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

Ronak Jain[*] Vatsal Khandelwal [†]

July 9, 2024

Abstract

Inaccurate beliefs about social norms can reduce useful social interactions and adversely affect an individual's ability to deal with negative shocks. We implement a randomized controlled trial with low-income workers in urban India who lack access to formal financial and healthcare support. We find that the majority of individuals underestimate their community's willingness to engage in dialogue around financial and mental health concerns. Belief correction leads to a large increase in the demand for network-based assistance. We show that the effects are driven by a reduction in the perceived costs of violating social norms arising due to concerns around reputation and insensitivity. We structurally estimate a network diffusion model and predict that our belief correction intervention will not lead to a shift in equilibrium engagement. Consistent with this, 2 years later, we find that the average beliefs of those exposed to the intervention are significantly more optimistic but still lower than the information delivered in the experiment. We compute the strength of counterfactual interventions needed to generate a sustained effect and find that belief correction can be used to 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.

[*]Harvard University. Email: rjain@g.harvard.edu

[†]University of Oxford. Email: vatsal.khandelwal@economics.ox.ac.uk

[1]We are very grateful to Stefan Dercon, Marcel Fafchamps, Simon Quinn, and Alexander Teytelboym for their guidance, support, and useful 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. Many thanks to Diana Beltekian, Binta Zahra Diop, Luke Milsom, Bassel Tarbush, Verena Weidemann, and Hannah Zillessen for their advice and feedback on previous versions of the paper and to seminar participants at various workshops at Harvard University and the University of Oxford for comments and discussions. Many thanks also 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 are deeply grateful to the NGO team, especially Bharati Chaturvedi, Ishani Palandukar, Shruti Sinha, and Balmukund Kumar for their collaboration. We also extend sincere thanks to our field team, especially Preetham Rodrigues and Abdul Latheef. P at the Inchara Foundation and to 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 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.

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. 2018, Angelucci et al. 2018, Munshi & Rosenzweig 2016, Munshi 2014, Banerjee et al. 2013, Munshi 2011, De Weerdt & Dercon 2006, Fafchamps & 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. This lack of demand can arise due to concerns about violating social norms. For example, individuals may hesitate to discuss financial matters if they believe that doing so can hurt their reputation. Such beliefs can prevent them from requesting financial support and consequently affect the quality of risk-sharing that takes place within the community. However, beliefs about others may not always be accurate (Bursztyn & 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 work with a sample of low-income informal sector workers and their families living in slums in urban India. Slums, characterized by poor living conditions, risk, low human capital, and limited policy-maker attention, can act as poverty traps (Marx et al. 2013).[1] Given that 860 million people live in such conditions globally (UN-Habitat 2013), it is crucial to understand how to strengthen social networks so that they can function as effective social safety nets. 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 implement multiple rounds of data collection. This includes a baseline survey with about 350 individuals in 2020 (henceforth called the "main sample"), a randomized controlled trial and an endline survey with 180 individuals in 2021 (henceforth called the "experiment sample"), and a replication exercise and survey experiments with 800 individuals in 2023 (henceforth called the "additional sample").[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 engage with their peers around these topics. Particularly, 63% reveal a willingness to engage with their community but 68% underestimate the willingness of others to do so.[3] Not only this, those who have pessimistic

[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 replication sample allows us to ensure that the results are not specific to the small sample or the pandemic.

[3]We run experiments with our additional sample to show that social desirability bias is not driving the high willingness to engage that we measure around these topics.

beliefs about others are less likely to express a willingness to engage, have a lower intensity of dialogue around these topics, and have fewer connections in their advice-taking networks relative to their total connections in the overall social network.[4]

Motivated by these patterns, we implemented a randomized control trial in which we provided the treatment group with accurate information about the baseline sample's willingness to engage in dialogue around mental health and financial concerns.[5]

The experiment yields three main results. First, the intervention significantly increases an individual's engagement with their community. Treated participants are 15 percentage points more likely to sign up for a potential savings group with their community members. Their willingness to sign up to train to become a listening volunteer to hear the anxieties of the members of their community also increases by 16 percentage points.[6] Moreover, we find that individuals in the treated group make a 29% higher contribution towards setting up this service. In addition, we also detect evidence of significant positive spillovers on willingness to engage in dialogue around physical health concerns that are also stigmatized in this setting. These results suggest that the correction of misperceptions can increase both the demand for network interactions and payments to set up informal avenues to interact.

