  and  $\xi_j$ . While it includes all the variables from Column (1), the sums and interactions with the sums are omitted due to the inclusion of individual FEs.

The rest of the sorting coefficients, absolute differences of household characteristics and their interactions with the Notification dummy, are reported in Table A7.

Table 2.12 Discussed well-sharing before the arsenic test

	Discussed well-sharing before the arsenic test		
	(1)	(2)	(3)
T1	0.144*** (0.0253)	0.0604** (0.0291)	0.120*** (0.0222)
T2	0.161*** (0.0257)	0.0470 (0.0291)	0.128*** (0.0232)
T3	0.155*** (0.0262)	0.0346 (0.0297)	0.127*** (0.0243)
Control Mean	0.29	0.28	0.29
Sample	Owner	Non-owner	Full
Observations	10,781	4,326	15,219
R-squared	0.056	0.070	0.056

Notes: The standard errors of the regression coefficients in this table are clustered at the village-community level. The regression includes the pre-specified controls and Upazila FEs, the strata dummies. An Upazila is a sub-district that serves as the second-level administrative unit in Bangladesh, below the districts. Our 135 villages belong to 6 different Upazilas. The outcome variable, whether the household discussed well-sharing before the arsenic test, was a Yes or No question asked among well users during the Endline survey. Therefore, the estimated coefficients represent the change in the probability of discussing well-sharing due to the treatments. T1 are villages where we only facilitate the ex-ante commitment by distributing coupons. T2 are villages where we distribute coupons and notify households about potential peer monitoring through the SMS campaign. T3 are villages where we distribute coupons and implement peer monitoring.

Table 2.13 The impacts of treatment on the arsenic consumption measured by the well testing in the endline

Arsenic concentration of the well used by the household in the Endline (ppb)						
Panel A. Overall treatment effect of ex-ante commitment						
	(1)	(2)	(3)	(4)	(5)	(6)
T1+T2+T3	-18.45* (10.09)	-15.67 (12.70)	-36.09*** (11.71)	-35.49** (13.91)	0.818 (2.317)	5.044 (12.26)
R-squared	0.403	0.256	0.329	0.274	0.043	0.105
Panel B. Treatment effect of each treatment						
	(1)	(2)	(3)	(4)	(5)	(6)
T1	-23.98* (14.47)	-21.07 (16.68)	-43.68*** (14.91)	-42.61** (16.31)	0.822 (2.835)	5.791 (16.91)
T2	-28.94** (12.21)	-34.41** (16.28)	-31.41** (15.14)	-30.48 (21.95)	2.558 (2.514)	11.96 (13.45)
T3	-3.812 (12.21)	5.024 (16.34)	-30.78** (14.74)	-28.64 (18.99)	-1.103 (2.755)	-7.648 (12.95)
R-squared	0.405	0.262	0.329	0.275	0.045	0.108
Control Mean	210.88	321.97	216.68	309.56	18.12	42.18
Well-Owner	✓	✓	✗	✗	✓	✗
Baseline Arsenic	ALL	HIGH	ALL	HIGH	LOW	LOW
Observations	8,349	5,343	2,251	1,393	3,006	858

Notes: This table reports the intention-to-treat estimates of the treatment effects of the interventions. Panel A shows the pooled impact of ex-ante commitment on arsenic consumption, namely comparing the 99 village communities that intervened with the coupon exchange with the 36 control village communities. Panel B shows each intervention separately. T1 are villages where we only facilitate the ex-ante commitment by distributing coupons. T2 are villages where we distribute coupons and notify households about potential peer monitoring through the SMS campaign. T3 are villages where we distribute coupons and implement peer monitoring. The regression coefficients are obtained by regressing the arsenic concentration of the well used by the household in the Endline on the treatment dummy (dummies), pre-specified controls, and the strata dummies (Upazila FEs). Standard errors are clustered at the village-community level. Therefore, the coefficients show the estimated impacts of treatments on arsenic mitigation. The control means shows the arsenic concentration of the well used in the Endline by the households in the 36 control villages. A well-owner is defined by whether the household privately owns the primary well in the baseline. Households were invited to the experiment if and only if they were well-owners. Baseline arsenic identifies if the household's baseline primary well contains arsenic greater or less than 50 ppb, which is the cutoff for arsenic-poisoned water determined by the Bangladeshi government.

Table 2.14 Switching

Switched	Unsafe to safe		To lower		Safe to unsafe		To higher	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Overall treatment effect of ex-ante commitment								
T1+T2+T3	0.00557 (0.00793)	-0.0145 (0.0256)	0.00527 (0.00674)	-0.00509 (0.0215)	-0.00217 (0.00485)	0.0459 (0.0314)	-0.0104** (0.00416)	0.0150 (0.0285)
R-squared	0.055	0.144	0.008	0.022	0.011	0.102	0.003	0.046
Panel B. Treatment effect of each treatment								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T1	0.00135 (0.00909)	-0.0283 (0.0272)	0.00630 (0.00988)	-0.00561 (0.0239)	-0.00556 (0.00689)	0.0273 (0.0381)	-0.0113** (0.00444)	0.00480 (0.0300)
T2	0.0249** (0.0121)	-0.000530 (0.0409)	0.0104 (0.00880)	0.000297 (0.0303)	-0.00207 (0.00585)	0.0791* (0.0411)	-0.00883* (0.00521)	0.0518 (0.0402)
T3	-0.00356 (0.0108)	-0.00483 (0.0325)	-0.000227 (0.00876)	-0.0106 (0.0305)	-0.000601 (0.00521)	0.00732 (0.0316)	-0.0109** (0.00484)	-0.0133 (0.0336)
R-squared	0.057	0.145	0.008	0.022	0.011	0.108	0.003	0.051
Control Mean	0.04	0.11	0.04	0.13	0.01	0.08	0.02	0.11
Well-Owner	✓	✗	✓	✗	✓	✗	✓	✗
Observations	4,863	1,184	7,869	2,042	3,006	858	7,869	2,042

Notes: This table reports the intention-to-treat estimates of the treatment effects of the interventions. A household is considered to have its well switched if it reports the primary well used in the Endline survey different from the well used in the Baseline survey. We consider four types of well switching: columns (1) - (2) show switching from unsafe to safe, that the household used a well that contains arsenic higher than 50 ppb in the Baseline but switched to a well that is lower than 50 ppb in the Endline; columns (3) - (4) show switching to a lower-arsenic-contaminated well, meaning that the well used in the Endline contains less arsenic than the well used in the Baseline; columns (5) - (6) show switching from safe to unsafe, that the household used a well that contains arsenic lower than 50 ppb in the Baseline but switched to a well that is higher than 50 ppb in the Endline; and columns (7) - (8) show switching to a higher-arsenic-contaminated well, meaning that the well used in the Endline contains higher arsenic than the well used in the Baseline. *Households that did not switch are assigned zero in all four types of switching.*

Panel A shows the pooled impact of ex-ante commitment on well switching, namely comparing the 99 village communities that intervened with the coupon exchange with the 36 control village communities. Panel B shows each intervention separately. T1 are villages where we only facilitate the ex-ante commitment by distributing coupons. T2 are villages where we distribute coupons and notify households about potential peer monitoring through the SMS campaign. T3 are villages where we distribute coupons and implement peer monitoring. The regression coefficients are obtained by regressing the switching status of the household on the treatment dummy (dummies), pre-specified controls, and the strata dummies (Upazila FEs). Standard errors are clustered at the village-community level. Therefore, the coefficients show the estimated impacts of treatments on the probability of switching.

The control means shows the arsenic concentration of the well used in the Endline by the households in the 36 control villages. A well-owner is defined by whether the household privately owns the primary well in the baseline. Households were invited to the experiment if and only if they were well-owners. Baseline arsenic identifies if the household's baseline primary well contains arsenic greater or less than 50 ppb, which is the cutoff for arsenic-poisoned water determined by the Bangladeshi government.

## CHAPTER 3

### PEER EFFECTS IN ADOPTION OF PREVENTIVE MEASURES: EVIDENCE FROM RURAL BANGLADESH

#### 3.1 Introduction

Covid-19 has highlighted the importance of adherence to precautionary and preventive health measures likely to an extent unseen so far. Adoption of preventive measures, such as wearing masks, social distancing, and washing hands with soap, has been shown to be effective in reducing cases and deaths (for example, see Howard et al., 2021). Yet policymakers and social scientists alike find it challenging to identify proper ways to encourage the adoption of such preventive measures.

The more general underlying question is what factor or series of factors motivate some people to change their behavior and others not. Even in some of the most severely affected regions, people differ widely in their views on taking relatively simple preventive measures such as social distancing (Gollwitzer et al., 2020). This evidence may suggest that people are making decisions not only based on the information but also rely on the social incentives underneath the preventive measures adoption decisions.

Using one in-person survey and two rounds of phone call survey towards round 3,000 households from 135 Bangladesh villages, we elicit the preventive measures adopted in the first wave and second wave of COVID-19 break out in Bangladesh. In companion with the rich network data we elicited through the survey, we detect a significant level of clustering on the adoption of preventive measures within the unit as small as para, the subdivision of village that usually resides around a hundred households. By creating an adoption index based on the preventive measures we recorded, we find households that are either geographically connected or socially connected score similar adoption indexes. In this paper, we study if the peer effects, which defined as the influence towards the individual from their peers, caused the clustering.

Identification of peer effects, however, is not straightforward. We rely on the Linear-in-Means model for its reasonable structure in tracking the peer effects. The model has the structure that a household's adoption decision depends on the mean adoption-level of the households' neighbors

and the mean characteristics of the neighbors<sup>1</sup>. Though appealing in its structure on matching the real decision process, consistently estimating the peer effects parameters has been shown difficult due to the two identification issues (Manski 1993). First, named by Manski as *correlated effects*, connected individuals presenting similar behavior may eventually due to they share some common (unobserved) characteristics and these characteristics also affect the tendency to connect. In our context, wealth level can be the factor that influences the adoption decision, and similarly, wealthy households may be more likely to interact with each other. Hence upon seeing both of the households wearing masks, we cannot conclude that they are necessarily influenced by each other's behavior.

Second, named by Manski as the *reflection problem*, the structure of the Linear-in-Means model permits simultaneity. Individuals are influenced by the mean adoption decisions of their neighbors while they are also influencing their neighbors' adoption decisions. Therefore, it is hard to disentangle the influences of neighbors' adoption decisions from their observable characteristics. To address these two issues, we use the identification strategy advanced by Bramoullé, Djebbari, and Fortin (2009) in which we assume the peers' network is randomly formed conditional on various household characteristics and geographical and time fixed effects. Implied by the model, we can use the neighbors' neighbors' mean characteristics as the instrument for the neighbors' mean adoption level as long as these neighbors' neighbors are not the direct connect to the households.

Peer effects are detected by all estimation strategies for mask-wearing decisions. We find that a 10 percentage point increase in mask wearing of geographical neighbors, who live within 50 meters from the households, will cause the household's likelihood to increase mask-wearing by 2 percentage points. A similar magnitude of influence has also been found in the knowledge of whether contaminated surfaces or objects can transmit the virus. The estimated peer effects on washing hands with soap are small and indistinguishable from 0. We do not find a significant impact from peers on social distancing and social gathering behavior or tendency to cover sneeze with hands or elbows. We find the influences of peers decrease as we expand the definition

---

<sup>1</sup>These effects are named as endogenous effects and contextual or exogenous effects by Manski.



of geographical neighbors. These findings mostly match the prediction of the social signaling model, hence providing us direct evidence on the existence of social incentives behind the adoption decision.

To explain these findings, we adopt the social signaling framework proposed by Benabou and Tirole (2006) to predict which types of preventive measures are more likely to be enforced socially and which types of neighbors are more likely to impose such social pressure (or benefit). The key point of the model is that adoption decisions depend not only on the health benefit and direct cost corresponds to the preventive measures, but also on the "visibility" of the preventive measures. As the result, high observable preventive measures such as wearing masks are easier to be socially enforced than the preventive measures that are less likely to be observed, such as washing hands with soap, which may more likely to be taken place in private. We also discuss how social learning can interplay with the social signaling model and to what extent the learning will influence the signaling model predictions.

Our study first contributes to the literature in estimating the social influence on health-related behavior. Though the subject has fairly well-documented in previous literature but as far as we know, has not been discussed extensively in the adoption of the preventive measures against the pandemic. Previous studies cover teenager smoking (Gaviria and Raphael, 2001; Nakajima, 2007 ), obesity (Christakis and Fowler, 2007; Trogdon, Nonnemaker, and Pais, 2008), alcohol use (Case and Katz (1991); Lundborg, 2006; Kremer and Levy, 2008), and many other health-related topics <sup>2</sup> However, fewer studies document the peer effects on health preventive measures. Godlonton and Thornton (2012) find that having neighbors learning their HIV results will increase the likelihood that individuals themselves learning their HIV results. However, unlike the addictive behavior, in which the peers' behavior usually positively correlates with the subject's own behavior, the preventive behavior can, oninining the contrary, can have a negative correlation. Kremer and Miguel (2007) find that people are less likely to take deworming if their direct contacts or second-order contacts

---

<sup>2</sup>For example, Gaviria and Raphael (2001) also discusses the peer effects in drug use in the paper. Duncan et al. (2005) cites the peer effect of a roommate in drug use and sexual behavior. In another important paper, Evans, Oates, and Schwab (1992) estimate the peer effects in teenage pregnancy.

(the direct contacts of the direct contacts) have taken the deworming and found the treatment was not effective, suggesting privatize such a health public goods provision is unsustainable. Hence, the decision problem may get complicated when individual potentially taking the peers' externalities into the equation. Our study adds another set of evidence of peer effects in health preventive measures by examining the context of COVID-19, which is undoubtedly one of the most important and urgent research topic in social science.

The second contribution lays in the estimation of peer effect using detailed network data. We compare the estimation result using the network generated by the real-life interactions and the network generated by the traditional proxies such as geographical proximity. Most of the previous research uses the classroom, school, and geographical subdivision such as villages as the proxy for the network, assuming that the real-life communications and social interactions happen within these social or geographical spheres. For example, classical papers like Evans, Oates, and Schwab (1992) and Gaviria and Raphael (2001) use all the students attending the same school as one networked group. Case and Katz (1991) use city blocks and Munshi (2003) defines the networked group as the migrants from the same origin community. This type of network definition may lead to one important assumption made by Charles Manski in his 1993 seminal paper, which discussed the reflection problem in the identification of peer effects. In the paper, he assumes that individual's peer groups form the partition of the social network, and hence every individual in the one peer group only interact with others in the same peer group. This setting may consider as unrealistic in many interesting questions. For example, networked defined by friends or geographical distance will violate the partition restriction as intransitive triples are likely to be found in these networks.

Many recent works consider the network that goes beyond this assumption by using richer and more detailed dataset <sup>3</sup>. And there is a growing number of studies using the network directly elicited by the costly network questions such as Conley and Udry (2010), Banerjee et al. (2014), Heß, Jaimovich, and Schündeln (2018), and Beaman et al. (2018)<sup>4</sup>. Having this set of information

---

<sup>3</sup>For example, Beaman (2012) uses the same nationality, Sojourner (2013) use the classroom randomized in the Star project, and both Krishnan and Patnam (2014) and Drago et al. (2020) uses the geographical distance to define the networked group.

<sup>4</sup>There are also a large number of empirical and methodological papers written using the National Longitudinal

can be important since we do not know how the network defined by the geographic or institutional structure overlaps the real network. Beaman et al. (2018) use a randomized controlled trial to show that the technology diffusion in the village in which the seeding agents are selected by the geographic-induced network is weaker than the counterpart villages in which the seeding agents are selected by the concrete network questions.

Last, of all, we contribute to the literature that seeks the potential mechanisms underlying the observed peer effects. Individual's adoption of preventive measures may be enforced by the norm and the pressure. The conformity due to social norms and pressure is also studied widely in both theory and empirical literature (see Young, 2015 and Burszty and Jensen, 2017 for review). The main point is that differential responses to a behavioral incentive that varied by its observability to the public have been supported by much empirical evidence. People are more likely to vote when voting is observed or discussed (Funk ,2010; Gerber, Green, and Larimer, 2008; DellaVigna et al., 2016 and 2017). Education investment can be either positively or negatively affected by the image concern when the investment can be observed (Burszty and Jensen, 2015). Not to mention the charitable giving (DellaVigna et al., 2012) and effort put into the workplace (Mas and Moretti, 2009). There seems lack of evidence on social pressure towards preventive measures and our study provides one piece of evidence to this important literature. Our study potentially offers a test to the existence of the social pressure channel.

One potential weakness in this study is that we use self-reported data on preventive health decisions. During COVID-19, fieldwork for data collection was neither feasible nor ethical. Many studies have relied on self-reported data from phone surveys during COVID-19 (for example, Banerjee et al., 2020 ). We of course can not rule out self-reporting bias in our data, but we note that it will be an issue only when such bias is correlated to our network-based measures, which is a higher bar.

The paper organized as follows: In Section 3.2, we describe the surveys, data, and the findings from summary statistics in detail. In Section 3.3, we discuss how to estimate the peer effects and

---

Study of Adolescent Health (Add Health), which captures the interpersonal relationship within US high schools as well as a rich set of demographic and academic variables.

the corresponding identification assumptions and the results will be summarized in Section 3.4. In Section 3.5, we utilize the theoretical model to explain the underlying social incentives that generate the peer effects. We conclude in Section 3.6.

## **3.2 Data**

To study the peer effects on preventive measures, we collected uniquely rich data, including the social networks, demographics, and adoptions of preventive measures of rural households from 135 Bangladesh villages. Innovations for Poverty Action Bangladesh (IPA Bangladesh) administered the surveys both through field visits and phone calls. The survey is in companion with an NSF-funded well switching experiment and consists of three rounds.

The first round of the survey, referred to as the census, was conducted in the field between January 15, 2020, and February 02, 2020, when the pandemic had not broken out in Bangladesh. As the name suggested, a large sample of 16,054 households from 135 Bangladesh villages or paras (subdivisions of villages) is contained. The second survey is conducted between May 8 and June 8, 2020, which we will call the May survey. The third survey, or will be referred to as the November survey, is conducted between October 27 and December 14, 2020. The May and November surveys were moved to phone call surveys due to health concerns and lockdown restrictions during the pandemic. The survey was requested to be finished by an available adult household member, preferably the household head. The phone numbers were collected during the census, and a backup number was also recorded in case the household could not be reached using the primary phone number.

