Silent Networks: The Role of Inaccurate Beliefs in Reducing Useful Social Interactions*

Ronak Jain†　　Vatsal Khandelwal ‡

April 4, 2025

Abstract

Inaccurate beliefs about social norms can reduce useful social interactions and adversely affect an individual's ability to smooth shocks. We implement an experiment with low-income workers in urban India who lack access to formal institutions and underestimate their community's willingness to engage in dialogue around financial and mental well-being. Belief correction leads to a large increase in the demand for network-based assistance, driven by a reduction in perceived costs due to reputation and insensitivity concerns. We structurally estimate a network diffusion model and predict that our intervention will not lead to a shift in equilibrium engagement. Consistent with this, two years later, we find that the beliefs of those exposed to the information are less optimistic than the information delivered in the experiment. We compute the strength of counterfactual interventions needed for a sustained effect and find that belief correction can generate both the demand and funding for such policies.

KEYWORDS: Social Networks, Social Norms, Beliefs, Risk Sharing, Mental Health.
JEL CLASSIFICATION: C93, D83, D91, I12, I31, Z13.

*We are very grateful to Stefan Dercon, Marcel Fafchamps, Simon Quinn, and Alexander Teytelboym for their guidance and suggestions. We also thank Johannes Abeler, Emily Breza, Banu Demir, James Duffy, Benjamin Enke, Asim Khwaja, Gabriel Kriendler, Pramila Krishnan, Edward Glaeser, Douglas Gollin, Matthew Jackson, Craig McIntosh, Mahreen Mahmud, Kate Orkin, and Gautam Rao for their helpful comments. We thank seminar participants at various conferences including PacDev, NEUDC, the Econometric Society European Winter Meeting, the Workshop on Networks and Development, CSAE Conference, CEPR-TCD TIME Conference; OxDev (Oxford), and department workshops at Harvard University and the University of Oxford for their comments. Many thanks to Diana Beltekian, Binta Zahra Diop, Luke Milsom, Bassel Tarbush, Verena Weidemann, and Hannah Zillessen for their feedback on previous versions of the paper. We are also deeply grateful to the NGO team, especially Bharati Chaturvedi, Ishani Palandukar, Shruti Sinha, and Balmukund Kumar for their collaboration and to Vikram Patel and Sadaf Vibha for sharing helpful information and resources regarding mental health and reviewing the appropriateness of the intervention in the local context. We extend sincere thanks to our field team, especially Preetham Rodrigues and Abdul Latheef at the Inchara Foundation, and Garima Tyagi, Swati Chauhan, Darshna Jadhav, Pushyami Chilakapati, Anshuman Singh, Dipti Pareek, and Aarti Malik, for help with surveying. Finally, we gratefully thank the Center for Study of African Economies, Oxford for funding the study. The project was approved by both the Harvard Institutional Review Board and CUREC, University of Oxford. The RCT was pre-registered at the AEA RCT registry: https://doi.org/10.1257/rct.6980-2.1.

†University of Zurich. Email: ronak.jain@econ.uzh.ch

‡University of Exeter. Email: v.khandelwal@exeter.ac.uk

1 Introduction

Individuals rely on their social networks to receive and provide support, especially in the absence of formal assistance. This includes using social ties to smooth consumption, gather information, or seek advice (Beaman et al., 2021; Breza et al., 2019; Banerjee et al., 2024; Angelucci et al., 2018; Munshi and Rosenzweig, 2016; Banerjee et al., 2013; Munshi, 2011; De Weerdt and Dercon, 2006; Fafchamps and Lund, 2003). However, the ability of the network to act as a social safety net can be severely constrained if individuals do not demand support in the first place. For example, individuals may hesitate to discuss financial matters if they believe doing so can hurt their reputation. Such beliefs can prevent them from requesting financial support and consequently affect the quality of risk-sharing within the community. Not only this, these beliefs may also prevent them from demanding formal sources of assistance. However, beliefs about others may not always be accurate (Bursztyn and Yang, 2022; Bursztyn et al., 2020).

In this paper, we study how inaccurate beliefs about peers can generate "silent networks" i.e., networks with limited useful interactions despite high potential benefits. We combine evidence from a field experiment, survey experiments, and a structural model to study how belief correction interventions can reduce information frictions, strengthen network ties, and improve socioeconomic outcomes. We work with a sample of low-income informal sector workers and their families living in slums in urban India. Slums can act as poverty traps as they are characterized by poor living conditions, high risk, low human capital, and limited attention from policymakers (Marx et al., 2013).[1] As a result, it is crucial for social networks to function as safety nets in such settings. Given that 860 million people live in such conditions globally (UN-Habitat, 2013), it is a particularly relevant context for this study.

We implemented multiple rounds of data collection. First, we conducted a baseline survey with about 350 individuals in 2020. This was followed by a randomized controlled trial and an endline survey with 180 individuals in 2021. Finally, we implemented larger-scale experiments with an additional sample of 800 individuals in 2023.[2] We document four important facts about our setting that hold across all waves and motivate our intervention. First, individuals face significant volatility in incomes and consumption. Second, they do not have access to credit markets to smooth financial shocks or access to professional help for the mental distress caused by these shocks. Third, despite a lack of access to formal assistance for mental health and financial concerns, the majority of individuals do not discuss these topics with their peers. In particular, 63% reveal a willingness to engage with their community but 68% underestimate the willingness of others to do so.[3] Moreover, underestimation is larger than 10 percentage points for the majority of the underestimators. Importantly, these beliefs are consequential— those who have pessimistic beliefs about others are less likely to express

[1]For example, Ivaschenko et al. (2018) find that only 1 in 5 individuals in the poorest quintile in low-income countries have access to social safety net programs.

[2]The experiment was conducted during the peak of the COVID-19 crisis in India. The additional sample allows us to check if network interactions are sparse and dependent on beliefs even outside this period.

[3]We provide survey evidence and run additional experiments to show that social desirability bias is not driving the high willingness to engage that we measure around these topics.

a willingness to discuss these topics, have a lower frequency of dialogue around these topics, and have fewer connections in their advice-taking networks relative to their total connections in the overall social network. Moreover, we also find that beliefs about peers' willingness to engage are correlated with whether individuals have volatile consumption and faced consumption crises in the last six months.

Motivated by these patterns, we implemented a randomized controlled trial in which we provided the treatment group with accurate information on others' willingness to engage in dialogue about mental health and financial concerns.[4] We find that the intervention significantly increases an individual's engagement with their community. First, treated participants are 15 percentage points more likely to sign up for a potential savings group with their community members. Second, their willingness to sign up for training to become a listening volunteer to hear the anxieties of their community members increases by 16 percentage points.[5] Third, they also make financial contributions that are 29% higher to establish this service. These results suggest that the correction of misperceptions can increase the demand for network interactions as well as payments to set up informal avenues to interact. Moreover, we find that the demand for support from formal sources of assistance also depends on the beliefs individuals hold about the social norm. For example, the treatment increases an individual's willingness to speak to a doctor about mental health concerns. In addition, we detect evidence of positive spillovers on willingness to engage in dialogue around physical health concerns, which are also stigmatized in this setting.

We conducted follow-up surveys two weeks and two years later. We find that the treated group has interacted significantly more with their peers regarding both mental and financial well-being concerns two weeks later. Specifically, treated individuals are 19 percentage points more likely to have reached out to any peer to discuss financial concerns and reach out to one more peer than the control group, on average. Two years later, those exposed to the treatment have significantly more optimistic beliefs about their peers' willingness to engage, higher dialogue intensity, and lower self-reported consumption volatility. We also continue to detect a causal link between beliefs about others and individual engagement in a larger-scale validation check with 800 individuals.[6]

To disentangle mechanisms, we use survey data and additional experiments. We first identify various types of social costs that may prevent individuals from engaging with others. To do so, we use a hypothetical choice experiment to uncover whether these are reputation costs due to gossip, signaling costs due to the inability to access information about jobs, or interaction costs due to being met with insensitivity. We ask individuals to predict whether an advice-taking link regarding financial matters and mental well-being would exist between any two randomly chosen individuals in their community, while exogenously varying the

[4] A prediction exercise with a separate sample in the same setting shows that this information is perceived as a strong information shock and the majority predict that this information will increase engagement with the network.

[5] This would be similar to the services that charities like Samaritans offer in the UK and US.

[6] The mechanisms behind this treatment effect are different from the main RCT as the information provided to treated participants then was about *other* communities as opposed to their community. More details are provided in Section 5.7 and Section F in the appendix.

characteristics of the hypothetical advisor. Using data on acceptances/rejections of about 4,740 links, we find that reputation and interaction-related costs are the primary concerns that hinder useful network interactions rather than signaling. Moreover, we find that these costs significantly impact those who are more pessimistic about others' willingness to engage, which shows that we can improve outcomes by implementing a belief correction intervention that targets precisely the subgroup for whom such concerns are active. We rule out methodological concerns such as social desirability and experimenter demand by implementing a list experiment and an experiment that randomizes the visibility of responses to the experimenter. We also present evidence that alternative mechanisms such as updating beliefs about the benefits of interacting or the incidence of stress, and social pressure are unlikely to explain our results.

Next, we adapt and structurally estimate a network model (Jackson and Yariv, 2007) that allows us to (a) illustrate why individuals can have inaccurate beliefs in the first place, (b) predict whether our treatment effects correspond to a persistent change in equilibrium beliefs and engagement, and (c) compare the effectiveness of belief correction with alternative policy instruments. Individuals 'engage' if they take a discrete action to interact with their network on topics that can impose punitive social costs.[7] This can include participation in savings groups or attendance in information sessions about mental health. They face a cost of engagement depending on how connected they are in the network and the extent to which they violate the social norm, i.e., the proportion of others unwilling to engage. We provide empirical evidence to support these assumptions of the model.

The key insight of the model is that naive agents may, on average, underestimate the willingness of others to engage around these topics. This is because they form incorrect beliefs about the norm after observing the level of engagement among their peers who, due to the "friendship paradox" (Jackson, 2019; Feld, 1991), are more connected than they are and face a higher cost of violating the norm.[8] As a result, societies can be stuck at a Bayes-Nash equilibrium of low dialogue, and belief correction interventions can have potentially positive short-run and long-run effects. Notably, these effects depend on the shape of the dynamic best response curve. Since the process is recursive, i.e., actions at time $t + 1$ depend on beliefs at time t which in turn depend on actions at time $t - 1$, a belief correction intervention will have only short-run effects if this process reverts to the pre-intervention equilibrium. Long-run effects arise if the process leads to a new equilibrium instead.

To structurally estimate the model, we use simulated method of moments along with an equilibrium selection criterion to estimate the model by leveraging the random variation in beliefs induced by the RCT.[9] We find that the equilibrium that best fits the data in our setting is indeed that of low engagement. Further, we predict that our credible belief-shifting intervention will only have positive short-run effects but will not alter equilibrium beliefs. Consistent with this prediction, we find that two years later, the average beliefs of those

[7]We borrow the term "engagement" from Jackson (2019) who describe the formation of misperceptions about behaviors that involve positive peer effects.

[8]The friendship paradox (Feld, 1991) states that an individual's friends, on average, have more friends than them.

[9]We use Quasi MCMC methods (Chen et al., 2018) to compute confidence intervals for our parameter estimates.

exposed to the information, while still significantly more optimistic than those who were not exposed, were less optimistic than the information that was delivered by us.

How easy would it be for alternative interventions to be able to translate the short-run effect of our belief correction into a persistent change? To answer this, we compare the effects of our belief-shifting intervention with two counterfactual interventions that either (1) increase the benefits of engagement around these issues (by conducting awareness sessions or providing explicit financial incentives, for example) or (2) reduce how much individuals care about the social norm (by setting up formal job information platforms, for example, so that individuals do not worry about reputation costs when asking for financial support).

We find that it would be very demanding for these alternative interventions to generate long-run effects that match the size of the short-run effects of belief correction. For instance, we find that the perceived benefits of interactions have to increase by about 50% of the estimated mean of the benefit distribution to lead to long-run effects of a magnitude at least as large as the short-run effect of the RCT. This suggests that belief correction can produce large positive short-run effects that might not be easy to sustain even using alternative policies. Such costly interventions may also be difficult to implement due to limited policy attention and funding. However, encouragingly, as our findings show, belief correction can be used to generate both the demand and funding for such costlier interventions.

We make three contributions to existing literature. First, by establishing a causal link between beliefs about social norms and the demand for network interactions, we contribute to the extensive literature on social networks (see (Breza et al., 2019; Breza, 2016) for a review of this literature. Specifically, we show how misperceptions about social norms can lead to dysfunctional networks. Existing work has documented that concerns such as lack of trust, limited commitment, and enforcement can adversely affect useful network interactions (Ambrus and Elliott, 2021; Möbius and Rozenblat, 2016; Jackson et al., 2012; Karlan et al., 2009; Fafchamps and Gubert, 2007; Ligon et al., 2002). Taking a step back, we show that inaccurate beliefs about social norms may lead people to not even demand support from their social ties, which is a necessary precursor for interactions.[10] Further, this literature has almost exclusively focused on networks in rural settings (e.g., (Morten, 2019; Banerjee et al., 2024; Munshi and Rosenzweig, 2016; Banerjee et al., 2013; Munshi and Rosenzweig, 2009)) where caste, ethnic, or religious affiliations typically provide an institutionalized platform for social interactions. By focusing on an urban setting, our results yield insights into contexts that lack an institution that organically facilitates network interactions.

Second, we contribute to the growing literature on inaccurate beliefs and information provision experiments designed to correct misperceptions (e.g., Delavande (2023); Bursztyn and Yang (2022); Haaland et al. (2020); Bursztyn et al. (2020); Jackson (2019); Perkins et al.

[10]By showing this, our study also contributes to the literature on the formation of efficient networks by presenting empirical evidence supporting key theoretical claims (Gilles, 2021; Gilles et al., 2012; Gilles and Sarangi, 2010; Van de Rijt and Buskens, 2005; Jackson and Watts, 2001; Jackson and Wolinsky, 1996). We show that network interactions can be costly, which can prevent the existence of improving paths from empty to non-empty, efficient networks. We document that most people overestimate these costs and are willing to pay to interact with others once the inaccuracies are corrected.

(2005, 1999)). We extend this literature by estimating a micro-founded model that explains why beliefs can be inaccurate in the first place. We also document misperceptions in a domain with substantial implications for the design of social protection policies, i.e., willingness to interact with others on sensitive topics. Using structural estimation, we can predict the equilibrium impact of our belief correction intervention and evaluate its effectiveness compared to alternative policies that do not target beliefs.

Third, recent studies show that concerns such as signaling or shame can reduce incentives to seek advice (Chandrasekhar et al., 2019; Banerjee et al., 2024) and that reputation costs can reduce incentives to share information (Chandrasekhar et al., 2022). We show that belief correction can reduce these perceived costs of violating social norms that arise due to reputation and insensitivity concerns. Our study therefore provides a potential solution to mitigate the detrimental effects of such concerns. In particular, we show that beliefs about others are malleable, and they can be easier to target as opposed to using costly psycho-social interventions. Belief correction can increase the demand and financing of such costlier interventions.

The rest of the paper proceeds as follows. In Section 2, we discuss our empirical setting and highlight stylized facts that motivate the experiment. We discuss belief elicitation and validation in Section 3 and the experiment design and results in Sections 4 and 5. Additional experiments and survey evidence to investigate mechanisms are presented in Section 6. In Section 7, we present a model that explains why inaccurate beliefs about peers might arise and reduce engagement in the network. We present structural estimation results and policy counterfactuals in Section 8. Section 9 concludes.

2 Context

Our sample includes low-income, informal sector workers and their family members living in Delhi, India. The individuals in our sample live in slums and primarily work in waste picking, sorting, and recycling. We collaborated with an NGO, Chintan, that promotes awareness of health and safety among these individuals and advocates for their rights. Globally, nearly 860 million people live in slums (UN-Habitat, 2013) and around 15 million people in developing countries earn a living by sorting and recycling waste (Medina, 2008). Our context is particularly relevant given the lack of policy recognition and formal assistance available to individuals, thereby requiring social networks to function as safety nets (Marx et al., 2013; Chaturvedi et al., 2018; Ivaschenko et al., 2018). We conducted two rounds of demographic surveys with our sample: a baseline survey with 352 individuals across 14 locations in and around Delhi (Figure A.1) in 2020 and an additional survey with 791 individuals in 2 such locations in 2023.

We use data from both waves to describe the context.[11] Tables 2 and 3 show descriptive statistics for our baseline sample. The average age of participants is around 34 years, 35% are female, and around 67% are currently employed. Individuals take advice from around three

[11] The additional sample also allows us to substantiate that any patterns we outline are not specific to the period of the pandemic.

peers on average. This is the average degree of their 'advice network' (the number of social ties with whom the respondent reports discussing personal concerns around mental well-being). They interact with around four peers on average in general, including for advice-taking, borrowing, lending, and working together. This is the average degree in the 'overall network' (the number of social ties with whom the respondent reports interacting in any capacity). Individuals have been living in these locations for an average of 20 years. Only 27% are migrants, out of which only 18% report talking often to those in their origin locations. 72% of the sample reports using the networks in these slums for information about jobs.

We now describe two key features of the context that motivate our intervention.

A. Financial Distress and Lack of Formal Assistance: Individuals in our sample have very low incomes: just under 45% earn between $\$2.5 - \$5$ a day and approximately 35% earn less than \$2 a day. Incomes are also very volatile. Figure A.2 plots the difference between the highest and lowest income earned by an individual in the last 6 months before being surveyed. We find that the average range of income over the last six months is 50% of the average income in the sample. Moreover, around 75% of the additional sample report having consumption crises in the last 6 months. These include not having enough monetary resources to maintain a healthy diet, incur necessary health expenditures, or spend on children's education. At the same time, we find that individuals also have limited access to formal sources of assistance. Around 50% of them do not have access to bank accounts and among those who do, 80% struggle to obtain a loan.

We also find evidence of distress in our main survey. Roughly 50% of individuals report often feeling that difficulties were piling up high and that they felt they could not overcome them. This is also indicated by the high value of the stress index in Table 2, which measures on a scale of 1–5 how often individuals have experienced stress-related feelings in the past two weeks.[12] This is also a setting where there is limited willingness and ability to access formal sources of assistance for such mental health concerns. For example, around 90% of individuals in our control group do not feel comfortable visiting a professional.

B. Low Levels of Interactions with Social Ties: Despite the limited access to formal assistance, we find evidence of low dialogue around mental health and financial concerns with peers in the network. Figure A.3 plots the frequency of dialogue around mental and financial well-being-related topics with peers in the last two weeks. Almost half the sample reports rarely speaking about mental health issues and financial concerns with their peers respectively.

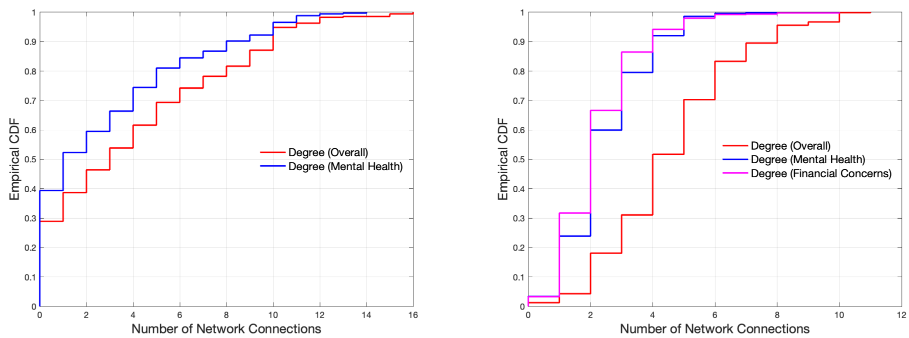
This lack of dialogue is also evident when we measure their social networks. Figure 1 shows the degree distributions of the overall networks and advice-taking networks. We can observe that the overall network degree distribution first-order stochastically dominates the degree distribution of the advice network. As we discuss in Section 3, the gaps between these networks are higher for those who have severely pessimistic beliefs about their peers'

[12] Stress-related feelings are measured using the 4-item perceived stress scale including questions such as "felt that you were unable to control the important things in your life" or "felt difficulties were piling up so high that you could not overcome them".

willingness to engage.[13] We also separately measure network links for borrowing and lending money in our additional sample and find similar patterns.

This is also surprising given the evidence that these networks can be helpful when support is demanded. To this effect, we find that those who do have more connections in their borrowing-lending networks are also those who face fewer consumption crises. Similarly, those who talk often to their peers about mental health concerns are less likely to report that their consumption is volatile. These correlations are shown in Table A.1.

Figure 1: Degree Distributions of Overall, Advice-Taking, and Financial Networks



Notes: The figure on the left plots the CDF of the number of connections in the overall and mental health-related advice networks in the main sample (in 2020) and the figure on the right additionally plots the CDF for borrowing-lending networks in the additional sample (in 2023). Individuals were asked to list up to ten other individuals in their community whom they interact with in any capacity i.e., to borrow/lend, take/give advice, work with etc. ("overall networks"), whom they only take advice from regarding mental health concerns ("advice-taking network"), and whom they contact for financial concerns ("financial concerns network").

3 Belief Elicitation and Validation

3.1 Belief Measurement

Measuring participant's willingness to engage: We first informed individuals what we mean by physical and mental health to standardize measurement. We then asked participants about their willingness to engage with others on physical health, mental health, and finance/employment-related issues respectively. 'Willingness to engage' was explained to the participants as the willingness to discuss ways to overcome concerns around these topics and how these concerns might be preventing them from achieving their goals. This is to ensure that we capture meaningful engagement. Our measure of engagement does not assume a direction in terms of who is helping whom. Importantly, this measure requires individuals to reflect on how their concerns might be preventing them and others from achieving their goals as opposed to simply being about checking in with their peers.

Incentivized elicitation of participant's beliefs about others' willingness to engage: Next, we clarified to respondents what we meant by 'community'. This was defined in terms of their geographic location i.e. individuals living around the NGO centre and affiliated to the

[13] We replicate the same exercise in 2023 and show that this feature is not specific to the time of the pandemic.