Second, we find that the demand for formal sources of assistance also depends on the beliefs individuals hold about the social norm. For example, we document significant positive treatment effects on an individual's willingness to speak to a doctor about mental health concerns. Third, we do not find any change in the demand for additional information about mental health or in financial self-efficacy (i.e. whether individuals believe they can manage their finances well in the coming months). The high treatment effect on the demand for savings groups together with the negligible effect on self-efficacy indicate that informal channels of borrowing may be weak or non-existent.

Two years later, we find that 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 conduct a larger-scale replication exercise with 800 individuals and continue to detect a causal link between beliefs about others and own engagement.[7]

Going further, we use survey evidence and implement additional experiments to disentangle mechanisms. We find that the treatment reduces the perceived costs of violating the social norm. In particular, we implement a hypothetical choice experiment to uncover whether

[4]Beliefs about peers' willingness to engage are also correlated with whether individuals have volatile consumption and have faced consumption crises in the last six months. This comes from the additional sample for which these consumption outcomes were measured.

[5]We also implement a participant prediction exercise with another sample in the same setting to show that this is indeed perceived to be a strong information shock. The majority predict that this information, if delivered, would increase engagement with the network. 38% predict that this increase will be large.

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

[7]The mechanisms behind this treatment effect are slightly different from the main RCT as the information provided to treated participants in 2023 was about other communities in 2021 as opposed to their own community in 2023. More details are provided in Section 4.7 and Section M in the appendix.

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 characteristics of the hypothetical advisor. Using data on acceptances/rejections of about 4740 links, we find that the reputation and interaction-related costs are the primary concerns rather than signaling. Moreover, we find that these costs significantly impact those who are more pessimistic about others' willingness to engage. This shows that we can improve outcomes by implementing a belief correction intervention that targets precisely the subgroup for whom such concerns are active. Finally, we rule out methodological concerns such as social desirability and experimenter demand by implementing a list experiment and an experiment that randomly hides or reveals the answers to the experimenter. We also present evidence against alternative mechanisms such as updating beliefs about the benefits of interacting, beliefs about the incidence of stress in the community, and social pressure.

Next, we adapt and structurally estimate a network model (Jackson & Yariv 2007, Jackson 2019) 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.[8] 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 and the extent to which they violate the endogenous social norm i.e. the proportion of others who are not willing to engage.

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 (Feld 1991, Jackson 2019), are more connected than they are and face a higher cost of violating the norm.[9] 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. Importantly, these effects depend on the shape of the dynamic best response curve. This is because the process is recursive — actions at time $t+1$ depend on beliefs at time t which in turn depend on actions at time $t-1$. The intervention will only have short-run effects if this process reverts to the pre-intervention equilibrium and long-run effects if this process leads to a new equilibrium.

We use simulated method of moments along with an equilibrium selection criterion to estimate the model by leveraging the random variation induced by the RCT. We use Quasi

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

[9] We provide empirical evidence to support the assumptions of the model that generate this inaccuracy in beliefs about the social norm.

MCMC methods (Chen et al. 2018) to compute confidence intervals for our parameter estimates. 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 2 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. How strong would alternative interventions have to be 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 information/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 signaling their type while asking for financial support).

We find that these alternative interventions have to be very strong for them to have long-run effects of a similar magnitude compared to the short-run effects of belief correction. For example, 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 are also infeasible to implement due to limited policy attention and funding. However, as the evidence from the RCT suggests, belief correction can be used to generate both the demand and funding for such costlier interventions.