### **General Survey**

Among the 16,054 households we reached in the initial census, we successfully recorded 11,933 households who completed all three survey rounds. A rich set of demographics, mortality, and migration status, as well as the composition of the household members for all the households and record their name, gender, age, education, as well as the GPS location of the households and their wells (if they have) were collected. A detailed section of asset questions is collected from households approached by the enumerators in the November survey. The asset question includes

whether the household has access to electricity, the possession of household appliances, the features of their houses, the acres of agricultural land the household owns, and etc. We also elicit the general risk and health risk tolerance through a five-scale question. Table 3.1 summarizes the main statistics we collected using these three survey rounds. We can see that the average size of a household is around 5 members with around 2 children. The households are relatively young, with an average age of 27, and only 30% of adults received at least primary education. We can also see that people, on average, are less willing to take risks on health than other things. The electricity is exceptionally prevalent but this may be because the households that can be reached through phone call surveys are almost surely connected to electricity. On average, each household has more than two phones, one being a smartphone. The ownership of televisions is also relatively high. On the other side, the ownership of transportation tools like bicycle, motorcycle, and car, however, is rather low.<sup>5</sup>

### **COVID-19 Preventive Measures and Transmission Knowledge**

COVID-19 preventive measures and knowledge were collected in the May and November survey, which coincided with the first and second wave of COVID-19 break out in Bangladesh. In the May survey, 20% of households are randomly selected to ask about their adoptions of preventive measures and COVID transmission knowledge. In addition to the initial 20% sample in the November survey, another 8% of the respondents were randomly selected to take this section. Six questions about the preventive measures were asked: (1) how many days the respondent was able to keep social distancing whenever they leave home in the past week; (2) number of days the respondent's household member attend the social gathering in the past week; (3) whether the respondent use hands or elbows to cover sneeze; (4) whether the respondent's household owns the mask; 5) whether the respondent's household uses the mask; (6) whether the respondent's household member always wash their hands with soap. Another two questions about the transmission knowledge were asked: (1) whether the respondent believes that asymptomatic patients can also transmit the virus; (2) whether the respondent believes that people can contract with the virus when touching virus-contaminated objects or surfaces. In the November survey, we also added whether the respondent

---

<sup>5</sup>We compute the first principal component of the asset variables and create the asset index. This method has been justified and adopted widely. For example, see Vyas and Kumaranayake (2006) .

believes the virus can transmit through aerosol to the knowledge list since the aerosol has been updated by US Centers for Disease Control and Prevention (CDC) as the "*The principal mode by which people are infected with SARS-CoV-2*".

Trends in adopting preventive measures and virus transmission knowledge can be found through the data collected from May and November surveys. Figure 3.1 shows that most respondents claimed that they have taken some level of prevention including social distance, mask-wearing, and washing hands with soap in both rounds of the survey. The exact numbers can be found in Table A12. There are two notable findings. First, there is a mixing trend in the adoption of preventive measures. On the one hand, comparing November to May, the number of days for social gatherings increases almost twofold. The ratio of households that reported using soaps decreases significantly by around 30 percentage points. Likewise, the proportion of people that own masks and use masks also decrease by around 10 percentage points. On the contrary, the social distance increase with the social gathering suggests the increase of the awareness for self-protection. The second finding is that there is a significant increase in the ratio of people believing the virus can transmit through the asymptomatic patients from around 60% in May to around 75% in November. At the same time, the ratio of people believing touching the contaminated surface or object transmits the virus remains almost the same at the level of 75%. As a result, two ratios converge.

These two findings suggest the possibility that the withdrawal of restrictions could significantly change the adoption of preventive measures such as more social gathering and less mask-wearing. Second, the convergence in knowledge may suggest some level of social learning, such that the gap between two related knowledge shrinks as the learning takes place during the social interaction between villagers. However, albeit interesting, testing these hypotheses is beyond the scope of the paper.

## **Networks**

To study the peer effects in the adoption of the preventive measures, we need to consider the social relationships between the villagers as through the interaction, people suppose to develop their beliefs on the value of taking preventive measures. We assume the interactions taken place in

a network setting where people communicate or simply observe and being observed by others in their networks. We study two types of networks. The first one is the social networks that capture the specific interactions like socializing, issue discussing, and borrowing and lending. We assume that people communicate within these networks. The second type of network is the geographical networks induced by the location where people live. Two households are regarded as geographical neighbors if they live close enough to each other. By living nearby, people may observe the behaviors of each other even no exact social interactions take place.

### **Social Networks**

In the November survey, we elicit a rich set of village social network data of all approached households by asking the respondents to report names about: (1) the households in the village they socialized with; (2) the households in the village they discussed farming issues; (3) the households in the village they discussed health issues; (4) the households in the village they discussed financial issues; (5) the households in the village they would approach when borrowing daily necessities and they would like to lend if asked; and (6) the household in the village they would approach when they need the money and they would like to lend if asked. In 17% of the sample, we also asked whom they discussed COVID-19 with in order to test a hypothesis about the information transmission in the network. We are able to construct fairly precise village social networks using these questions. Table 3.2 summarizes the average number of connections (degree) of each question. Not surprisingly, households on average listed most of the names in the socialization question with on average they reported socializing with 3 other households. The number of named households are statistically the same across all three issue discussing questions and across all two borrow or lending questions. Notably, the total number of connections shown as *Total Degree* is 4, suggesting a high level of overlaps of names across all 6 questions. Another thing worth noticing is the relatively high standard deviation. This is due to the fact that the distribution of degree usually contains "fat" right tail feature with a nontrivial amount of households connect to many other households <sup>6</sup>.

We took multiple measures to reduce the cost of eliciting network data at the same time

---

<sup>6</sup>The highest degree reported is 56 in our data.

guarantee the accuracy of the elicitation. First, we restrict the social network section to six questions as previous studies document high repetition of nomination across different network questions (for example, see Cheng, Huang, and Xing ,2019). Second, to expedite the social network data collection, we developed a new procedure by taking advantage of the digital survey through the tablet. Figure A11 presents an in-survey drop-down menu contains the name of every household in the village we designed. Once the respondent replied a name, the enumerator can search the name by spelling and the drop-down menu will appear and precisely match the name to the household. Additionally, each enumerator was assigned with maps of the villages he surveyed. The household locations are printed on the map using the GPS data collected in the first round survey. The enumerator could geographically locate the correct household based on the description from the respondent if the spelling of the name was wrong so that the name was not presented in the drop-down menu. Breza et al. (2020) calculate that the data entry and matching cost accounts for almost 10% of the survey cost in a social network survey. Our procedure should significantly increase the efficiency by reducing the matching labor subsequently to the data collection and by increasing the accuracy of matching.

### **Geographical Networks**

Another important set of networks studied in previous literature is the geographical networks, in which two households are assumed to be connected if they live close enough to each other. This set of networks can be as well important in the village context as living near to each other may have strong implications of the social interaction. Using the GPS data, we are able to calculate the distance between every pair of households in the data. This means that we are able to know every household's *geographical neighbors* defined as the other households living within a distance. Panel B of Table 3.2 shows that geographical networks are essentially denser than social networks. Even in a small radius of 30 meters, there are on average more than 6 households live within.

The geographical neighbors are also important in the sense that even these households do not necessarily have direct interactions with the household, they are more likely to monitor the household's behaviors, hence may more likely to spread the rumors through the networks. Households



may value the evaluation from the geographical neighbors as much as the social network neighbors as they could interact with the geographical neighbors in the future.

### **Clustering of Preventive Measures and Knowledge**

Combining the preventive measures data and the network data will help to examine the existence of clustering on adoptions, that is, connected individuals are more likely to achieve similar adoption decisions. We use the traditional normalizing practice to construct an adoption index<sup>7</sup>. When the index assigned to each household is higher than the average, we regard the household takes a high-level of preventive measures and regard the household takes a low-level of preventive measures if the index is lower than the average. The network is defined as the combination of communication networks and geographical networks. It is visually clear that there is a significant level of clustering shown by Figure 3.2, in which two typical villages are selected from the sample of 135 villages. The blue nodes are the households with a level of adoption lower than the average and the red nodes are the households with a level of adoption higher than the average. The pooling pattern of same-colored nodes suggests the existence of clustering of adoption decisions in our sample. To test the correlation, we also calculate a preventive and a knowledge index using the standardizing practice and regress the household indexes to the neighbors' mean indexes. According to Table A15, we find that having 1 standard deviation of increase in neighbors mean indexes associates with almost 0.2 standard deviation of increase in household's indexes.

Several potential mechanisms drive the clustering phenomenon<sup>8</sup>. First, the common backgrounds shared by the connected households may also lead to similar adoption decisions. For example, connected households who are also close in the wealth level indicate the same ability to buy the masks and soaps. Connected households could also share a similar level of knowledge as they obtain information from the same source like the local radio station or the poster attached on the streets. They may be influenced by the same local policies that govern their social interactions and force them to take certain preventive measures.

---

<sup>7</sup>First, we standardized each preventive measure. Then we sum up the standardized values and standardized the sum again

<sup>8</sup>Or sociologists referred as *Homophily*

Second, social incentives, or peers' influences, can cause the clustering through two channels: social learning and social pressure. It is very likely that people who are connected interact more often, and hence social learning takes place during the interaction. As for the villagers in our context, they may go to the same Mosque or taking the water from the same well. The communication and learning are taking place during the interactions will lead villagers to agree on the health benefits and financial or physical costs of taking the preventive measures. And such an agreement lead to the same adoption decision. Social pressure, on the other hand, influences the decision indirectly by affecting the future payoff of villagers, possibly due to the reputation effect. Being less conforming will be more costly in the village setting, where the network is rather denser with a higher level of connections between households. For the villagers, the network serves as an important platform for informal risk-sharing so that deviation from the cooperation may result in high costs such as ostracism. Thus, the decision to adopt preventive measures can go beyond the simple calculation of health benefits. Taking preventive measures may signal the type of the households and the recognition of others will be essential for facilitating the future gains from the relation with other households.

Eventually, by studying the social incentives lying beneath the adoption, we explore the potential effective enforcement policy <sup>9</sup>. Some preventive measures features high signaling value such as wearing masks as they are more easily observed. However, other preventive measures such as washing hands with soaps are not as easily observed, hence attached with less signaling value. In these cases, more clever policy needs to be "engineered" to trigger the social incentives.

### **3.3 Estimation**

In this section, we verify if the peers' influences indeed take a significant role in influencing the households' preventive measure adoption and knowledge acquirement. We apply the traditional Linear-in-Means model and achieve the identification of peer effects using the network data and an instrumental variable strategy.

---

<sup>9</sup>This lead to growing literature in social engineering. For example Karing (2018) used differently colored bracelet that signal the number of vaccines a child took and found that the signaling value attached to wearing the bracelet significantly increases the ratio of children that took full set of vaccines in Sierra Leone.

## Notations

Consider  $\mathbf{Y}$  be the vector of a preventive measure with  $i$ 's entry be household  $i$ 's preventive measure  $y_i$ .  $\mathbf{X}$  be the  $N \times K$  matrix in which the  $X_{ik}$  represents the characteristics  $k$  of household  $i$ . The network is represented by adjacency matrix  $\mathbf{A}$ .  $\mathbf{A}_{ij} = 1$  if either  $i$  reports that she interacts with  $j$  or  $j$  reports that she interacts with  $i$  and  $\mathbf{A}_{ij} = 0$  otherwise. Hence the adjacency matrix  $\mathbf{A}$  we are considering is undirected and unweighted. Follow the tradition of the network literature, we assume that the individuals are not connected with themselves, which make the  $\mathbf{A}$  has a diagonal of 0s. Hence the  $i$ -th row of  $\mathbf{A}$  describes the social interactions of household  $i$ . Denote  $N_i$  be the set of households that have direct connection with household  $i$ , that is  $\mathbf{A}_{ij} = \mathbf{A}_{ji} = 1$ . Then the degree of household  $i$ , defined as the total number of direct connections household  $i$  possesses, has the formulation that  $|N_i| = \sum_{j \neq i} \mathbf{A}_{ij}$ . Last of all, a row-normalized adjacency matrix  $\mathbf{G}$  is a adjacency matrix that has its entry normalized by the degree of the corresponding household. So that  $\mathbf{G}_{ij} = \frac{1}{|N_i|}$  if  $\mathbf{A}_{ij} = 1$  and  $\mathbf{G}_{ij} = 0$  if  $\mathbf{A}_{ij} = 0$ .

## Linear-in-Means Model

The Linear-in-Means model has been used widely in the peer effects literature. According to Manski (1993), the model states that three effects steering the individual's decision to adopt a certain behavior: (1) endogenous effects: the effects of neighbors' average adoption level on the household's adoption decision; (2) contextual or exogenous effects: the effects of neighbors' average exogenous characteristics on the household's adoption decision; (3) correlated effects: the effects of the uncommon background or institution that affects the adoption decision. Peer effects are integrated by the endogenous effects and contextual effects. In general, the model takes the form:

$$y_i = \alpha + \beta \frac{\sum_{j \in N_i} y_j}{|N_i|} + \mathbf{X}_i \gamma + \frac{\sum_{j \in N_i} \mathbf{X}_j}{|N_i|} \delta + \epsilon_i. \quad (3.1)$$

Despite the fact that Equation 3.1 is already intuitive in the interpretation of peer effects, this model can also be structurally interpreted as the best response function towards individual maximizing her behavior  $y_i$  given a quadratic utility function which incorporates both the individual outcome and preference to conform to the peers' mean behavior. The conformity term, or the endogenous

effects in the model, reflects the potential existence of two mechanisms, social learning, and social norm, discussed in the previous section.

Though the Linear-in-Means model is appealing for its intuition, Manski discussed the identification failure of these three effects in his paper. First, it is hard to distinguish the endogenous and contextual effects from the correlated effect, and second, the bi-direction of peer effects produces simultaneity: while the household's decision is influenced by neighbors, the household is also influencing the neighbor's decision.

### Estimation Strategies

With the notations defined in the previous section, we are able to write Equation 3.1 in matrix form to help to understand the identification strategy:

$$\mathbf{Y} = \alpha + \beta\mathbf{G}\mathbf{Y} + \mathbf{X}\gamma + \mathbf{G}\mathbf{X}\delta + \epsilon. \quad (3.2)$$

There are two major difficulties to credibly estimate the peer effects parameters  $(\beta, \delta)$ . First, the correlated effects defined as the common backgrounds that affect the connection formations of households and also directly affect the adoption decision. For example, households connect to each other may because they have the similar SES status measured by their wealth level, and wealthier households incline to better protect themselves than poorer households as they are able to purchase the mask, the soap, and are able to keep feeding their children with their savings rather than keep working for food during the pandemic. Households may also obtain the COVID information from the same information source like the local radio station, posters outside of the Mosque, and someone in the village that spread the news or rumors about the pandemic. These information sources will also affect the adoption decision. That is to say, the peer effects can be spurious in the sense that it actually captures the effects of the same characteristics and omitted backgrounds on the preventive measures. To account for such an issue, we control various household characteristics such as the household size, children ratio, and asset index. The identification assumption is that after controlling the household characteristics, the village-level fixed effects  $\mathbf{u}_v$ , and time fixed effects  $\mathbf{u}_t$  we have:

$$\mathbb{E}[\epsilon|\mathbf{X}, \mathbf{G}, \mathbf{u}_v, \mathbf{u}_t] = 0, \quad (3.3)$$

which states that conditional on the households' characteristics and the village-level fixed effect and time fixed effects, the formation of network  $\mathbf{G}$  is strictly exogenous. With this assumption, Bramoullé, Djebbari, and Fortin (2009) show that the correlated effects no longer being an issue to the estimation. This is likely given our setting. Village-level fixed effects account for the unobserved source of information shared by all the villagers that shape the adoption decisions. Time fixed effects, on the other hand, capture the COVID-19 trend that varies across time. And after we control various household characteristics, the network that influencing the adoption decision can be seen as randomly generated.

The second issue is that even we correctly managed the correlated effects, it is still hard to disentangle the influences from the endogenous effects ( $\mathbf{GY}$ ) and from the contextual effects ( $\mathbf{GX}$ ) due to the clear simultaneity structure of the Linear-in-Means model. Fortunately, tools have been developed following the tradition in the econometric literature of the spatial autoregressive model, in which a unit's outcome is assumed to be influenced by the mean outcome of neighbor units located spatially close to it. The common solution involves using the mean characteristics of the neighbor units as the instrument for the mean outcome of the neighbor units. We take the advantages of the network data by following Kelejian and Prucha (1998) and Bramoullé, Djebbari, and Fortin (2009). From the data, we not only see who are the neighbors that influencing the households' decision directly, but also the set of a household's neighbors' neighbors, neighbor's neighbors' neighbors, and so on so forth. These higher-order neighbors are indirectly influencing the households by first influencing the households' neighbors' decision makings and their mean characteristics are likely not correlates with the households' unobserved characteristics after we control the village backgrounds. That is to say, the mean adoption decision of neighbors can be instrumented by the mean characteristics of neighbors' neighbors who are not connected to the household directly. Notation wise, we instrument  $\mathbf{GY}$  by  $\mathbf{G}^2\mathbf{X}$ , where  $\mathbf{G}^2$  is the second-order adjacency matrix where  $\mathbf{G}_{ij}^2 > 0$  implies that household  $i$  and  $j$  are not directly connected with  $\mathbf{G}_{ij} = 0$  but has at least one common neighbor  $k$  such that  $\mathbf{G}_{ik} > 0$  and  $\mathbf{G}_{jk} > 0$ .

Further, We extend the current estimation of the Linear-in-Means model by adding the hetero-

generosity of networks to the model. Specifically, we consider the potential differential influences from the *social networks* and *geographical networks*. Define household  $i$ 's *geographical neighbors* as all the other households living within 50 meters from  $i$ . The distance is calculated using the GPS data we collected from the survey. 50 meters is selected as on average, there are about 3 households from the sampled networks from which we collected the preventive measures. Such distinction induces three types of neighbors for each household: socially connected but not geographically connected (will be referred to as *SN neighbors*), geographically connected but not socially connected (will be referred as *Geo neighbors*), both geographically and socially connected (will be referred as *Geo+SN neighbors*). Such distinction not only leverages the advantages of our data but also to some extent explores the heterogeneity in link strength, that different types of neighbors are likely to have different levels of influences towards the households' decision making. In terms of notation, we write:

$$\mathbf{GY} = [\mathbf{GY}^{sn+geo} \quad \mathbf{GY}^{sn} \quad \mathbf{GY}^{geo}], \text{ and } \mathbf{GX} = [\mathbf{GX}^{sn+geo} \quad \mathbf{GX}^{sn} \quad \mathbf{GX}^{geo}],$$

and the corresponding coefficients will capture the heterogeneous peer effects. Likewise, we instrument  $\mathbf{GY}$  by  $[\mathbf{G}^2\mathbf{X}^{sn+geo} \quad \mathbf{G}^2\mathbf{X}^{sn} \quad \mathbf{G}^2\mathbf{X}^{geo}]$ .

We compare the estimates of parameters in Eq.3.1 using multiple methods. We first show the OLS method to serve as the benchmark reference which assesses the correlation between the own adoption decision and peers' adoption decisions, which at the same time, tests the clustering hypothesis observed from the data. Then we add village-level fixed effects and time fixed effects to the regression to control the potential sorting and hence the correlated effect. Next, we apply the instrumental variable strategy proposed by Bramoullé, Djebbari, and Fortin (2009) which utilize the characteristics of peers' peers as the instrument for the endogenous effect to disentangle the endogenous effects and contextual effects.

We use the data from the November survey for two reasons. First, more households were surveyed in November giving an effective sample size of 3,028 households versus 2,768 households from the May survey. Second, the networks were collected in the November survey so the networks are more likely to reflect the real influences of networks on the adoption decision. Initially, 3,028

households are included in the sample as they complete the survey on the adoption of preventive measures. The isolated households, defined as the household that does not have any connection with other households given the defined networks, are automatically dropped as we can not observe the adoption decisions, **GY**, and mean characteristics, **GX**, of their neighbors. In all three methods, we include household size, the ratio of child household member, the ratio of the male household member, the ratio of the adult household member that finished at least primary school, the level of risk tolerance, and the asset index as the households' characteristics. Summary statistics for the demographics, assets, and network statistics of this smaller sample can be found in Table A13 and Table A14. The standard error is clustered at the village level.

### **3.4 Results**

We estimate the peer effects on various COVID-19 preventive measures and knowledge from three types of neighbors:(1) the set of neighbors that live within a circle with a radius of 50 meters and reported to have social interactions with the household, denoted as *Geo+SN neighbors*; (2) the set of neighbors that live outside of the circle with a radius of 50 meters but reported to have social interactions with the household, denoted as *SN neighbors*; (3) the set of neighbor that lives within the circle with a radius of 50 meters but reported no interaction with the household, denoted as *Geo neighbors*. We find strong positive correlations in all of the measures and knowledge. The correlations disappear when fixed effects are included. Only the peer effects on mask-wearing and knowledge on contaminated surface or object can transmit the virus are detected from the geographical neighbors through IV estimation.