NGO were to be interpreted as a community.[14] Following this, we elicited their beliefs about their community's willingness to engage with three matters – physical health, mental health, and financial issues. For example, to identify beliefs around engagement with financial concerns, we asked them how many individuals among any 10 participants in their community associated with the NGO would be willing to share and discuss financial and/or employment-related concerns with their friends, discuss ways to overcome them, and how these problems might be preventing them from achieving their goals. These beliefs were elicited in an incentive-compatible way. Individuals were informed that they would receive a prize of Rs. 50 (a third of their total participation incentive) if their guess was accurate within +/- 1 (on a range of 0-10). The choice to measure beliefs by asking about any 10 individuals rather than asking in terms of percentages was made to ensure simplicity and ease of understanding.[15]

Incentivized elicitation of beliefs about stress and stigma: In addition to this, and particularly to get at mechanisms that we shall discuss later, we also asked individuals to predict two additional statistics. First, we asked them to predict the average level of stress in their social networks. Specifically, we asked them to predict how many individuals among any 10 participants in their community would 'often' or 'very often' feel that "difficulties were piling up so high that they could not overcome them." Comparing this prediction with the actual number that reports feeling this way allows us to infer whether individuals overestimate or underestimate the incidence of mental health distress in their communities. Second, to get at stigma, we asked if the individual agreed with the statement: "People should stay away from individuals who have mental health issues" and then asked them how many among any 10 individuals in their community would agree with the statement.[16]

3.2 Evidence of Misperceptions

Table 3 provides summary statistics about the elicited beliefs. Approximately 71% of individuals report being willing to engage in dialogue around financial well-being, but the average engagement expected from the community around this topic is around 60%. Similarly, approximately 63% of individuals report being willing to engage in conversations about mental health but the average engagement expected from others around this topic was roughly equal to 50%. Figure 2 plots the distribution of individual beliefs about their community's willingness to engage in dialogue around mental health and financial concerns. We find evidence of substantial misperceptions in individual beliefs about others' willingness to engage in a dialogue about mental health and financial concerns.

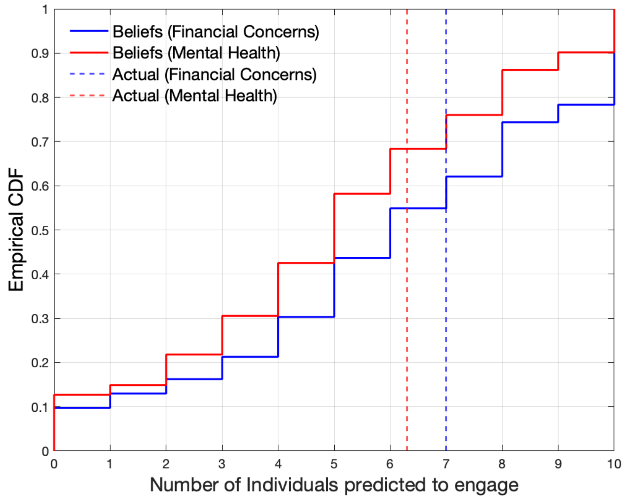
The proportions of under-estimators by type of dialogue, i.e., mental health, physical health,

[14] Our participants did not have any difficulty being clear on this as the NGO has been quite active in engaging with these communities.

[15] We found from the pilot conducted in January 2020 that many individuals did not understand what a percentage meant unless it was translated into these words.

[16] We find high levels of stigma against mental health. 50% of the surveyed sample agrees with the statement "People should stay away from individuals suffering from mental health issues". This contrasts with 63% of them being willing to share mental health-related concerns with their peers. After having spoken to the NGO staff and a psychologist, we realized that the phrasing and translation of this question may have led participants to think about severe mental health disorders.

Figure 2: Beliefs about Community Engagement



Notes: The solid lines in the figure plot the CDF of how many community members out of a randomly chosen 10 individuals believe will be willing to engage around financial concerns and mental health issues respectively. The dotted lines indicate the actual sample willingness to engage.

financial well-being, and stress, are shown in Figure A.5. Around 68% of the participants underestimate the percentage of individuals in their own community who are willing to discuss mental health-related issues. Moreover, underestimation is larger than 10 percentage points for the majority of the underestimators. We find that the proportion of individuals who underestimate mental health-related engagement is higher than those who underestimate engagement around financial concerns, even though the latter is also very high at around 58%. The majority of individuals overestimate the level of stress in the community.

3.3 Validation of Elicited Beliefs

The accuracy of elicited beliefs can be gauged by correlating them with the individual's behavior in the past (Delavande et al., 2011). We therefore validate our belief measures in the baseline sample by showing that they are correlated with self-reported behavior, network characteristics, and economic outcomes of interest, and the signs of these correlations are in line with our priors.

3.3.1 Correlations of Individual Willingness to Engage and Beliefs about Community

First, we find that participants' beliefs about their community's willingness to engage on mental, financial, and physical health are correlated with each other (Table A.2). This implies that optimism about the openness of the community on one dimension implies optimism around other dimensions as well. Notably, these beliefs are also correlated with an individual's willingness to engage. It is worth noting that individuals are not equally pessimistic or optimistic about their community's eagerness to engage across different topics; beliefs about the community's willingness to engage across different conversation topics are not perfectly correlated. Accordingly, individuals did not always report a willingness to engage on all topics. For example, 42% of those who do not agree to engage with mental health concerns agree to engage with concerns around financial well-being. Both of these facts strongly indicate that individuals carefully reflected on these questions and these beliefs are not cheap talk.

3.3.2 Correlations with Dialogue Intensity

Next, we find that these beliefs are correlated with the individual's dialogue intensity. We regress an individual's self-reported dialogue frequency around mental health concerns and financial concerns (in the two weeks before the survey) with their beliefs about their community's willingness to engage with these topics respectively. Table A.3 shows that optimistic beliefs about peers are associated with higher self-reported dialogue intensity around both kinds of topics. The correlations are significant and positive. Panel B of Table A.3 additionally shows that individuals who underestimate the sample-average engagement in their community are significantly less willing to engage in dialogue around mental health.

3.3.3 Correlations with Network Gaps

We also document a negative correlation between an individual's beliefs and the gaps between their overall and advice network degrees. This suggests that pessimistic beliefs are associated with the structure of the observed social networks, i.e., individuals who are more pessimistic about their community's openness have fewer connections in their advice networks relative to their overall network links (Table A.3). These degree gaps are measured by subtracting the size of one's advice network from the size of their overall network and computing this difference as a proportion of the size of their overall network. We find that degree gaps are significantly larger for individuals categorized as severe underestimators, i.e., those for whom the difference between their beliefs and the average in their community lies in the bottom 25th percentile.

3.3.4 Correlations with Consumption Outcomes

Using data from the additional sample, we find that beliefs about peers are also correlated with self-reported variance in consumption and whether they have faced consumption crisis events. These correlations are reported in Table A.5.

3.3.5 Large-scale validation in 2023

We also measured beliefs using the additional sample in 2023 to check that the misperceptions were not specific to the COVID-19 pandemic. We find even higher levels of underestimation of the community's willingness to engage and a high willingness to engage around both financial concerns and mental well-being. As shown in Table A.5, beliefs are also correlated with network connections as before, in that more optimistic individuals have more connections in their mental health and financial networks respectively. These results are presented in Appendix Section F.

4 Experiment

4.1 Timeline and Design

We conducted the main experiment and endline surveys with 180 individuals from February to April 2021; 92 individuals were in the control group and 88 in the treatment group. We also conducted a larger scale experiment with an additional sample of around 800 individuals in 2023.[17]

The treatment group received two pieces of information: (i) the true average sample willingness to engage in dialogue about mental health and (ii) information that the sample's willingness to engage in a dialogue about financial well-being was high.[18] More specifically, the statements used are as follows:

Information 1: *"Just like we surveyed you, we also surveyed other people and we have found from their responses that out of any 10 individuals in your community and communities similar to yours (affiliated with the NGO), X individuals are willing to discuss their mental health concerns with their peers, how they can be preventing them from achieving their goals, and how they can be overcome."*

Information 2: *"Just like we surveyed you, we also surveyed other people and we have found from their responses that the proportion of individuals willing to discuss their financial/work-related concerns with their peers, how they can be preventing them from achieving their goals, and how they can be overcome is high."*[19] [20]

All other components of the survey were identical for the control and treatment groups.

Out of Sample Participant Predictions about the Effect of the Intervention:

We anticipate the treatment to have been an information shock to the participants due to the lack of dialogue around these issues and the stigma associated with talking about mental well-being. To this effect, we asked participants in our additional sample to predict how they think others would respond if Information 1 were delivered to them. 67% of the participants anticipated an increase in engagement with savings groups after hearing the above information, and 42% thought that this increase would be large.

4.2 Balance Tests

We show that the sample is balanced across control and treatment groups across a host of baseline variables including demographic information, network connections, dialogue intensity, beliefs about mental health, physical health, financial concerns, participant

[17]We discuss this in more detail in Section F in the appendix.

[18]We provided information about sample-level averages computed using data on willingness to engage from our initial surveys of about 350 individuals rather than community-level averages. We did this because some communities had low response rates and we did not wish to shift beliefs about their community based on information obtained from small samples. Importantly, since the communities are very similar to each other, information from the entire sample is informative and helpful.

[19]Note that we do not provide a precise estimate here (the precise estimate of the proportion of individuals willing to share these concerns is 70%) because unlike mental health, where there is a high proportion of under-estimators, in this case, only 58% underestimate and we do not wish to make them pessimistic.

[20]The second piece of information was provided closer to the end of the survey (with only a few questions remaining) before the questions related to financial outcomes.

well-being, and own willingness to engage. In particular, we test balance on 44 variables and find that the sample is unbalanced only for 2 variables. These results are presented in Tables B.1 and B.2. We also regress treatment status on these baseline covariates and find that the F statistic is 0.68 and the corresponding p-value is 0.89. This suggests that the baseline characteristics are balanced across treatment and control. However, for robustness, we also run a specification that controls for any unbalanced covariates.

4.3 Specification

We run a simple regression where we regress each outcome on the treatment indicator for the individual, with robust standard errors. We additionally report p-values of wild bootstrapped t-statistics in line with Cameron et al. (2008) where we cluster the standard errors at the level of the NGO center. Further, we also conduct correction for tests of multiple hypotheses as per Benjamini et al. (2006); Anderson (2008). These q-values are computed at the level of the outcome families. We also run an additional robustness exercise where we take a very conservative approach and treat all the outcomes as one family before computing the q-values.

Finally, we also account for the fact that participants may not have always completed the endline survey as a result of which some outcomes may have received more responses than others. To address any balance-related concerns that may arise due to this, we adopt the following additional strategy as a robustness check in the appendix. We run balance tests for *each* sub-sample for which the outcome variable is non-missing. Then, we include the unbalanced controls (at 5% significance) in a conservative, robust specification. All the main results are robust to this alternative specification unless specified otherwise.

5 Results

We now present the results on pre-specified outcomes measured immediately after the treatment.

5.1 Community Engagement

As Table 4 shows, we find a 14 percentage point increase in community engagement (significant at 1%).[21] This variable measures the willingness to have useful interactions with the community and is the mean of three binary variables: (a) willingness to train as a listening volunteer for their community (i.e., getting trained to listen to the anxieties of other community members)[22], (b) willingness to contribute to setting up this listening service, and (c) willingness to participate in a savings group.

Figure 3 plots the treatment effects on these engagement measures separately. We find that the treatment increases the probability of being willing to enlist as a listening volunteer and participate in a listening service by about 16 percentage points (significant at 5%). Treated

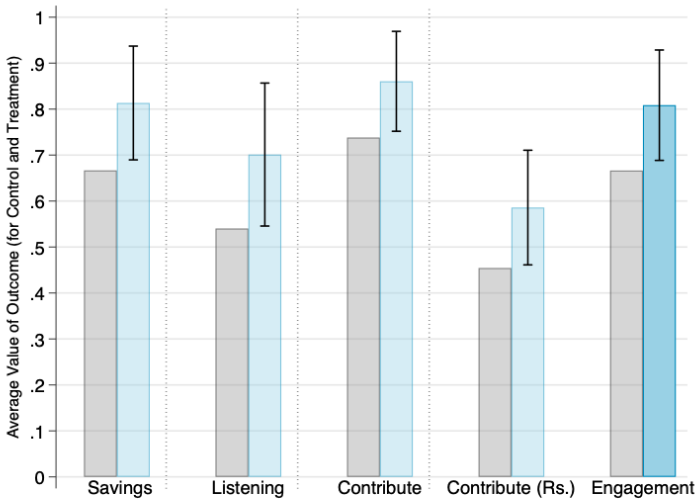
[21] The q-values for all community engagement outcomes are also significant at 5%.

[22] This would have been similar to the services that charities like Samaritans offer in the UK and US.

participants are also about 12% more likely to wish to financially contribute to set up training sessions for this and donate roughly Rs. 6.6 more, which is a 29% increase compared to the control group (significant at 5%).[23] We also find that about 67% of the control group is willing to participate in savings groups, and the treatment increases this further by approximately 15 percentage points (significant at 5%).[24]

The large increase in the demand for savings groups and listening services shows that beliefs about peers can affect the extent to which individuals are willing to interact with their social networks and pay to set up informal avenues for interactions. Our robustness exercise also shows that the effect of the treatment on these various outcomes is still significant when unbalanced controls are included.

Figure 3: Treatment Effects on Engagement with the Network



Notes: The above figure shows the average value of each outcome for the control group (in gray) and the treatment group (in blue) with 95% confidence intervals for the difference between the two values. "Savings", "Listening", and "Contribute" refer to binary variables indicating whether the individual is willing to participate in savings groups, train and volunteer for a listening service, and actually contribute to set it up. "Engagement" is an average of these three variables. "Contribute (Rs.)" is the actual monetary contribution made by the individuals, normalized to be between 0 and 1.

5.2 Demand for Formal Support

Panel B in Table 4 shows that the treatment also affects the participant's investment in their own health. We find a reduction in self-reported hesitation to speak to a doctor which implies that becoming informed about their community's higher willingness to engage reduces the stigma associated with visiting a doctor and likely lowers the perceived costs of violating the social norm. At the same time, we also find that treatment reduces participation in an experimenter-run depression scoring (using standard questions from the PHQ-9 questionnaire used to measure depression) and the likelihood of wanting to listen to helpline

[23]Individuals were informed that this money would be deducted from their prize of Rs. 50 if they won based on their guesses and would be informed about it at the end of the endline survey. Participant beliefs about their community are balanced across treated and control groups so there is no reason to anticipate that the treatment group would be selectively more pessimistic about their chances of winning this prize.

[24]This question, unlike the other measures of community engagement, was asked after the participants received the two treatment statements about mental health concerns and financial concerns.

numbers. We find that these results are robust to the inclusion of unbalanced controls. The impact of the treatment on these measures suggests that an individual may consider support from their community and support from formal sources (assisted by the experimenter in the form of depression scoring and helpline numbers) as substitutes. While the willingness to talk to a doctor is purely indicative of lower hesitation, the actual decision to not participate in depression scoring or listen to helpline numbers suggests a tension between formal and informal sources of assistance.

5.3 Other Types of Dialogue

Panel C in Table 4 provides evidence that correcting beliefs about others' willingness around mental health can have significant positive spillover effects on other kinds of dialogue. In particular, the treated group is 21% more likely to report a willingness to talk to their friends and family about physical health-related issues. We know from the baseline survey that individuals have high levels of stigma associated with talking about physical health-related concerns. This result suggests that correcting beliefs about a community's stance on particular forms of network interactions can have positive impacts on dialogue around other stigmatized topics as well. We find that the treatment effect is also robust to the addition of unbalanced controls.

5.4 Demand for Additional Information

Table C.1 shows that the treatment does not affect a participant's demand for additional information about mental health. It does not have any effect on an individual's willingness to participate in a potential mental health-related information session and their willingness to listen to good practices about having mental health-related conversations. Admittedly, our ability to detect effects on the take-up of the information session is low as 67% of the control group agrees to participate in the information session anyway. This is despite the time costs involved in this in the future and potential costs in terms of stigma. We realized that individuals may have interpreted this session as one of the regular training sessions (unrelated to mental health) that the NGO organizes, which is why the stigma associated with attending information sessions was perceived to be low.

5.5 Other Outcomes

We find no impacts on other outcomes measuring financial self-efficacy and concerns around stigma. These results are reported in the supplementary appendix in Table C.1. Contrary to our priors, 92% of the individuals in the control group were willing to allow their names to be revealed to encourage others to participate in the information session by mentioning that they agreed. Given the baseline finding of considerable reported stigma around mental health, we realized that this was a weak measure of stigma. As information sessions are commonly held by the NGO, individuals likely did not incur any stigma costs due to it.

We also do not find any effect on an individual's belief in their ability to manage their finances, afford their children's education (if applicable), or start a business if they wanted to.

Our prior was that correction of beliefs about community willingness to engage with *both* mental health and financial concerns would have improved financial self-efficacy. However, the finding provides suggestive evidence that traditional mechanisms of favor-exchange and/or risk-sharing may not already be in place within the community so being more optimistic about others in terms of dialogue doesn't necessarily mean that one can also borrow from them in times of need. The fact that individuals wish to engage in useful interactions is evident from a significant increase in the willingness to participate in savings groups and willingness to pay to set them up.

5.6 Two-week and Two-year Follow up

5.6.1 Effects after two weeks

We conducted a follow-up survey two weeks after the intervention in 2021.

The list of pre-specified outcomes measured during the follow-up survey is provided in the appendix in Section E.1. Around 90% of our follow-up surveys were conducted during or after April 2021. This was during the peak of the COVID crisis in India. This affected our follow-up survey as we could only contact 112 people, 57 of which were in our treatment group. However, we test balance on 44 baseline variables and find that the sample is unbalanced only for 2 variables. We also present robustness checks controlling for these unbalanced variables and show that attrition in this survey is not correlated with treatment status or with baseline characteristics.

We find in Table 5 that individuals who received the treatment are $11-19\%$ more likely to have reported that they reached out to any peer to discuss their own or their peers' mental health or financial concerns respectively in the two weeks following the treatment. We also find that the treated group has reached out to 1 additional peer about both mental health (p-value<0.05) and employment-related concerns (p-value<0.01). It is worth noting that the effect on network interactions is stronger when we ask whether individuals are reaching out to others rather than whether others reached out to them. This is consistent with the fact that we did not deliver the information intervention to everyone in the community, which is why the response is stronger for those who were treated. Additional results of the follow-up survey on other outcomes such as well-being and dictator games, along with concerns about how COVID-19 could have affected these results are discussed in Section E in the appendix.

5.6.2 Effects after two years

In addition to the two-week follow-up, we conducted a demographic survey in 2023 where we classified individuals in terms of the following measures- (1) 'Previous Participant' i.e., whether they had participated in the baseline survey and/or experiment in 2020-21, (2) 'Heard about Beliefs of Community' i.e., whether they reported having heard about the views of members of their community, and (3) 'Heard about Beliefs of Community and Previous Participant' i.e., whether they reported having heard about the views of members of their community and were a part of the previous surveys and experiments i.e., directly contacted by the survey team and (4) None of the above.[25] We cannot track and merge individuals across the two waves but can identify previous participants and those exposed to the treatment information using self-reported measures. We find that out of a total of 469 individuals, 60 report being previous participants (Category 1), 37 report having heard information about their community's views on talking about mental health and finances shared in the main experiment (Category 2), and 24 were previous participants who report having heard this information (Category 3). The remainder had neither participated nor heard about the previous experiment.

We acknowledge that this categorization of capturing exposure to the treatment is imperfect in that (a) they rely on self-reported information and (b) they may capture differences in characteristics apart from exposure to treatment. While we cannot tackle the first issue, we tackle the second by using post-double selection Lasso (Belloni et al., 2014) to account for the fact that individuals who are exposed to the intervention or who report being previous participants might have different characteristics in terms of their gender, income, age, and number of connections in the overall network.

We report the effect of being in one of the above categories on beliefs, dialogue intensity, willingness to engage, and volatility of consumption in 2023 in Tables 6 - 7. We find evidence that individuals who report being in categories 1-3 have more optimistic beliefs about others' willingness to engage around mental health and financial concerns, even in 2023, with the effects being significantly positive for those who report having heard the information. Moreover, these individuals are more likely to be willing to engage with others around mental health and financial concerns and are more likely to report engaging in higher (i.e., above-median) levels of dialogue around these topics with their peers. We find that these effects are significant across most specifications. Finally, these individuals also report facing significantly lower volatility of consumption and report facing fewer consumption crises. This suggests that beliefs about the community can increase network interactions and consequently improve important economic outcomes.

While the evidence is only suggestive, it is reassuring that many effects are still significant even for the most conservative measure of exposure i.e., whether the respondent was a previous participant or not. However, it is still not clear if we have shifted beliefs and engagement permanently. We find, for example, that the average beliefs of those who report having heard information about their community's willingness to engage are still lower than the belief that was delivered to them. We will explore whether the intervention may or may not have shifted equilibrium beliefs when we discuss the model and structural estimation.

5.7 Larger-scale Validation

Before proceeding, we briefly present the results of an additional experiment conducted with around 800 individuals in 2 NGO centers in 2023. This exercise was conducted to test whether

[25] This survey was conducted as part of the additional surveys done in 2023 where we had an overlap of 1 center from the previous survey. Individuals who said yes to any of the above measures are excluded from the results of the experiment.

(a) the patterns of low dialogue and underestimation of peer willingness to engage and (b) the causal effect of beliefs about others' willingness to engage on an individual's own engagement with the network, is not specific to the timing when the experiment was conducted. We find that the baseline patterns of low dialogue and pessimistic beliefs also hold in this larger sample. In addition, we implemented a slight variation of the intervention, where we provided information about the willingness to engage of *other* similar communities, as opposed to their own.[26] We find that this information continues to have a significant causal impact on willingness to engage.

However, the underlying mechanisms are different in this case. Specifically, we find that those who are pessimistic about their community are significantly less likely to engage in a savings group when they are told that *other* communities are more willing to engage. At the same time, those who are optimistic about their own community behave in the opposite manner. This pattern also holds for volunteering for listening services. We continue to find that treated participants give financial contributions to set up informal avenues for interactions such as savings groups and the payments are significantly higher for those who are pessimistic about the social norm. What we can infer from this is that the choice of reference group significantly affects individual responses. Specifically, individuals respond differently depending on whether we disclose the willingness to engage of their own community versus that of other communities. While the former leads them to learn about their peers, the latter encourages relative comparisons i.e., learning about their peers in relation to others. A more detailed discussion of these results is in the appendix (Section F).

5.8 Heterogeneity

We analyze heterogeneous effects by baseline willingness to engage, dialogue intensity, number of network connections, stigma, and beliefs.[27] The results are shown in Tables C.2-C.7 in the appendix. We highlight the following key findings.