We make three contributions to existing literature. First, we contribute to the literature on social networks (eg: Breza et al. (2019), Möbius & Rozenblat (2016), Breza (2016), Munshi (2014), Jackson et al. (2012), Munshi (2011), Karlan et al. (2009), Fafchamps & Gubert (2007)) by establishing a link between beliefs about social norms and the demand for network interactions i.e. how misperceptions about social norms can lead to dysfunctional networks. Existing work has documented concerns such as lack of trust, limited commitment, and enforcement that can adversely affect useful network interactions (Ambrus & Elliott 2021, Möbius & Rozenblat 2016, Jackson et al. 2012, Karlan et al. 2009, Fafchamps & Gubert 2007, Ligon et al. 2002). Such concerns arise if individuals demand support from their social ties in the first place. Taking a step back, we show that individuals may not even demand support from their social ties due to inaccurate beliefs about social norms.[10] Further, the bulk of this literature has almost exclusively focused on networks in rural settings (eg: Morten (2019), Banerjee et al. (2018), Munshi & Rosenzweig (2016), Banerjee et al. (2013), Munshi & Rosenzweig (2009)) where caste, ethnic, or religious affiliation typically provide an

[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 & Sarangi 2010, Van de Rijt & Buskens 2005, Jackson & Watts 2001, Jackson & Wolinsky 1996). We provide evidence to show that network interactions can indeed be costly and these costs can prevent the existence of improving paths from empty to non-empty, efficient networks. We show that most individuals overestimate these costs and are willing to pay to interact with others once the inaccuracies are corrected.

institutionalized platform for social interactions. By focusing on an urban setting, our results yield insights into contexts that lack an institution that facilitates network interactions.

Second, we contribute to the large literature on inaccurate beliefs and information provision experiments designed to correct misperceptions (eg: Delavande (2023), Bursztyn & Yang (2022), Haaland et al. (2020), Bursztyn et al. (2020), Jackson (2019), Perkins et al. (2005, 1999)). We add to this literature by using a theoretical model to show how beliefs can be inaccurate in the first place. We are also able to predict the equilibrium impact of our belief correction intervention and evaluate its effectiveness compared to alternative policies that do not target beliefs, by structurally estimating the model. Additionally, using survey evidence and additional experiments, we are able to provide evidence on the mechanisms through which belief correction interventions generate positive treatment effects. Finally, we document misperceptions in a domain that has substantial implications on the design of policies around social protection i.e. willingness to interact with others on sensitive topics.

Third, recent studies show that behavioral concerns such as signaling or shame can reduce incentives to seek advice (Chandrasekhar et al. 2019, Banerjee et al. 2018) or that reputation costs can reduce incentives to share useful information (Chandrasekhar et al. 2022). We contribute to this literature by providing a potential solution to mitigate the detrimental effects of such concerns. In particular, we show that individuals can have inaccurate beliefs about others and beliefs can act as a malleable statistic that can be corrected to increase interactions, without having to address the underlying behavioral concerns using costly psycho-social interventions. We also show that addressing these concerns is possible via costlier interventions that can be financed by belief-correction interventions such as ours.

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 Section 4. Additional experiments and survey evidence to investigate mechanisms are presented in Section 5. In Section 6, 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 7. Section 8 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.[11] Globally, nearly 860 million people live in slums(UN-Habitat 2013) and about 15 million people in developing countries earn by sorting and recycling waste (Medina 2008). Our context is also 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

[11] We collaborated with an NGO, Chintan, that promotes awareness of the health and safety of these individuals and advocates for their rights.

two rounds of demographic surveys with our sample: a 'main' survey with 352 individuals across 14 locations in and around Delhi, India (in 2020) and an 'additional' survey with 791 individuals in 2 such locations (in 2023). Figure F.1 shows these locations.