#### **OLS and FE**

In the basic OLS practice, we estimate the correlation between the household's adoption decision and the mean adoption decision of three types of neighbors: who live nearby and had social interactions, who does not live nearby but had social interactions, and who live nearby but does not have social interactions. As shown in each section of Table A16, Table A17, and Table A18, we find strong positive correlations across all measures and knowledge and all types of neighbors. The smallest influence is already around a magnitude of 10 percentage points on own mask from

the geographical neighbors with the largest influence over 30 percentage points on always use soap from social network neighbors. Even for the social network neighbors who do not live close to the households, correlations remain high for the preventive measures and knowledge that are hard to be observed. Another surprising finding is that the highest correlation usually arises from the geographical neighbors who did not report having social interactions with the household in the past. This may suggest that the stay home order, though voluntary in the study area, indeed decrease the mobility of the households and the neighbors in the smaller geographical radius now have a higher influence on the household's decision makings <sup>10</sup>.

Notably, most of high correlations disappear after the fixed effects are included. The OLS estimation with fixed effects is included as the first column of each section of Table 3.3, Table 3.4, and Table 3.5. This suggests that a large component of clustering results from the village-level characteristics like the health posters posted near the local Mosque or schools or the time-related characteristics like the local COVID-19 trend. The change is especially remarkable in the knowledge section. The highly significant positive correlation in knowledge mostly reduces to zero correlation, with only the social network neighbors and geographical neighbors' influences on knowledge of whether touching the contaminated surface or object will transmit the virus remain marginally significantly. This change supports the hypothesis that a similar level of knowledge may be contributed by sharing the same information source. Villagers may get to know about the virus transmission through the local radio station, health posters posted on the streets or near the Mosque, or even through someone who spread the news or rumors about COVID-19. Though such a social learning channel does not explain why there is a knowledge update in transmission from the asymptomatic patient and from aerosol on the village and time dimensions while the knowledge update in transmission from touching the contaminated surface or object remains local.

The influences on social distance and social gathering remain marginally significant from the first type of household who live nearby and had social interactions. The correlation on always use soap when washing hands also decrease significantly, which matches the prediction of the signaling

---

<sup>10</sup>Moreover, the high correlations deliver a more detailed examination to the clustering phenomenon presented by Figure 3.2.



model that the peer effects are stronger for measures that are easily observed.

The decision of owning masks and wearing masks, on the contrary, remains mostly significant. This suggests that even we control various household characteristics and the potential village-level influences like common information sources, there is still a strong local influence on the mask wearing behaviors. However, we cannot conclude that the observed conformity is generated from mimicking the neighbor's adoption decision, which may be caused by the social incentives explained in the model. The simultaneity lies inside the Linear-in-Means model eventually convolute the interpretation of the coefficients, even it remains significant and large after including the fixed effects.

### **Instrumental Variables**

To consistently estimate the coefficients of endogenous effects, which is the influence of neighbors' adoption decision, and coefficients of contextual effects, which is the influence of neighbors' mean characteristics, we use the mean characteristics of neighbors' neighbors to instrument the mean adoption of neighbor. According to Table 3.3, Table 3.4, and Table 3.5, we find wearing masks when leaving home still remains significant with a high effect size. The 10 percentage points increase in the mask wearing of geographical neighbor will increase the household's mask-wearing by 2.3 percentage points. The influences from other types of the neighbor on mask-wearing remain large but insignificant. This suggests that, indeed, not wearing masks may incur the household a costly punishment from the network. Not wearing a mask may not only be associated with a loss of health benefits but also with a negative social image that is detrimental to future relational transactions and risk-sharing. The mask owning, however, drops to zero influence which suggests the observed contextual effects may generate the positive correlations found through OLS regressions. For example, the connected households are similar in their SES status and hence have similar abilities to buy masks.

The IV estimation of peer effects of neighbors on social distance, social gathering, and covering sneezes are noisy, even though the estimated coefficients are large for social gatherings. Not surprisingly, we do not find evidence that households are mimicking the soap usage decision as

washing hands with soap is possibly hard to observe by others and hence (1) households are not likely to find out whether their neighbors are using soaps, and (2) the reputational effect from using soap is essentially zero.

Surprisingly, the knowledge spillover related to whether touching the contaminated surface or object will transmit the virus remains marginally significant with a large magnitude of 20%. The cause of the knowledge spillover of this particular knowledge is unclear, but it indeed casts doubt on the knowledge convergence channel: if there is a knowledge convergence, why do we not observe that the knowledge on all the potential transmission methods converges uniformly?

### **How Sensitive are the Estimates to the Different Definitions of Geographical Neighbors?**

It is reasonable to doubt that instrumental variable estimation of the peer effects may be highly sensitive to the definition of geographical neighbor as there is no hard evidence that households are necessarily influenced by the neighbors of a certain geographical radius. The set of isolated households varies with the radius, with the smaller radius associated with the most data-dropping exercises. We reestimate the IV regression of mask usage and knowledge on surface and object transmission as these are the two variables that we find significantly large endogenous effects. As shown by Figure 3.3, we find that the estimates of influence from Geo+SN neighbors remain stable, though the effects are not statistically distinguishable from 0. However, the influences of geographical neighbors decrease as distance increases. For example, the estimates on the geographical neighbors' influences on the surface or object transmission converge to 0 as the radius expands to 70 meters, while the influence is significant at least 10% level when the radius is no greater than 50 meters. This implies the different levels of geographical neighbors have different levels of influence on different kinds of measures and knowledge. A possible explanation is that the influence of mask-wearing is wider as people may care about most people's views in the village, while the influence on knowledge is constrained to a smaller circle where the communications take place.

### Lagged Peer Behavior

Equation 3.1 assumes that the individual  $i$ 's current behavior is affected by individual  $j$ 's current behavior, while in reality,  $i$ 's behavior may be more likely to be affected by  $j$ 's previous behavior. One idea suggested by previous literature (e.g., Frank and Xu, 2020) is to use lagged behavior to reduce such a bias. We estimate a variation of Equation 3.1 by collaborating  $j$ 's behavior collected from the May survey. In Equation 3.4, the behavior  $y$  is further sub-scripted with the time  $t$ .

$$y_{i,t} = \alpha + \beta \frac{\sum_{j \in N_i} y_{j,t-1}}{|N_i|} + \mathbf{X}_i \gamma + \frac{\sum_{j \in N_i} \mathbf{X}_j}{|N_i|} \delta + \epsilon_i. \quad (3.4)$$

In the estimation,  $y_{i,t}$  stands for the preventive measure adopted by household  $i$  in *November*, and  $y_{j,t}$  stands for the preventive measure adopted by household  $j$  in *May*. The household level characteristics  $\mathbf{X}_i$  come from the census, so they are the same as Equation 3.1. The peers' mean behavior  $\frac{\sum_{j \in N_i} y_{j,t-1}}{|N_i|}$  is instrumented by the second-order neighbors' mean characteristics.

Table A19 shows the FE and IV estimates of peers' prior mean behavior influences,  $\beta$ , on an individual's current adoption of social distancing and social gathering. Table A20 shows peers' prior mean behavior influences an individual's adoption of masks, soap usage, and covering sneezing. Table A21 shows peers' prior mean behavior influences an individual's knowledge about COVID-19 transmission. Unfortunately, we cannot replicate the results obtained from the main specification for at least two reasons. First, given that only a few households were interviewed in both surveys, the sample sizes are not large enough to obtain precise estimates. The sample size was further reduced in the IV estimation due to the sparsity of the sampled networks. Second, the gap between the two surveys is over half a year, while the influence of the pandemic changed rapidly, so the influence across periods may be small.

### 3.5 Framework

Why does geographical neighbors' adoption have large peer effects on mask-wearing? We explore how social incentives lead to peer effects through a social signaling model initially proposed by Benabou and Tirole (2006). The model not only considers the direct health benefit villagers obtained from adopting preventive measures but also captures the reputation effect that is attached

to the adoption behavior. We further discuss how social learning, another important channel, interplays with the signaling process.

### **Social signaling**

In the presence of social norms or pressure, households incur reputational punishment when not abide by the norm, which could possibly drive the peer effects observed from the data. For example, others may regard the household as irresponsible when observing the household members not wearing masks when going outside especially in the village that wearing masks has become a norm. The punishment could involve ostracism that is more significant for the rural people in developing countries, as the absence of formal insurance pushes villagers to rely on each other when encountering risk. Then, signaling a high-type could be beneficial for future transactions and risk-sharing.

But who are the high-type? A high-type could be the person that be more responsible in terms of health or has a greater sense of prosociality. Specific to the COVID setting, a high-type could be the person who intrinsically values the direct benefit of adopting preventive measures over the cost, revealing a greater valuation for the health benefit of themselves, as well as a greater valuation ("warm glow") for positive externalities toward their families, and other villagers. Consider a society with a unit mass of households. Denote household's direct valuation for a preventive measure be  $b \in \mathbb{R}$ . The direct valuation  $b$  absorbs both the health benefit gained from taking the preventive measure and the positive utility from the sense of "warm glow" when taking the preventive measure. It also absorbs the cost of adoption.

$b$  is randomly drawn from a commonly known joint distribution, but the realized values are privately observed only by the household. Consistent with our definition, the high-types are the households with realized cost and benefit satisfied  $b > 0$  and low-types are the households with  $b < 0$ . In other words,  $b$  serves as a latent value that determine the type. The corresponding behavioral interpretation would be even though the set of types can be infinitely large in reality, people usually regarding others within a small number of categories of types, such as high or low.

Households, upon observing the realized benefit, make the adoption decision  $a \in \{0, 1\}$ , with

$a = 1$  stands for adopt and  $a = 0$  stands for not adopt. So a household  $i$ 's direct benefit can be written as:

$$u_i = b_i \cdot a_i, \quad a_i \in \{0, 1\}.$$

In addition to the direct benefit, the household also values the reputation gain linking to the adoption decision. We adopt the formulation from Benabou and Tirole (2006) in which the reputation term has a component of the conditional expectation of household type given their action. In our setting, since we only have two types and two actions, the conditional expectation is the conditional probability of the household being high-type given the household's adoption decision has been observed by others. The adoption decision is only observed with probability  $x$ . This is a realistic assumption as (1) people are assumed to travel less frequently during the pandemic, so it is not likely that a behavior will be observed by all the people; (2) the preventive measures are also varied by their observability. For example, if people wash their hands using a tubewell located inside of their houses, then others most likely will not be able to know whether they are using the soap. At the same time, wearing masks can be observed easily. Furthermore, the household values the reputation at  $\mu$ . Thus, a household with  $\mu = 0$  never bothered by the outsider's views and concern for the outsider's view becomes more heavily when  $\mu$  gets larger. For tractability, assume that all households have the same valuation for the reputation  $\mu$ . To sum up, the household's adoption decision is essentially a utility maximization problem with a reputation component generated by the assessment from others from observing the household's adoption decision:

$$\max_{a_i \in \{0,1\}} b_i \cdot a_i + x_i \mu \Pr(b_i > 0 \mid a_i).$$

One immediate hypothesis of the model is that some low-type households with  $b = \epsilon < 0$  may as well adopt the preventive measures to pool with high-types and earn the reputation as long as the magnitude of reputation is enough to compensate the direct loss  $\epsilon$ . The second hypothesis is the magnitude of the  $\epsilon$  correlates negatively with the observability  $x$ , so that when the decision is easier to be observed, there is a higher incentive for the low-types to mimic the high-types. Because the higher chance of being observed means a higher chance to receive the reputation. To test these

hypotheses, assume that

$$b_i \sim \mathcal{N}(\bar{b}, \sigma_b^2)$$

First, consider the simplest case that all the actions are taken in private, that no one observes except those who took the action. Then under this private setting with  $x = 0$  for all households, we can calculate the adoption ratio, which is empirically interested:

$$\Pr(\text{Adoption}) = \Pr(b > 0) = \Phi\left(\frac{\bar{b}}{\sigma_b}\right),$$

In the public setting, where  $x > 0$ , adopt the preventive measure earns the individual the direct benefit as well as the reputation from the recognition of others. In this case, some low-type households are willing to adopt preventive measures when the reputation gain surpasses the direct loss. As more low-type adopts, the signal of adoption becomes noisier so that reputation gain shrinks. In the Perfect Bayesian Equilibrium, the reputation gain for mimicking should exactly equal to the direct loss due to adoption for the low-type households. We can derive the threshold cutoff  $\bar{\Delta}$  such that low-type households with cost and benefit parameters satisfy  $b \in [\bar{\Delta}, 0)$  are willing to adopt the preventive measures. The threshold  $\bar{\Delta}$  can be found through Bayes' Rule, and we have:

**Proposition 1** The threshold  $\bar{\Delta}$  solves

$$\bar{\Delta} + x\mu\Sigma = 0,$$

where

$$\Sigma = \frac{\Phi(\bar{b}/\sigma_b)}{\Phi((\bar{b} - \bar{\Delta})/\sigma_b)}.$$

Thus

- (1)  $\bar{\Delta} \leq 0$ , thus when having signaling concerns, more people will take preventive measures;
- (2)  $\frac{\partial \bar{\Delta}}{\partial x} < 0$ , indicating that higher observability creates higher incentive for low-type households to adopt;
- (3)  $\frac{\partial \bar{\Delta}}{\partial \bar{b}} < 0$ , indicating the higher expected benefit of the preventive measure associated with a higher level of adoption.

Intuitively, given that  $x\mu\Sigma \geq 0$ ,  $\bar{\Delta} \leq 0$  indicating a weakly larger set of households will take preventive measure. Second, when  $\bar{\Delta}$  becomes smaller (more negative and lower type now adopts),  $\Sigma$  becomes smaller as  $\Phi((\bar{b} - \bar{\Delta})/\sigma_b)$  becomes larger. To maintain the balance,  $x$  becomes larger since we assume  $\mu$  to be fixed. That is to say, as the observability increases, a wider gap between the cost and benefit is permitted for preventive measures. Third, a higher  $\bar{b}$  increases  $\Sigma$ . This is because even though  $\Phi((\bar{b} - \bar{\Delta})/\sigma_b)$  increases with  $\bar{b}$ , the CDF of standard normal distribution is concave after zero hence the increase of  $\Phi((\bar{b} - \bar{\Delta})/\sigma_b)$  is less than the increase of  $\Phi(\bar{b}/\sigma_b)$  as  $\bar{b} - \bar{\Delta} \geq \bar{b}$ . Thus  $\bar{\Delta}$  decreases in order to maintain the equality.

The first prediction shows that whenever a preventive measure can signal the individual types, some low-type individuals will adopt it. The second prediction shows that when a preventive measure is easier to observe, it becomes a better signaling device, which induces more low-type individuals to adopt it. The third prediction tells that in a cluster with a higher valuation of the preventive measure, the people who live inside the cluster will be more likely to adopt the preventive measure.

### 3.6 Conclusion

In this paper, we studied the peer effects in the context of preventive measures and knowledge against COVID-19. The differential level of peer effects on preventive measures eventually suggests that the policy enforcement should design specific to each preventive measure. Measures like mask-wearing are easier to enforce because wearing masks may attach to a high social incentive due to their high visibility. On the contrary, the social incentives for measures like wash hands with soap are smaller as these behaviors are more private than wearing masks, leading these behaviors less likely to be enforced through social pressure. Such a result highlights the importance of education and the spread of knowledge on how to effectively protect self from the virus.

We also find the considerably localized knowledge spillover in the virus transmission through touching the contaminated surface or object, but not the virus transmission through asymptomatic patients and aerosol. Even though the ratio of people knowing these three transmission modes remains at similar levels of 75% - 80%, we are yet to examine why the magnitude of geographical

correlations vary across different knowledge. Hence, the differential level of knowledge spillover could be an interesting research topic in the future.

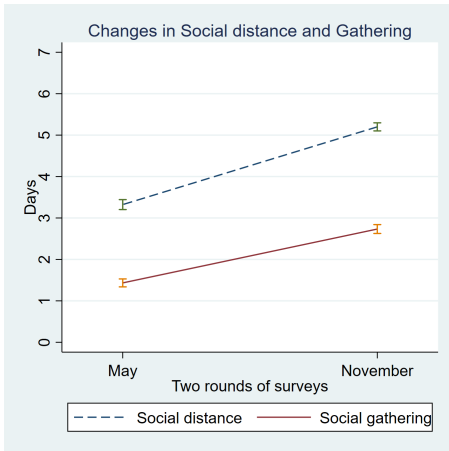
There are still couple of extension can be done to better address our question. First, though widely used during the pandemic, the phone call survey has its limitation in credibility. On the one hand, people may not trust the caller who are not familiar with. On the other hand, there could be a selected attrition due to the phone call survey can only conducted with the households that have a phone. We address the second problem as we collected 2 phone number during the census that maximize the possibility of reaching the households in the following up surveys. Second, the substitability between different preventive measures may be considered in both the empirical analysis and model. For example, mask-wearing may decrease the social distancing as people believe that they are already well-protected. In theory, as each preventive measure is associated with a cost and clearly people are subject to some budget constraint, the rational individual finds the optimal adoption by trading off the direct benefit and the reputation of all of the possible preventive measures as one optimization problem instead of separate and independent problems.

In general, the existence of social incentives further suggests the importance of correcting the public beliefs on the return of preventive measures. This is because people earn social benefits like reputation from acting similarly to others, while misinformation may cause taking less to no preventive measures to become the norm. As a result, actively taking preventive measures incurs the social cost instead of the benefit in this case. It would be important to study how to correct the misbelief and create positive social incentives that encourage preventive measures against COVID-19.