5.8.1 Heterogeneity by Baseline Engagement

First, Table C.2 shows that the treatment effect on community engagement was significantly higher among those individuals who were not willing to engage with their peers regarding mental health. They are more likely to express a willingness to (a) engage with their community, (b) participate in a mental health-related information session, and (c) listen to good practices to have conversations with others.[28] This suggests that the intervention, as expected, helps those who need it most. This is further corroborated by the results in Table C.3 where we perform heterogeneity by baseline dialogue around mental health and financial concerns. We find that the treatment effect on community engagement is significantly higher

[26] This was done to avoid an additional round of belief-elicitation in these communities due to financial and logistical constraints. Participants were informed that the information provided to them about similar communities was collected in 2021.

[27] We also discuss heterogeneity by baseline beliefs about stress in Section 6 where we discuss mechanisms.

[28] While the p-value for the interaction term is not significant for the last outcome, the bootstrapped p-value is still significant at 5%.

for those whose dialogue around mental and financial well-being was less than or equal to the median in the baseline.

5.8.2 Heterogeneity by Baseline Networks

Next, Table C.4 shows heterogeneity in treatment effects by the size of an individual's overall network. Individuals who are more connected in the network are more likely to respond to the treatment by reporting to be willing to participate in the savings group. In fact, the treatment effect is zero for those who have no peers in the network and every additional peer increases the probability that a treated individual says "Yes" to participation by 3 percentage points. More connected individuals also contribute significantly more to setting up the listening service. Moreover, individuals who have more peers are on average significantly less likely to make positive contributions to the listening service in the community but the treatment increases this probability. We will revisit this result once we introduce the theoretical setup whose core assumption directly implies that those who are more connected in the network must benefit more from the intervention.

5.8.3 Heterogeneity by Baseline Stigma

Third, Table C.5 shows that conditional on saying "Yes" to financially contributing to the listening service, individuals who express stigma towards mental health (i.e., who report they wish to "stay away" from those suffering from mental health concerns) donate Rs. 7 less, on average, compared to those who didn't explicitly report stigma. This is expected since the listening service is directly associated with mental health. Interestingly, we find that individuals who express stigma and are in the treatment group donate significantly more (i.e., Rs. 9 more) than those who are in the control group and express stigma. While receiving optimistic beliefs about community engagement may not address stigma, this suggests it can significantly increase community engagement among those who exhibit it. One potential explanation is that the stigma revealed by the participant to the experimenter might itself depend on their beliefs about the norm and what is considered acceptable, i.e., those who are pessimistic about the social norm can also be those who express stigma. For those with prior stigma, once the information shock reduces pessimism about others' willingness to engage, it might also increase their inclination to engage and contribute towards doing so.

5.8.4 Heterogeneity by Baseline Beliefs

Finally, we also check for heterogeneity by whether individuals underestimate community engagement in Table C.6. We are not powered to detect heterogeneity by whether individuals underestimated engagement with concerns around mental health or finances. However, most effects go in the expected direction and are larger in magnitude for underestimators.

6 Mechanisms

6.1 Main Mechanism: Reduction in the Cost of Violating the Norm

6.1.1 Network Prediction Experiment

To disentangle the mechanisms, we first implement an additional experiment to causally identify the impact of various costs that can reduce network interactions in this setting. We run this experiment with around 800 waste-pickers and their family members in the same setting. Each individual is asked to predict whether they think a link would exist between two randomly chosen, hypothetical agents A and B in their community, where A is the potential advisor and B is the person who needs A's support with their financial or mental health-related concerns respectively.

Crucially, we exogenously vary the advisor A's characteristics along three distinct dimensions: (1) whether A is central (i.e., well connected) in their community, (2) whether A has contacts in private jobs, and (3) whether A has attended a mental health sensitivity training organized by the NGO to talk sensitively about these issues. These three binary characteristics are varied to test the impact of the following perceived costs respectively: (1) reputational cost as A is central and can gossip, (2) signaling cost if A can make negative inferences about B's type and be less likely to inform B about jobs or recommend them, and (3) interaction cost if A is not sensitive and might mock/mistreat B. For a randomly chosen vector of characteristics, the respondent was asked to predict whether B would approach A for mental health-related support and financial support respectively. We varied the vector 3 times for each respondent, giving us about $2,370$ predictions for each randomly chosen vector of characteristics, leading to a total of $4,740$ predictions for mental health and financial well-being related interactions, respectively.

We find that 23% of links are rejected in the case of mental health-related support and 24% in the case of financial support.[29] Table 8 presents the results of the experiment.[30] We find that, as the respondent becomes more pessimistic about their community's willingness to engage with them, they become significantly less likely to report the existence of a link between B and A when A is network central or A has attended sensitivity training. This pattern holds for both mental health and finance-related support.[31]

Why are individuals who are more pessimistic about the social norm of engaging more likely to reject these links? Regarding advisor A's centrality, we find that out of those who reject links when A is central, 42% say that it is because they think A may not have enough time to speak to B. At the same time, 39% say it is because they think B would fear being gossiped about. We also find that while A having contacts in private jobs does not have a causal impact on whether individuals think B would seek help from A, signaling for jobs is a dominant concern for many individuals in this setting. In particular, we find that 30% of individuals who reject

[29] Given that this is a setting with very low dialogue and advice-taking around these issues, we speculate that it is the hypothetical nature of the exercise that results in fewer rejections than expected.

[30] We also employ a probit model instead of a linear probability model. The results are similar across both specifications.

[31] The results are also similar across OLS and probit specifications.

links when A has contacts in private jobs say that this is because they fear that A will think B is not a capable candidate to recommend for jobs. Interestingly, and contrary to our expectations, we find that attendance at the sensitivity training acts as a negative signal about A, potentially leading participants to worry that there may be something wrong if the individual had to be trained to act sensitively in the first place.

The evidence from this hypothetical network formation experiment suggests that the reputation and interaction-related costs (both induced by a social norm of low dialogue) are significant reasons why links are not formed in these communities. We find that these perceived costs are significantly higher for those who are more pessimistic about the social norm. As a result, we posit that, by shifting beliefs, the intervention would have reduced the perceived costs of violating social norms. This is also reinforced by our qualitative interviews with a few participants. When asked if they feel comfortable talking about their concerns, respondents report "feeling ashamed if the other person refuses their request", worried that the person "will make fun [of them] later", and feeling "suffocated/trapped because others can make fun of their poverty".

6.2 Alternative Mechanisms

Next, we use additional experiments and survey evidence to rule out alternative mechanisms.

6.2.1 Social Desirability and Experimenter Demand

We show that experimenter demand or social desirability concerns are unlikely to drive our results using (a) a list experiment and (b) randomly varying whether individuals' decisions to participate in a savings group or as a listening volunteer were visible to the enumerator in our additional sample of 800 individuals. In addition, we also discuss how the survey responses collected during and before the RCT show that it is unlikely that social desirability concerns are driving participant responses.

List Experiment

In this exercise, we randomly divided individuals into one control and two treatment groups. The control group receives three statements and is asked to report how many of these opinions they agree with. The treatment groups are given the same three opinions as the control group and additionally receive either one or two more statements around mental health-related and financial well-being-related engagement with peers respectively. Importantly, all groups are only asked how many statements they agree with and not whether they agree/disagree with each statement. This allows them to mask their response to individual statements. The statements are as follows: (1) Only individuals who have received formal education should enter into politics. (2) The Delhi government is taking the required steps to deal with air pollution. (3) Teachers should be paid more remuneration than film actors. (4) Individuals should take time to listen to the mental health concerns of their peers (Only to Treatment Groups 1 and 2) (5) Individuals should take time to listen to the employment or money-related concerns of their peers. (Only to Treatment Group 2)

If individuals only reported favorable views about engaging with peers around these topics due to experimenter demand or social desirability, we would expect to find no difference between the number of statements that the control and treatment groups agree with. Instead, we find that the control group agrees with 2.21 statements on average, Treatment Group 1 agrees with 2.97 statements on average, and Treatment Group 2 agrees with 3.71 statements on average. All pairwise comparisons suggest that the average agreements are significantly different across groups.

Randomly varying whether participant response is visible to the enumerator

In this exercise, we randomly varied whether participants were asked to report their willingness to sign up for savings groups and training to be a listener to the enumerator or could enter this privately on their screen without the enumerator being able to see their responses. We find that there was no difference between their responses in the two conditions (Table D.2). Moreover, as shown in the table, it also does not correlate with whether or not they were randomly provided with information about others' willingness to engage, mitigating concerns around experimenter demand.

Additional Survey Evidence against Social Desirability

We now outline how the survey responses also indicate that social desirability concerns are unlikely to affect our results. First, in our baseline survey, around 49% of the participants agreed that one must stay away from those with mental health concerns. This suggests that social desirability concerns did not prevent a significant proportion of individuals in both groups from expressing their opinions to the surveyor. Second, we find that individuals do not always agree to engage on all the different dimensions we ask them about. For instance, 42% of the individuals who do not agree to engage with mental health-related concerns, agree to engage with concerns around financial matters. Third, as we show in Section 3.3, the choice to engage is correlated with baseline beliefs about peers' willingness to engage. Third, we find a significant treatment effect on an incentivized/costly outcome (contribution to finance listening services) implying that the treatment effect cannot entirely be attributed to costless signals to appear socially likable. Finally, our findings that the treatment differently (and positively) affects the subgroup that (a) had low dialogue frequency around mental and financial well-being in the baseline and (b) reported that they do not wish to discuss mental health concerns with peers (Tables C.2 and C.3), further adds to our confidence that the results are not driven by experimenter demand effects.

6.2.2 Updating Beliefs about the Incidence of Stress

It can be argued that the treatment information about others' willingness to engage led individuals to update their beliefs about the incidence of mental health concerns in their community. We do not think this is the case for the following reasons.

Firstly, we measured individual beliefs about stress in their community during the baseline and found that over 60% of individuals overestimate how stressed their community is. This suggests that knowledge about the incidence of stress was not the binding constraint preventing individuals from sharing financial or mental health concerns with their network ties. Additionally, as shown in Table C.7, we find that individuals who underestimate stress (computed using baseline beliefs about stress in the community and actual community-level average) are not more likely to respond to the treatment than those who overestimate it.

Similarly, we might also wonder if the treatment might have led individuals to update their beliefs about the incidence of stress downward if they take others' willingness to engage to mean that others are not as stressed and can therefore make time for them. If this was true, we could expect the overestimators of the incidence of stress at baseline to benefit the most from the treatment. However, we do not find this to be the case as the interaction term for overestimators of stress and treatment is negative and insignificant as well (Table C.7).

6.2.3 Altruism or Social Pressure

While it could be the case that the treatment increases community engagement due to altruism or social pressure after being informed about the community's willingness to engage, we do not think these drive our results. Our reasoning is as follows.

First, as outlined above, dialogue frequency was low before the intervention even though the vast majority of individuals overestimate levels of stress in their community. Altruistic concerns alone are therefore unlikely to drive engagement.

Second, if individuals only engaged for the benefit of others and not themselves (i.e., they selflessly reciprocate), then it is not clear why we would find that treated individuals are significantly (20 percentage points) less likely to listen to helpline numbers. Substitution between help from the community and help from outside the network likely explains this result instead.

Third, individuals were informed that their decisions to contribute to setting up a listening service for the community and to train for volunteering for this will not be visible to others. Finally, we also asked individuals if they would like their names to be included in the list of potential participants in an information session around mental health which can be used to motivate others to participate. We do not find any significant differences between the treatment and control groups in this outcome which lends additional support that social pressure to conform or wanting to appear supportive in front of others may not be the relevant mechanism.

6.2.4 Updating Beliefs about Benefits of Interacting

Lastly, it can also be that the treatment led individuals to update their beliefs about the benefits of engaging. We do not think this is the main mechanism at work as the majority of participants reveal a high willingness to engage at baseline, indicating that they are aware that there is a benefit to engaging with the network. The fact that this willingness doesn't materialize into actual engagement suggests that it is not that participants are not aware of the benefits of interacting but rather that they are worried about the costs of doing so.

Our setting is also not one where there are no financial or mental-health-related interactions whatsoever. While the frequency of dialogue is low, such interactions are not absent altogether and as we have outlined, those who are more connected are also more likely to have better consumption outcomes. Therefore, it seems less plausible that those who do not interact are unaware of the potential benefits others in their community are receiving from interacting. In fact, in qualitative interviews, many respondents also say that they are aware that it is important to interact with their community around these issues.

7 Theoretical Framework

We adapt the model in Jackson and Yariv (2007) and Jackson (2019) to illustrate how individuals may form inaccurate beliefs about others' willingness to engage and how this may affect their decision to engage.

Let N be the set of individuals in a society connected in a social network represented by the matrix G where $g_{ij} = 1$ if an undirected link exists between i and j and 0 otherwise. An agent's degree d_i is the number of connections they have, i.e., $d_i = \sum_j g_{ij}$. The distribution of degrees of all agents is denoted by $P(d)$. Individuals choose whether to 'engage', $e_i \in \{0, 1\}$, which can be interpreted as the choice to take a visible action to engage with their community on matters related to mental health and financial well-being. For example, this can include directed interactions such as contributing to a community listening service (to hear the mental health concerns of other community members) or undirected interactions such as participating in a savings group (to share own financial concerns and hear the concerns of others). [32]

Choosing $e_i = 1$ provides individual i with a private benefit $b_i > 0$, independently drawn from an atomless distribution function H. The individual also faces a cost of engagement $c(d_i, 1 - E[e_{j:j\in N}])$ proportional to the expected disengagement in society $(1 - E[e_{j:j\in N}])$ and their network degree (d_i). We assume that c is weakly increasing in both its arguments, such that $\frac{\partial c}{\partial d_i (1-E[e_{j:j\in N})} > 0$ and that $c(:, 0) = 0$.

Given these assumptions, agent i's utility can be written as follows:

$$U(e_i, d_i; e_{j:j\in N_i}) = [b_i - c(d_i, 1 - E[e_{j:j\in N}])]e_i \tag{1}$$

An agent i will engage if $c(d_i, 1 - E[e_{j:j\in N}]) < b_i$ i.e., the probability that an agent i engages should be equal to $1 - H(c(d_i, 1 - E[e_{j:j\in N}]))$. We assume that agents know their degree d_i, the degree distribution in the network $P(d)$, benefits of engagement b_i, and the distribution of benefits H. Before proceeding to the implications of this model, we present empirical evidence supporting the assumptions of the model.

7.1 Empirical Evidence for Model Assumptions

Let us consider how the costs of engagement depend on the expected non-engagement in the network. As discussed in Section 6, these costs can include reputational costs due to gossip,

[32] As we are interested in illustrating how misperceptions about social norms can arise within a network, we abstract from modeling the directed link-by-link choice to interact with a specific peer. We do not incorporate network formation here but discuss the implications of this in the appendix in Section G.2.

interaction costs if met with insensitivity and signaling costs if it hurts job prospects. We have assumed that costs of engaging are rising in network degree, and more specifically that $\frac{\partial c}{\partial d_i(1-E[e_{j:j\in N}])} > 0$ i.e., that engaging is more costly for more connected individuals when most others are not engaging. In line with this assumption, we find that more connected individuals in our sample are also those who are less likely to engage. In particular, using data from our additional sample, we show that degree centrality is negatively correlated with wanting to participate in savings groups (see Figure G.1 in the appendix). While this evidence is not causal, existing literature on social norms has also suggested that while more connected agents can create and challenge norms, it can be more costly for them to do so precisely because of the losses they can incur if they do not succeed (Young, 2015). Additionally, costs can also be higher for more connected individuals as they may not have as much time to engage. We do find substantial evidence for this line of reasoning in our network prediction experiment (outlined in Section 6) where 42% of individuals do not think that advice-taking links will be formed when the advisor is more connected because the advisor will not have time.

Furthermore, we find that more well-connected agents respond more to the information intervention. This is in line with the strictly positive cross derivative that we have assumed $\frac{\partial c}{\partial d_i(1-E[e_{j:j\in N}])} > 0$. To corroborate this further, we find that when we regress the decision to participate in savings groups and listening services on the interaction between an agent's degree centrality and their beliefs about average non-engagement around mental or financial well-being, the coefficient on the interaction term is almost always negative.[33]

7.2 Equilibrium

As in Jackson and Yariv (2007), we can solve for a symmetric Bayesian Nash equilibrium such that every agent will input the equilibrium engagement i.e., $E[e_{j:j\in N}] = \hat{e}$ into their utility function and agents with the same degree d_i and benefit b_i would choose the same action.

Recall that the probability that an agent i with degree d_i engages is equal to $1 - H(c(d_i, 1 - E[e_{j:j\in N}]))$ where H is the CDF of the benefit distribution. The Bayes-Nash equilibria $\hat{e}$ in this setting can then be defined using the condition below:

$$\hat{e} = \sum_d P(d)(1 - H(c(d, 1 - \hat{e}))). \tag{2}$$

The equation specifies that the proportion of individuals who engage in equilibrium (i.e., $\hat{e}$) must be equal to the probability that a person with degree d would engage, summed across all degrees, weighted by the degree distribution.

We proceed as in Jackson (2019), asking what if individuals gain utility as per this game but best respond to the actions taken among their peers? This leads agents to naively use the expected proportion of their friends in their network neighborhood $N(i)$ who do not engage

[33] The coefficient is weakly positive (but close to zero) when we use listening services as a measure of engagement and beliefs about others' willingness to engage around mental health as the measure of beliefs. It is negative in the remaining cases, i.e., when we either use beliefs about others' willingness to engage around financial well-being as a measure of beliefs or when we use savings groups as the measure of engagement.

i.e., $1 - E[e_{j:j\in N(i)}]$ as a proxy for the society-wide average $1 - E[e_{j:j\in N}]$. This is because individuals may not be aware of the norm and try to estimate it using what they observe among their peers. The distribution of the degree of an individual's neighbors is equal to $\widetilde{P}(d) = \frac{d}{E[d]}P(d)$ as per the 'friendship paradox' (Feld, 1991; Jackson, 2019). We know from Feld (1991) and Jackson (2019) that $\widetilde{P}(d)$ first order stochastically dominates $P(d)$. This is because peers, by virtue of being connected to someone, by definition, are more connected than a random person in the network. Then, in this case, we must have that:

$$\widetilde{e} = \sum_d \widetilde{P}(d)(1 - H(c(d, 1-\widetilde{e}))) = 1 - \sum_d \widetilde{P}(d)(H(c(d, 1-\widetilde{e}))) \tag{3}$$

where $\tilde{e}$ denotes the equilibrium engagement when individuals use $\widetilde{P}(d)$ instead of $P(d)$. We rewrite the equation (as in the second term above) to avoid having to write the difference terms (i.e., terms that subtract the probabilities and expected actions from 1), and for ease of interpretation of the interaction effects.

Let us denote the probability that a randomly chosen peer does not engage by a. Clearly $\widetilde{a} + \widetilde{e} = 1$. Then, the above equation implies that:

$$\widetilde{e} = 1 - \sum_d \widetilde{P}(d)(H(c(d, 1-\widetilde{e}))) = 1 - \widetilde{a} \implies \widetilde{a} = \sum_d \widetilde{P}(d)(H(c(d, \widetilde{a})))$$

Given that peers have a higher degree on average (compared to a random person in the network) and the costs of engagement are weakly rising in the agent's degree, the probability that a randomly chosen peer does not engage is always higher than the probability that a randomly chosen person from the network does. This is why individuals overestimate the costs of engagement when they proxy non-engagement using their peers. The following Lemma highlights this result.

Lemma 1 $E[H(c(d,a)] \leq \widetilde{E}[H(c(d,a)] \; \forall a \in [0,1]$

The proof for Lemma 1 is provided in the appendix in Section G.4.

The following proposition compares equilibria with and without the friendship paradox in the simple example where there exist three equilibria of non-engagement out of which there is one stable equilibrium at zero and two stable and unstable strictly positive equilibria respectively – as shown in Figure 4. This follows from standard assumptions about the shape of the CDF H of the cost function that we will assume for now but will estimate empirically later on.

Proposition 1 *Let $E[a_{j:j\in N}]$ be the society-wide expected non-engagement and a_h and a_l be the non-zero symmetric Bayes Nash equilibrium non-engagement that solves Equation 2 such that $a_h > a_l$. Let $E[a_{j:j\in N(i)}]$ be the expected non-engagement among the agent's friends and let $\widetilde{a}_l$ and $\widetilde{a}_h$ be the symmetric Bayes Nash equilibria that solve Equation 3 such that $\widetilde{a}_l < \widetilde{a}_h$. Then,*

$$\widetilde{a}_l \leq a_l \text{ and } \widetilde{a}_h \geq a_h$$

This directly follows from Lemma 1. The proof is provided in Section G.5 in the Appendix.

Figure 4 illustrates this result. The proposition implies that in a society where there are three possible equilibrium values of non-engagement out of which two are non-zero, one stable, and one unstable, the highest stable equilibrium non-engagement is even higher when individuals observe the actions of their peers and not the entire society. The lower unstable equilibrium is even lower, implying that a larger shift in expected engagement is required to cross the tipping point and reach the zero equilibrium non-engagement. The inequality is strict if the cost function is strictly rising in the agent's degree. While we have assumed a shape for H for illustrative purposes here, we will estimate this shape using the data and leveraging the shift in a that is induced by the RCT. To understand this, let us consider the dynamic version of the equilibrium equation.

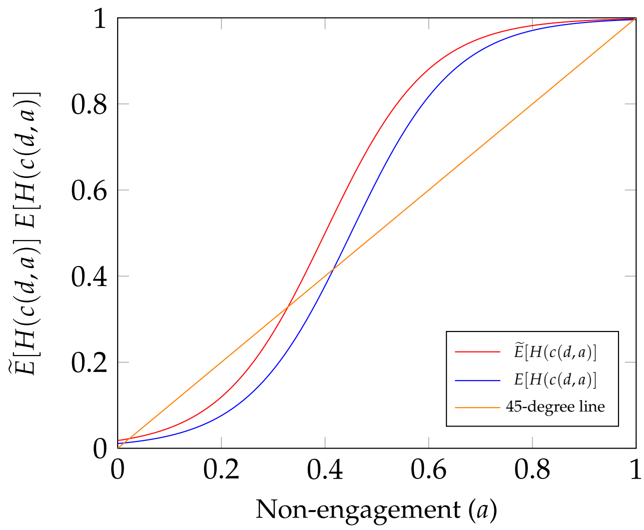


Figure 4: Illustration for Proposition 1 under standard assumptions about H.