We use data from both waves to describe the context.[12] Tables A.1 and A.2 show descriptive statistics for our main sample. The average age of participants is around 34 years, 35% are female, and around 67% are currently employed. Individuals take advice from around 3 peers on average — this is the average degree of the 'advice network' (i.e. the number of social ties with whom the respondent reports discussing personal concerns around mental well-being). They interact with around 4 peers on average in general including advice-taking, borrowing, lending, and working together — this is the average degree in the 'overall network' (i.e. the number of social ties with whom the respondent reports interacting with 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: Our sample has 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 F.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 faced consumption crises in the last 6 months – these include crises such as 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 out of the ones who do, 80% do not find it easy to take 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 A.1. 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 F.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

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

the degree distributions of the overall networks and advice-taking networks. We can observe that the overall network degree distribution first-order stochastically dominates to advice network degree distribution. As we will shortly discuss in Section 3, the gaps between these networks are higher for those who have severely pessimistic beliefs about their peers' 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 F.1.

3 Belief Elicitation and Validation

3.1 Belief Measurement

To provide an explanation for why there is low dialogue around mental health and financial concerns despite limited access to sources of assistance, we document inaccurate beliefs about the community's willingness to engage. We measure an individual's beliefs about their community's willingness to engage with finance/employment-related issues, mental health concerns, and physical health concerns; the latter is elicited to act as a point of reference.

We first informed individuals what we mean by physical and mental health to standardize measurement. Following this, we asked them about their own willingness to engage with others on physical health, mental health, and finance/employment-related issues respectively. Willingness to engage is explained to the participants as willingness to discuss ways to overcome concerns around these topics and how these concerns might be preventing them from achieving their goals.[14] This is to ensure that we capture meaningful engagement. Crucially, our measure of engagement does not assume a direction in terms of who is helping whom. Further, being willing to engage is a costly action in that it requires individuals to think about how their concerns might be preventing them and others from achieving their goals as opposed to simply being about checking in with their peers.

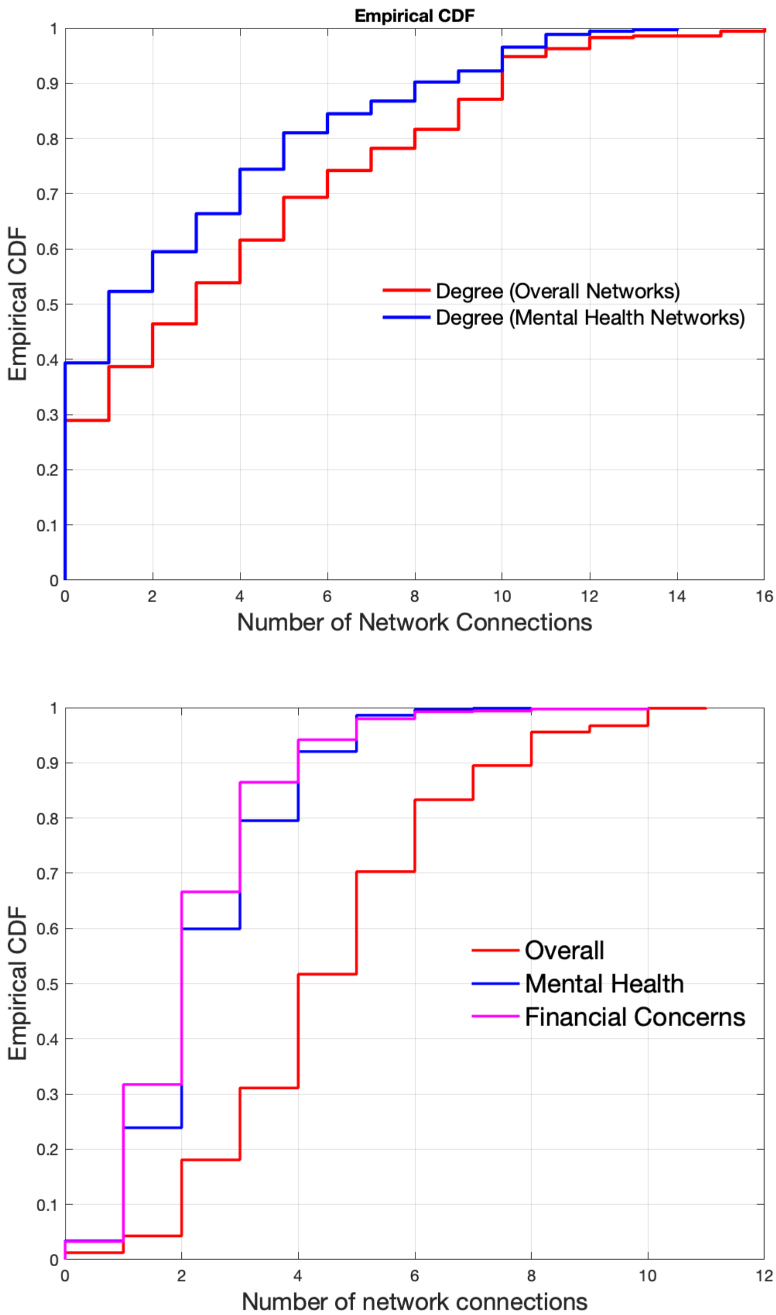
Next, we clarified to respondents what we meant by 'community'. This was defined in terms of their geographic location i.e. individuals living around the NGO centre and affiliated to the NGO were to be interpreted as a community.[15] After this, we elicited their beliefs about their community's willingness to engage with three matters – physical health, mental health, and financial issues. For example, to get at beliefs around engagement with financial concerns, we