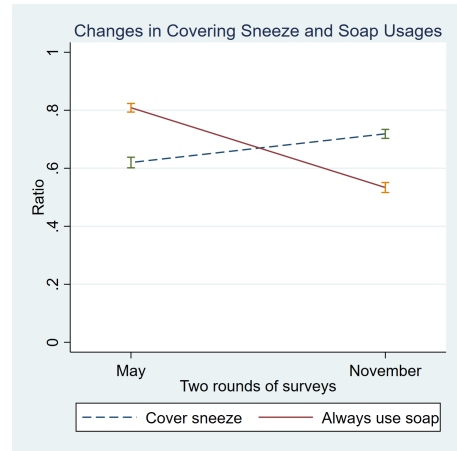


### 3.7 Figures

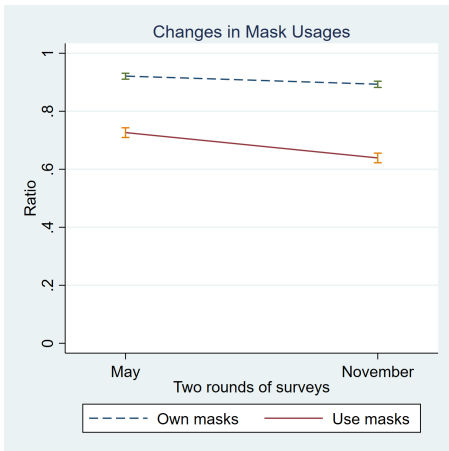
Figure 3.1 Changes in COVID-19 Preventive Measures and Transmission Knowledge



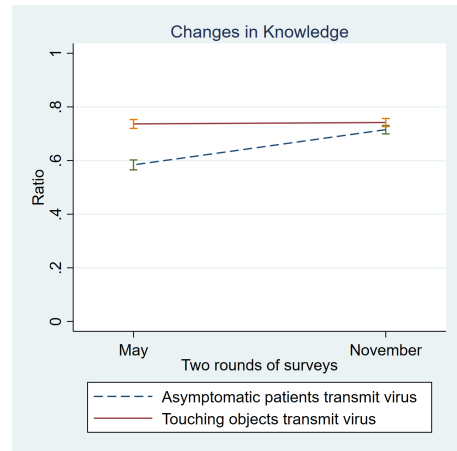
(a) Changes in Social Distance and Gathering



(b) Changes in Covering Sneeze and Soap Usages



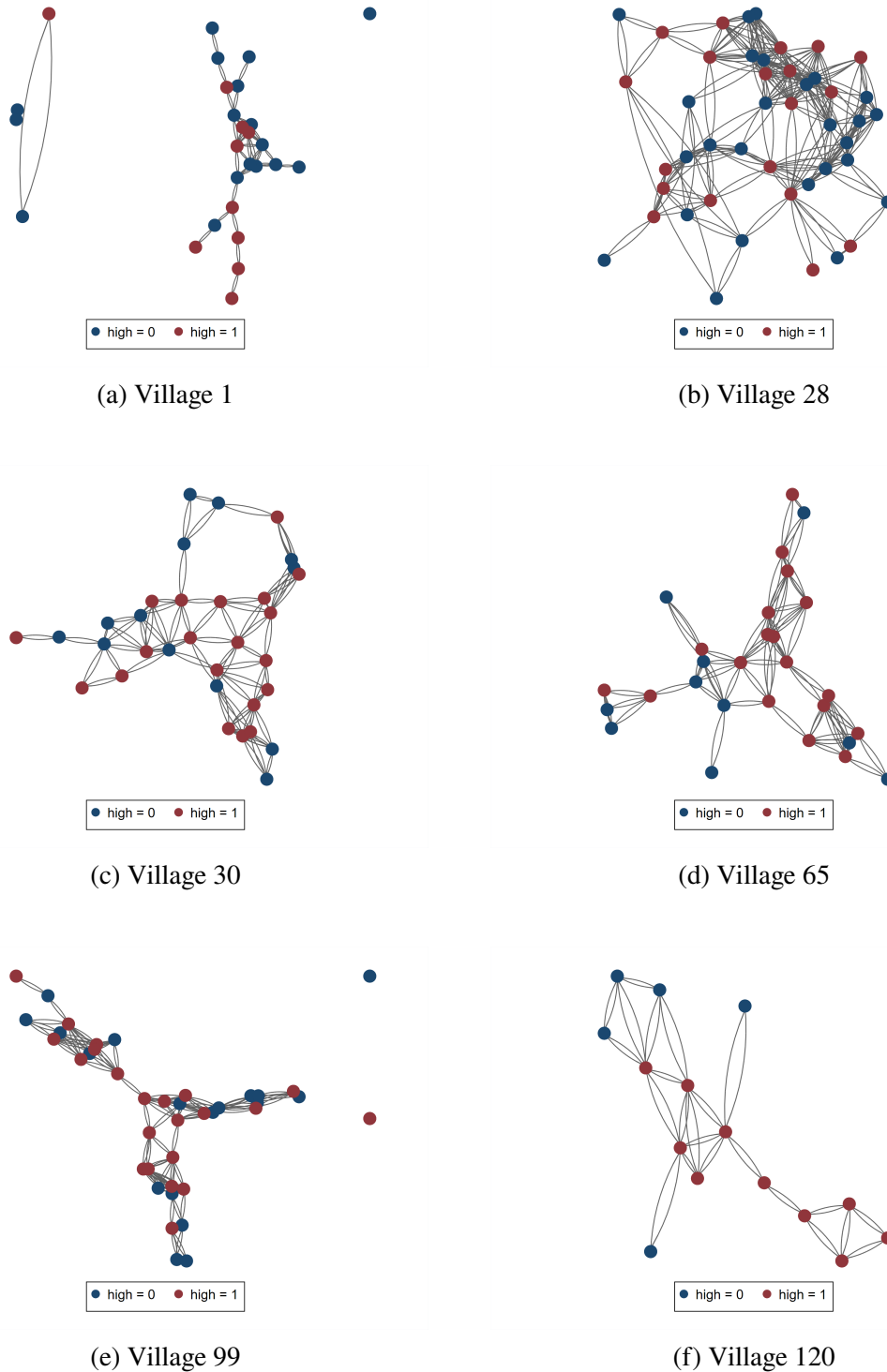
(c) Changes in Mask Ownership and Mask Usages



(d) Changes in COVID-19 Transmission Knowledge

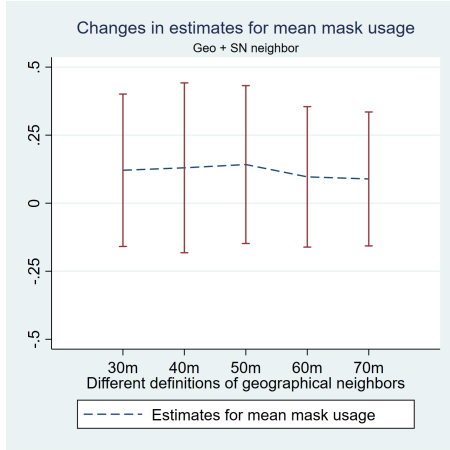
Note: Even though the survey is named as a May or November survey, some of the respondents are approached in the previous or subsequent months due to the large sample size. Except for social distance and social gathering, which were recorded by number of days practiced in the previous week, all other preventive measures and knowledge are binary choices

Figure 3.2 Clustering of Preventive Measure Adoptions in Villages Subnetworks

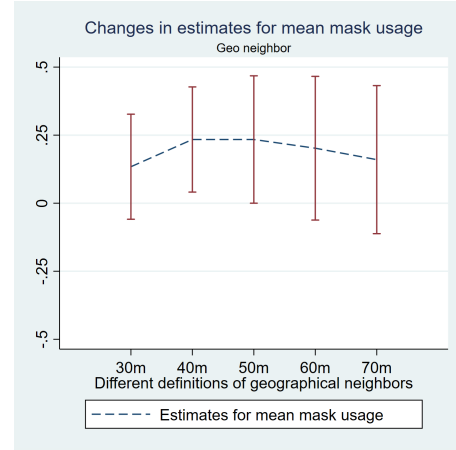


Note: Blue (Red) nodes stand for the households with the adoption index higher (lower) than average. The networks contain fewer nodes and fewer connections as we only surveyed 28% of households on preventive measures, leading to having subnetworks instead of full networks.

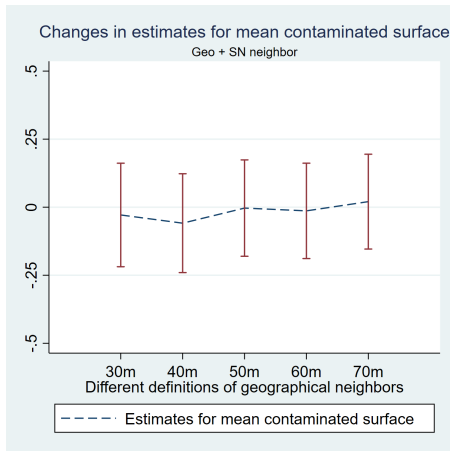
Figure 3.3 Changes in IV Estimates Under Different Geographical Neighbors



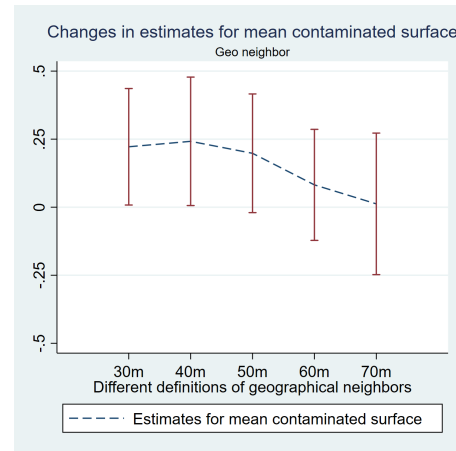
(a) Changes in Estimates for Geo+SN Neighbors' Influences on Mean Mask Usages



(b) Changes in Estimates for Geo Neighbors' Influences on Mean Mask Usages



(c) Changes in Estimates for Geo+SN Neighbors' Influences on Contaminated Surface Transmission



(d) Changes in Estimates for Geo Neighbors' Influences on Contaminated Surface Transmission

Note: The confidence intervals are constructed at 5% significance level. The effective sample size varies with the definition of geographical neighbors because a different radius will include different amount of households. For 30 meters, we have 2,478 households. For 40 meters, we have 2,677 households. For 60 meters, we have 2,874 households, and for 70 meters, we have 2,911 households.

### 3.8 Tables

Table 3.1 Demographics and Assets

	N	mean	sd
<b>Panel A: Demographics</b>			
Household size	11,933	4.62	1.947
Average age	11,933	27.364	11.585
Male ratio	11,933	0.459	0.197
Child ratio	11,933	0.401	0.228
Primary education ratio	11,933	0.297	0.266
HH head is female	11,933	0.254	0.435
HH head education (yrs)	11,917	3.345	4.041
HH head age	11,924	45.37	14.36
Tolerance of health risk	10,854	1.855	1.165
Tolerance of general risk	11,391	2.055	1.188
<b>Panel B: Assets</b>			
Rooms	11,930	2.764	1.259
Whether access to electricity	11,933	0.992	0.0879
Fans	11,931	2.372	1.183
Mobilephones	11,930	1.938	1.107
Smartphones	11,919	0.867	0.959
Cycles or rickshaws	11,929	0.111	0.364
Motorcycles	11,930	0.0632	0.272
Vehicles	11,909	0.0154	0.201
TVs	11,930	0.489	0.531
Computers	11,930	0.0248	0.181
Refrigerators	11,930	0.525	0.524
Acres of agricultural land	11,706	3.180	24.31
Private wells	11,933	0.817	0.520

Table 3.2 Social Networks

	N	mean	sd
<b>Panel A: Social Networks</b>			
Socialization	11,933	3.017	2.495
Discuss farming issues	11,933	0.788	1.433
Discuss health issues	11,933	0.879	1.433
Discuss financial issues	11,933	0.84	1.387
Borrow or lend daily necessities	11,933	1.336	1.706
Borrow or lend money	11,933	1.162	1.532
<i>Total Degree</i>	11,933	4.086	3.970
<b>Panel B: Geographical Networks</b>			
<= 30m	11,933	6.031	3.933
<= 50m	11,933	13.405	7.551
<= 100m	11,933	34.321	17.159
<= 200m	11,933	69.692	29.781

Notes: Degree is defined as the number of connections each household have. This is to say that a household can have a higher degree than the number of names it reports as the non-listed households may list the household.

Table 3.3 FE and IV Estimations of Peer Effects on Social Distance, Social Gathering, and Cover Sneeze

	Social distance		Social gathering		Cover sneeze	
	(1) FE	(2) IV	(3) FE	(4) IV	(5) FE	(6) IV
<i>Social Distance</i>						
Geo+SN	0.0778*	0.0328				
	(0.0406)	(0.0936)				
SN	0.0632	-0.0768				
	(0.0642)	(0.173)				
Geo	0.118**	0.0392				
	(0.0451)	(0.0696)				
<i>Social Gathering</i>						
Geo+SN			0.0827**	0.210		
			(0.0404)	(0.283)		
SN			0.0709	0.599		
			(0.0578)	(0.507)		
Geo			0.0565	0.279		
			(0.0412)	(0.340)		
<i>Cover Sneeze</i>						
Geo+SN					0.0247	-0.0724
					(0.0459)	(0.0795)
SN					0.175***	0.295
					(0.0654)	(0.274)
Geo					0.153***	0.110
					(0.0458)	(0.109)
Observations	2,815	2,815	2,815	2,815	2,815	2,815
Village FE	NO	YES	NO	YES	NO	YES
Date FE	NO	YES	NO	YES	NO	YES

Table 3.4 FE and IV Estimations of Peer Effects on Own Masks, Wear Masks, and Wash Hands with Soap

	Own Mask		Use Mask		Always Soap	
	(7) FE	(8) IV	(9) FE	(10) IV	(11) FE	(12) IV
<i>Own Mask</i>						
Geo+SN	0.0903** (0.0441)	0.0518 (0.0581)				
SN	0.209*** (0.0670)	0.0846 (0.143)				
Geo	0.0160 (0.0320)	0.0600 (0.0574)				
<i>Use Mask</i>						
Geo+SN			0.0763** (0.0343)	0.142 (0.145)		
SN			0.185*** (0.0512)	0.263 (0.203)		
Geo			0.0971** (0.0463)	0.234** (0.117)		
<i>Always Soap</i>						
Geo+SN					0.0154 (0.0397)	0.0679 (0.138)
SN					0.114* (0.0686)	-0.122 (0.416)
Geo					0.00272 (0.0453)	-0.0492 (0.144)
Observations	2,815	2,815	2,815	2,815	2,815	2,815
Village FE	NO	YES	NO	YES	NO	YES
Date FE	NO	YES	NO	YES	NO	YES

Table 3.5 FE and IV Estimations of Peer Effects on COVID-19 Virus Transmission Knowledge

	Asymptomatic		Surface or Object		Aerosol	
	(13) FE	(14) IV	(15) FE	(16) IV	(17) FE	(18) IV
<i>Asymptomatic</i>						
Geo+SN	0.0481 (0.0457)	-0.0200 (0.0914)				
SN	-0.0232 (0.0674)	-0.00768 (0.171)				
Geo	0.0652 (0.0424)	0.159 (0.112)				
<i>Surface or Object</i>						
Geo+SN			0.0334 (0.0391)	-0.0033 (0.0885)		
SN			0.116* (0.0664)	0.0242 (0.217)		
Geo			0.0692* (0.0385)	0.198* (0.109)		
<i>Aerosol</i>						
Geo+SN					-0.00668 (0.0415)	-0.0616 (0.0797)
SN					0.0664 (0.0670)	0.0308 (0.176)
Geo					-0.0166 (0.0453)	0.0109 (0.0887)
Observations	2,815	2,815	2,815	2,815	2,815	2,815
Village FE	NO	YES	NO	YES	NO	YES
Date FE	NO	YES	NO	YES	NO	YES



## BIBLIOGRAPHY

- Abdul, K. S. M., Jayasinghe, S. S., Chandana, E. P., Jayasumana, C., & De Silva, P. M. C. (2015). Arsenic and human health effects: A review. *Environmental toxicology and pharmacology*, 40(3), 828–846.
- Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110(3), 629–676.
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211–236.
- Ashraf, N., Karlan, D., & Yin, W. (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *The Quarterly Journal of Economics*, 121(2), 635–672.
- Attanasio, O., Barr, A., Cardenas, J. C., Genicot, G., & Meghir, C. (2012). Risk pooling, risk preferences, and social networks. *American Economic Journal: Applied Economics*, 4(2), 134–67.
- Banerjee, A., Alsan, M., Breza, E., Chandrasekhar, A. G., Chowdhury, A., Duflo, E., Goldsmith-Pinkham, P., & Olken, B. A. (2020). *Messages on covid-19 prevention in india increased symptoms reporting and adherence to preventive behaviors among 25 million recipients with similar effects on non-recipient members of their communities* (tech. rep.). National Bureau of Economic Research.
- Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144), 1236498.
- Barnwal, P., van Geen, A., von der Goltz, J., & Singh, C. K. (2017). Demand for environmental quality information and household response: Evidence from well-water arsenic testing. *Journal of Environmental Economics and Management*, 86, 160–192.
- Barnwal, P., Yao, Y., Wang, Y., Juy, N. A., Raihan, S., Haque, M. A., & van Geen, A. (2021). Assessment of excess mortality and household income in rural bangladesh during the covid-19 pandemic in 2020. *JAMA network open*, 4(11), e2132777–e2132777.
- Barr, A., & Genicot, G. (2008). Risk sharing, commitment, and information: An experimental analysis. *Journal of the European Economic Association*, 6(6), 1151–1185.
- Bazzi, S., Gaduh, A., Rothenberg, A. D., & Wong, M. (2019). Unity in diversity? how intergroup contact can foster nation building. *American Economic Review*, 109(11), 3978–4025.
- Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2018). Can network theory-based targeting increase technology adoption?

- Beaman, L. A. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the us. *The Review of Economic Studies*, 79(1), 128–161.
- Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American economic review*, 96(5), 1652–1678.
- Braghieri, L., Levy, R., & Makarin, A. (2022). Social media and mental health. *American Economic Review*, 112(11), 3660–3693.
- Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of econometrics*, 150(1), 41–55.
- Breza, E., & Chandrasekhar, A. G. (2019). Social networks, reputation, and commitment: Evidence from a savings monitors experiment. *Econometrica*, 87(1), 175–216.
- Breza, E., Chandrasekhar, A. G., McCormick, T. H., & Pan, M. (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, 110(8), 2454–84.
- Bursztyn, L., Callen, M., Ferman, B., Gulzar, S., Hasanain, A., & Yuchtman, N. (2020). Political identity: Experimental evidence on anti-americanism in pakistan. *Journal of the European Economic Association*, 18(5), 2532–2560.
- Bursztyn, L., Egorov, G., Enikolopov, R., & Petrova, M. (2019). *Social media and xenophobia: Evidence from russia* (tech. rep.). National Bureau of Economic Research.
- Bursztyn, L., Egorov, G., Haaland, I., Rao, A., & Roth, C. (2023). Justifying dissent. *The Quarterly Journal of Economics*, 138(3), 1403–1451.
- Bursztyn, L., & Jensen, R. (2015). How does peer pressure affect educational investments? *The quarterly journal of economics*, 130(3), 1329–1367.
- Bursztyn, L., & Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, 9, 131–153.
- Cardenas, J. C., Stranlund, J., & Willis, C. (2000). Local environmental control and institutional crowding-out. *World development*, 28(10), 1719–1733.
- Carson, R. T., Koundouri, P., & Nauges, C. (2011). Arsenic mitigation in bangladesh: A household labor market approach. *American Journal of Agricultural Economics*, 93(2), 407–414.
- Case, A. C., & Katz, L. F. (1991). The company you keep: The effects of family and neighborhood on disadvantaged youths.
- Cheng, C., Huang, W., & Xing, Y. (2019). A theory of multiplexity: Sustaining cooperation with

multiple relationships.

- Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4), 370–379.
- Coate, S., & Ravallion, M. (1993). Reciprocity without commitment: Characterization and performance of informal insurance arrangements. *Journal of development Economics*, 40(1), 1–24.
- Conley, T. G., & Udry, C. R. (2010). Learning about a new technology: Pineapple in Ghana. *American economic review*, 100(1), 35–69.
- Corno, L., La Ferrara, E., & Burns, J. (2022). Interaction, stereotypes, and performance: Evidence from South Africa. *American Economic Review*, 112(12), 3848–3875.
- Dahl, G. B., Kotsadam, A., & Rooth, D.-O. (2021). Does integration change gender attitudes? The effect of randomly assigning women to traditionally male teams. *The Quarterly journal of economics*, 136(2), 987–1030.
- DellaVigna, S., List, J. A., & Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The quarterly journal of economics*, 127(1), 1–56.
- DellaVigna, S., List, J. A., Malmendier, U., & Rao, G. (2016). Voting to tell others. *The Review of Economic Studies*, 84(1), 143–181.
- Dickinson, D., & Villeval, M.-C. (2008). Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic Behavior*, 63(1), 56–76.
- Drago, F., Mengel, F., & Traxler, C. (2020). Compliance behavior in networks: Evidence from a field experiment. *American Economic Journal: Applied Economics*, 12(2), 96–133.
- Duncan, G. J., Boisjoly, J., Kremer, M., Levy, D. M., & Eccles, J. (2005). Peer effects in drug use and sex among college students. *Journal of abnormal child psychology*, 33(3), 375–385.
- Ederer, F., Pinkham-Goldsmith, P., & Jensen, K. (2023). Anonymity and identity online.
- Enikolopov, R., Makarin, A., & Petrova, M. (2020). Social media and protest participation: Evidence from Russia. *Econometrica*, 88(4), 1479–1514.
- Evans, W. N., Oates, W. E., & Schwab, R. M. (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy*, 100(5), 966–991.
- Fafchamps, M., & Gubert, F. (2007). The formation of risk sharing networks. *Journal of development Economics*, 83(2), 326–350.