7.3 Dynamics

Consider the dynamic version of Equation 3:[34]

$$\widetilde{a}_t = \sum_d \widetilde{P}(d)(H(c(d, \widetilde{a}_{t-1})))$$

What will happen when we provide individuals with information about the proportion of others in the community who are willing to engage? The provision of society-level willingness to engage will be useful in the case where individuals overestimate non-engagement because they naively proxy for the norm using the actions taken by their peers. While society will be in equilibrium as per Equation 3 before the information is provided, providing them with information about a in period t will affect everyone's choices and influence a_{t+1}. This will then affect how much engagement they expect in t_2, which will then affect actual engagement in $t+2$, and so on. Depending on the location of the original equilibrium and the shape of H, this intervention will either lead to engagement reverting

[34]We can think of the previous period's non-engagement a_{t-1} as the relevant 'belief' that individuals have about others' willingness to not engage while deciding whether or not to engage in period t.

back to the same equilibrium (i.e., achieve only short-run effects) or reach a new one (i.e., achieve both short-run and long-run effects). Note that we do not suggest that the provision of this information will change the naive manner in which individuals choose their actions i.e., they will still continue to be affected by the friendship paradox and will update as per Equation 3. Instead, this information shock will shift the norm based on which they make their choice in period $t+1$ and the dynamics will continue thereon.

8 Structural Estimation

We now proceed with structurally estimating the model using data from the randomised controlled trial. The exercise helps achieve two objectives. First, it allows us to compute the equilibrium that these communities are at and evaluate the scope for the success of belief correction in our setting both in the short and long run. Second, it allows us to compare the effects of our belief-shifting intervention with alternative policy instruments that do not target beliefs.

8.1 Assumptions

We make a few simplifying assumptions before proceeding. First, we assume that the cost of engagement takes the form $c(d, 1 - E[e_{j:j\in N(i)}]) = \theta * d * E[a_{j:j\in N(i)}]$ where $a_i = 1 - e_i$ is the choice to disengage. The benefit of engagement b is drawn from a log-logistic distribution H. We assume that the probability that a randomly chosen peer of degree d disengages under equilibrium disengagement, a, follows a log-logistic distribution, i.e. $H(c(d,a)) = (1 + (\theta da/\alpha)^{-\kappa})^{-1}$ where we set $\alpha = 1$. We show below that this functional form allows for complete flexibility in the number and type of possible equilibria.[35] Figure H.1 plots the log-logistic function for different values of θ and κ showing that we do not impose a specific shape. Assuming a log-logistic distribution implies that $H(0) = 0$, so agents always engage if everyone else in their network is engaging thereby implying that $b_i > 0 \; \forall \; i \in N$. In a robustness exercise, we will also show that the estimation results are similar even if we assume a logistic distribution.

Based on the evidence of a positive relationship between optimistic beliefs about peers' willingness to engage and own engagement detected from the experiment results, we restrict the strategic complementarity parameter θ to be strictly positive. Further, we restrict κ to be non-negative because the log-logistic distribution only allows a non-negative domain.

We use the estimation exercise to recover estimates of the two parameters that affect every agent's predicted choice and the magnitude, type, and number of equilibria: θ (the extent of complementarities in disengagement) and κ (the parameter which affects the shape of the distribution H of the benefit of engagement).

[35] The log-logistic distribution has fatter tails compared to the log-normal distribution. We use this instead of a log-normal distribution since its cumulative distribution function can be written in closed form and this helps with accuracy and speed in our simulations. This distribution is also a more suitable choice than the logistic distribution since it has a positive domain and $H(0) = 0$. The parameters of the logistic distribution only allow $H(0) = 0$ in the limiting case.

8.2 Estimation Strategy

We use simulated method of moments and proceed in the following steps. First, we empirically compute the degree distribution $P(d)$ using the reported degrees in the overall network in our baseline sample.[36] We construct the degree distribution for peers $\widetilde{P}(d)$ as follows:

$$\widetilde{P}(d) = P(d) * (d/E(d)).$$

where we compute the sample average degree as an estimate of the expected degree $E(d)$.[37]

The estimation algorithm begins by choosing random starting values for θ and κ. For this initial choice, we numerically solve for the Bayes-Nash equilibria that would arise as per Equation 3. Any equilibrium a^* must solve the following equation:

$$a^*(\theta, \kappa) = \sum \widetilde{P}(d) H(c(d, a^*(\theta, \kappa)).$$

All solutions to this equation are potential equilibria that the individuals in the control group could use to decide whether to engage or not. To choose an equilibrium from this set, we first compute the probability that each individual in the control group disengages depending on each of the computed equilibria, their degree, and the proposed value of θ and κ. This is used to compute the likelihood of each equilibrium a^* for the control group:

$$L(a^*) = \sum_i \{1\{a_i = 1\} ln[H(c(d_i, a^*(\theta, \kappa)] + 1\{a_i = 0\} ln[1 - H(c(d_i, a^*(\theta, \kappa)]\}.$$

We pick the equilibrium a_c^* that maximizes the likelihood for the control group. This equilibrium selection method is discussed in Bisin et al. (2011) and De Paula (2013). We use a_c^* and d_i to compute the expected disengagement for each agent i in the control group. This is used to construct the first moment of interest: expected disengagement in the control group.

Next, we compute the probability that each individual in the treatment group disengages. This probability depends on the exogenously delivered belief that $a_t^* = 0.4$, the individual's degree d_i, and the proposed value of θ and κ. Note that in this case, we do not have to solve for the equilibrium and pick the one that maximizes the likelihood, as we know that the individuals in the treatment group were provided with the expected disengagement as a part of the treatment. We use $a_t^* = 0.4$ and d_i to compute the expected disengagement for each agent i in the treated group. This is the second moment of interest.

The algorithm updates θ and κ, repeats the steps above, and iterates to find θ^* and κ^* that minimize the squared deviation between the moments estimated from the model and the moments computed from the data, with an equal weight placed on both. We use the *pattern search* solver with multiple starting points to run these iterations and find the optimal θ and κ.

[36] Recall that the overall network contains a link between two agents if they interact in any capacity. Degree in this network captures how connected a person is in general, and how much they might worry about violating the social network.

[37] We do not compute degree distributions for each NGO center due to small samples in each center. There is no reason to believe that the networks should be systematically different across centers. As a result, we pool the sample to compute the degree distribution.

It is important to note that small changes in θ and κ can change the equilibrium configuration, the chosen equilibrium for the control group, and the actions chosen by individuals. Pattern search with multiple randomly chosen starting points allows us to find the global optimum for such objective functions (Audet and Dennis Jr, 2002).[38] We perform this exercise by using the willingness to participate in the listening service as the measure of engagement. As the intervention delivered precise information about average engagement with mental health, individuals are more likely to interpret the revealed average engagement with mental health in terms of expected participation in the listening service (as opposed to expected participation in the savings group, which is another measure of engagement we have). Nonetheless, we also show robustness to alternative measures of engagement discussed in more detail in section 8.3.1.

8.2.1 Confidence Intervals

We use Quasi MCMC methods outlined in Chen et al. (2018) to compute confidence intervals for our parameter estimates. The process is discussed in more detail in the appendix in Section H.1. Briefly, we perturb θ and κ around the computed θ^* and κ^*, re-compute a weighted loss function, choose the perturbed value if the loss function reduces and choose it with some probability α even if the loss function increases. Crucially, the magnitude of α depends on how far away the current loss function is from that in the previous iteration. This process continues until the estimates converge in distribution. We then sample various θ and κ from the converged distribution to construct the confidence interval.

8.3 Estimation Results

The first row of Table 1 shows the minimized value of the objective function and compares the predicted moments (computed at θ^*, κ^*) with the actual moments in the data. The table indicates that the estimated parameters fit the data very well. We also find that the model predicts well those moments that were not targeted, namely, the standard deviation and skewness of engagement.

Figure 5 plots the equilibrium equation $a = \sum_d \widetilde{P}(d)(H(c(d,a);\theta,\kappa))$ evaluated at the estimated θ^* and κ^*. Figure H.3 plots the estimated curve with the confidence set and shows that the confidence set is tight and the shape of the best response curve is similar for all the θ and κ values that lie in this set. We make the following observations. First, we find that based on the estimated curve, there exist two stable equilibria at $a^* = 0$ and $a^* = 0.89$ and one unstable equilibrium at $a^* = 0.23$. Second, we find that the equilibrium that maximizes the likelihood is the high, stable level of disengagement at $a^* = 0.89$. This is expected in our setting as we measured low levels of dialogue around these issues in the baseline survey. These two

[38] The algorithm searches for optima by considering 'mesh points' that are a specified distance ('mesh size') away from the user-supplied starting point. If a mesh point results in a lower value of the objective function, the mesh size is doubled and if the original point is better than all considered mesh points, the distance is halved. The solver converges when the mesh size is very small. In addition to using multiple starting points, we also start with a mesh size five times higher than the default to ensure that the solver doesn't search for local optima close to the starting values. The algorithm is repeated for several randomly chosen starting values to ensure that we find the lowest possible value of the objective function.

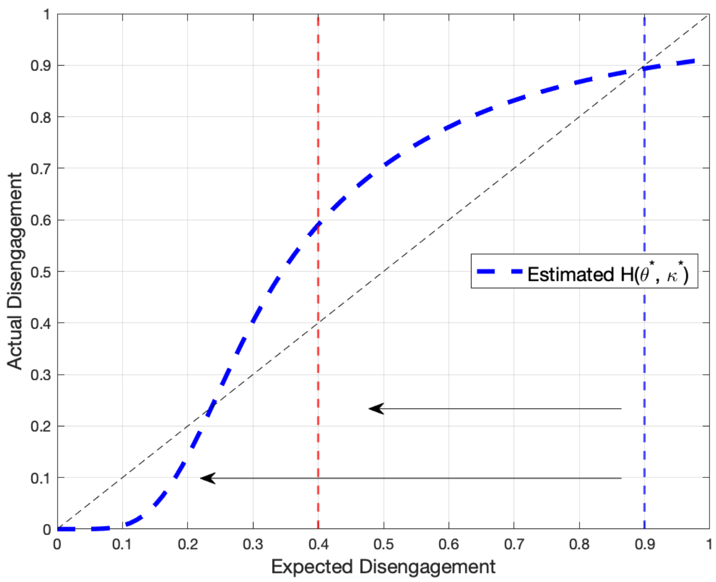
Table 1: Structural Estimation Model Fit

	Targetted Moments				Untargetted Moments	
Objective Function	Predicted Mean (Treated)	Actual Mean (Treated)	Predicted Mean (Control)	Actual Mean (Control)	Standard Deviation (Squared Gap)	Skewness (Squared Gap)
0.0005	0.309	0.299	0.440	0.460	0.003	2.6e-05

Notes: This table shows the model fit for the estimated values of θ^* and κ^* for the main measure of engagement. The objective function is computed at the estimated parameters and is equal to the sum of the squared gap between the mean of the treatment and control groups predicted by the model and in the data respectively.

findings suggest that under the assumptions of our model and estimation strategy, there exists a threshold beyond which beliefs have to be shifted to lead to a long-run change.

Figure 5: Structural Estimation: Actual and Expected Disengagement



Notes: The figure plots actual disengagement a and expected disengagement $\sum_d \widetilde{P}(d)(H(c(d,a)))$ at the estimated θ^* and κ^*. The blue dotted line shows the equilibrium that maximises the likelihood while the red dotted line shows the predicted short-run effect of the intervention.

The structural estimation exercise allows us to compare the belief intervention we delivered with one that would cause a long-run change in beliefs. We can see that shifting perceived engagement to 0.4, as we do in our experiment, is not likely to change long-run beliefs. This is because the community is stuck at a high stable level of disengagement and the delivered beliefs do not cross the threshold at $a^* = 0.23$. In other words, shifting the initial equilibrium and initiating a long-run change in engagement would only be possible via a belief-correction intervention if the belief about the proportion of others willing to engage is shifted to a magnitude greater than or equal to 80%. This is a large increase and it may not always be possible to credibly implement such a change. Consistent with this prediction, we find that two years later, the average beliefs of those exposed to the information, while still significantly more optimistic than those who were not exposed, were less optimistic than the information that was delivered by us.

We will now compare this to counterfactual interventions that instead increase the benefits of engagement or reduce the individual's concerns about violating the social norm. Crucially, both of these counterfactuals do not target beliefs.

8.3.1 Robustness to other measures of engagement and distributional assumptions

We show robustness by using two alternative measures of engagement: (1) Willingness to participate in savings groups and (2) Community engagement i.e., the average of community engagement measures as described in the reduced form analysis. Note that the information delivered to the treated individual is the proportion of individuals in the community who are willing to engage about mental health-related concerns. In this sense, both these outcomes can still be interpreted as proxies for engagement but using willingness to participate in a listening service as the main measure is ideal. We find that there exists a similar threshold in beliefs for these alternative outcomes as well and the shape of the curve is also similar i.e., the estimated curves are S-shaped with three equilibria. Under the assumptions of the model, this suggests that belief-shifting interventions, when strong enough to cross the threshold, can lead to a shift in long-run beliefs in almost all cases. Table H.1 shows the parameter estimates and value of the objective function for the various outcomes.

We also test whether the shape of the best response function is different if benefits are assumed to be distributed as per a logistic distribution instead of a log-logistic. We re-estimate the model under this assumption using the same procedure as before. However, we now estimate three parameters θ, κ, and σ where θ is defined as before and κ and σ are the mean and standard deviation of the logistic distribution respectively. We use an additional moment condition i.e., the standard deviation of engagement in the data to identify σ. The results are shown in Figure H.4 in the appendix. The shape of the curve and the resultant equilibria are remarkably similar to that estimated under the log-logistic distribution. We now proceed with the counterfactuals.

8.3.2 Counterfactual 1: Shifting the perceived benefits of interactions.

We first consider a shift in perceived benefits of engagement which could happen via information or awareness sessions informing individuals about the benefits of network interactions or a cash transfer that incentivised individuals to interact. The top panel of Figure 6 illustrates the intuition behind the first counterfactual that we run. The figure plots the estimated expected disengagement $\sum_d \widetilde{P}(d)(H(c(d,a)))$ in the listening service, evaluated at the estimated θ^* and κ^*. The dotted line in orange shows how much the curve would have to be shifted in order to lead to a long-run effect that is at least as large as the short-run effect of the belief correction intervention.

Shifting the curve is equivalent to increasing the benefits of engagement by a fixed constant $b > 0$. As we increase the mean of the benefit distribution by $b > 0$, we shift the curve downwards until the benefit increment $b = b^* > 0$ where the only equilibria is that of zero disengagement. When the benefit increment is slightly lower than b^*, there exists a stable level of disengagement that is lower than the current level predicted by the model but still higher than the shift caused by the intervention. In other words, we must keep increasing the value of b until we reach a point such that equilibrium disengagement is less than or equal to 0.4 i.e., the short-run effect of the RCT. This tells us how large the benefit increment has to be to lead to a long-run effect at least as large as the short-run effect of the RCT.

We find that this benefit increment must be as high as 48% of the mean of the benefit distribution H to have a persistent long-run impact of a magnitude that is at least as large as the short-run effect of the RCT. It is also important to highlight that a push beyond a 48% increment in benefits would lead to a reduction in disengagement to zero.

8.3.3 Counterfactual 2: Reducing how much agents care about the social norm.

Instead of shifting benefits, we could also think of policy instruments that reduce θ i.e., how much the violation of social norms contributes to individual costs of engagement. We can think of reducing θ as making individuals care less about violating the social norm. This can take the form of different psychological interventions geared at reducing the individual's costs of violating the social norm by helping them deal with concerns such as gossip or insensitivity. Alternatively, setting up a formal job referral service via the NGO may make individuals care less about signalling their type to others while approaching them for financial assistance.

As before, the question of interest is: how much do we have to lower θ to achieve a persistent, long-run reduction in disengagement that is comparable to the short-run effect achieved by the belief correction intervention? This is shown in the bottom panel of Figure 6. We keep reducing the value of θ until we reach a point such that the equilibrium disengagement is less than or equal to 0.4 i.e., the effect of this intervention is at least as large as the short-run effect of belief correction. We find that θ has to be lower by 33% of the estimated θ^* to be able to have this effect. This is also a large shift and would require a strong and potentially costly intervention to be able to achieve the required result.

Summary of Counterfactual Analysis: Counterfactual analysis shows how demanding it can be to initiate long-run change using alternative policies that shift the benefits or costs of interacting. Our empirical results show that in the absence of such a big push, a belief-shifting intervention is a cheaper alternative that could be used to finance policies which could help one. Further, the short-run effect can be long-lasting since we do not know how many periods of updating it will take for individuals to move back to the stable equilibrium of high disengagement. As discussed in the section on long-run effects, we have suggestive evidence that exposure to the treatment still has significant effects on dialogue and consumption outcomes two years later.

9 Conclusion

Social networks may not function as effective social safety nets if individuals do not demand these interactions in the first place. This is especially critical in environments where networks must function as social safety nets due to high-income volatility, financial distress, and lack of any formal assistance. In one such setting, we show that inaccurate beliefs about others' willingness to engage in dialogue around financial and mental health-related concerns can reduce useful social interactions. Unlike the existing literature on social networks that focuses on contexts in which caste, ethnic, or religious affiliations promote social interactions, our results generate insights into contexts (such as urban settlements) that may lack such cohesive institutions.

Using a randomized controlled trial, we causally show that correcting misperceptions about peers affects the demand for network interactions. Belief correction leads to an increase in the demand for network-based assistance and community engagement. Additional experiments and survey evidence help us understand that the treatment effects arise due to a reduction in the perceived costs of violating the social norm. We also rule out concerns about social desirability. The target group of our intervention i.e., those who are pessimistic about the norm are precisely those who are significantly more likely to think that reputation and interaction-related costs matter while forming advice-taking links. While interventions that address such concerns can be difficult to implement, we show that we can cover a lot of ground by correcting misperceptions.

Estimation of a network diffusion model suggests that our intervention is not likely to have long-run effects. Using this exercise, we compare interventions that target beliefs and other policy instruments such as setting up information sessions, savings groups, or job referral platforms. Counterfactual analysis shows how demanding it would be to initiate long-run change using these alternative policies that shift the benefits or costs of interacting. However, belief-correction interventions can be used to increase individual demand for network interactions in the short run and generate funds for costlier interventions. This is especially an advantage in underfunded environments like ours that suffer from policy neglect.

There are three natural extensions to this work. First, while we focus on shifting beliefs about the social norm, our model suggests that inaccurate beliefs exist in equilibrium because individuals inaccurately estimate the social norm from a biased sample composed of their network peers. It would be interesting to consider interventions that can help avoid this error in inference and belief updating. Second, while the focus in this paper has primarily been on addressing demand-side constraints, future work can study whether individuals have accurate beliefs about supply-side features such as the benefits of engaging. Finally, the relative effectiveness of providing information to correct one misperception versus another and how these affect long-run outcomes is also an exciting avenue for future work.

Tables and Figures

A Summary Statistics

Table 2: Summary Statistics

	Mean	SD
A. Demographics		
Age	33.75	(9.371)
Female	0.352	(0.478)
Monthly HH Income ($<$ Rs 2,500)	0.157	(0.364)
Monthly HH Income (Rs 2,500-5,000)	0.189	(0.392)
Monthly HH Income (Rs 5,000-10,000)	0.444	(0.498)
Monthly HH Income ($>$ Rs 10,000)	0.210	(0.408)
Employed (Yes/No)	0.668	(0.472)
B. Well-Being		
Stress (Index; Scale 1-5)	3.076	(0.864)
Unable to manage difficulties (Often/Very Often)	0.503	(0.501)
Life Satisfaction (Scale 1-4)	2.851	(0.951)
Happiness (Scale 1-4)	2.479	(0.981)
C. Networks		
Number of network connections (Advice)	2.787	(3.378)
Number of network connections (Overall)	3.954	(3.945)
Observations	352	

Notes: The above table shows the summary statistics (mean and standard deviation) for various demographic characteristics of interest from the baseline sample.

Table 3: Summary Statistics

	Mean	SD
A. Willingness to Engage		
Willingness to Talk (Financial Concerns)	0.707	(0.456)
Willingness to Talk (Mental Health)	0.631	(0.483)
Willingness to Talk (Physical Health)	0.618	(0.487)
Stigma (Physical Health)	0.555	(0.498)
Stigma (Mental Health)	0.503	(0.501)
B. Beliefs		
Beliefs (Mental Health)	4.985	(3.012)
Beliefs (Physical Health)	5.126	(3.029)
Beliefs (Financial Concerns)	5.960	(3.196)
Beliefs -Stigma (Mental Health)	4.802	(3.312)
Beliefs -Stigma (Physical Health)	4.812	(3.253)
Beliefs -Stress (Mental Health)	5.641	(3.313)
C. Frequency of Conversations		
Dialogue (Physical Health; 1-5)	2.180	(1.300)
Physical Health talk (Never)	0.430	(0.496)
Physical Health talk (Rarely)	0.223	(0.417)
Physical Health talk (Sometimes)	0.162	(0.369)
Physical Health talk (Often)	0.110	(0.313)
Physical Health talk (Very Often)	0.0762	(0.266)
Dialogue (Mental Health; 1-5)	2.511	(1.304)
Mental Health talk (Never)	0.292	(0.456)
Mental Health talk (Rarely)	0.252	(0.435)
Mental Health talk (Sometimes)	0.191	(0.394)
Mental Health talk (Often)	0.182	(0.386)
Mental Health talk (Very Often)	0.0831	(0.276)
Dialogue (Financial Concerns; 1-5)	2.847	(1.470)
Financial Concerns talk (Never)	0.256	(0.437)
Financial Concerns talk (Rarely)	0.200	(0.401)
Financial Concerns talk (Sometimes)	0.181	(0.386)
Financial Concerns talk (Often)	0.166	(0.372)
Financial Concerns talk (Very Often)	0.197	(0.398)
Observations	334	

Notes: The above table shows summary statistics (mean and standard deviation) for additional baseline variables such as willingness to engage, dialogue intensity, stigma, and beliefs about peers. All variables in Panel A are binary. We measure stigma by asking if individuals think people should "stay away" from those with specific health concerns. Beliefs about peers are measured in terms of 0-10 individuals in the community. Dialogue intensity is measured on a scale from 1-5 where 1 is 'Never' and 5 is 'Very Often'.