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

[14] For example, in the case of financial concerns, we asked individuals: "Would you be willing to share and discuss financial and/or employment-related concerns with your friends, discuss ways to overcome them, and how these problems might be preventing you from achieving your goals?"

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

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



Notes: The figure on the top 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 bottom 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 ("borrowing network").

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) in case 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.[16]

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 feel often or very often 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 individuals among any 10 individuals in their community associated with the NGO would agree with the statement.[17]

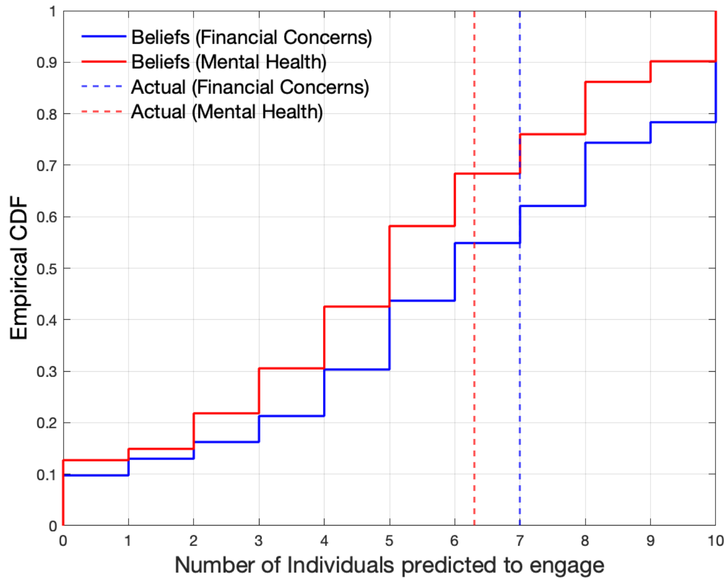
3.2 Evidence of Misperceptions

Table A.2 provides information on these 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 individual beliefs about their community's willingness to engage in dialogue around mental health and financial concerns. We can infer if an individual underestimates engagement on a particular topic by comparing the average willingness to engage in the sample and the individual's beliefs around the average willingness to engage. We find evidence of substantial misperceptions in individual beliefs about others' willingness to engage in a dialogue about mental health and financial concerns.

[16]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.

[17]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 figure plots 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.

The proportion of under-estimators by type of dialogue i.e. mental health, physical health, financial well-being, and stress is shown in Figure F.5. Around 68% of the individuals underestimate the percentage of individuals in their own community who are willing to discuss mental health-related issues. 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.