- Fafchamps, M., & Lund, S. (2003). Risk-sharing networks in rural philippines. *Journal of development Economics*, 71(2), 261–287.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4), 1645–1692.
- Falk, A., Becker, A., Dohmen, T. J., Huffman, D., & Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences.
- Feigenberg, B., Field, E., & Pande, R. (2013). The economic returns to social interaction: Experimental evidence from microfinance. *Review of Economic Studies*, 80(4), 1459–1483.
- Fendorf, S., Michael, H. A., & van Geen, A. (2010). Spatial and temporal variations of groundwater arsenic in south and southeast asia. *Science*, 328(5982), 1123–1127.
- for Disease Control, C., Prevention, C. f. D. C., & Prevention. (n.d.). Scientific brief: Sars-cov-2 and potential airborne transmission.
- Frank, K. A., & Xu, R. (2020). Causal inference for social network analysis. *The Oxford handbook of social networks*, 288–310.
- Funk, P. (2010). Social incentives and voter turnout: Evidence from the swiss mail ballot system. *Journal of the European Economic Association*, 8(5), 1077–1103.
- Gabore, S. M. (2020). Western and chinese media representation of africa in covid-19 news coverage. *Asian Journal of Communication*, 30(5), 299–316.
- Gaviria, A., & Raphael, S. (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics*, 83(2), 257–268.
- Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3), 447–456.
- Gelman, A., Trevisani, M., Lu, H., & Van Geen, A. (2004). Direct data manipulation for local decision analysis as applied to the problem of arsenic in drinking water from tube wells in bangladesh. *Risk Analysis: An International Journal*, 24(6), 1597–1612.
- George, C. M., Zheng, Y., Graziano, J. H., Rasul, S. B., Hossain, Z., Mey, J. L., & Van Geen, A. (2012). Evaluation of an arsenic test kit for rapid well screening in bangladesh. *Environmental science & technology*, 46(20), 11213–11219.
- Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American political Science review*, 33–48.
- Giné, X., Karlan, D., & Zinman, J. (2010). Put your money where your butt is: A commitment

- contract for smoking cessation. *American Economic Journal: Applied Economics*, 2(4), 213–35.
- Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from# brexit and# uselection. *European Economic Review*, 136, 103772.
- Hermes, N., Lensink, R., & Mehrteab, H. T. (2005). Peer monitoring, social ties and moral hazard in group lending programs: Evidence from eritrea. *World development*, 33(1), 149–169.
- Heß, S., Jaimovich, D., & Schündeln, M. (2018). Development projects and economic networks: Lessons from rural gambia. *The Review of Economic Studies*.
- Howard, J., Huang, A., Li, Z., Tufekci, Z., Zdimal, V., van der Westhuizen, H.-M., von Delft, A., Price, A., Fridman, L., Tang, L.-H., et al. (2021). An evidence review of face masks against covid-19. *Proceedings of the National Academy of Sciences*, 118(4).
- Jamil, N. B., Feng, H., Ahmed, K. M., Choudhury, I., Barnwal, P., & Van Geen, A. (2019). Effectiveness of different approaches to arsenic mitigation over 18 years in araihar, bangladesh: Implications for national policy. *Environmental Science & Technology*, 53(10), 5596–5604.
- Karing, A. (2018). Social signaling and childhood immunization: A field experiment in sierra leone. *University of California, Berkeley*.
- Kaur, S., Kremer, M., & Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, 123(6), 1227–1277.
- Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), 99–121.
- King, G., Pan, J., & Roberts, M. E. (2013). How censorship in china allows government criticism but silences collective expression. *American political science Review*, 107(2), 326–343.
- King, G., Pan, J., & Roberts, M. E. (2017). How the chinese government fabricates social media posts for strategic distraction, not engaged argument. *American political science review*, 111(3), 484–501.
- Kinnan, C. (2022). Distinguishing barriers to insurance in thai villages. *Journal of Human Resources*, 57(1), 44–78.
- Kocherlakota, N. R. (1996). Implications of efficient risk sharing without commitment. *The Review of Economic Studies*, 63(4), 595–609.
- Kremer, M., & Levy, D. (2008). Peer effects and alcohol use among college students. *Journal of Economic perspectives*, 22(3), 189–206.

- Kremer, M., & Miguel, E. (2007). The illusion of sustainability. *The Quarterly journal of economics*, 122(3), 1007–1065.
- Krishnan, P., & Patnam, M. (2014). Neighbors and extension agents in ethiopia: Who matters more for technology adoption? *American Journal of Agricultural Economics*, 96(1), 308–327.
- Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review*, 111(3), 831–870.
- Ligon, E. (1998). Risk sharing and information in village economies. *The Review of Economic Studies*, 65(4), 847–864.
- Ligon, E., Thomas, J. P., & Worrall, T. (2002). Informal insurance arrangements with limited commitment: Theory and evidence from village economies. *The Review of Economic Studies*, 69(1), 209–244.
- Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, 111(6), 1807–1844.
- Lundborg, P. (2006). Having the wrong friends? peer effects in adolescent substance use. *Journal of health economics*, 25(2), 214–233.
- Madajewicz, M., Pfaff, A., Van Geen, A., Graziano, J., Hussein, I., Momotaj, H., Sylvi, R., & Ahsan, H. (2007). Can information alone change behavior? response to arsenic contamination of groundwater in bangladesh. *Journal of development Economics*, 84(2), 731–754.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3), 531–542.
- Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1), 112–45.
- Meghir, C., Mobarak, A. M., Mommaerts, C., & Morten, M. (2020). Migration and informal insurance. Available at SSRN 3684510.
- Mosquera, R., Odunowo, M., McNamara, T., Guo, X., & Petrie, R. (2020). The economic effects of facebook. *Experimental Economics*, 23, 575–602.
- Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the us labor market. *The Quarterly Journal of Economics*, 118(2), 549–599.
- Nakajima, R. (2007). Measuring peer effects on youth smoking behaviour. *The Review of Economic Studies*, 74(3), 897–935.
- Pfaff, A., Schoenfeld Walker, A., Ahmed, K. M., & van Geen, A. (2017). Reduction in exposure to arsenic from drinking well-water in bangladesh limited by insufficient testing and awareness.

- Journal of Water, Sanitation and Hygiene for Development*, 7(2), 331–339.
- Pitt, M. M., Rosenzweig, M. R., & Hassan, M. N. (2021). Identifying the costs of a public health success: Arsenic well water contamination and productivity in bangladesh. *The Review of Economic Studies*, 88(5), 2479–2526.
- Qin, B., Strömberg, D., & Wu, Y. (2017). Why does china allow freer social media? protests versus surveillance and propaganda. *Journal of Economic Perspectives*, 31(1), 117–140.
- Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in delhi schools. *American Economic Review*, 109(3), 774–809.
- Schindler, D., & Westcott, M. (2021). Shocking racial attitudes: Black gis in europe. *The Review of Economic Studies*, 88(1), 489–520.
- Singer, P. (2011). Visible man: Ethics in a world without secrets. *Harper's Magazine*, 34, 47.
- Sojourner, A. (2013). Identification of peer effects with missing peer data: Evidence from project star. *The Economic Journal*, 123(569), 574–605.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of personality and social psychology*, 69(5), 797.
- Stiglitz, J. E. (1990). Peer monitoring and credit markets. *The world bank economic review*, 4(3), 351–366.
- Tarozzi, A., Maertens, R., Ahmed, K. M., & Van Geen, A. (2021). Demand for information on environmental health risk, mode of delivery, and behavioral change: Evidence from sonargaon, bangladesh. *The World Bank Economic Review*, 35(3), 764–792.
- Townsend, R. M. (1994). Risk and insurance in village india. *Econometrica: journal of the Econometric Society*, 539–591.
- Trogdon, J. G., Nonnemaker, J., & Pais, J. (2008). Peer effects in adolescent overweight. *Journal of health economics*, 27(5), 1388–1399.
- Udry, C. (1994). Risk and insurance in a rural credit market: An empirical investigation in northern nigeria. *The Review of Economic Studies*, 61(3), 495–526.
- Van Geen, A., Ahmed, E. B., Pitcher, L., Mey, J. L., Ahsan, H., Graziano, J. H., & Ahmed, K. M. (2014). Comparison of two blanket surveys of arsenic in tubewells conducted 12 years apart in a 25 km<sup>2</sup> area of bangladesh. *Science of the Total Environment*, 488, 484–492.
- van Geen, A., Ahsan, H., Horneman, A. H., Dhar, R. K., Zheng, Y., Hussain, I., Ahmed, K. M., Gelman, A., Stute, M., Simpson, H. J., et al. (2002). Promotion of well-switching to mitigate

- the current arsenic crisis in bangladesh. *Bulletin of the World Health Organization*, 80(9), 732–737.
- Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. *Health policy and planning*, 21(6), 459–468.
- Wu, A. H. (2020). Gender bias among professionals: An identity-based interpretation. *Review of Economics and Statistics*, 102(5), 867–880.
- Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P.-Y. (2023). Supporting qualitative analysis with large language models: Combining codebook with gpt-3 for deductive coding. *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, 75–78.
- Young, H. P. (2015). The evolution of social norms. *economics*, 7(1), 359–387.
- Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2023). Can large language models transform computational social science? *arXiv preprint arXiv:2305.03514*.



## APPENDIX A

### CHAPTER 1

#### VPN in China

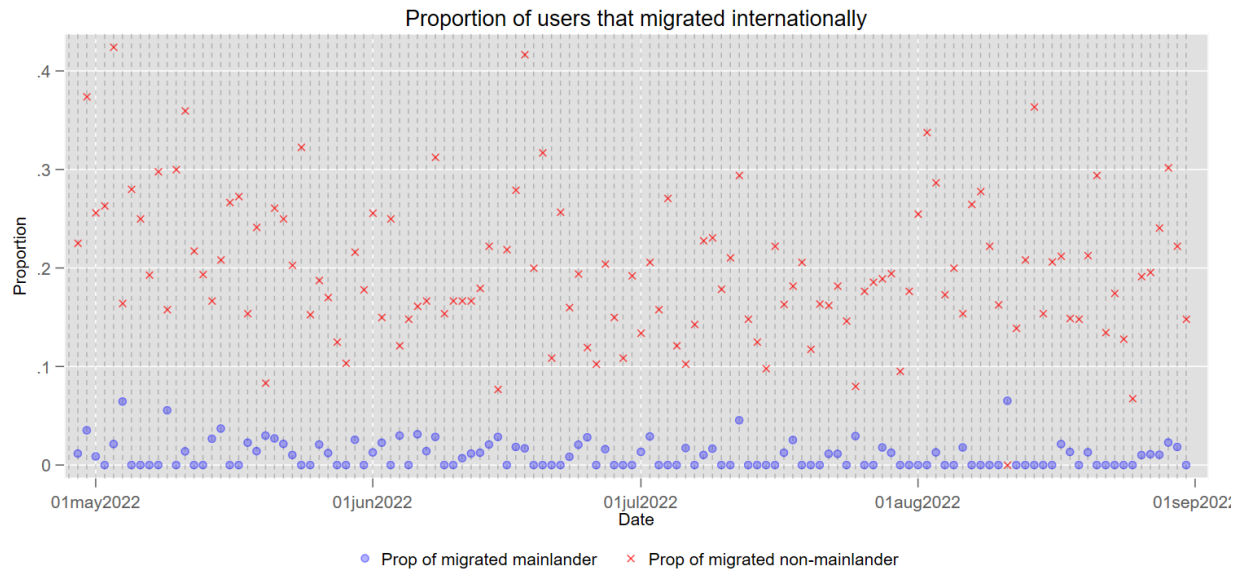
Due to the expansion of the Great Firewall project, through which the Chinese government barred mainland Chinese access to a large set of foreign platforms, including YouTube, Facebook, X (formerly Twitter), and many other highly visited websites, Chinese netizens anecdotally increased their use of VPN to bypass the restriction. In the following discussion, we discuss whether the usage or manipulation of VPN has influenced our key results.

First, most contemporary VPN applications have configured with *Dynamic Split Tunneling* in default, automating users to selectively route certain types of traffic through the VPN while letting other traffic go directly to the internet. In our case, visiting domestic websites such as Sina Weibo will route the traffic directly to the website without first changing the IP address. Therefore, the IP location exhibited on Weibo should be mostly correct unless the users intentionally terminate the split tunneling and disguise the trace entirely through VPN.

Second, we take a small sample of random commenters who made the comment between April 28 and August 31, 2022. This sample includes 12,580 mainland users and 8,154 non-mainland users. The location of these commenters is identified based on the tag in their comments. We obtain a second source of their IP locations by scraping their personal Weibo profile page in early 2023. The IP location in the profile can be different and more accurate than the IP location in the comment, as the IP location in the profile uses the user's most frequently used IP address in the past month, while IP location in the comment uses the instantaneous IP address.

Figure A1 provides suggestive evidence against the potential of VPN manipulation as a driving force. This figure shows the proportion of commenters with an IP location in the profile that differs internationally from the IP location in the comment. We find the proportion of migration does not substantially change over time. Suppose that the policy motivates mainland users to terminate the VPN and stop disguising themselves as non-mainland users, we would expect commenters with non-mainland comment-IP locations who commented immediately after April 28th are more likely

Figure A1 "International Migration" of IP Location



to 'return' as mainlanders according to their profile. However, we do not observe a salient trend of reduction of migration rate according to the distribution of red crosses in Figure A1.

## Proof of Proposition 1

In this section, we prove the uniqueness of the threshold for the unobserved identity case.

In the unobserved case,  $L$  and  $F$  share the same threshold  $\hat{b} = \kappa S$ . Define  $\tilde{b} = E(b_i | b \geq \hat{b})$  and  $\bar{b} = E(b_i)$ , then observe that:

$$\begin{aligned}
S &= \mu \int_{b: d_L^S(b)=0} |E(b_i | d_i = 1) - b| dF_{bL} + (1 - \mu) \int_{b: d_F^S(b)=0} |E(b_i | d_i = 1) - b| dF_{bF} \\
&= Pr(i \in L \cap b_i \leq \hat{b}) \tilde{b} - Pr(i \in L \cap b_i \leq \hat{b}) E(b_i | i \in L \cap b_i \leq \hat{b}) \\
&\quad + Pr(i \in F \cap b_i \leq \hat{b}) \tilde{b} - Pr(i \in F \cap b_i \leq \hat{b}) E(b_i | i \in F \cap b_i \leq \hat{b}) \\
&= Pr(b_i \leq \hat{b}) \tilde{b} - Pr(b_i \leq \hat{b}) E(b_i | b_i \leq \hat{b}) \\
&= \tilde{b} - [Pr(b_i \geq \hat{b}) \tilde{b} + Pr(b_i \leq \hat{b}) E(b_i | b_i \leq \hat{b})] \\
&= \tilde{b} - \bar{b}
\end{aligned}$$

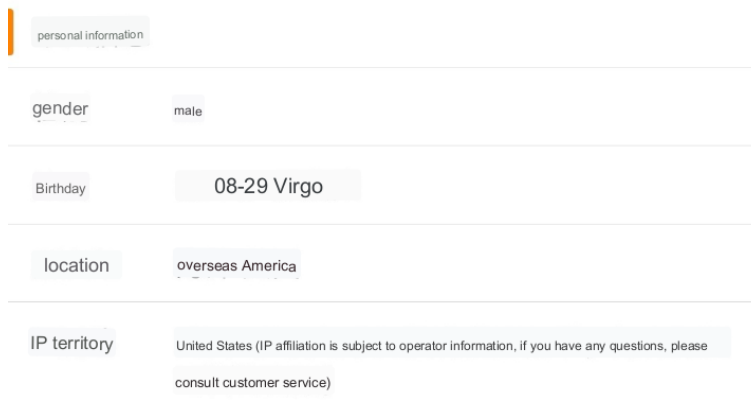
Thus we have:

$$\hat{b} = \kappa(\tilde{b} - \bar{b})$$

Note that  $\frac{d}{db} \tilde{b} = \frac{1}{2}$  and  $\kappa \in (0, 1)$ , we have a unique solution for  $\hat{b}$ .

Figure A2 Examples of IP Tag in Profile and in Comment

(a) In Profile



personal information

---

gender male

---

Birthday 08-29 Virgo

---

location overseas America

---

IP territory United States (IP affiliation is subject to operator information, if you have any questions, please consult customer service)

Note: The *IP territory* is the IP location

(b) In Comment



 : I hope the old man is safe.  
22-4-27 23:20

 : Now there are no friends in the United States who have the new crown, so there are really no friends.  
22-4-28 00:36 from America

Table A1 Summary Statistics of Posts, Posters, and Comments

	Mean	SD	Min	p25	p75	Max
<b>Panel A. Post</b>				<b>N = 17,889</b>		
# of likes	7214.7	28678	0	235	3357	944,024
# of reposts	620.4	3365	0	14	280	310,457
# of comments	635	2303.1	1	61	463	85,617
<b>Panel B. Poster</b>				<b>N = 509</b>		
# of followers (10,000)	181.4	262.6	0.7	23.9	23.9	2,475.1
# of likes	9,049.6	15,678.4	63.9	1,173.8	9,965.2	135,160.3
# of reposts	827.3	1,470.2	0.1	87.5	848.3	848.3
# of comments	703.1	1,123.4	17.6	183.0	775.0	11,330.9
<b>Panel C. Comment</b>				<b>N=2,795,583</b>		
# of likes	11.4	338.2	0	0	1	127,683
# of first-stage comments likes	16.5	436.5	9	0	1	127,683

Notes: This table summarizes the main characteristics of Posts, Posters, and Comments in our data. The post statistics show the distribution of the number of likes, reposts, and comments received by the posts. The poster statistics show the distribution of the *total* number of likes, reposts, and comments received by summing up the statistics of posts written by each poster. The comment statistics show the distribution of the number of likes, reposts, and comments received by the comments. The first-stage comments are the comments that replied directly to the post, not the comments that replied to some other comments.

Table A2 Keyword for Testing the Bias of COVID-Oriented Search

Category	Keywords for search
<b>COVID-19</b>	COVID-19
Technology	New Energy, Space Travel, BYD (A Chinese automotive company), Data Leak, Microchip, Tesla
Economy	Exchange Rate, GDP, Mortgage, Job Market, Stock Market
Entertainment+Sports	Winter Olympic, Short Videos, User-Generated Media, Soccer
Social Affairs	Population, Birth Rate, Human Trafficking, Scam, Plagiarism
Politics	Corruption, Taiwan, Hong Kong, Ukraine, Xinjiang, Territory
Education	Gaokao (Chinese college entrance exam), Peking University, Study Abroad

Notes: This table presents the keywords that are used in each category of news topic. These topics and corresponding keywords are summarized from Wikipedia China News (<https://zh.wikipedia.org/wiki/2022中国大陆>) from January 1 to August 31, 2022. They are used to find the influential posters. We compare the influential posters found through these keywords to the influential posters found through COVID-19. *BYD* is a Chinese automotive company that is one of the largest electronic vehicle producers, second to Tesla. *Gaokao* is the Chinese college entrance exam.