B Endline Results

Table 4: Treatment Effect on Endline Outcomes

Panel A: Treatment Effects on Community Engagement

Variables	Savings Group	Listening Volunteer	Listening Contribution	Contribution (in Rupees)	Community Engagement
Treatment	0.147**	0.161**	0.122**	6.577**	0.138***
	(0.0711)	(0.0730)	(0.0612)	(2.959)	(0.0516)
Bootstrap p-value	0.0831	0.0511	0.0561	0.0931	0.0340
q-values	0.046	0.046	0.046	0.046	0.046
Constant	0.667***	0.540***	0.738***	22.72***	0.662***
	(0.0548)	(0.0537)	(0.0483)	(2.090)	(0.0393)
Observations	150	174	170	163	150
R-squared	0.028	0.027	0.023	0.030	0.046

Panel B: Treatment Effects on Own Health Outcomes

	Speaking to the Doctor (MH)	Depression Scoring (Immediate)	Depression Scoring	Listening to Helpline Numbers
Treatment	0.0995*	-0.272***	-0.136*	-0.197**
	(0.0600)	(0.0809)	(0.0747)	(0.0844)
Bootstrap p-value	0.232	0.0300	0.469	0.0511
q-values	0.164	0.008	0.224	
Constant	0.118***	0.606***	0.789***	0.591***
	(0.0373)	(0.0584)	(0.0488)	(0.0610)
Observations	154	143	143	137
R-squared	0.018	0.074	0.023	0.039

Panel C: Treatment Effects on Other Endline Outcomes

VARIABLES	Memory (Numbers Remembered)	Physical Health Dialogue (with Family)
Treatment	-0.271	0.208***
	(0.205)	(0.0711)
Bootstrap p-value	0.387	0.0210
Constant	1.053***	0.184***
	(0.151)	(0.0448)
Observations	153	155
R-squared	0.011	0.053

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Wild bootstrap *p*-values reported (reps=999) using the method outlined in Cameron et al. (2008) where we treat the NGO centre as the cluster unit.

C Longer Term Effects

Table 5: Effects on Network Interactions after 2 weeks

	Financial Concerns			Mental Health Concerns		
	Reached Out (Yes/No)	Number of Friends Reached Out to	Friend Reached Out (Yes/No)	Reached Out (Yes/No)	Number of Friends Reached Out to	Friend Reached Out (Yes/No)
Treatment	0.188**	1.212***	0.146	0.113	0.911**	0.0414
	(0.0930)	(0.427)	(0.0914)	(0.0863)	(0.386)	(0.0833)
q-values	0.052	0.034	0.054	0.034	0.008	0.139
Constant	0.321***	0.877***	0.283***	0.226***	0.500***	0.226***
	(0.0647)	(0.206)	(0.0625)	(0.0580)	(0.158)	(0.0580)
Observations	110	109	109	109	109	109
R-squared	0.036	0.068	0.023	0.016	0.048	0.002

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. The *q*-values (Benjamini et al., 2006) reported in each table treat all the outcomes in this table as multiple hypotheses being tested together. Standard errors are robust.

Table 6: Effects on Beliefs and Willingness to Engage after 2 years

Panel A: Effects on Beliefs about Peers after 2 years

VARIABLES	Beliefs (MH)	Beliefs (MH)	Beliefs (MH)	Beliefs (FC)	Beliefs (FC)	Beliefs (FC)
Previous Participant	0.193			0.00819		
	(0.299)			(0.261)		
Heard Information		0.899***			0.751***	
		(0.324)			(0.259)	
Heard Information and was previous participant			1.442***			0.754**
			(0.387)			(0.307)
Constant	3.607***	3.560***	3.558***	3.658***	3.600***	3.621***
	(0.0869)	(0.0866)	(0.0850)	(0.0930)	(0.0915)	(0.0898)
Observations	467	467	467	467	467	467

Panel B: Effects on Willingness to Engage after 2 years

VARIABLES	Willingness to Engage (MH)	Willingness to Engage (MH)	Willingness to Engage (MH)	Willingness to Engage (FC)	Willingness to Engage (FC)	Willingness to Engage (FC)
Previous Participant	0.0185			0.0611*		
	(0.0448)			(0.0362)		
Heard Information		0.0561			0.0715*	
		(0.0478)			(0.0405)	
Heard Information and was previous participant			0.140***			0.126***
			(0.0165)			(0.0158)
Constant	0.865***	0.863***	0.860***	0.872***	0.874***	0.874***
	(0.0169)	(0.0166)	(0.0165)	(0.0165)	(0.0160)	(0.0158)
Observations	467	467	467	467	467	467

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. We implement post double selection lasso (Belloni et al., 2014) with robust standard errors accounting for income, age, gender, and number of network connections. Previous participant is a binary variable equal to 1 if the respondent in 2023 reports being contacted for the previous study in 2020-21. Heard information is equal to 1 if they report having heard anything about their community's willingness to engage from their peers and 0 otherwise. "MH" refers to mental health concerns and "FC" refers to financial concerns.

Table 7: Effects on Network Interactions and Consumption Outcomes after 2 years

Panel A: Effects on Network Interactions after 2 years

VARIABLES	(1) MH Talk (Above Median)	(2) MH Talk (Above Median)	(3) MH Talk (Above Median)	(4) FC Talk (Above Median)	(5) FC Talk (Above Median)	(6) FC Talk (Above Median)
Previous Participant	0.123*			0.0814		
	(0.0689)			(0.0685)		
Heard Information		0.183**			0.263***	
		(0.0841)			(0.0830)	
Heard Information and was previous participant			0.210**			0.171
			(0.102)			(0.104)
Constant	0.410***	0.412***	0.415***	0.369***	0.358***	0.370***
	(0.0244)	(0.0237)	(0.0234)	(0.0239)	(0.0231)	(0.0229)
Observations	467	467	467	467	467	467

Panel B: Effects on Consumption Outcomes after 2 years

VARIABLES	Consumption Crisis (Yes/No)	Consumption Crisis (Yes/No)	Consumption Crisis (Yes/No)	Consumption varies a lot	Consumption varies a lot	Consumption varies a lot
Previous Participant	0.0124			-0.0649**		
	(0.0690)			(0.0275)		
Heard Information		-0.157*			-0.0683**	
		(0.0833)			(0.0302)	
Heard Information and was previous participant			-0.155			-0.0948***
			(0.102)			(0.0139)
Constant	0.521***	0.535***	0.530***	0.0983***	0.0953***	0.0948***
	(0.0248)	(0.0241)	(0.0237)	(0.0148)	(0.0142)	(0.0139)
Observations	467	467	467	467	467	467

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. We implement post double selection lasso (Belloni et al., 2014) with robust standard errors accounting for income, age, gender, and number of network connections. *Previous participant* is a binary variable equal to 1 if the respondent in 2023 reports being contacted for the previous study in 2020-21. *Heard information* is equal to 1 if they report having heard anything about their community's willingness to engage from their peers and 0 otherwise. "MH" refers to mental health, and "FC" refers to financial concerns.

D Mechanisms

Table 8: Hypothetical Network Prediction Experiment (OLS)

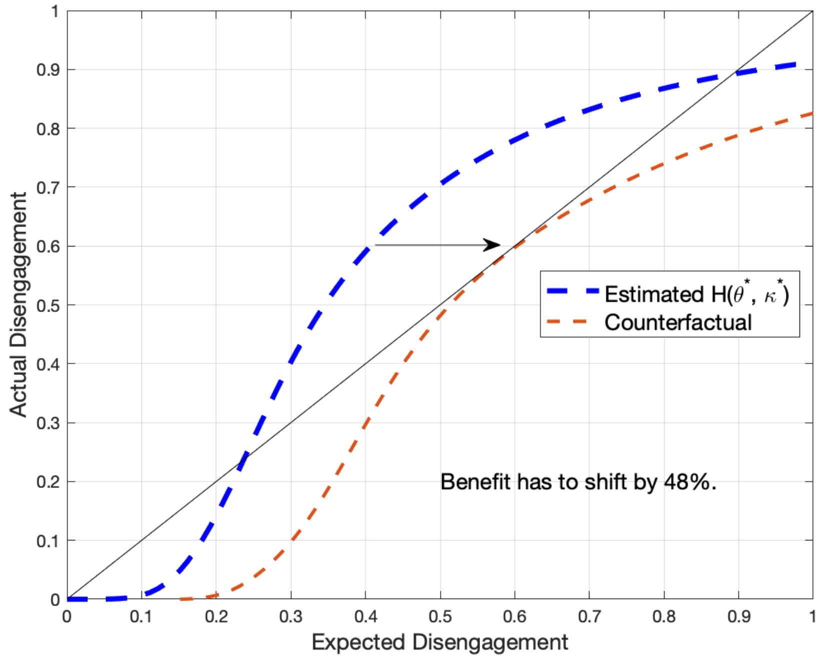
VARIABLES	(1) B takes advice (MH)	(2) B takes advice (MH)	(3) B takes advice (FC)	(4) B takes advice (FC)
Advisor has job contacts	0.00851	0.00687	0.0249	0.0240
	(0.0174)	(0.0173)	(0.0177)	(0.0176)
Advisor is network central	0.0100	0.00740	0.0222	0.0233
	(0.0174)	(0.0173)	(0.0177)	(0.0176)
Advisor has attended training	-0.0181	-0.0182	-0.00929	-0.00972
	(0.0174)	(0.0173)	(0.0177)	(0.0176)
Advisor has job contacts X Own Beliefs		0.0226		0.0158
		(0.0183)		(0.0173)
Advisor is network central X Own Beliefs		0.0347*		0.0354**
		(0.0184)		(0.0173)
Advisor has attended training X Own Beliefs		0.0312*		0.0482***
		(0.0183)		(0.0173)
Beliefs (MH)		-0.00548		
		(0.0184)		
Beliefs (FC)				-0.0635***
				(0.0176)
Constant	0.766***	0.767***	0.737***	0.737***
	(0.0173)	(0.0172)	(0.0178)	(0.0176)
Observations	2,372	2,372	2,370	2,370
R-squared	0.001	0.013	0.002	0.008

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. This table presents the results the link prediction experiment where individuals are asked to predict if a link exists between a hypothetical advisor (A) and a randomly chosen person (B). "MH" refers to mental health and "FC" refers to financial concerns. "Beliefs" refers to the respondent's prediction (standardized) about how many community members would be willing to engage around mental or financial concerns. The interaction terms interact the advisor characteristics with Beliefs (MH) in Column 2 and Beliefs (FC) in Column 4. We report robust standard errors.

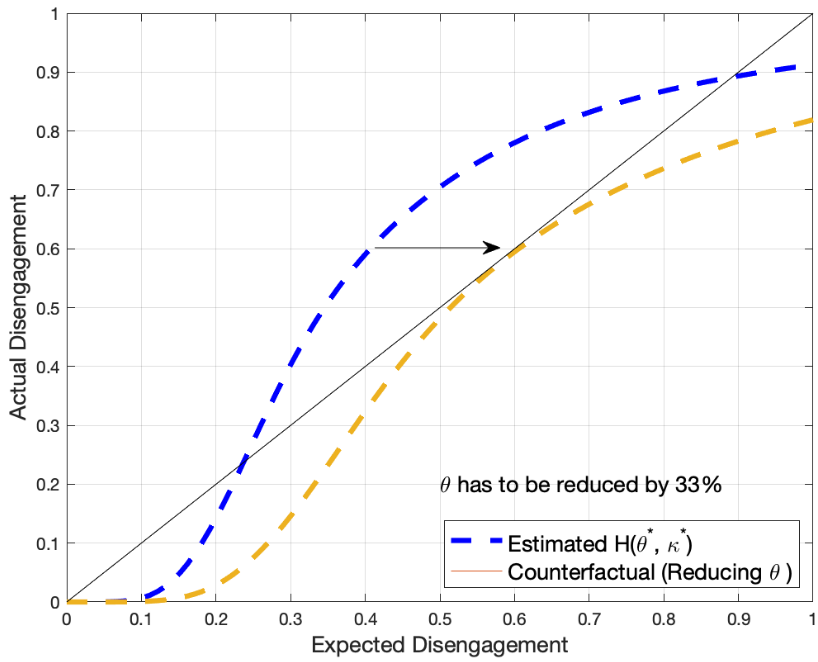
E Structural Estimation: Counterfactual Results

Figure 6: Counterfactual Interventions

(A) Increasing the Benefit of Engagement.



(B) Reducing how much agents care about the Social Norm.



Notes: The figures plot actual disengagement a and expected disengagement $\sum_d \widetilde{P}(d)(H(c(d,a)))$ at the estimated θ^* and κ^* in blue. In addition to this, the figures also plot the proposed counterfactuals that increase benefits (top) and reduce θ (bottom) in the dotted line in orange/yellow.

References

Ambrus, A. and Elliott, M. (2021), 'Investments in social ties, risk sharing, and inequality', *The Review of Economic Studies* **88**(4), 1624–1664.

Anderson, M. L. (2008), 'Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects', *Journal of the American statistical Association* **103**(484), 1481–1495.

Angelucci, M., De Giorgi, G. and Rasul, I. (2018), 'Consumption and investment in resource pooling family networks', *The Economic Journal* **128**(615), 2613–2651.

Audet, C. and Dennis Jr, J. E. (2002), 'Analysis of generalized pattern searches', *SIAM Journal on optimization* **13**(3), 889–903.

Banerjee, A., Breza, E., Chandrasekhar, A. G. and Golub, B. (2024), 'When less is more: Experimental evidence on information delivery during india's demonetisation', *Review of Economic Studies* **91**(4), 1884–1922.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O. (2013), 'The diffusion of microfinance', *Science* **341**(6144), 1236498.

Beaman, L., BenYishay, A., Magruder, J. and Mobarak, A. M. (2021), 'Can network theory-based targeting increase technology adoption?', *American Economic Review* **111**(6), 1918–1943.

Belloni, A., Chernozhukov, V. and Hansen, C. (2014), 'Inference on treatment effects after selection among high-dimensional controls', *Review of Economic Studies* **81**(2), 608–650.

Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006), 'Adaptive linear step-up procedures that control the false discovery rate', *Biometrika* **93**(3), 491–507.

Bisin, A., Moro, A. and Topa, G. (2011), The empirical content of models with multiple equilibria in economies with social interactions, Technical report, National Bureau of Economic Research.

Breza, E. (2016), *Field experiments, social networks, and development*, Oxford University Press Oxford, UK.

Breza, E., Chandrasekhar, A., Golub, B. and Parvathaneni, A. (2019), 'Networks in economic development', *Oxford Review of Economic Policy* **35**(4), 678–721.

Bursztyn, L., González, A. L. and Yanagizawa-Drott, D. (2020), 'Misperceived social norms: Women working outside the home in saudi arabia', *American economic review* **110**(10), 2997–3029.

Bursztyn, L. and Yang, D. Y. (2022), 'Misperceptions about others', *Annual Review of Economics* **14**, 425–452.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008), 'Bootstrap-based improvements for inference with clustered errors', *The review of economics and statistics* **90**(3), 414–427.

Chandrasekhar, A. G., Duflo, E., Kremer, M., Pugliese, J. F., Robinson, J. and Schilbach, F. (2022), Blue spoons: Sparking communication about appropriate technology use, Technical report, National Bureau of Economic Research.

Chandrasekhar, A. G., Golub, B. and Yang, H. (2019), 'Signaling, shame, and silence in social learning', *Available at SSRN: https://ssrn.com/abstract=3261632 or http://dx.doi.org/10.2139/ssrn.3261632* .

Chaturvedi, B., Varma, A., Mukherjee, C., Chaturvedi, R. and Khan, I. (2018), *Wastepickers: Delhi's Forgotten Environmentalists?*, Chintan Environmental Research and Action Group.

Chen, X., Christensen, T. M. and Tamer, E. (2018), 'Monte carlo confidence sets for identified sets', *Econometrica* **86**(6), 1965–2018.

De Paula, A. (2013), 'Econometric analysis of games with multiple equilibria', *Annu. Rev. Econ.* **5**(1), 107–131.

De Weerdt, J. and Dercon, S. (2006), 'Risk-sharing networks and insurance against illness', *Journal of development Economics* **81**(2), 337–356.

Delavande, A. (2023), Expectations in development economics, *in* 'Handbook of Economic Expectations', Elsevier, pp. 261–291.

Delavande, A., Giné, X. and McKenzie, D. (2011), 'Measuring subjective expectations in developing countries: A critical review and new evidence', *Journal of development economics* **94**(2), 151–163.

Fafchamps, M. and Gubert, F. (2007), 'The formation of risk sharing networks', *Journal of development Economics* **83**(2), 326–350.

Fafchamps, M. and Lund, S. (2003), 'Risk-sharing networks in rural philippines', *Journal of development Economics* **71**(2), 261–287.

Feld, S. L. (1991), 'Why your friends have more friends than you do', *American Journal of Sociology* **96**(6), 1464–1477.

Gilles, R. P. (2021), 'Building social networks under consent: A survey', *Game Theory and Networks: New Perspectives and Directions* pp. 203–258.

Gilles, R. P., Chakrabarti, S. and Sarangi, S. (2012), 'Nash equilibria of network formation games under consent', *Mathematical Social Sciences* **64**(2), 159–165.

Gilles, R. P. and Sarangi, S. (2010), 'Network formation under mutual consent and costly communication', *Mathematical social sciences* **60**(3), 181–185.

Haaland, I., Roth, C. and Wohlfart, J. (2020), 'Designing information provision experiments'.

Ivaschenko, O., Rodriguez Alas, C. P., Novikova, M., Romero Robayo, C., Bowen, T. V. and Zhu, L. (2018), *The state of social safety nets 2018*, number 124300, The World Bank.

Jackson, M. O. (2019), 'The friendship paradox and systematic biases in perceptions and social norms', *Journal of Political Economy* **127**(2), 777–818.

Jackson, M. O., Rodriguez-Barraquer, T. and Tan, X. (2012), 'Social capital and social quilts: Network patterns of favor exchange', *The American Economic Review* **102**(5), 1857–1897.

Jackson, M. O. and Watts, A. (2001), 'The existence of pairwise stable networks', *Seoul journal of economics* **14**(3), 299–321.

Jackson, M. O. and Yariv, L. (2007), 'Diffusion of behavior and equilibrium properties in network games', *American Economic Review* **97**(2), 92–98.

Jackson, M. and Wolinsky, A. (1996), 'A strategic model of economic and social networks', *Journal of Economic Theory* **71**(1), 44–74.

Karlan, D., Mobius, M., Rosenblat, T. and Szeidl, A. (2009), 'Trust and social collateral', *The Quarterly Journal of Economics* **124**(3), 1307–1361.

Ligon, E., Thomas, J. P. and Worrall, T. (2002), 'Informal insurance arrangements with limited commitment: Theory and evidence from village economies', *The Review of Economic Studies* **69**(1), 209–244.

Marx, B., Stoker, T. and Suri, T. (2013), 'The economics of slums in the developing world', *Journal of Economic perspectives* **27**(4), 187–210.

Medina, M. (2008), 'The informal recycling sector in developing countries: organizing waste pickers to enhance their impact'.

Möbius, M. M. and Rozenblat, T. S. (2016), 'Informal transfers in networks'.

Morten, M. (2019), 'Temporary migration and endogenous risk sharing in village india', *Journal of Political Economy* **127**(1), 1–46.

Munshi, K. (2011), 'Strength in numbers: Networks as a solution to occupational traps', *The Review of Economic Studies* **78**(3), 1069–1101.

Munshi, K. and Rosenzweig, M. (2009), Why is mobility in india so low? social insurance, inequality, and growth, Technical report, National Bureau of Economic Research.

Munshi, K. and Rosenzweig, M. (2016), 'Networks and misallocation: Insurance, migration, and the rural-urban wage gap', *The American Economic Review* **106**(1), 46–98.

Myerson, R. B. (1991), *Game theory: analysis of conflict*, Harvard university press.

Perkins, H. W., Haines, M. P. and Rice, R. (2005), 'Misperceiving the college drinking norm and related problems: a nationwide study of exposure to prevention information, perceived norms and student alcohol misuse.', *Journal of studies on alcohol* **66**(4), 470–478.

Perkins, H. W., Meilman, P. W., Leichliter, J. S., Cashin, J. R. and Presley, C. A. (1999), 'Misperceptions of the norms for the frequency of alcohol and other drug use on college campuses', *Journal of American College Health* **47**(6), 253–258.

UN-Habitat (2013), *State of the world's cities 2012/2013: Prosperity of cities*, Routledge.

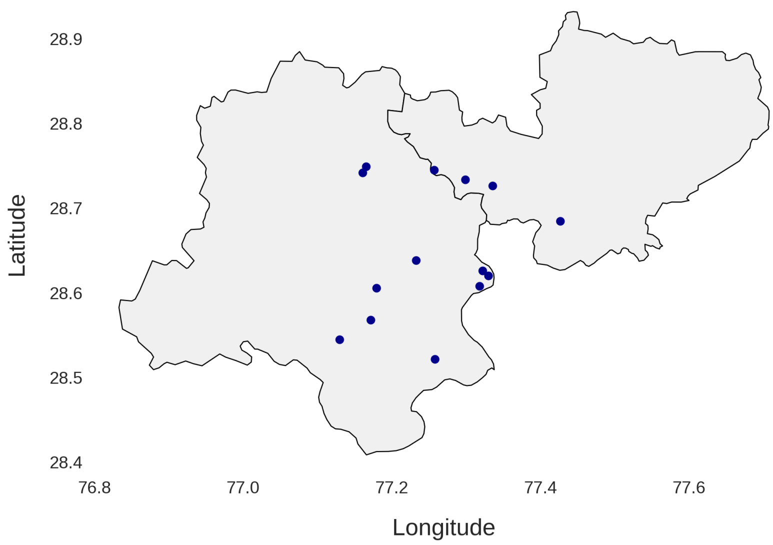
Van de Rijt, A. and Buskens, V. (2005), 'A nash equilibrium refinement for myerson's network formation game', *ISCORE paper* **230**.

Young, H. P. (2015), 'The evolution of social norms', *economics* **7**(1), 359–387.