We also plot the average willingness to engage around these issues across all communities in Figure 2 to indicate the proportion of individuals who think that their community is less willing to interact than the actual willingness to interact across the entire sample. The difference between an individual's prediction for their community, the community level actual willingness, and the sample level willingness for mental health concerns and financial concerns is plotted in Figure F.6 respectively. This shows that even though the majority of individuals are under-estimators, there is substantial heterogeneity in the extent to which they underestimate their willingness to engage. Before proceeding further, we validate these measures of beliefs using survey-based evidence. We also address concerns around social desirability using evidence from the survey in the next section and using data from additional survey experiments in Section 5.

3.3 Validation of Elicited Beliefs

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

3.3.1 Correlations with Willingness to Engage and Beliefs about other Topics

First, we find that individual beliefs about their community's willingness to engage around various topics are correlated with each other and with their own willingness to engage. Table F.2 reports correlations between beliefs about community engagement with mental, financial, and physical health and the individual's own willingness to engage with these topics. The table shows that beliefs about the community's willingness to engage are positively correlated with each other along different topics of engagement. This implies that optimism on one dimension implies optimism around other dimensions as well. Further, beliefs are additionally correlated with own willingness to engage as one would expect. It is important to note that there does not exist a one-to-one correspondence between beliefs along various dimensions even though the correlations are large and positive. This shows that individuals were not equally pessimistic or optimistic about their community's eagerness to engage with them along various topics of dialogue. Moreover, 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.

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 F.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. Table F.4 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 suggesting that pessimistic beliefs are associated with the structure of the observed social networks i.e. more pessimistic individuals have fewer connections in their advice networks relative to their overall network links (Table F.3). These degree gaps are measured by subtracting an individual's degree in their advice network from their degree in their overall network and computing this difference as a proportion of their degree in the overall network. We find that degree gaps are significantly larger for individuals categorised 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 F.6.

3.3.5 Replication in 2023

Finally, we re-measured beliefs in 2023 to ensure that the misperceptions are not specific to the COVID-19 pandemic. We find even higher levels of underestimation of community's willingness to engage and a high willingness to engage around both financial concerns and mental well-being. 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.

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 replication exercise with a larger sample of around 800 individuals in 2023.[18]

The treatment group received two pieces of information: (i) the true average sample willingness to engage in dialogue about mental health and (ii) information about the sample's willingness to engage in a dialogue about financial well-being.[19] 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."*[20] [21]

[18]We discuss this in more detail in Section M in the appendix.

[19]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.

[20]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.

[21]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.

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

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 especially 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.[22]

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 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 G.1 and G.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 centre. 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 (discussed in the outcomes list) since we assume that the hypotheses under which the treatment affects the outcomes in different families are inherently different. The q-values are reported in each table for the respective family of outcomes. 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 entire 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.

[22]This includes all individuals in 2023 who were not included in the replication treatment.

4.4 Results

4.4.1 Effects on Network Engagement

First, as Table B.1 shows, we find a 14 percentage point increase in community engagement (significant at 1%).[23] 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)[24], (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 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 (a 29% increase compared to control) significant at 5%.[25] 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%). 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.

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. The robustness exercise in the appendix in Table H.12 shows that the effect of the treatment on these various outcomes is still significant when unbalanced controls are included. Moreover, as we will shortly discuss, we also find in our replication exercise in 2023 that those who are pessimistic about their peers and receive information about others' willingness to engage contribute significantly more for setting up savings groups and listening services.