Figure A3 Distribution of Post Topic for Each Geo Location

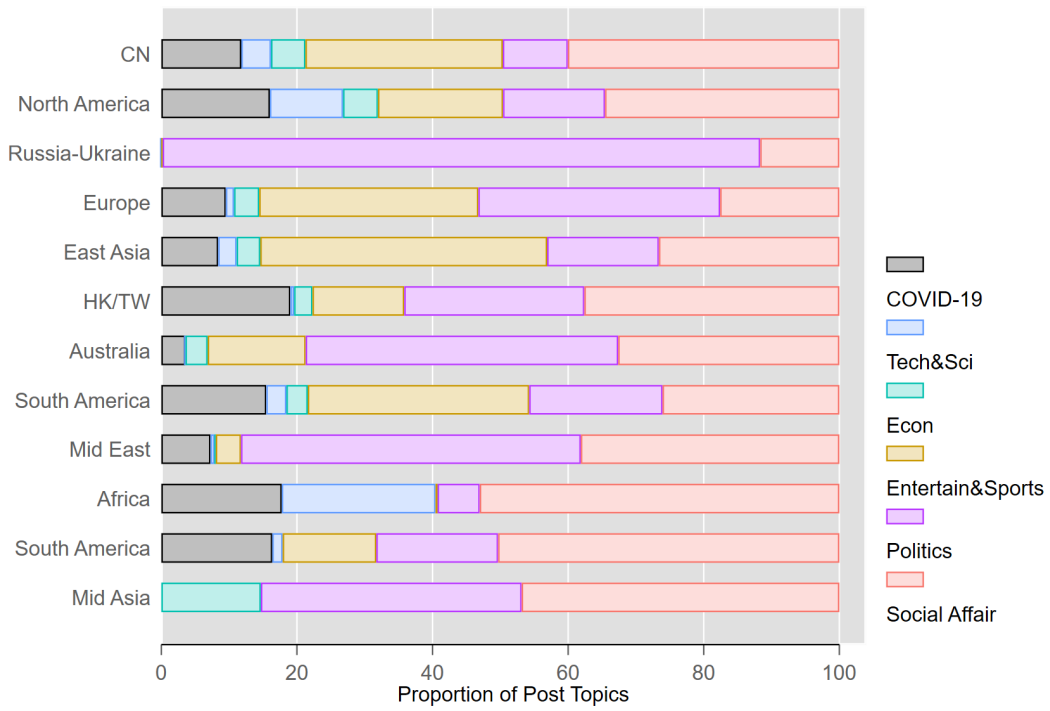


Figure A4 Number of Comments vs. Population

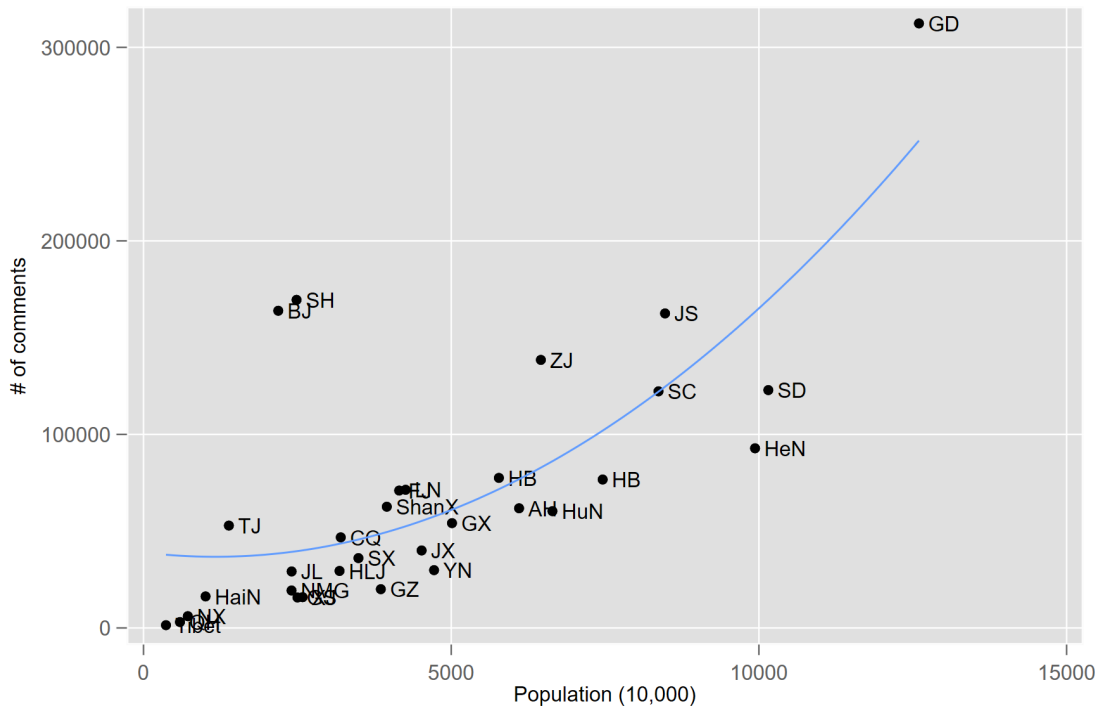




Table A3 RD Balance

	(1)		(2)		(3)		(4)		(5)		(6)		(7)									
	Western	profile	comment	COVID	profile	comment	Tech&Sci	profile	comment	Economy	profile	comment	Entertain	profile	comment	Politics	profile	comment	Social	profile	comment	
Revelation	-0.00903	0.00956	0.0418	-0.0284	-0.0108	-0.000865	-0.0369*	-0.0188	0.0610**	-0.0120	-0.080***	0.0111	0.0462	0.0188								
	(0.0251)	(0.0211)	(0.0275)	(0.0268)	(0.0166)	(0.0150)	(0.0203)	(0.0175)	(0.0259)	(0.0240)	(0.0217)	(0.0195)	(0.0284)	(0.0272)								
Observations	4,657	4,312	4,657	4,312	4,657	4,312	4,657	4,312	4,657	4,312	4,657	4,312	4,657	4,312								
Control Mean	0.2346	0.1661	0.2346	0.1661	0.2346	0.1661	0.2346	0.1661	0.2346	0.1661	0.1460	0.1401	0.2346	0.1661								

Notes: This table presents the balance test for the regression discontinuity specification. The estimates are obtained by regressing the post type on the treatment variables defined in the main RD equation with a 30-day bandwidth, controlling the linear time trend, previously defined set of controls, and the day-of-week fixed effect. The coefficient represents the change in the ratio of the outcome type of post due to the policy. The insignificant estimate suggests no discontinuous change happened due to the policy change. Heteroskedasticity robust standard errors are reported below point estimates.

Figure A5 Policy Effect Estimated Using Narrow and Wide Bandwidth

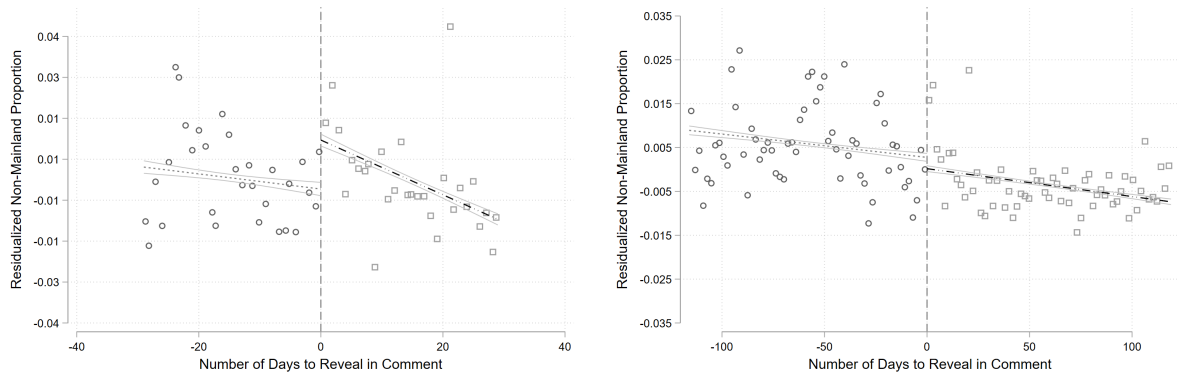


Table A4 Impact of Reveal IP Location in Comment on Non-mainland Participation, Extended Bandwidth

Bandwidth Days =	Non-mainland Comment			
	<b>120</b>	60	80	100
IP in Comment	-0.00707*** (0.000631)	0.00476*** (0.000888)	-0.00168** (0.000765)	-0.00303*** (0.000679)
Observations	1,891,324	995,425	1,304,044	1,639,899
Control Mean	0.0515	0.0523	0.0518	0.0521

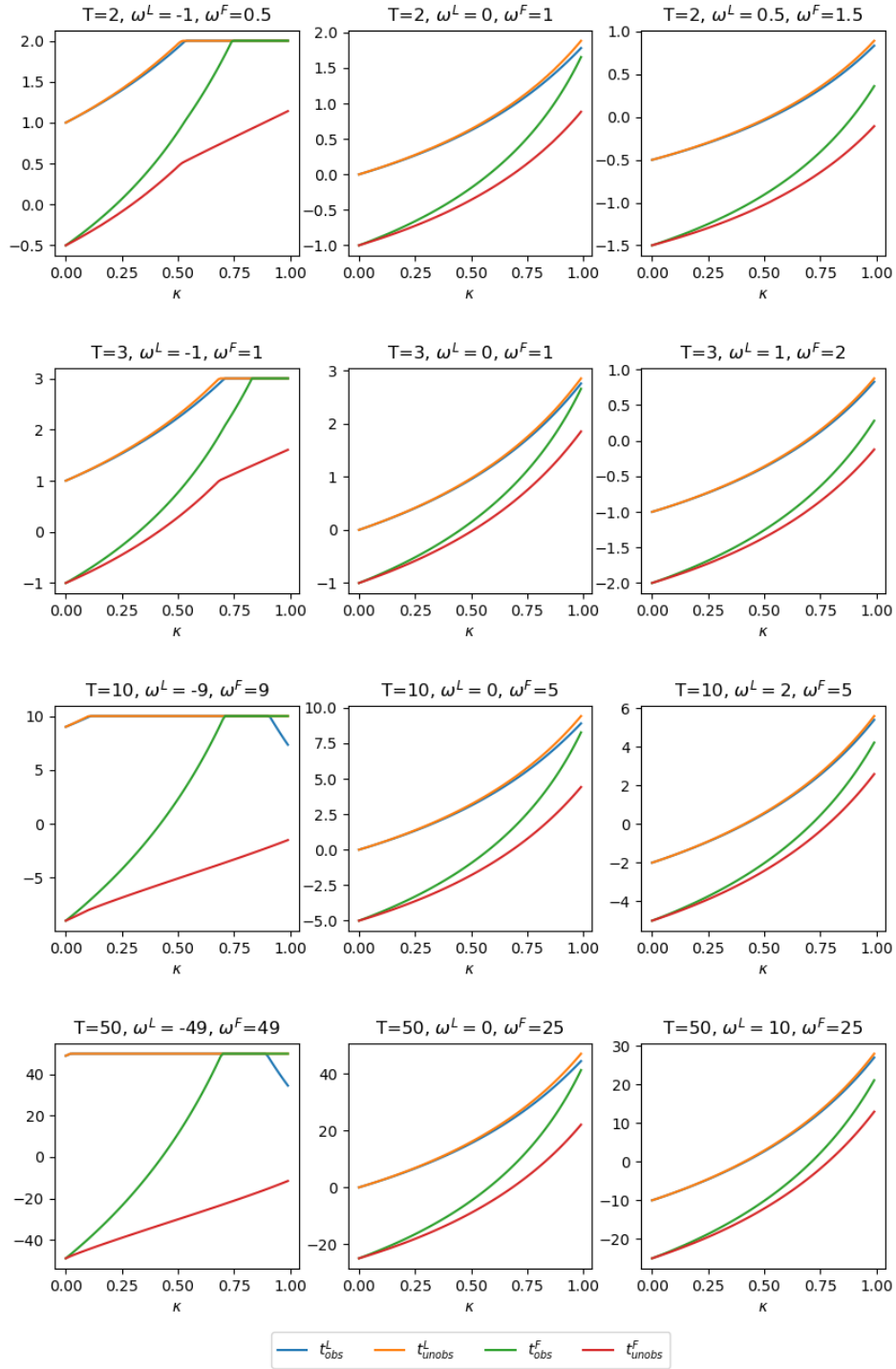
Notes: This table provides the additional result to the main table by extending the bandwidth. The main results are derived by estimating the main regression discontinuity equation, with the outcome being the presence of non-mainland comments. The control mean is the proportion of non-mainland comments within the bandwidth days before the policy implementation. The regression includes the set of controls defined in the previous section, as well as day-of-week and post-type fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Table A5 Balance of Posts and Poster Characteristics Before and After De-anonymization

	Anonymized			De-anonymized			diff
	n	mean	sd	n	mean	sd	
<b>Panel A. Post basic characteristics</b>							
# of likes	842353	14510.34	38821.94	1953230	14563.80	38297.54	53.465
# of reposts	842353	1493.22	8390.69	1953230	1113.44	3262.95	-379.780*
# of comments	842353	1018.61	2277.61	1953230	1146.61	2237.59	128.001**
Length	842353	488.77	611.50	1953230	481.75	665.70	-7.021
<b>Panel B. Post geographic identity</b>							
CN	842353	0.79	0.41	1953230	0.87	0.34	0.076***
nAm	842353	0.15	0.36	1953230	0.12	0.33	-0.028***
sAm	842353	0.00	0.06	1953230	0.00	0.04	-0.002
HKTW	842353	0.05	0.23	1953230	0.06	0.23	0.004
eAsia	842353	0.07	0.26	1953230	0.06	0.24	-0.009
sAsia	842353	0.02	0.14	1953230	0.01	0.12	-0.004
mAsia	842353	0.00	0.05	1953230	0.00	0.03	-0.001
midEast	842353	0.02	0.12	1953230	0.01	0.09	-0.008***
EU	842353	0.09	0.29	1953230	0.06	0.24	-0.029***
Aus	842353	0.03	0.18	1953230	0.02	0.13	-0.017***
AFRica	842353	0.00	0.05	1953230	0.00	0.06	0.001
RU	842353	0.14	0.34	1953230	0.04	0.20	-0.097***
<b>Panel C. Post topic</b>							
COVID	842353	0.15	0.36	1953230	0.20	0.40	0.045***
Tech&Sci	842353	0.07	0.25	1953230	0.07	0.26	0.001
Econ	842353	0.10	0.30	1953230	0.10	0.30	0.001
Entertain&Sports	842353	0.33	0.47	1953230	0.28	0.45	-0.050***
Politics	842353	0.23	0.42	1953230	0.17	0.38	-0.061***
Social affairs	842353	0.28	0.45	1953230	0.35	0.48	0.061***
<b>Panel D. Poster characteristics</b>							
Non-mainland poster	842353	0.15	0.36	1953230	0.14	0.35	-0.012
# of followers	842353	252.65	308.83	1953230	249.22	299.89	-3.431

Note: The differences are calculated by regressing the balance variables on the de-anonymization indicator

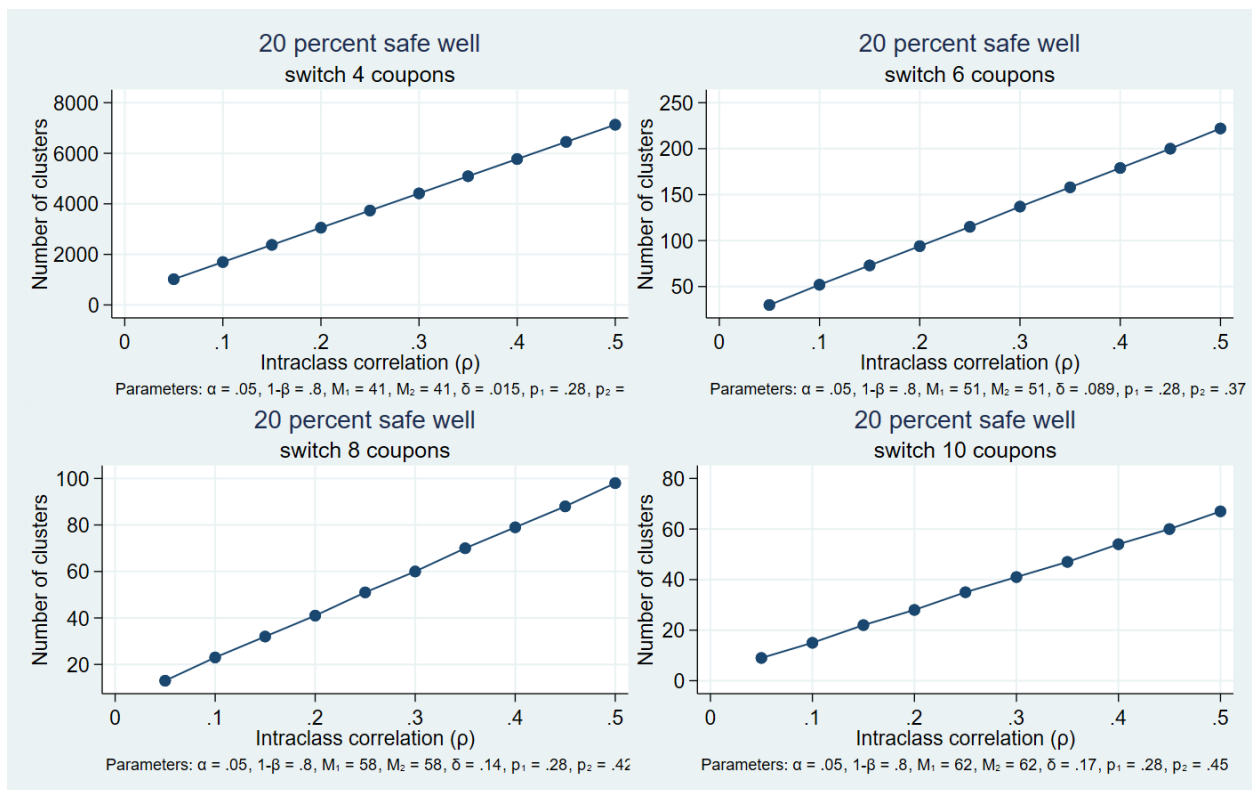
Figure A6 Simulation of Induced Private Benefit Thresholds



## APPENDIX B

### CHAPTER 2

Figure A7 Minimal number of treatment clusters with 20% of safe wells



Notes:  $M_1$  and  $M_2$  are the predicted number of populations that have at least one safe well to switch to.  $p_1$  is the baseline switching rate.  $p_2$  is the predicted switching rate. Henceforth  $\delta$  is the predicted treatment effect.