Supplementary Appendix

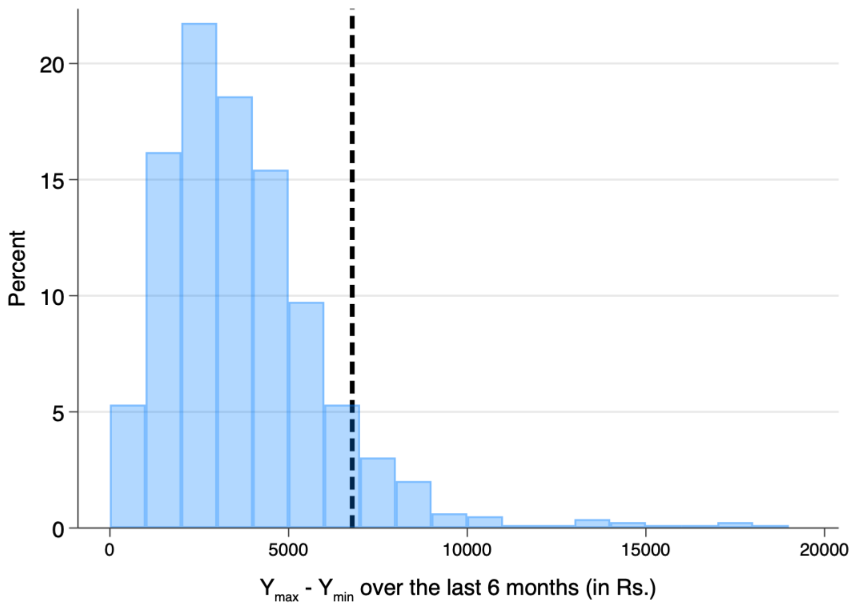
A Baseline

Figure A.1: Survey Locations in the National Capital Region



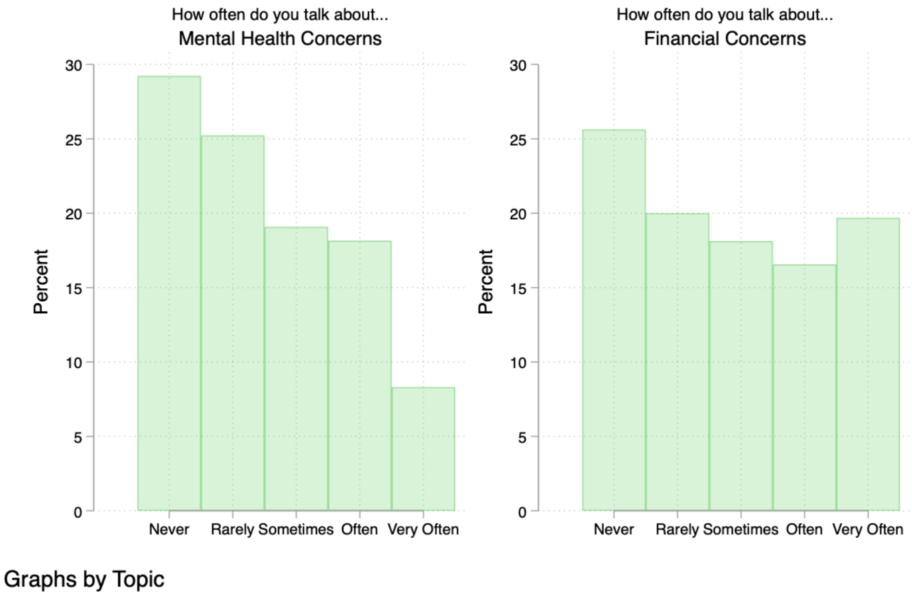
Notes: The figure plots the locations of various centers where we conducted the surveys (in blue) on the map of the national capital region.

Figure A.2: Volatility of Incomes across 6 months



Notes: The figure plots the difference between the maximum and minimum income an individual has earned in the last six months (in Rs). The dashed line is the average current income in the sample. This data was collected for the additional sample in 2023.

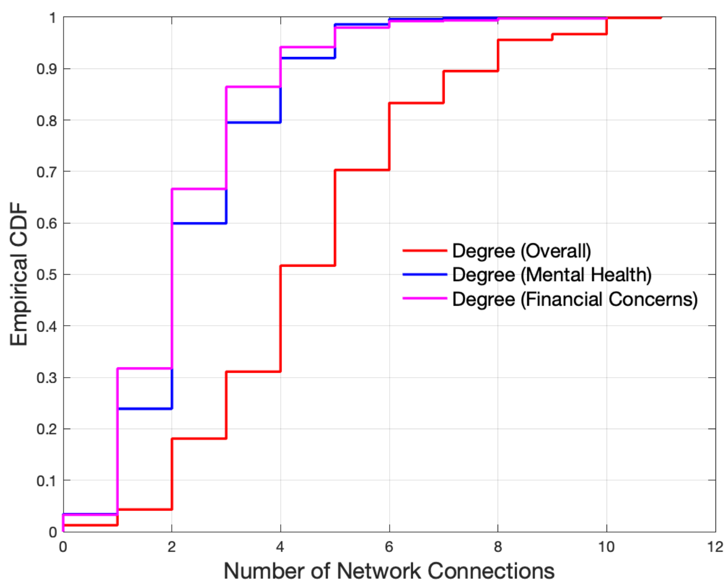
Figure A.3: Intensity of Dialogue in the last two weeks



Notes: The figure plots the baseline intensity of dialogue around mental health and financial concerns with peers in the last two weeks.

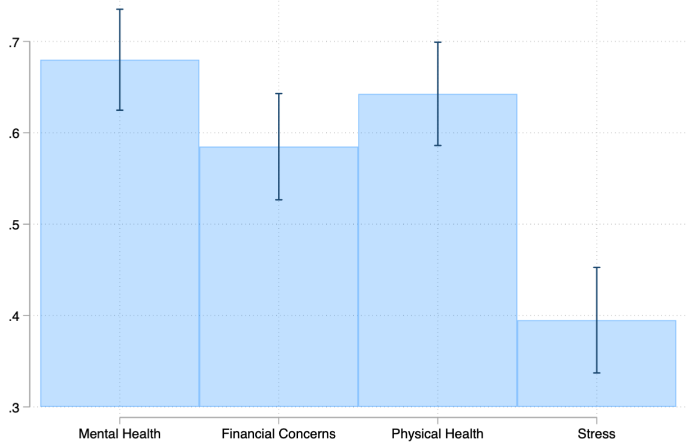
Click to go back.

Figure A.4: Degree Distributions for Overall Networks, Mental Health Advice Networks, and Financial Networks in 2023



Notes: This figure plots the CDF of the number of connections in the overall networks, advice networks, and financial networks for borrowing and lending. Individuals were asked to list up to ten other individuals in their community whom they interact with to borrow/lend, take/give advice, work with etc. ("Overall"), whom they only take advice from regarding mental health issues ("Mental Health"), and whom they contact for financial support ("Financial Concerns").

Figure A.5: Percentage of Underestimators



Notes: This figure plots proportion of underestimators by type of dialogue and additionally by "Stress" where individuals are asked to predict how many others do they think would say that their "difficulties were piling up so high that they could not overcome them" often or very often. Underestimators are identified by comparing their beliefs about community's willingness to engage with actual average in the community (i.e., NGO center).

Table A.1: Correlations between Consumption Volatility, Demographic Characteristics, Network Characteristics, and Beliefs about Peers in 2023

VARIABLES	(1) Volatile Consumption	(2) Consumption Crisis
Male	0.0272	-0.0445
	(0.0380)	(0.0361)
Age	-0.00224	0.00241*
	(0.00142)	(0.00140)
Income	2.32e-05***	-2.99e-05***
	(6.12e-06)	(5.82e-06)
Degree (FC)	-0.0150	-0.0297**
	(0.0140)	(0.0132)
Degree (MH)	0.0473***	0.00126
	(0.0137)	(0.0136)
Talks to peers (MH)	-0.144***	0.218***
	(0.0444)	(0.0430)
Talks to peers (FC)	-0.0213	0.0857**
	(0.0439)	(0.0428)
Migrant	-0.0266	0.0378
	(0.0398)	(0.0380)
Beliefs (MH)	-0.0319***	0.0213**
	(0.0107)	(0.0101)
Beliefs (FC)	0.00689	-0.00654
	(0.0101)	(0.00964)
Constant	0.560***	0.406***
	(0.0875)	(0.0843)
Observations	789	789
R-squared	0.079	0.138

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results when the variables "Volatile Consumption" and "Consumption Crisis" are regressed on demographic characteristics, network characteristics, dialogue intensity, and beliefs about peers. Volatile consumption is a binary variable equal to 1 when the individual reports that their consumption has fluctuated a little/a lot over the last 6 months as opposed to not at all. Consumption crisis is a binary variable equal to 1 if the individual reports having faced consumption crisis events very often (i.e., not having enough financial resources to meet basic dietary needs, education expenditures, or health expenditures)."MH" refers to mental health and "FC" refers to financial concerns. Talks to peers (MH)/ Talks to peers (FC) is a binary variable equal to 1 if the individual reports having spoken about mental health or financial concerns to their peers often/very often in the last 2 weeks.

Table A.2: Correlations among Beliefs and Engagement along various dimensions

	Beliefs (FC)	Beliefs (PH)	Beliefs (MH)	Willing to engage (FC)	Willing to engage (PH)	Willing to engage (MH)
Beliefs (FC)	1					
Beliefs (PH)	0.591***	1				
Beliefs (MH)	0.588***	0.496***	1			
Willing to engage (FC)	0.151*	0.172**	0.206***	1		
Willing to engage (PH)	0.0586	0.0860	0.181**	0.446***	1	
Willing to engage (MH)	0.0516	0.137*	0.106	0.440***	0.550***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This table presents correlations between beliefs and willingness to engage along various topics. The entry in the i'th row and j'th column reports the correlation coefficient between the i'th and j'th variable. "MH" refers to mental health, "PH" refers to physical health, and "FC" refers to financial concerns. Beliefs refer to how many community members out of any 10 individuals believe would be willing to engage around a topic. Willing to engage is a binary variable equal to 1 if they themselves are willing to engage with others around a topic.

Table A.3: Correlations of Network Gaps and Dialogue Intensity

Panel A: Correlations with Beliefs

VARIABLES	Degree Gap	Dialogue (MH)	Dialogue (PH)	Dialogue (FC)
Beliefs (MH)	-0.015	0.064*		
	(0.011)	(0.033)		
Beliefs (PH)			0.021	
			(0.024)	
Beliefs (FC)				0.085***
				(0.021)
Observations	210	271	275	263
R-squared	0.015	0.022	0.002	0.034

Panel B: Correlations with Underestimation

VARIABLES	Degree Gap	Degree Gap	Dialogue (MH)	Dialogue (MH)
Underestimator	-0.015		-0.337*	
	(0.065)		(0.169)	
Severe Underestimator		0.178**		-0.373*
		(0.070)		(0.183)
Observations	210	210	271	271
R-squared	0.000	0.039	0.015	0.015

Note: This table reports regression results where we regress degree gaps and dialogue intensity on (a) various measures of beliefs and (b) whether an individual underestimates engagement with mental health. *Underestimator* is equal to 1 if an individual underestimates how many others in their community are willing to engage. *Severe Underestimator* is equal to 1 if the difference between actual willingness to engage and their belief is greater than the 75th percentile. We construct "Degree Gap" as a network-based measure of an individual's disengagement using differences in degrees in overall and advice networks. "MH" refers to mental health, "PH" refers to physical health, and "FC" refers to financial concerns. Standard errors are robust and clustered at the center level.

Table A.4: Correlations of Willingness to Engage with Beliefs about Community.

VARIABLES	(1) Willing to have MH Dialogue	(2) Willing to have MH Dialogue	(3) Willing to have PH Dialogue	(4) Willing to have PH Dialogue
Beliefs (MH)	0.016			
	(0.012)			
Beliefs -Stigma (MH)		-0.012*		
		(0.006)		
Beliefs (PH)			0.014	
			(0.010)	
Beliefs -Stigma (PH)				-0.023**
				(0.008)
Observations	274	277	275	276
R-squared	0.010	0.007	0.007	0.023

Notes: This table reports results where we regress willingness to engage in dialogue on beliefs about peers willingness to engage and beliefs about stigma among peers. 'MH" refers to mental health, "PH" refers to physical health, and "FC" refers to financial concerns. Standard errors are robust and clustered at the level of the center.

Table A.5: Correlation between Beliefs and Demographic/Network Characteristics (in 2023)

VARIABLES	(1) Belief (MH)	(2) Belief (FC)
Male	-0.109	0.0199
	(0.132)	(0.139)
Age	-0.000661	0.00437
	(0.00599)	(0.00607)
Income	4.22e-05*	1.15e-05
	(2.32e-05)	(2.33e-05)
Degree (Overall)	-0.150***	-0.171***
	(0.0510)	(0.0510)
Degree (MH)	0.265***	0.129*
	(0.0664)	(0.0681)
Degree (FC)	0.351***	0.312***
	(0.0609)	(0.0591)
Talks to peers (MH)	0.0708	0.208
	(0.153)	(0.182)
Talks to peers (FC)	-0.177	-0.0771
	(0.146)	(0.174)
Years in Location	-0.00832	0.00474
	(0.00532)	(0.00583)
Consumption Crisis	0.277*	0.167
	(0.142)	(0.147)
Volatile Consumption	-0.405***	-0.113
	(0.134)	(0.138)
Happiness	0.204***	0.342***
	(0.0608)	(0.0632)
Constant	2.512***	1.798***
	(0.393)	(0.393)
Observations	775	775
R-squared	0.109	0.085

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: This table presents the regression results where we regress beliefs about community members' willingness to engage around mental health concerns and financial concerns on demographic characteristics, network characteristics, dialogue intensity, and economic indicators. "MH" refers to mental health concerns and "FC" refers to financial concerns. Talks to peers (MH)/ Talks to peers (FC) is a dummy equal to 1 if the individual reports having spoken about mental health or financial concerns to their peers often/very often in the last 2 weeks. Degree refers to number of connections. Happiness is defined on a scale of 1-5. Volatile consumption is a dummy equal to 1 when the individual reports that their consumption has fluctuated a little/a lot over the last 6 months. Consumption crisis is a dummy equal to 1 if the individual reports having faced consumption crisis events very often (i.e. not having enough financial resources to meet basic dietary needs, education expenditures, or health expenditures).

B Design

B.1 Endline Balance

Table B.1: Endline Balance: Demographics and Health

	Control	Treatment	p-value
Age	34.69	34.66	0.98
Female	0.30	0.34	0.60
Income (Category)	2.81	2.83	0.89
Monthly HH Income (< Rs 2,500)	0.14	0.11	0.48
Monthly HH Income (Rs 2,500-5,000)	0.16	0.22	0.30
Monthly HH Income (Rs 5,000-10,000)	0.44	0.41	0.65
Monthly HH Income (Rs 10,000+)	0.26	0.27	0.89
Employed	0.67	0.68	0.91
Stress (Index; Scale 1-4)	3.10	3.07	0.82
Satisfaction (Scale 1-4)	2.84	2.94	0.47
Happiness (Scale 1-4)	2.41	2.48	0.65
Degree (Advice)	2.59	3.14	0.28
Degree (Overall)	3.84	4.57	0.24
Degree (Advice Network) >0	4.67	4.76	0.88
Degree (Overall Network) >0	5.79	5.91	0.85
N	92	88	

Notes: This table presents the results of balance tests between the treatment and control group for the endline sample where the first two columns show the means of the variables for the Control and Treatment Group and the third column shows the *p*-values for the two-sided test that the two means are equal using robust standard errors. *Degree* refers to the number of connections and *Degree* (> 0) only includes strictly positive entries.

Table B.2: Endline Balance: Willingness to Talk and Beliefs

	Control	Treatment	p-value
Willingness to Talk(Financial)	0.70	0.74	0.59
Willingness to Talk (Mental Health)	0.67	0.56	0.12
Willingness to Talk (Physical Health)	0.60	0.63	0.68
Stigma (Physical Health)	0.63	0.53	0.21
Stigma (Mental Health)	0.49	0.50	0.94
Beliefs (Mental Health)	4.70	5.18	0.33
Beliefs (Physical Health)	5.08	5.26	0.73
Beliefs (Financial Concerns)	5.59	6.71	0.03
Beliefs -Stigma (Mental Health)	4.84	5.05	0.70
Beliefs -Stigma (Physical Health)	4.95	5.05	0.84
Beliefs -Stress (Mental Health)	5.67	6.36	0.20
Dialogue (Physical Health; 1-5)	2.22	2.08	0.51
Physical Health talk (Never)	0.45	0.47	0.78
Physical Health talk (Rarely)	0.18	0.19	0.88
Physical Health talk (Sometimes)	0.16	0.19	0.59
Physical Health talk (Often)	0.11	0.07	0.34
Physical Health talk (Very Often)	0.09	0.07	0.64
Dialogue (Mental Health; 1-5)	2.50	2.39	0.58
Mental Health talk (Never)	0.34	0.33	0.87
Mental Health talk (Rarely)	0.17	0.30	0.05
Mental Health talk (Sometimes)	0.21	0.13	0.19
Mental Health talk (Often)	0.21	0.14	0.27
Mental Health talk (Very Often)	0.07	0.10	0.53
Dialogue (Financial Concerns ; 1-5)	2.81	3.01	0.37
Financial Concerns talk (Never)	0.27	0.20	0.28
Financial Concerns talk (Rarely)	0.18	0.21	0.56
Financial Concerns talk (Sometimes)	0.18	0.18	1.00
Financial Concerns talk (Often)	0.20	0.18	0.70
Financial Concerns talk (Very Often)	0.17	0.23	0.33
N	92	88	

Notes: This table presents additional results of balance tests between the treatment and control group for the endline sample where the first two columns show the means of the variables for the Control and Treatment Group and the third column shows the *p*-values for the two-sided test that the two means are equal using robust standard errors. All variables are binary except beliefs and dialogue intensity. Beliefs are measured in terms of 0-10 individuals in the community. Dialogue intensity is measured on a scale from 1-5 where 1 is 'Never' and 5 is 'Very Often'.

C Endline Results

Table C.1: Effects on Additional Endline Outcomes

Panel A: Effects on Demand for Information

	Information Session	Listening to Good Practices (Immediate)	Listening to Good Practices
Treatment	0.0534	-0.0507	0.0263
	(0.0685)	(0.0846)	(0.0636)
Bootstrap p-value	0.430	0.511	0.697
q-values	1	1	1
Constant	0.674***	0.551***	0.821***
	(0.0491)	(0.0603)	(0.0472)
Observations	180	141	139
R-squared	0.003	0.003	0.001

Panel B: Effects on Self Efficacy

VARIABLES	Goals (Finance)	Goals (Education)	Goals (Business)	Self Efficacy
Treatment	0.154	-0.200	0.0436	-0.0268
	(0.264)	(0.243)	(0.278)	(0.222)
Bootstrap p-value	0.647	0.468	0.852	0.893
q-values	1	1	1	1
Constant	2.219***	2.729***	2.548***	2.481***
	(0.176)	(0.183)	(0.190)	(0.151)
Observations	148	140	144	139
R-squared	0.002	0.005	0.000	0.000

Panel C: Effects on Stigma

VARIABLES	Stigma (Information Session)	List Experiment	Depression Score Revelation
Treatment	0.0313	-0.0763	-0.0917
	(0.0430)	(0.211)	(0.0881)
Bootstrap p-value	0.373	0.781	0.451
q-values	1	1	1
Constant	0.922***	3.507***	0.925***
	(0.0338)	(0.157)	(0.0423)
Observations	128	154	64
R-squared	0.004	0.001	0.020

Robust standard errors in parentheses. Wild bootstrap p-value reported, reps=999.
*** p<0.01, ** p<0.05, * p<0.1

Notes: We report robust standard errors. We also report wild bootstrap p values using the method outlined in Cameron et al. (2008) where we treat the NGO center as the cluster unit. The q-values (Benjamini et al., 2006) reported in each table treat the outcomes in the table as multiple hypotheses being tested together.

C.1 Heterogeneity

Table C.2: Treatment Effects by Baseline Willingness to Talk about Mental Health

Panel A: Effect on Engagement Outcomes

	Savings Group	Listening Volunteer	Listening Contribution	Contribution (in Rupees)	Contribution (>0)	Community Engagement
Treatment	0.374***	0.434***	0.204*	11.86**	7.817	0.315***
	(0.137)	(0.119)	(0.119)	(4.964)	(5.063)	(0.104)
Willingness to talk (Mental Health)	0.466***	0.375***	0.186	10.19**	6.190	0.295***
	(0.122)	(0.112)	(0.114)	(4.441)	(4.641)	(0.0949)
Interaction	-0.265*	-0.355**	-0.106	-6.677	-5.553	-0.235**
	(0.154)	(0.151)	(0.138)	(6.223)	(6.075)	(0.117)
Constant	0.350***	0.280***	0.625***	15.87***	26.07***	0.467***
	(0.108)	(0.0909)	(0.100)	(3.656)	(4.083)	(0.0876)
Observations	138	160	156	150	121	138
R-squared	0.185	0.103	0.053	0.069	0.029	0.152
Bootstrap p-value	0.0390	0.101	0.345	0.114	0.187	0.0561

Panel B: Effect on Other Outcomes

VARIABLES	Information Session	Listening to Good Practices (Immediate)	Listening to Good Practices
Treatment	0.444***	0.0292	0.168
	(0.116)	(0.152)	(0.129)
Willingness to talk (Mental Health)	0.488***	0.0389	0.182
	(0.105)	(0.138)	(0.120)
Interaction	-0.510***	-0.144	-0.184
	(0.142)	(0.187)	(0.150)
Constant	0.333***	0.526***	0.684***
	(0.0918)	(0.116)	(0.108)
Observations	164	132	131
R-squared	0.143	0.010	0.027
Bootstrap p-value	0.0120	0.471	0.0340

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows heterogeneous effects on the demand for additional information by baseline willingness to engage with mental health related concerns. We report robust standard errors. We also report wild bootstrap *p*-values for the interaction term using the method outlined in Cameron et al. (2008) where we treat the NGO center as the cluster unit.

Click to go back.

Table C.3: Heterogeneous Treatment Effects by Baseline Dialogue

Panel A: Baseline Dialogue Around Mental Health

	Savings Group	Listening Volunteer	Listening Contribution	Contribution (in Rupees)	Contribution (>0)	Community Engagement
Treatment	0.0227	0.0581	0.0808	8.129*	5.643	0.0437
	(0.0948)	(0.107)	(0.0651)	(4.226)	(3.770)	(0.0665)
Baseline MH Dialogue (Below Median)	-0.305***	-0.195*	-0.264***	-12.53***	-6.565*	-0.243***
	(0.107)	(0.106)	(0.0943)	(3.957)	(3.702)	(0.0753)
Interaction	0.282**	0.206	0.119	-0.164	-2.510	0.217**
	(0.142)	(0.146)	(0.117)	(5.632)	(5.189)	(0.100)
Constant	0.805***	0.636***	0.864***	28.90***	33.86***	0.772***
	(0.0627)	(0.0734)	(0.0524)	(2.951)	(2.661)	(0.0458)
Observations	150	174	170	163	129	150
R-squared	0.088	0.048	0.093	0.137	0.078	0.118
Bootstrap p-value	0.0931	0.185	0.296	0.979	0.456	0.108

Panel B: Baseline Dialogue Around Financial Concerns

	Savings Group	Listening Volunteer	Listening Contribution	Contribution (in Rupees)	Contribution (>0)	Community Engagement
Treatment	0.0368	-0.0950	-0.0150	0.887	1.296	0.00919
	(0.0771)	(0.1000)	(0.0847)	(4.532)	(3.980)	(0.0600)
Baseline FC Dialogue (Below Median)	-0.363***	-0.453***	-0.204**	-9.458**	-4.641	-0.317***
	(0.0974)	(0.0951)	(0.0912)	(4.170)	(3.852)	(0.0673)
Interaction	0.183	0.436***	0.237**	9.400	3.410	0.220**
	(0.130)	(0.138)	(0.119)	(5.948)	(5.408)	(0.0937)
Constant	0.875***	0.806***	0.857***	28.44***	33.70***	0.844***
	(0.0593)	(0.0667)	(0.0599)	(3.246)	(2.848)	(0.0422)
Observations	150	174	170	163	129	150
R-squared	0.133	0.133	0.056	0.059	0.022	0.176
Bootstrap p-value	0.154	0.00100	0.0811	0.0280	0.312	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows heterogeneous effects on community engagement by a binary variable which indicates whether baseline frequency of dialogue with peers around mental health financial concerns respectively was less than/equal to the median. We report robust standard errors. We also report wild bootstrap *p*-values for the interaction term using the method outlined in Cameron et al. (2008) where we treat the NGO center as the cluster unit.

Table C.4: Heterogeneous Treatment Effects by Degree (Overall)

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) Contribution (>0)	(6) Community Engagement
Treatment	0.0117	0.159	0.150*	2.992	-3.434	0.117
	(0.110)	(0.107)	(0.0901)	(4.303)	(3.632)	(0.0774)
Degree (Overall)	-0.00736	0.0104	0.00258	-0.464	-0.893*	0.00156
	(0.0137)	(0.0126)	(0.0128)	(0.500)	(0.464)	(0.00846)
Interaction	0.0307*	-0.00116	-0.00625	0.838	1.532**	0.00432
	(0.0183)	(0.0172)	(0.0160)	(0.733)	(0.632)	(0.0123)
Constant	0.696***	0.499***	0.728***	24.59***	34.88***	0.656***
	(0.0755)	(0.0746)	(0.0707)	(3.052)	(2.466)	(0.0572)
Observations	150	174	170	163	129	150
R-squared	0.052	0.034	0.024	0.038	0.050	0.049
Bootstrap p-value	0.0450	0.910	0.566	0.153	0.0651	0.633

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows heterogeneous effects on various measures of community engagement by the number of connections in the overall network. We report robust standard errors. We also report wild bootstrap p-values for the interaction term using the method outlined in Cameron et al. (2008) where we treat the NGO center as the cluster unit.

Table C.5: Treatment Effects by Stigma (Mental Health)

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) Contribution (>0)	(6) Community Engagement
Treatment	0.202**	0.146	0.125	3.224	-0.930	0.149**
	(0.0982)	(0.105)	(0.0866)	(4.171)	(3.726)	(0.0703)
Stigma (Mental Health)	0.110	-0.148	-0.0453	-7.107*	-7.071*	-0.0175
	(0.102)	(0.106)	(0.0884)	(4.262)	(4.033)	(0.0732)
Interaction	-0.0907	0.0921	-0.00473	7.291	9.349*	-0.0119
	(0.144)	(0.150)	(0.124)	(6.001)	(5.470)	(0.103)
Constant	0.632***	0.610***	0.775***	26.25***	33.87***	0.684***
	(0.0685)	(0.0742)	(0.0613)	(2.912)	(2.695)	(0.0491)
Observations	139	161	157	151	122	139
R-squared	0.045	0.053	0.028	0.050	0.040	0.056
Bootstrap p-value	0.527	0.451	0.951	0.274	0.00300	0.854

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows heterogeneous effects on various measures of community engagement by whether the individual has stigma against mental health during the baseline in that they believe that "people should stay away from those with mental health issues". We report robust standard errors. We also report wild bootstrap p-values for the interaction term using the method outlined in Cameron et al. (2008) where we treat the NGO center as the cluster unit.

Click to go back.

Table C.6: Treatment Effects by Underestimators

Panel A: Underestimation of Willingness to Engage around Mental Health

	Savings Group	Listening Volunteer	Listening Contribution	Contribution (in Rupees)	Contribution (>0)	Community Engagement
Treatment	0.0804	0.0833	0.117	11*	8.636	0.110
	(0.116)	(0.130)	(0.127)	(5.723)	(5.587)	(0.0759)
Underestimator (MH)	-0.131	-0.287**	0.0192	5.667	6.329	-0.104
	(0.122)	(0.130)	(0.125)	(5.316)	(5.263)	(0.0820)
Interaction	0.0799	0.128	0.00541	-6.310	-7.626	0.0215
	(0.149)	(0.163)	(0.147)	(6.923)	(6.588)	(0.101)
Constant	0.813***	0.750***	0.750***	19.33***	26.36***	0.771***
	(0.0992)	(0.110)	(0.110)	(4.610)	(4.724)	(0.0650)
Observations	126	146	144	138	113	126
R-squared	0.045	0.091	0.024	0.036	0.024	0.080
Bootstrap p-value	0.632	0.429	0.974	0.402	0.224	0.848

Panel B: Underestimation of Willingness to Engage around Financial Concerns

	Savings Group	Listening Volunteer	Listening Contribution	Contribution (in Rupees)	Contribution (>0)	Community Engagement
Treatment	0.107	0.200	0.100	7.347	5.196	0.118
	(0.122)	(0.135)	(0.114)	(5.352)	(4.983)	(0.0874)
Underestimator (FC)	-0.0295	0.0294	0	1.562	1.786	-0.00128
	(0.126)	(0.134)	(0.117)	(5.032)	(4.698)	(0.0857)
Interaction	0.0387	0.0206	0.0389	-0.694	-2.689	0.0325
	(0.154)	(0.169)	(0.141)	(6.706)	(6.161)	(0.109)
Constant	0.737***	0.500***	0.750***	21.84***	29.64***	0.684***
	(0.103)	(0.113)	(0.0982)	(4.209)	(3.967)	(0.0689)
Observations	126	147	144	138	111	126
R-squared	0.026	0.045	0.024	0.032	0.015	0.053
Bootstrap p-value	0.824	0.838	0.851	0.952	0.705	0.737

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: This table shows heterogeneous effects on community engagement by whether the individual underestimates community willingness to engage around mental health (MH) and financial concerns (FC respectively. We report robust standard errors. We also report wild bootstrap *p*-values for the interaction term using the method outlined in Cameron et al. (2008) where we treat the NGO center as the cluster unit.

Click to go back.

Table C.7: Heterogeneous Treatment Effects by Baseline Beliefs about Stress

Panel A: Underestimation of Stress in the Community

	Savings Group	Listening Volunteer	Listening Contribution	Contribution (in Rupees)	Contribution (>0)	Community Engagement
Treatment	0.102	0.192*	0.0901	4.487	1.750	0.110
	(0.0958)	(0.102)	(0.0899)	(4.328)	(3.940)	(0.0706)
Underestimator (Stress)	-0.0210	-0.00905	0.0444	0.590	-1.700	-0.00560
	(0.116)	(0.119)	(0.104)	(4.609)	(4.214)	(0.0814)
Interaction	0.0534	-0.0134	0.0870	5.146	2.677	0.0400
	(0.148)	(0.164)	(0.124)	(6.308)	(5.773)	(0.102)
Constant	0.735***	0.538***	0.737***	23.47***	32.50***	0.696***
	(0.0769)	(0.0809)	(0.0724)	(3.274)	(2.998)	(0.0534)
Observations	128	149	146	140	113	128
R-squared	0.024	0.039	0.035	0.036	0.011	0.046
Bootstrap p-value	0.550	0.940	0.290	0.436	0.607	0.541

Panel B: Beliefs about Stress in the Community

	Savings Group	Listening Volunteer	Listening Contribution	Contribution (in Rupees)	Contribution (>0)	Community Engagement
Treatment	0.143	0.200	0.140	7.454	3.895	0.141
	(0.131)	(0.140)	(0.104)	(5.039)	(4.644)	(0.0891)
Stress Belief (>4)	0.0965	0.0837	-0.0192	2.335	4.387	0.0665
	(0.120)	(0.122)	(0.107)	(4.668)	(4.312)	(0.0839)
Interaction	-0.0393	-0.0248	-0.0380	-2.293	-1.933	-0.0305
	(0.157)	(0.170)	(0.133)	(6.484)	(5.881)	(0.110)
Constant	0.667***	0.481***	0.769***	22.31***	29***	0.653***
	(0.0978)	(0.0975)	(0.0838)	(3.568)	(3.407)	(0.0676)
Observations	128	149	146	140	113	128
R-squared	0.031	0.044	0.023	0.027	0.022	0.051
Bootstrap p-value	0.679	0.871	0.758	0.784	0.770	0.721

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Panel A shows shows heterogeneous effects on various measures of community engagement by whether the individual underestimates the level of stress in their community i.e. their belief is less than the true community-level stress. Panel B shows heterogeneous effects on various measures of community engagement by whether the individual's belief about the level of stress in their community is above the proportion who are actually not willing to engage (i.e. 40%). This is to check if those who think more than 40% are stressed and might assume that these individuals are not willing to engage change their behaviour when they are told that majority are willing to engage. We report robust standard errors. We also report wild bootstrap p-values for the interaction term using the method outlined in Cameron et al. (2008) where we treat the NGO center as the cluster unit.

Click to go back.

D Mechanisms

Table D.1: List Experiment to detect Social Desirability Concerns

	3 statements + Statement about engagement with MH	3 statements	Average Difference between 4 and 3 statements
Mean Agreements	2.971	2.217	0.754***
Observations			670

	3 statements + 2 statements about engagement with MH and FC	3 statements	Average Difference between 5 and 3 statements
Mean Agreements	3.717	2.217	1.499***
Observations			504

Notes: These tables present the results of the list experiments where individuals were randomly divided into groups and either asked how many of the three statements they agree with (i.e., control) or asked how many of 4 and 5 statements they agree with (i.e., treated). The additional statements were relating to engagement with mental health and financial concerns respectively. "MH" refers to mental health and "FC" refers to financial concerns. The first two columns show the mean number of agreements when different number of statements are provided. The third column computes the difference between treated and control conditions.

Table D.2: Additional Experiment to detect Social Desirability Concerns

VARIABLES	(1) Savings Group	(2) Listening Service	(3) Savings Group	(4) Listening Service
Increased Distance between Enumerator and Respondent	0.00398	0.0326	0.0241	0.0497
	(0.0280)	(0.0324)	(0.0365)	(0.0453)
Treatment			-0.0609	-0.0305
			(0.0397)	(0.0467)
Increased Distance x Treated			-0.0352	-0.0311
			(0.0555)	(0.0647)
Constant	0.807***	0.690***	0.838***	0.706***
	(0.0199)	(0.0233)	(0.0263)	(0.0326)
Observations	791	791	791	791
R-squared	0.000	0.001	0.011	0.004

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: This table presents the results of an experiment where we randomly vary whether the enumerator enters the respondent's answer (*Increased Distance*=0) or gives them the device to answer it privately (*Increased Distance*=1). We also interact it with the treatment to see if the response being visible to the enumerator affects responses differently across the two groups. We report robust standard errors.

E Follow up Survey

E.1 Follow up Survey Outcomes

The follow up survey outcomes are listed below.

1. Well-being: Happiness, Life-Satisfaction, Stress
2. Self-reported Dialogue
   (a) Indicator whether they have initiated a conversation with their peers to discuss matters relating to either of their mental and financial concerns.
   (b) Number of peers they have initiated conversations with to discuss matters relating to either of their mental and financial concerns.
   (c) Indicator whether their peers have initiated a conversation with them to discuss matters relating to their own mental and financial concerns.
3. Physical Health Safety
   (a) How often did they follow COVID rules (distancing, wearing masks)?
   (b) How often did they wear safety gear (masks, gloves, jacket) while working in hazardous conditions?
4. Altruism towards community members (Dictator Game)
5. Self-reported contact with a doctor/helpline for mental or physical health concerns

E.2 Follow up Survey: Main Results

The results on network interactions have been presented in the main text in Table 5. Next, we find in Table E.1 that individuals in the treatment group are likely to donate Rs. 31 more than the control group in a hypothetical dictator game where they can split Rs. 200 with a randomly chosen member of their community. This result is robust to the inclusion of unbalanced controls as well. However, we also document that individuals in the treatment group are more likely to report higher levels of unhappiness, stress, and lower life satisfaction. At the same time, we also find that the treatment group wears work safety gear significantly more often than the control group, conditional on having worked. Given that we find an (a) increase in dialogue around mental health and (b) investments in self-care (in terms of greater adoption of safety gear), a potential mechanism explaining lower well-being could be that the role of dialogue in alleviating stress becomes ineffective in the presence of correlated risks such as COVID-19.

However, in the absence of real-time data on the location of these individuals, we acknowledge that we cannot rule out the possibilities that (a) the treatment group has lower well-being because they have migrated to an area of low COVID risk and hear about the impact of the pandemic on their peers via increased dialogue or (b) the treatment group is located in an area of high COVID risk and well being is consequently lower. We have checked balance on several baseline variables and do not ex-ante see any reason why the treatment group will be more or less likely to migrate. However, if this were to be the case, we believe that mechanism (a) seems to be more likely. This is because the result in Tables E.1 shows that the treatment group wears COVID-19 masks less often than the control group. 92% of our sample reports to have *always* worn a mask in the past two weeks before the call and the negative effect on the treatment only driven by the remaining 8%. Individuals in low risk-areas may have been less likely to wear COVID-19 safety masks. Additionally, the proportion of individuals for whom the question on safety gear adoption is missing, is higher for the treatment group after controlling for unbalanced baseline variables. This may imply that they are away from their place of work.

Thus, while we document positive effects on dialogue, altruism, and safety gear adoption, we also document negative effects on happiness, life-satisfaction, and stress. These results highlight that the correlated nature of risk can affect the success of belief-shifting interventions.

Table E.1: Other Outcomes

Panel A: Effects on Other Outcomes

VARIABLES	Masks (How often)	Safety Gear (How often)	Maintain Distance (How often)	Mental Health-Doctor (Made Call)
Treatment	-0.179	0.783***	-0.0351	-0.0229
	(0.108)	(0.223)	(0.145)	(0.0534)
q-values	0.024	0.008	0.19	0.14
Constant	3.943***	2.977***	3.604***	0.0943**
	(0.0419)	(0.180)	(0.0907)	(0.0405)
Observations	104	69	104	109
R-squared	0.027	0.120	0.001	0.002

Panel B: Effects on Well-being

VARIABLES	Happiness	Life Satisfaction	Stress	Altruism (Dictator Game)
Treatment	-0.568***	-0.434**	0.240	30.66**
	(0.176)	(0.168)	(0.167)	(13.68)
q-values	0.024	0.024	0.054	0.048
Constant	3.208***	3.434***	2.684***	54.90***
	(0.115)	(0.109)	(0.125)	(9.793)
Observations	103	103	102	104
R-squared	0.094	0.062	0.020	0.047

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The q-values (Benjamini et al., 2006) reported in each table treat all the outcomes in this table as multiple hypotheses being tested together. Standard errors are robust.

F Larger Scale Experiment

We conducted a demographic survey with a sample of $\sim$ 800 individuals in 2 NGO centers in Delhi in 2023 to alleviate the concern that the baseline patterns are specific to the sub-sample or because the survey was conducted during the pandemic. Table F.1 shows that the baseline patterns of low engagement both in terms of low dialogue and gaps in the number of connections in advice networks and overall networks continue to hold. The proportion of individuals who have never or rarely had a conversation about mental health and financial concerns is lower than before but still as high as 30%. We also separately measured the number of links for financial and mental health-related advice-taking and find that the average number of links in the former is even lower than that of advice-taking around mental health. We find that those who are more optimistic about their community's willingness to engage are more likely to have fewer gaps in connections between overall and financial networks. We also continue to find that individuals report a high willingness to engage and an even larger proportion underestimate how willing their community is to engage with them. This reassures us that the baseline patterns measured before the experiment continue to hold.

Next, we present the results from the implementation of a similar information treatment. Half of these individuals were chosen to be treated and given the same information as the treated individuals in the RCT with two crucial differences that make this a weaker replication of the original experiment. First, we provided individuals with information about the average willingness to engage in 2021. They were told that this is not information from 2023 and were asked to answer questions assuming this were true today.[39] Further, unlike the original

[39]This is because we were logistically constrained to contact individuals again after eliciting their willingness to engage and wished to run this additional experiment in the same round. As a result, we could only provide them with information collected previously.

treatment where individuals were told that the information we are providing is from their own community, in the additional sample, we told individuals that the information is collected from other similar communities of waste-pickers managed by the NGO in Delhi. This is because this center was not a part of our original sample and even if it were, the center composition would have changed due to COVID-induced migration.

Table F.2 shows the balance checks for this sample. We find that while there is balance in most variables, treatment and control groups are not similar in terms of the number of overall network connections, willingness to engage in mental health related dialogue, and the proportion whose belief about community's willingness to engage is lower than the delivered information. Since this is a larger sample, we can use post double selection Lasso (Belloni et al., 2014) to estimate the treatment effects. The effect on willingness to participate in savings groups and listening services is shown in Tables F.4 and F.6. We continue to find evidence that beliefs about community willingness to engage significantly affect network engagement. We also present robustness checks where we exclude those who were a part of the previous study or report having heard about their community's views on engagement via their peers in Tables F.5 and F.7. This is to ensure that the previous information delivery does not affect our findings. We find that the results are similar in direction, magnitude, and statistical significance whether these individuals are included or not.

The mechanisms behind this treatment are different from the original treatment in that individuals are provided information about other communities like theirs from two years ago. As a result, we find negative average effects on willingness to participate in savings groups. Disentangling this further, we find that those who are pessimistic about their community, are significantly less likely to engage when they are told that other communities are more willing to engage. Those who are optimistic about the community behave in the opposite manner. This pattern also holds for listening services even though the negative average effect is not significant. These effects can arise if those who believe that their community is not very willing to engage do not treat the new information as a positive signal about their own community but instead compare their community with other seemingly more supportive communities and draw a negative inference. This highlights the importance of the type of reference group whose beliefs are conveyed, as beliefs about peers can be updated either directly or through relative comparisons.

Regardless of how the information affects individual engagement, these results confirm that information about own community's willingness to engage (relative to others, in this case) can affect own willingness to engage with the network. Not only this, despite the weaker treatment, as Table F.8 shows, we detect a large increase in financial contributions made by the participants for setting up savings groups and listening services. These contributions are significantly higher than the control group for treated participants who have lower initial predictions about their community's willingness to engage with them i.e., those who think that not many people from their community would be willing to engage with them.[40] This suggests that while they may be less willing to engage upon hearing the information, they are more willing to finance avenues for such interactions to be set up in future. We are planning to use these collected funds to set up avenues for informal interactions (such as savings groups) with the help of the NGO.

[40]Note that this outcome is only accurately collected for the second center which is why the subsample here is close to half of the entire sample.

F.1 Results from the Larger Scale Experiment

Table F.1: Summary Statistics

	(1)	
Age	34.03	(12.35)
Female	0.549	(0.498)
Income	6780.9	(3142.1)
Degree (Overall)	4.655	(2.077)
Degree (MH Advice)	2.431	(1.310)
Degree (FC Advice)	2.216	(1.324)
Happiness (1-5)	3.326	(1.140)
Willingness to Talk (FC)	0.898	(0.303)
Willingness to Talk (MH)	0.880	(0.325)
Beliefs (MH; 0-10)	3.753	(1.800)
Beliefs (FC; 0-10)	3.571	(1.864)
Dialogue (MH; 1-4)	2.930	(1.011)
MH talk (Never)	0.113	(0.316)
MH talk (Rarely)	0.210	(0.407)
MH talk (Sometimes)	0.312	(0.464)
MH talk (Often)	0.365	(0.482)
Dialogue (FC; 1-4)	2.889	(1.010)
FC talk (Never)	0.119	(0.324)
FC talk (Rarely)	0.215	(0.411)
FC talk (Sometimes)	0.325	(0.469)
FC talk (Often)	0.341	(0.474)
Observations	791	

mean coefficients; sd in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The above table shows the summary statistics (mean and standard deviation) for various demographic characteristics of interest for the additional sample in 2023. "MH" refers to mental health concerns and "FC" refers to financial concerns.