Figure A8 Minimal number of treatment clusters with 40% of safe wells

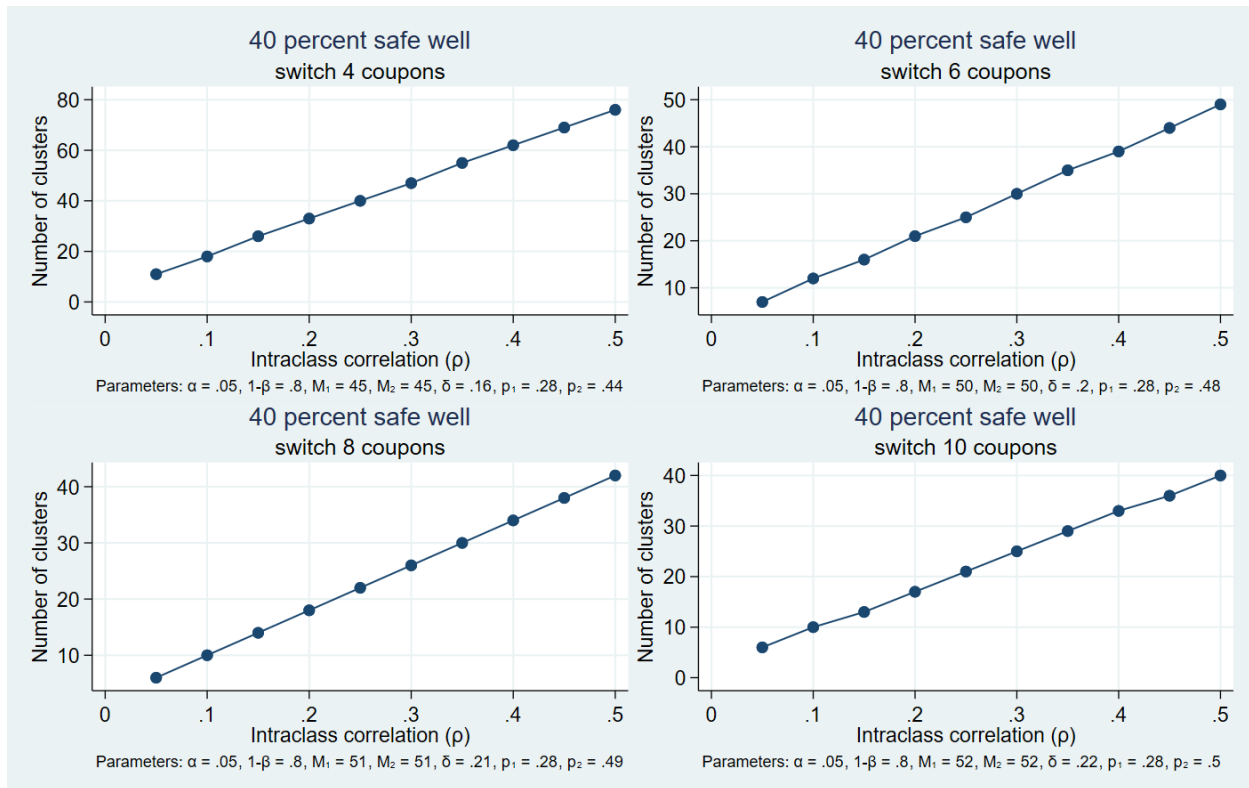
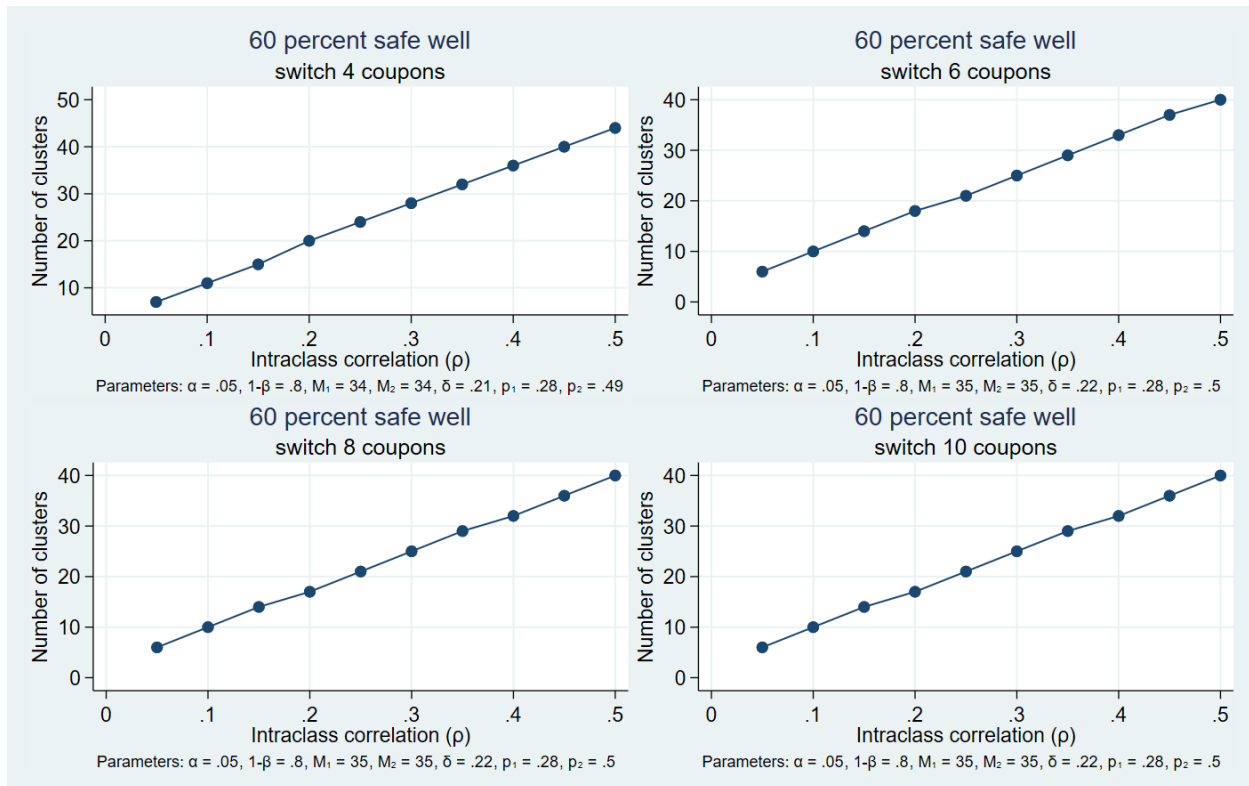


Figure A9 Notes:  $M_1$  and  $M_2$  are the predicted number of population that have at least one safe well to switch to.  $p_1$  is the baseline switching rate.  $p_2$  is the predicted switching rate. Henceforth  $\delta$  is the predicted treatment effect.

Figure A10 Minimal number of treatment clusters with 60% of safe wells



Notes:  $M_1$  and  $M_2$  are the predicted number of population that have at least one safe well to switch to.  $p_1$  is the baseline switching rate.  $p_2$  is the predicted switching rate. Henceforth  $\delta$  is the predicted treatment effect.



Table A6 Dyad regression of coupon exchange between pairs of households

	Exchanged coupon
T2+T3 villages	-0.0153** (0.00741)
Social connect	0.263*** (0.0225)
Geo distance	-0.0265*** (0.00261)
Diff in:	
HH size	-0.000815* (0.000440)
Average age	-5.38e-06 (0.000151)
Male ratio	-0.00496 (0.00508)
Child ratio	-0.00437 (0.00504)
Education ratio	-0.0245*** (0.00429)
Asset index (PCA)	-0.00266** (0.00122)
Altruism index	-0.00935*** (0.00144)
Trust index	-0.0137*** (0.00168)
Positive reciprocity index	-0.0134*** (0.00182)
Negative reciprocity index	-0.00943*** (0.00193)
Observations	171,666
R-squared	0.128

Notes: This table shows the coefficients from a set of dyad regressions in which the outcome variable indicates whether two households from the same village exchanged coupons. Total number of dyads is 171,666. The table reports the coefficients of a dyad variable, which indicates whether the pair of households living in a village was notified by peer monitoring. The coefficients are obtained by regressing whether two households from the same village exchanged coupons on the set of dyad variables, the strata dummies, and the Upazila FEs. The standard errors are clustered in the village community level.

Table A7 Assort to coupon groups, HH characteristics and preferences

	Whether households $i$ and $j$ exchanged coupons	
	(1)	(2)
HH size diff	-0.000451 (0.000776)	-0.000387 (0.000708)
HH size diff × Notification	-0.000617 (0.000991)	-1.10e-05 (0.00109)
Age diff	0.000239 (0.000242)	0.000254 (0.000218)
Age diff × Notification	-0.000414 (0.000309)	-0.000381 (0.000283)
Male ratio diff	-0.00986 (0.00924)	0.00262 (0.00796)
Male ratio diff × Notification	0.00915 (0.0112)	0.00377 (0.00956)
Child ratio diff	-0.00223 (0.00601)	0.00505 (0.00734)
Child ratio diff × Notification	-0.00351 (0.00915)	-0.00832 (0.0102)
Edu ratio diff	-0.0250*** (0.00621)	-0.0262*** (0.00582)
Edu ratio diff × Notification	0.00119 (0.00842)	0.0105 (0.00782)
Altruism index diff	-0.00989*** (0.00302)	-0.0128*** (0.00313)
Altruism diff × Notification	0.00121 (0.00340)	0.00422 (0.00388)
Trust index diff	-0.0128*** (0.00309)	-0.0155*** (0.00354)
Trust index diff × Notification	-0.00117 (0.00366)	0.000831 (0.00423)
Positive reciprocity index diff	-0.0127*** (0.00334)	-0.0150*** (0.00370)
Positive reciprocity index diff × Notification	-0.000736 (0.00397)	0.00332 (0.00452)
Negative reciprocity index diff	-0.0102*** (0.00319)	-0.0160*** (0.00275)
Negative reciprocity index diff × Notification	0.00133 (0.00401)	0.00166 (0.00346)
Observations	171,666	171,666
R-squared	0.130	0.246
FEs	Village	Village+Individual

Notes: This table is the second part of Table 2.11. This table shows the coefficients from a dyad regression in which the outcome variable indicates whether two households from the same village exchanged coupons. The table reports the coefficients of household characteristics, social preferences, and their interactions with a treatment dummy of the peer monitoring notification (T2+T3 villages), **Notification**. Standard errors are clustered at the village community level.

The regression in both columns includes three sets of variables: (1) social connection, geographic distance, asset level, and their interactions with the Notification dummy; (2) absolute differences of household characteristics and their interactions with the Notification dummy; (3) sums of household characteristics and their interactions with the Notification dummy. Column (1) additionally includes Village FEs, and Column (2) additionally includes Village FEs and two-way-individual FEs.

Table A8 The impacts of treatment on the arsenic consumption measured by the well testing in the Endline, among switchers

Arsenic concentration of the well used by the household in the Endline (ppb)						
Panel A. Overall treatment effect of ex-ante commitment						
	(1)	(2)	(3)	(4)	(5)	(6)
T1+T2+T3	-41.91*** (14.85)	-43.41*** (16.15)	-42.62*** (15.89)	-41.26** (18.60)	-5.466 (19.55)	-23.42 (18.97)
R-squared 0.029	0.233 0.038	0.289	0.019	0.036	0.185	0.208
Panel B. Treatment effect of each treatment						
	(1)	(2)	(3)	(4)	(5)	(6)
T1	-41.68** (19.12)	-39.92** (19.35)	-37.33* (20.64)	-37.74* (20.85)	-19.94 (34.50)	-15.72 (30.62)
T2	-51.29*** (18.01)	-55.12*** (19.75)	-38.58* (20.01)	-36.39 (27.41)	-8.522 (19.57)	-18.61 (18.94)
T3	-30.25* (17.87)	-32.96 (20.13)	-58.32*** (19.21)	-54.66** (26.42)	5.546 (21.17)	-46.49* (23.77)
R-squared	0.330	0.323	0.298	0.295	0.156	0.228
Control Mean	172.24	204.75	196.65	235.67	45.41	105.21
Well-Owner	✓	✓	✗	✗	✓	✗
Baseline Arsenic	ALL	HIGH	ALL	HIGH	LOW	LOW
Observations	1,148	944	955	596	204	359

Notes: This table reports the intention-to-treat estimates of the treatment effects of the interventions *among households who were identified that switched the wells*. A household is considered to have its well switched if it reports the primary well used in the Endline survey as different from the well used in the Baseline survey. Panel A shows the pooled impact of ex-ante commitment on arsenic consumption, namely comparing the 99 village communities that intervened with the coupon exchange with the 36 control village communities. Panel B shows each intervention separately. T1 are villages where we only facilitate the ex-ante commitment by distributing coupons. T2 are villages where we distribute coupons and notify households about potential peer monitoring through the SMS campaign. T3 are villages where we distribute coupons and implement peer monitoring. The regression coefficients are obtained by regressing the arsenic concentration of the well used by the household in the Endline on the treatment dummy (dummies), pre-specified controls, and the strata dummies (Upazila FEs). Standard errors are clustered at the village-community level. Therefore, the coefficients show the estimated impacts of treatments on arsenic mitigation.

The control means shows the arsenic concentration of the well used in the Endline by the households in the 36 control villages. A well-owner is defined by whether the household privately owns the primary well in the baseline. Households were invited to the experiment if and only if they were well-owners. Baseline arsenic identifies if the household's baseline primary well contains arsenic greater or less than 50 ppb, which is the cutoff for arsenic-poisoned water determined by the Bangladeshi government.

Table A9 The impacts of treatment on the arsenic consumption measured by the well testing in the endline, no controls

Arsenic concentration of the well used by the household in the Endline (ppb)						
Panel A. Overall treatment effect of ex-ante commitment						
	(1)	(2)	(3)	(4)	(5)	(6)
T1+T2+T3	-19.18* (10.35)	-16.90 (12.64)	-30.78*** (11.42)	-25.31* (14.19)	1.114 (2.246)	2.982 (10.42)
R-squared	0.401	0.260	0.330	0.275	0.043	0.084
Panel B. Treatment effect of each treatment						
	(1)	(2)	(3)	(4)	(5)	(6)
T1	-25.85* (15.00)	-22.76 (17.11)	-41.58*** (14.50)	-34.95** (16.79)	0.610 (2.709)	1.140 (14.44)
T2	-29.43** (12.52)	-34.57** (16.11)	-24.68 (15.93)	-15.94 (21.48)	2.982 (2.491)	9.593 (10.96)
T3	-4.058 (12.34)	2.766 (16.30)	-23.21 (14.19)	-18.64 (18.77)	-0.697 (2.614)	-7.260 (10.93)
R-squared	0.403	0.265	0.331	0.276	0.045	0.087
Control Mean	212.40	322.96	215.04	305.90	17.91	40.50
Well-Owner	✓	✓	✗	✗	✓	✗
Baseline Arsenic	ALL	HIGH	ALL	HIGH	LOW	LOW
Observations	9,384	6,050	3,067	1,907	3,334	1,160

Notes: This table reports the intention-to-treat estimates of the treatment effects of the interventions *without the pre-specified controls*. A household is considered to have its well switched if it reports the primary well used in the Endline survey as different from the well used in the Baseline survey. Panel A shows the pooled impact of ex-ante commitment on arsenic consumption, namely comparing the 99 village communities that intervened with the coupon exchange with the 36 control village communities. Panel B shows each intervention separately. T1 are villages where we only facilitate the ex-ante commitment by distributing coupons. T2 are villages where we distribute coupons and notify households about potential peer monitoring through the SMS campaign. T3 are villages where we distribute coupons and implement peer monitoring. The regression coefficients are obtained by regressing the arsenic concentration of the well used by the household in the Endline on the treatment dummy (dummies) and the strata dummies (Upazila FEs). Standard errors are clustered at the village-community level. Therefore, the coefficients show the estimated impacts of treatments on arsenic mitigation.

The control means shows the arsenic concentration of the well used in the Endline by the households in the 36 control villages. A well-owner is defined by whether the household privately owns the primary well in the baseline. Households were invited to the experiment if and only if they were well-owners. Baseline arsenic identifies if the household's baseline primary well contains arsenic greater or less than 50 ppb, which is the cutoff for arsenic-poisoned water determined by the Bangladeshi government.

Table A10 Switching, among switchers

Switched	Unsafe to safe		To lower		Safe to unsafe		To higher	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Overall treatment effect of ex-ante commitment								
T1+T2+T3	0.0747 (0.0536)	-0.0471 (0.0646)	0.0754 (0.0487)	-0.0163 (0.0523)	-0.0110 (0.0786)	0.0272 (0.0488)	-0.122*** (0.0441)	0.0306 (0.0626)
R-squared	0.233	0.289	0.019	0.036	0.185	0.208	0.029	0.038
Panel B. Treatment effect of each treatment								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T1	0.0582 (0.0567)	-0.106 (0.0707)	0.0533 (0.0603)	-0.0218 (0.0582)	-0.110 (0.102)	0.0220 (0.0580)	-0.152*** (0.0467)	0.00695 (0.0712)
T2	0.132* (0.0689)	0.0126 (0.0870)	0.0973 (0.0622)	-0.0344 (0.0674)	-0.0237 (0.0894)	0.0525 (0.0585)	-0.110** (0.0550)	0.0884 (0.0828)
T3	0.0381 (0.0744)	0.0168 (0.0736)	0.0726 (0.0627)	0.0191 (0.0824)	0.0539 (0.0811)	-0.0368 (0.0533)	-0.105** (0.0491)	-0.0142 (0.0840)
R-squared	0.237	0.299	0.020	0.038	0.204	0.213	0.031	0.044
Control Mean	0.40	0.33	0.48	0.37	0.13	0.23	0.28	0.32
Well-owner	✓	✗	✓	✗	✓	✗	✓	✗
Observations	464	389	668	748	204	359	668	748

Notes: This table reports the intention-to-treat estimates of the treatment effects of the interventions *among households who were identified that switched the wells*. A household is considered to have its well switched if it reports the primary well used in the Endline survey different from the well used in the Baseline survey. We consider four types of well switching: columns (1) - (2) show switching from unsafe to safe, that the household used a well that contains arsenic higher than 50 ppb in the Baseline but switched to a well that is lower than 50 ppb in the Endline; columns (3) - (4) show switching to a lower-arsenic-contaminated well, meaning that the well used in the Endline contains less arsenic than the well used in the Baseline; columns (5) - (6) show switching from safe to unsafe, that the household used a well that contains arsenic lower than 50 ppb in the Baseline but switched to a well that is higher than 50 ppb in the Endline; and columns (7) - (8) show switching to a higher-arsenic-contaminated well, meaning that the well used in the Endline contains higher arsenic than the well used in the Baseline. *Households that did not switch are assigned zero in all four types of switching.*

Panel A shows the pooled impact of ex-ante commitment on well switching, namely comparing the 99 village communities that intervened with the coupon exchange with the 36 control village communities. Panel B shows each intervention separately. T1 are villages where we only facilitate the ex-ante commitment by distributing coupons. T2 are villages where we distribute coupons and notify households about potential peer monitoring through the SMS campaign. T3 are villages where we distribute coupons and implement peer monitoring. The regression coefficients are obtained by regressing the switching status of the household on the treatment dummy (dummies), pre-specified controls, and the strata dummies (Upazila FEs). Standard errors are clustered at the village-community level. Therefore, the coefficients show the estimated impacts of treatments on the probability of switching.

The control means shows the arsenic concentration of the well used in the Endline by the households in the 36 control villages. A well-owner is defined by whether the household privately owns the primary well in the baseline. Households were invited to the experiment if and only if they were well-owners. Baseline arsenic identifies if the household's baseline primary well contains arsenic greater or less than 50 ppb, which is the cutoff for arsenic-poisoned water determined by the Bangladeshi government.