Table F.2: Balance Tests for the Full Sample in 2023

	(1) Control	(2) Treatment	(3) (1) vs. (2), p-value
Female	0.558	0.539	0.591
Age	33.901	34.144	0.783
Income	6911.429	6657.143	0.256
Degree (FC Advice)	2.275	2.160	0.222
Degree (MH Advice)	2.410	2.451	0.666
Degree (Overall)	4.491	4.810	0.031
Dialogue (MH; 1-4)	2.919	2.941	0.767
Dialogue (FC; 1-4)	2.894	2.884	0.898
Willingness to Talk (MH)	0.857	0.901	0.056
Willingness to Talk (FC)	0.881	0.914	0.125
Beliefs (MH; 0-10)	3.678	3.825	0.251
Beliefs (FC; 0-10)	3.564	3.579	0.909
Underestimators (MH)	0.823	0.771	0.067
N	385	406	

Notes: This table presents the results of balance tests between the treatment and control group for the entire replication sample. The first two columns show the means of the variables for the Control and Treatment Group and the third column shows the p values for the two-sided test that the two means are equal. "MH" refers to mental health concerns and "FC" refers to financial concerns. Underestimators is a binary variable equal to 1 if the individual's belief about proportion of community willing to engage is less than six i.e. the information provided to the treatment group. *Degree* refers to the number of connections

Table F.3: Balance Tests for Center 2 in 2023

	(1) Control	(2) Treatment	(3) (1) vs. (2), p-value
Female	0.437	0.427	0.856
Age	30.934	29.653	0.266
Income	8166.887	8370.000	0.593
Degree (FC Advice)	2.318	2.153	0.308
Degree (MH Advice)	2.616	2.567	0.757
Degree (Overall)	4.954	4.767	0.478
Dialogue (MH; 1-4)	2.722	2.693	0.807
Dialogue (FC; 1-4)	2.755	2.660	0.419
Willingness to Talk (MH)	0.874	0.927	0.129
Willingness to Talk (FC)	0.914	0.927	0.684
Beliefs (MH; 0-10)	3.954	3.960	0.975
Beliefs (FC; 0-10)	3.411	3.547	0.527
Underestimator (MH)	0.755	0.733	0.668
N	156	156	

Notes: This table presents the results of balance tests between the treatment and control group for the second center in the replication sample. We present these balance tests separately as the mental health fund outcome was only accurately measured for this center. The first two columns show the means of the variables for the Control and Treatment Group and the third column shows the p values for the two-sided test that the two means are equal. "MH" refers to mental health concerns and "FC" refers to financial concerns. Underestimators is a binary variable equal to 1 if the individual's belief about proportion of community willing to engage is less than six i.e. the information provided to the treatment group. *Degree* refers to the number of connections

Table F.5: Replication Exercise: Effect on Willingness to Participate in Savings Groups (Robust)

VARIABLES	(1) Savings Group	(2) Savings Group	(3) Savings Group
Treatment	-0.0759**	-0.0748**	-0.0806***
	(0.0297)	(0.0297)	(0.0290)
Beliefs (MH)		-0.0400	
		(0.0255)	
Treatment x Beliefs (MH)		-0.00215	
		(0.0277)	
Beliefs (FC)			0.0709***
			(0.0195)
Treatment x Beliefs (FC)			0.0573**
			(0.0289)
Constant	0.944***	0.968***	0.950***
	(0.0327)	(0.0586)	(0.0443)
Observations	696	696	696
Number of groups	0	0	0

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: This table shows regression results from the replication sample in 2023, excluding those who were a part of the previous study or report having heard about their community's views on engagement via their peers. We use post double selection Lasso (Belloni et al., 2014) with robust standard errors to estimate the treatment effects where we select from a range of baseline variables such as gender, age, income, number of network connections, dialogue intensity, and willingness to engage. Column 1 shows the main treatment effects while Columns 2 and 3 interact the treatment indicator with beliefs. "Beliefs" refers to the respondent's prediction (standardized) about how many community members would be willing to engage around mental or financial concerns. "MH" refers to mental health concerns and "FC" refers to financial concerns.

Table F.4: Replication Exercise: Effect on Willingness to Participate in Savings Groups

VARIABLES	(1) Savings Group	(2) Savings Group	(3) Savings Group
Treatment	-0.0693**	-0.0726***	-0.0712***
	(0.0270)	(0.0272)	(0.0264)
Beliefs (MH)		-0.0311	
		(0.0221)	
Treatment x Beliefs (MH)		-0.0107	
		(0.0247)	
Beliefs (FC)			0.0616***
			(0.0175)
Treatment x Beliefs (FC)			0.0592**
			(0.0266)
Constant	0.882***	0.976***	0.891***
	(0.0334)	(0.0545)	(0.0451)
Observations	789	789	789
Number of groups	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table shows regression results from the replication sample in 2023. We use post double selection Lasso (Belloni et al., 2014) with robust standard errors to estimate the treatment effects where we select from a range of baseline variables such as gender, age, income, number of network connections, dialogue intensity, and willingness to engage. Column 1 shows the main treatment effects while Columns 2 and 3 interact the treatment indicator with beliefs. "Beliefs" refers to the respondent's prediction (standardized) about how many community members would be willing to engage around mental or financial concerns. "MH" refers to mental health concerns and "FC" refers to financial concerns.

Table F.6: Replication Exercise: Effect on Willingness to Participate in Listening Service

VARIABLES	(1) Listening Service	(2) Listening Service	(3) Listening Service
Treatment	-0.0442	-0.0444	-0.0430
	(0.0323)	(0.0324)	(0.0319)
Beliefs (MH)		-0.0677***	
		(0.0253)	
Treatment x Beliefs (MH)		0.0287	
		(0.0304)	
Beliefs (FC)			0.0213
			(0.0230)
Treatment x Beliefs (FC)			0.0953***
			(0.0324)
Constant	0.730***	0.789***	0.695***
	(0.0226)	(0.0609)	(0.0523)
Observations	789	789	789
Number of groups	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table shows regression results from the replication sample in 2023. We use post double selection Lasso (Belloni et al., 2014) with robust standard errors to estimate the treatment effects where we select from a range of baseline variables such as gender, age, income, number of network connections, dialogue intensity, and willingness to engage. Column 1 shows the main treatment effects while Columns 2 and 3 interact the treatment indicator with beliefs. "Beliefs" refers to the respondent's prediction (standardized) about how many community members would be willing to engage around mental or financial concerns. "MH" refers to mental health concerns and "FC" refers to financial concerns.

Table F.7: Replication Exercise: Effect on Willingness to Participate in Listening Service (Robust)

VARIABLES	(1) Listening Service	(2) Listening Service	(3) Listening Service
Treatment	-0.0290	-0.0270	-0.0316
	(0.0345)	(0.0345)	(0.0342)
Beliefs (MH)		-0.0747***	
		(0.0283)	
Treatment x Beliefs (MH)		0.0241	
		(0.0326)	
Beliefs (FC)			0.0250
			(0.0254)
Treatment x Beliefs (FC)			0.0957***
			(0.0348)
Constant	0.721***	0.818***	0.718***
	(0.0248)	(0.0673)	(0.0437)
Observations	696	696	696
Number of groups	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table shows regression results from the replication sample in 2023, excluding those who were a part of the previous study or report having heard about their community's views on engagement via their peers. We use post double selection Lasso (Belloni et al., 2014) with robust standard errors to estimate the treatment effects where we select from a range of baseline variables such as gender, age, income, number of network connections, dialogue intensity, and willingness to engage. Column 1 shows the main treatment effects while Columns 2 and 3 interact the treatment indicator with beliefs. "Beliefs" refers to the respondent's prediction (standardized) about how many community members would be willing to engage around mental or financial concerns. "MH" refers to mental health concerns and "FC" refers to financial concerns.

Table F.8: Replication Exercise: Effect on Contribution to Set up Savings Groups and Listening Services

VARIABLES	(1) Contribution (in Rupees.)	(2) Contribution (in Rupees.)	(3) Contribution (in Rupees.)	(4) Contribution (in Rupees.)
Treatment	2.468	2.916	2.334	2.632
	(2.299)	(2.415)	(2.272)	(2.330)
Beliefs (MH)		1.020		
		(1.281)		
Treatment x MH Beliefs		-4.470*		
		(2.520)		
Beliefs (FC)			0.414	
			(0.901)	
Treatment x FH Beliefs			-3.409**	
			(1.698)	
Beliefs (Average)				1.030
				(1.349)
Treatment x Avg. Beliefs				-5.158**
				(2.456)
Constant	16.96***	16.84***	17.00***	16.95***
	(1.364)	(1.359)	(1.362)	(1.366)
Observations	312	312	312	312
R-squared	0.004	0.019	0.014	0.023

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

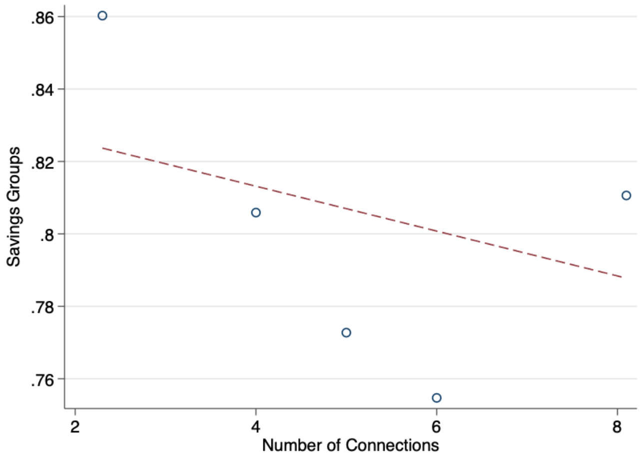
Notes: This table shows regression results from one center in the replication sample in 2023 for which this outcome was accurately measured, with robust standard errors. Column 1 shows the main treatment effects while Columns 2, 3, and 4 interact the treatment indicator with individual beliefs. "Beliefs" refers to the respondent's prediction (standardized) about how many community members would be willing to engage around mental or financial concerns. "MH" refers to mental health concerns and "FC" refers to financial concerns, and "Average" is the average of the two.

G Theory

G.1 Evidence for Model Assumptions

The following figures show correlations between degree centrality and survey measures of engagement with the network.

Figure G.1: Correlation between Degree and Willingness to Participate in Savings Groups



Notes: The figure shows the correlation between willingness to participate in savings group and the degree centrality of the agent using a binscatter with 10 quantiles.

G.2 Implications of Network Formation

Unlike Jackson (2019), we have assumed that individuals cannot choose to make or break links. Let us briefly discuss the implications of allowing individuals to form their networks. Jackson (2019) discuss that agents who benefit more from an action will choose to be more connected as they can leverage the positive strategic complementarities. In our case, the opposite holds. Agents who have a high benefit of engaging (with mental health related conversations, for instance) will choose to be less connected as higher number of connections will increase their costs of engagement. Under network formation, we will expect to see a positive relationship between private benefits from engaging and social isolation.

G.3 Two Period Model of Network Formation

Consider briefly a model with two time periods that can explain how a support network (such as an advice taking network or a risk sharing network) can form given an overall network. Consider the case where individuals myopically choose e_i in period 1 given their idiosyncratic benefit and the costs of engaging as per the social norm in their overall network. This decision to engage, visible to all other agents, is interpreted as the stated willingness to form a support network link in period 2. An agent j who sees agent i having chosen $e_i = 1$ in period 2 will interpret this as i's consent to form a mental/financial concerns related support-taking link. This is because i has sent a potentially costly message where the cost is proportional to how much they violate the social norm i.e., the proportion of others who are not willing to form these links. On the other hand, agent j will interpret $e_i = 0$ as i's wish to not form such a support link as the cost of approaching someone to suggest a link can be very high, especially when they have indicated that they are not willing to engage. This is similar to the models of costly consent in Myerson (1991); Gilles et al. (2012); Gilles (2021) where a network is formed among dyads that consent but an individual who expresses a wish to form a link with an agent who doesn't express this wish can face a cost. In this way, we can relate the individual choice to engage in period 1 with the support network that emerges in period 2. In particular, the number of agents n in the support network will be equal to the number of agents who engage in period 1 i.e., $n = \sum_{i \in N} e_i^*$. As a result, it will depend on the equilibrium in period 1 as that affects engagement decisions e_i^*. In other words, interventions that alter beliefs about the social norm i.e., about the proportion of others who are willing to engage can also instigate changes in the size of the advice network. The size of the support network is also then, by definition, constrained by the misperception of the social norm caused due to the friendship paradox.

G.4 Proof for Lemma 1

Consider the distribution function of the probability mass function $\widetilde{P}(d)$

$$\widetilde{\mathbf{P}}(\bar{\mathbf{d}}) = \sum_{d=1}^{d=\bar{d}} \widetilde{P}(d)$$

We first prove that the CDF corresponding to P is first order stochastically dominated by the CDF corresponding to $\widetilde{P}$. Consider $\bar{d} \leq E[d]$. In this case, $\sum_{d=1}^{d=\bar{d}} \widetilde{P}(d) = \sum_{d=1}^{d=\bar{d}} \frac{d}{E[d]} P(d) \leq \sum_{d=1}^{d=\bar{d}} P(d)$ since $\frac{d}{E(d)} \leq 1$.

When $\bar{d} > E[d]$, then $\sum_{d=1}^{d=\bar{d}} \widetilde{P}(d) = 1 - \sum_{\bar{d}}^{d=d_{max}} \widetilde{P}(d)$. The second term can be written as $\sum_{\bar{d}}^{d=d_{max}} \widetilde{P}(d) = \sum_{\bar{d}}^{d=d_{max}} \frac{d}{E(d)} P(d) > \sum_{\bar{d}}^{d=d_{max}} P(d)$ since $\frac{d}{E(d)} > 1$. So, $\sum_{d=1}^{d=\bar{d}} \widetilde{P}(d) < \sum_{d=1}^{d=\bar{d}} P(d)$ in this case as well.

Hence, $\widetilde{\mathbf{P}}(\bar{\mathbf{d}})$ first order stochastically dominates $\mathbf{P}(\bar{\mathbf{d}})$.

We can rewrite

$$\sum_d \widetilde{P}(d) H(c(d,a)) = H(c(d_{max},a)) - \sum_{d=1}^{d=d_{max}-1} \widetilde{\mathbf{P}}(\mathbf{d})(H(c(d+1,a)) - H(c(d,a)))$$

and similarly rewrite

$$\sum_d P(d) H(c(d,a)) = H(c(d_{max},a)) - \sum_{d=1}^{d=d_{max}-1} \mathbf{P}(\mathbf{d})(H(c(d+1,a)) - H(c(d,a)))$$

Given that H is weakly increasing in degree d and $\widetilde{\mathbf{P}}(\mathbf{d}) < \mathbf{P}(\mathbf{d})$, this implies that-

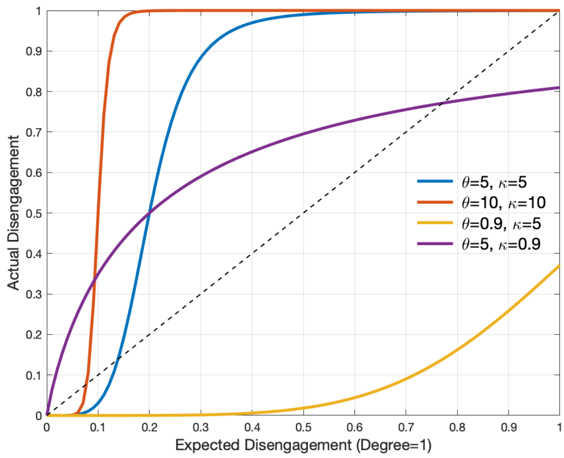
$$\sum_d \widetilde{P}(d) H(c(d,a)) \geq \sum_d P(d) H(c(d,a))$$

G.5 Proof for Proposition 1

If c and consequently H is strictly increasing in degree, Lemma 1 implies that $\widetilde{E}[H(c(d,z'))] > E[H(c(d,z')] \; \forall z' \in (0,1]$. This means that for any fixed point z that satisfies Equation 2, $\widetilde{E}[z] > z$. Given that we have assumed H to be such that there are three equilibria out of which 1 is unstable, this implies that the fixed points a_l and a_h solving equation 2 will not solve equation 3 and (b) the fixed points $\widetilde{a}_h$ and $\widetilde{a}_l$ that solve 3 will be such that $\widetilde{a}_l < a_l$ and $\widetilde{a}_h > a_h$. The intuition for Case 2 is analogous.
Click to go back.

H Structural Estimation

Figure H.1: Different shapes of $H(\theta, \kappa)$ for various values of θ and κ



Notes: The figure plots the benefit distribution H distributed log-logistic with shape parameter κ and strategic complementarity parameter θ. Degree is normalised to be 1 in this example.

Table H.1: Structural Estimation Prediction Checks for various measures of Engagement

Variable	Objective Function	Predicted Mean (Treated)	Actual Mean (Treated)	Predicted Mean (Control)	Actual Mean (Control)
Listening Service	0.0005	0.309	0.299	0.440	0.460
Savings Groups	0.0012	0.217	0.187	0.318	0.333
Community Engagement	0.0001	0.201	0.200	0.332	0.338

Notes: This table shows the model fit for the estimated values of θ^* and κ^* for three different choice of outcomes. The objective function is computed at the estimated parameters and is equal to the sum of the squared percentage gap between the mean of the treatment and control groups predicted by the model and in the data respectively.

H.1 Constructing Confidence Sets using Quasi MCMC Methods

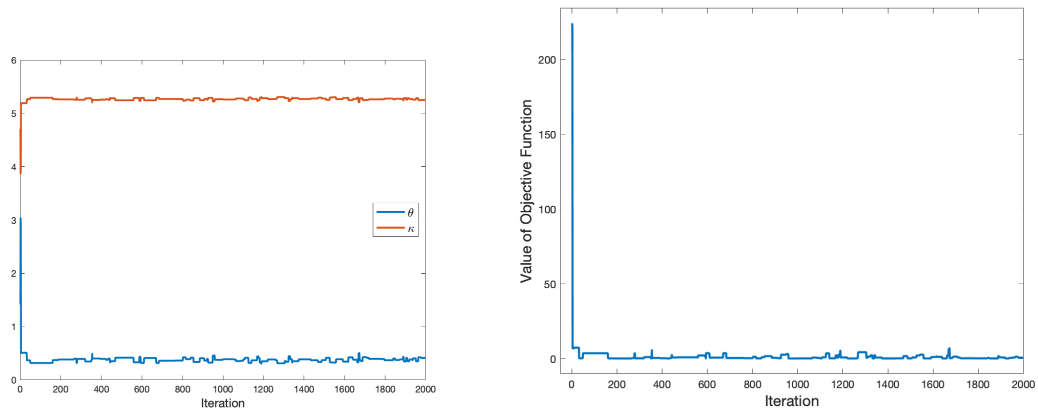
We estimate the confidence intervals for θ an κ as follows. We first compute the covariance matrix for the moment conditions (i.e., the mean of the treatment group and the control group) using the variance-covariance matrix returned via the OLS regression on treatment and control indicators and no constant. We denote this covariance matrix as $\hat{cov}$. We redefine our weighted loss function in terms of $\hat{cov}$ by placing weights on the moments using the inverse of the covariance matrix. Next, we compute the loss function for a perturbed value of θ and κ around θ^* and κ^* computed via pattern-search. We use a step-size of 0.001 and compute a 2x2 Jacobian matrix indicating how much the two moment conditions change due to a change in each parameter. This Jacobian matrix J is used to construct another covariance matrix cov_{shock} as follows: $cov_{shock} = inv(J' * inv() * J)$.

Now, we proceed with the MCMC chain. We first select an intial value of θ_0 and κ_0 that is equal to θ^* and κ^* plus an additional shock coming from a multivariate normal distribution whose variance-covariance matrix is specified by cov_{shock}. We compute the weighted loss function for this chosen value of θ and κ. Then, select another value θ_1 and κ_1 which is equal to θ and κ plus another shock. We compute the loss function again. Now, we choose the value θ_1 and κ_1 if the loss function is lower than the previous one. If not, the new θ_1 and κ_1 are still chosen with a probability α_1 where α_t is specified as follows:

$$\alpha_t = e^{\frac{1}{2}(Loss_t - Loss_{t-1})}$$

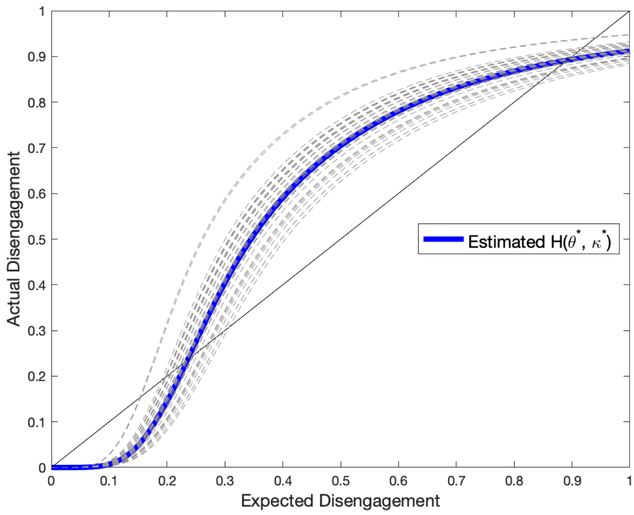
Crucially, α_t is close to zero if $Loss_t$ is very high compared to $Loss_{t-1}$ but is positive if the difference is not very large. If θ_1 and κ_1 are selected, then θ_2 and κ_2 are constructed as before by adding a shock to θ_1 and κ_1. If not, θ_2 and κ_2 are constructed by adding a shock to θ_0 and κ_0. The process continues until the chain converges in distribution. Once the chain converges, we draw θ and κ from the converged distribution to construct the confidence set. We find that the chain converges quickly. This is shown in Figures H.2 that plot the chosen values in each iteration and the loss associated with each values respectively. We draw θ and κ from the last 500 iterations and plot the best response curve for these values in Figure H.3

Figure H.2: Convergence of θ and κ over 2000 iterations



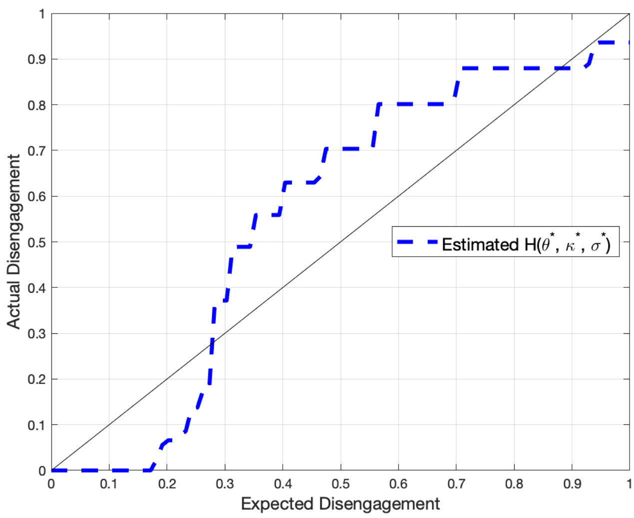
Notes: The figure on the left plots θ and κ over various iterations. In each iteration t, a candidate value of θ and κ is chosen with probability 1 if $Loss_t$ is less than $Loss_{t-1}$ and with a probability $e^{-0.5(Loss_t - Loss_{t-1})}$ if it is higher than before. The figure on the right plots the objective function over various iterations.

Figure H.3: Actual and Estimated Disengagement (Confidence Set)



Notes: The figure plots actual disengagement a and expected disengagement $\sum_d \widetilde{P}(d)(H(c(d,a)))$ at the estimated θ^*, κ^* (in blue) and at various θ and κ pairs computed using Quasi-MCMC methods.

Figure H.4: Actual and Estimated Disengagement (assuming a Logistic Distribution)



Notes: The figure plots actual disengagement a and expected disengagement $\sum_d \widetilde{P}(d)(H(c(d,a)))$ at the estimated θ^*, κ^*, and σ^* in blue. Here, we assume a logistic distribution for H during the estimation so that $H(c(d,a)) = 1/(1+e^{((-\theta * d * a)+\kappa))/\sigma})$.