Table A11 Switching, no controls

Switched	Unsafe to safe		To lower		Safe to unsafe		To higher	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Overall treatment effect of ex-ante commitment								
T1+T2+T3	0.00329 (0.00772)	-0.00156 (0.0223)	0.00432 (0.00675)	-0.00474 (0.0213)	-0.000476 (0.00422)	0.0260 (0.0272)	-0.00785** (0.00382)	0.0118 (0.0266)
R-squared	0.048	0.135	0.004	0.021	0.007	0.072	0.002	0.017
Panel B. Treatment effect of each treatment								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T1	0.00241 (0.00903)	-0.0171 (0.0219)	0.00636 (0.00962)	-0.00668 (0.0241)	-0.00633 (0.00619)	0.00806 (0.0346)	-0.00892** (0.00390)	0.00700 (0.0291)
T2	0.0205* (0.0116)	0.0218 (0.0353)	0.00914 (0.00865)	0.00562 (0.0271)	0.000189 (0.00522)	0.0533 (0.0339)	-0.00667 (0.00489)	0.0457 (0.0393)
T3	-0.00773 (0.0103)	0.00268 (0.0301)	-0.00169 (0.00839)	-0.0146 (0.0294)	0.00175 (0.00451)	-0.00739 (0.0272)	-0.00794* (0.00464)	-0.0228 (0.0299)
R-squared	0.050	0.137	0.004	0.021	0.007	0.077	0.002	0.022
Control Mean	0.04	0.10	0.04	0.13	0.01	0.08	0.02	0.11
Well-owner	✓	✗	✓	✗	✓	✗	✓	✗
Observations	5,443	1,600	8,777	2,760	3,334	1,160	8,777	2,760

Notes: This table reports the intention-to-treat estimates of the treatment effects of the interventions *without the pre-specified controls*. A household is considered to have its well switched if it reports the primary well used in the Endline survey different from the well used in the Baseline survey. We consider four types of well switching: columns (1) - (2) show switching from unsafe to safe, that the household used a well that contains arsenic higher than 50 ppb in the Baseline but switched to a well that is lower than 50 ppb in the Endline; columns (3) - (4) show switching to a lower-arsenic-contaminated well, meaning that the well used in the Endline contains less arsenic than the well used in the Baseline; columns (5) - (6) show switching from safe to unsafe, that the household used a well that contains arsenic lower than 50 ppb in the Baseline but switched to a well that is higher than 50 ppb in the Endline; and columns (7) - (8) show switching to a higher-arsenic-contaminated well, meaning that the well used in the Endline contains higher arsenic than the well used in the Baseline. *Households that did not switch are assigned zero in all four types of switching.*

Panel A shows the pooled impact of ex-ante commitment on well switching, namely comparing the 99 village communities that intervened with the coupon exchange with the 36 control village communities. Panel B shows each intervention separately. T1 are villages where we only facilitate the ex-ante commitment by distributing coupons. T2 are villages where we distribute coupons and notify households about potential peer monitoring through the SMS campaign. T3 are villages where we distribute coupons and implement peer monitoring. The regression coefficients are obtained by regressing the switching status of the household on the treatment dummy (dummies), pre-specified controls, and the strata dummies (Upazila FEs). Standard errors are clustered at the village-community level. Therefore, the coefficients show the estimated impacts of treatments on the probability of switching.

The control means shows the arsenic concentration of the well used in the Endline by the households in the 36 control villages. A well-owner is defined by whether the household privately owns the primary well in the baseline. Households were invited to the experiment if and only if they were well-owners. Baseline arsenic identifies if the household's baseline primary well contains arsenic greater or less than 50 ppb, which is the cutoff for arsenic-poisoned water determined by the Bangladeshi government.

## APPENDIX C

### CHAPTER 3

#### Proof of Proposition 1.

First, notice that by assumption, the agent with  $b \geq 0$  always adopt. This is to say, upon observing  $a = 0$ , others immediately know that the agent must have  $b < 0$ , and thus  $\Pr(b \geq 0|a = 0) = 0$ .

To derive  $\Pr(b \geq 0|a = 1)$ , recall the fact that agent adopts when  $b \geq \bar{\Delta}$ . Apply Bayes' Rule, for any agent  $i$ :

$$\begin{aligned}
 \Pr(b \geq 0|a_i = 1) &= \frac{\Pr(a_i = 1|b \geq 0) \cdot \Pr(b \geq 0)}{\Pr(a_i = 1|b \geq 0) \Pr(b \geq 0) + \Pr(a_i = 1|b < 0) \cdot \Pr(b < 0)} \\
 &= \frac{\Pr(b \geq 0)}{\Pr(b \geq 0) + \frac{\Pr(\bar{\Delta} \leq b < 0)}{\Pr(b < 0)} \cdot \Pr(b < 0)} \\
 &= \frac{\Pr(b \geq 0)}{\Pr(b \geq 0) + \Pr(b < 0) - \Pr(b < \bar{\Delta})} \\
 &= \frac{\Pr(b \geq 0)}{1 - \Pr(b < \bar{\Delta})} \\
 &= \frac{\Phi(\bar{b}/\sigma_b)}{\Phi((\bar{b} - \bar{\Delta})/\sigma_b)}.
 \end{aligned}$$

At the margin, type  $\bar{\Delta}$  agent is indifferent between adopt or not, thus when  $a = 1$

$$\bar{\Delta} + x\mu\Sigma = 0,$$

where

$$\Sigma = \frac{\Phi(\bar{b}/\sigma_b)}{\Phi((\bar{b} - \bar{\Delta})/\sigma_b)} \tag{C.1}$$

**Prediction 1:**  $\bar{\Delta} \leq 0$ .

Since  $\Sigma > 0$  for all the parameter values and  $x$  and  $\mu$  by their nature are at least 0,  $\bar{\Delta} \leq 0$ , indicating that low-type individuals start adopting under reputation concern.

**Prediction 2:**  $\frac{\partial \bar{\Delta}}{\partial x} < 0$ .

Denote

$$F(x, b, \bar{\Delta}) = \bar{\Delta} + x\mu\Sigma$$

The at  $F(x, b, \bar{\Delta}) = 0$ ,

$$F_x(x, b, \bar{\Delta}) = \frac{\partial}{\partial x} (\bar{\Delta} + x\mu\Sigma) = 0,$$

where

$$\begin{aligned} 0 &= \frac{\partial}{\partial x} (\bar{\Delta} + x\mu\Sigma) \\ 0 &= \frac{\partial \bar{\Delta}}{\partial x} + \frac{\mu\Phi\left(\frac{\bar{b}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)} + x\mu \frac{\partial}{\partial x} \frac{\Phi\left(\frac{\bar{b}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)}. \end{aligned} \quad (\text{C.2})$$

Given that

$$\frac{\partial}{\partial x} \left\{ \frac{\Phi\left(\frac{\bar{b}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)} \right\} = \frac{\Phi\left(\frac{\bar{b}}{\sigma_b}\right) \phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right) \frac{\partial \bar{\Delta}}{\partial x}}{\sigma_b \left[\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)\right]^2},$$

where  $\phi(\cdot)$  is the standard normal p.d.f., we rearrange (5) and derive

$$\frac{\partial \bar{\Delta}}{\partial x} = - \frac{\mu\sigma_b \Phi\left(\frac{\bar{b}}{\sigma_b}\right) \Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)}{\sigma_b \left[\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)\right]^2 + x\mu\Phi\left(\frac{\bar{b}}{\sigma_b}\right) \phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)} < 0, \quad (\text{C.3})$$

as the denominator and nominator of the right-hand side of (6) are both positive.

**Prediction 3:**  $\frac{\partial \bar{\Delta}}{\partial b} < 0$ .

We apply the same strategy: partially differentiating  $\bar{\Delta} + x\mu\Sigma = 0$  with respect to  $\bar{b}$ , so that

$$\frac{\partial \bar{\Delta}}{\partial \bar{b}} + x\mu \frac{\partial}{\partial \bar{b}} \frac{\Phi\left(\frac{\bar{b}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)} = 0. \quad (\text{C.4})$$

We can show that

$$\frac{\partial}{\partial \bar{b}} \left\{ \frac{\Phi\left(\frac{\bar{b}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)} \right\} = \frac{\phi\left(\frac{\bar{b}}{\sigma_b}\right) \Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right) - \Phi\left(\frac{\bar{b}}{\sigma_b}\right) \phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right) \left(1 + \frac{\partial \bar{\Delta}}{\partial \bar{b}}\right)}{\sigma_b \left[\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)\right]^2} \quad (\text{C.5})$$

Plug (8) back to (7) we get

$$\frac{\partial \bar{\Delta}}{\partial \bar{b}} = \frac{x\mu \left[ \phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right) \Phi\left(\frac{\bar{b}}{\sigma_b}\right) - \phi\left(\frac{\bar{b}}{\sigma_b}\right) \Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right) \right]}{\sigma_b \left[\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)\right]^2 + x\mu\Phi\left(\frac{\bar{b}}{\sigma_b}\right) \phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)}. \quad (\text{C.6})$$



To show (9) is negative, note that given

$$\bar{\Delta} = -x\mu \frac{\Phi\left(\frac{\bar{b}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)}, \quad (\text{C.7})$$

we can write

$$\frac{\partial \bar{\Delta}}{\partial \bar{b}} = -\frac{x\mu\phi\left(\frac{\bar{b}}{\sigma_b}\right) + \bar{\Delta}\phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)}{\sigma_b\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right) - \bar{\Delta}\phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)}.$$

Since  $\bar{\Delta} < 0$ , the denominator is positive. To show that the nominator is also positive, notice that

$$x\mu\phi\left(\frac{\bar{b}}{\sigma_b}\right) + \bar{\Delta}\phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right) = x\mu\phi\left(\frac{\bar{b}}{\sigma_b}\right) - x\mu \frac{\Phi\left(\frac{\bar{b}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)}\phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right) \quad (\text{C.8})$$

$$= x\mu\Phi\left(\frac{\bar{b}}{\sigma_b}\right) \left[ \frac{\phi\left(\frac{\bar{b}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}}{\sigma_b}\right)} - \frac{\phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)} \right]. \quad (\text{C.9})$$

Then, given the symmetry of standard normal distribution:

$$\frac{\phi(x)}{\Phi(x)} = \frac{\phi(-x)}{\Phi(x)} = \frac{\phi(-x)}{1 - \Phi(-x)} = H(-x), \quad (\text{C.10})$$

which satisfies *monotone hazard rate property* given we assumed standard normal distribution<sup>1</sup>.

This means that  $H(z)$  increasing with  $z \in \mathbb{R}$ . Therefore,  $H(-x)$  is monotonic decreasing with  $x$ .

Since  $\bar{b} - \bar{\Delta} > \bar{b}$ ,

$$\frac{\phi\left(\frac{\bar{b}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}}{\sigma_b}\right)} > \frac{\phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}-\bar{\Delta}}{\sigma_b}\right)}. \quad (\text{C.11})$$

This shows that the nominator is also positive. Thus  $\frac{\partial \bar{\Delta}}{\partial \bar{b}} < 0$ .

---

<sup>1</sup>See Footnote 5 of Egorov et al. (2023).

Table A12 Self-Reported Preventive Measures and COVID-19 Knowledge

	N	mean	sd
<i>Panel A: In the previous week, have you:</i>			
<b>May Survey</b>			
Social distance (days)	2,768	3.68	3.2
Social gathering (days)	2,768	1.43	2.6
Cover sneeze	2,768	0.62	
Own mask	2,768	0.92	
Use mask	2,768	0.73	
Always use soap	2,768	0.84	
<b>November Survey</b>			
Social distance (days)	3,028	5.2	2.8
Social gathering (days)	3,028	2.8	3.1
Cover sneeze	3,028	0.74	
Own mask	3,028	0.92	
Use mask	3,028	0.66	
Always use soap	3,028	0.55	
<i>Panel B1: What can cause virus transmission?</i>			
<b>May Survey</b>			
Asymptomatic	3,028	0.60	
Touching object <sup>1</sup>	3,028	0.75	
<b>November Survey</b>			
Asymptomatic	3,028	0.74	
Touching object	3,028	0.77	
Aerosol	3,028	0.81	

Table A13 Demographics and Assets

	N <sup>1</sup>	mean	sd
<b>Panel A: Demographics</b>			
Household size	3,028	4.647	1.943
Average age	3,028	27.233	11.381
male ratio	3,028	0.463	0.196
child ratio	3,028	0.404	0.227
primary education ratio	3,028	0.304	0.262
HH head is female	3,028	0.258	0.438
HH head education (yrs)	3,023	3.395	4.070
HH head age	3,023	45.43	14.35
Tolerance of health risk	2,701	1.818	1.193
Tolerance of general risk	3,028	2.089	1.203
<b>Panel B: Assets</b>			
Rooms	3,028	2.726	1.249
Whether access to electricity	3,028	0.988	0.107
Fans	3,028	2.371	1.182
Mobilephones	3,028	1.950	1.123
Smartphones	3,028	0.873	0.964
Cycles or rickshaws	3,028	0.116	0.410
Motorcycles	3,028	0.0657	0.291
Vehicles	3,023	0.0159	0.188
TVs	3,028	0.483	0.528
Computers	3,028	0.0188	0.138
Refrigerators	3,028	0.527	0.515
Acres of agricultural land	3,028	3.148	21.65
Private wells	3,028	0.826	0.520

Notes: The full sample is 3,028 but some respondents may choose not to answer some of the questions.

Table A14 Social Networks

	N	mean	sd
<b>Panel A: Social Networks</b>			
Socialization	3,028	0.724	0.917
Discuss farming issues	3,028	0.272	0.646
Discuss health issues	3,028	0.290	0.665
Discuss financial issues	3,028	0.253	0.558
Borrow or lend daily necessities	3,028	0.392	0.684
Borrow or lend money	3,028	0.364	0.660
Total Degree	3,028	1.106	1.463
<b>Panel B: Geographical Networks</b>			
<= 30m	3,028	1.518	1.466
<= 50m	3,028	3.314	2.441
<= 100m	3,028	8.589	4.889
<= 200m	3,028	17.742	8.285

Notes: Degree is defined as the number of connections each household have. This is to say that a household can have a higher degree than the number of names it reports as the non-listed households may list the household.

Table A15 OLS Testing For Clustering

	Preventive Index	Knowledge Index
<i>Neighbors' Mean:</i>		
Preventive Index	0.167*** (0.0473)	
Knowledge Index		0.156*** (0.0521)
Observations	3,028	3,028
R-squared	0.160	0.193
Village FE	YES	YES

Notes: Neighbors are defined as either geographically connected or socially connected. The indexes are calculated by first normalizing each preventive or knowledge variable and then normalizing the sum. The test is defined as regressing the household's preventive or knowledge index on the neighbors' means. The village-fixed effects are added to address the village-level heterogeneity.

Table A16 OLS Estimations of Peer Effects on Social Distance, Social Gathering, and Cover Sneeze

	Social distance	Social gathering	Cover sneeze
	(1)	(2)	(3)
<i>Social Distance</i>			
Geo+SN	0.173*** (0.0401)		
SN	0.182*** (0.0588)		
Geo	0.263*** (0.0368)		
<i>Social Gathering</i>			
Geo+SN		0.216*** (0.0385)	
SN		0.242*** (0.0579)	
Geo		0.302*** (0.0364)	
<i>Cover Sneeze</i>			
Geo+SN			0.156*** (0.0424)
SN			0.321*** (0.0600)
Geo			0.360*** (0.0344)
Observations	2,815	2,815	2,815
Village FE	NO	NO	NO
Date FE	NO	NO	NO

Table A17 OLS Estimations of Peer Effects on Own Masks, Wear Masks, and Wash Hands with Soap

	Own Mask	Use Mask	Always Soap
	(4)	(5)	(6)
<i>Own Mask</i>			
Geo+SN	0.138*** (0.0453)		
SN	0.253*** (0.0606)		
Geo	0.093** (0.0371)		
<i>Use Mask</i>			
Geo+SN		0.178*** (0.0318)	
SN		0.293*** (0.0482)	
Geo		0.227*** (0.0390)	
<i>Always Soap</i>			
Geo+SN			0.189*** (0.0383)
SN			0.337*** (0.0590)
Geo			0.253*** (0.0410)
Observations	2,815	2,815	2,815
Village FE	NO	NO	NO
Date FE	NO	NO	NO

Table A18 OLS Estimations of Peer Effects on COVID-19 Virus Transmission Knowledge

	Asymptomatic	Surface or Object	Aerosol
	(7)	(8)	(9)
<i>Social Distance</i>			
Geo+SN	0.210*** (0.0438)		
SN	0.150*** (0.0603)		
Geo	0.281*** (0.0395)		
<i>Social Gathering</i>			
Geo+SN		0.153*** (0.0396)	
SN		0.245*** (0.0627)	
Geo		0.239*** (0.0343)	
<i>Cover Sneeze</i>			
Geo+SN			0.131*** (0.0395)
SN			0.250*** (0.0615)
Geo			0.192*** (0.0370)
Observations	2,815	2,815	2,815
Village FE	NO	NO	NO
Date FE	NO	NO	NO



Table A19 FE and IV Estimations of Peer Effects with Lagged Peer Behavior, Social Interaction

	Social distance		Social gathering	
	(1) FE	(2) IV	(3) FE	(4) IV
<i>Social Distance</i>				
Geo+SN	0.124 (0.0959)	0.937 (0.681)		
SN	0.165* (0.0972)	-0.958 (0.917)		
Geo	0.0575 (0.0850)	-0.198 (1.030)		
<i>Social Gathering</i>				
Geo+SN			0.0871 (0.209)	0.177 (0.851)
SN			0.131 (0.208)	0.427 (0.827)
Geo			0.0379 (0.130)	0.403 (0.477)
Observations	604	155	604	155
Village FE	YES	YES	YES	YES
Date FE	YES	YES	YES	YES
IV	NO	YES	NO	YES

Table A20 FE and IV Estimations of Peer Effects with Lagged Peer Behavior, Personal Hygiene

	Use Mask		Use Soap		Cover Sneeze	
	(1) FE	(2) IV	(3) FE	(4) IV	(5) FE	(6) IV
<i>Use Mask</i>						
Geo+SN	0.164 (0.147)	-0.128 (0.737)				
SN	0.355* (0.150)	0.570 (0.689)				
Geo	0.408 (0.102)	0.0559* (0.288)				
<i>Use Soap</i>						
Geo+SN			-0.0428 (0.185)	-0.121 (0.522)		
SN			0.195 (0.162)	0.275 (0.795)		
Geo			-0.116 (0.153)	-0.188 (0.738)		
<i>Cover Sneeze</i>						
Geo+SN					0.166 (0.102)	0.238 (0.414)
SN					-0.128 (0.0928)	-0.192 (0.354)
Geo					-0.0609 (0.0628)	0.154 (0.382)
Observations	604	155	604	155	604	155
Village FE	YES	YES	YES	YES	YES	YES
Date FE	YES	YES	YES	YES	YES	YES
IV	NO	YES	NO	YES	NO	YES

Table A21 FE and IV Estimations of Peer Effects with Lagged Peer Behavior, Knowledge

	Asymptomatic		Surface or Object	
	(1) FE	(2) IV	(3) FE	(4) IV
<i>Use Mask</i>				
Geo+SN	-0.0835 (0.102)	-0.0618 (0.407)		
SN	0.168* (0.0911)	0.382 (0.365)		
Geo	0.0243 (0.0918)	0.0279 (0.459)		
<i>Use Mask</i>				
Geo+SN			-0.0415 (0.0888)	0.370 (0.607)
SN			0.0801 (0.0953)	-0.0223 (0.412)
Geo			-0.0733 (0.0933)	-0.087 (0.356)
Observations	604	155	604	155
Village FE	YES	YES	YES	YES
Date FE	YES	YES	YES	YES

Figure A11 The Social Network Drop Down Menu