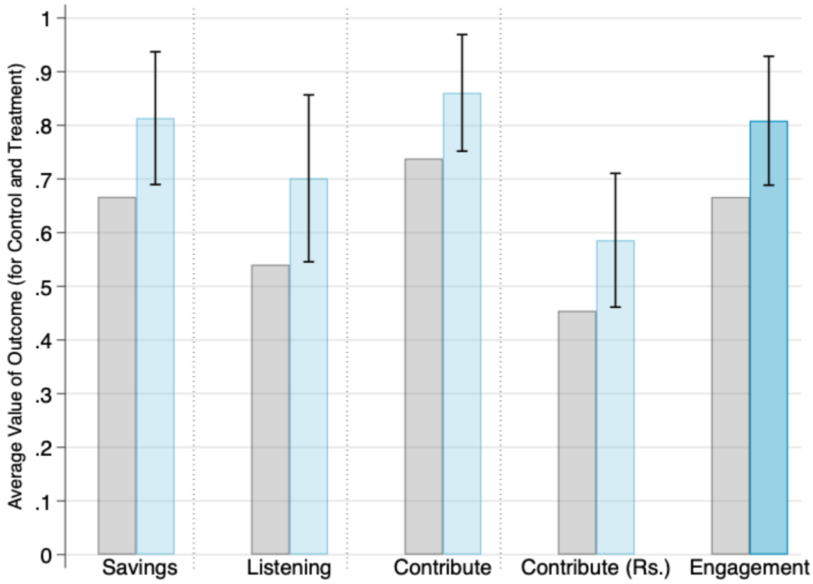
---

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

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

[25] 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.

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, normalised to be between 0 and 1.

4.4.2 Effects on Demand for Formal Support

Table B.2 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 found that treatment reduced participation in an experimenter-run depression scoring (using standard questions from PHQ-9) and the likelihood of wanting to listen to helpline numbers. We find that these results are robust to the inclusion of unbalanced controls. The impact of these treatments 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.

4.4.3 Effects on Other Types of Dialogue

Table B.3 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 as well as shown in the appendix in Table H.17.

4.4.4 Effects on the Demand for Additional Information

Table H.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 participants agree 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. Discussing this with the NGO, we realized that this might have been because individuals may have interpreted this session as one of the regular training sessions (unrelated to mental health) that the NGO organizes with them which is why the stigma associated with attending information sessions in general would be perceived to be low.

4.4.5 Effects on Other Outcomes

We find no impacts on other outcomes measuring financial self-efficacy and concerns around stigma. These results are reported in the appendix in tables H.2 and H.3. 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 favour-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 them and willingness to pay to set them up.

4.5 Heterogeneity

Before proceeding to discuss the replication exercise and additional experiments for mechanisms, we analyze heterogeneous effects by baseline willingness to engage, dialogue intensity, number of network connections, stigma, and beliefs. We will also discuss heterogeneity by baseline beliefs about stress in the section on mechanisms. The results are shown in Tables H.4-H.11 in the appendix. We highlight the following key findings.

First, Tables H.4 and H.5 show 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.[26] This suggests that the intervention, as expected, helps those who need it most. This is further corroborated by the results in Tables H.6 and H.7 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 for those whose dialogue around mental and financial well-being was less than or equal to the median in the baseline.

Next, Table H.8 shows heterogeneity in treatment effects by an individual's degree in their 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.

Third, Table H.9 shows that conditional on saying 'Yes' to financially contributing for 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 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 express stigma in the control group. While receiving optimistic beliefs about community engagement may not address stigma, this suggests it can significantly increase community engagement among those who exhibit it. Note that this might also be because the stigma that is 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. Once the information shock reduces pessimism about others' willingness to engage, it might also increase their inclination to engage and contribute towards doing so.

Finally, we also check for heterogeneity by whether individuals underestimate community

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

engagement in Tables H.10 and H.11. 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.

4.6 2-week and 2-year Follow up

4.6.1 Effects after 2 weeks

We conducted a follow-up survey more than 2 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 L.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. We test balance on 44 baseline variables as before and find that the sample is unbalanced for 2 variables.[27] We also show that attrition in this survey is not correlated with treatment status and not correlated with various baseline characteristics. However, it is important to note that we do not have information on the participant's or their peers' exposure to COVID-19 and more importantly, information about their location if they had migrated out of Delhi. Differential exposure of the control or treatment group to COVID-induced risk and migration decisions can affect the validity of these results. As a result, we do not proceed to discuss these results or the underlying mechanisms in more detail here and present them in the appendix in Section L.

4.6.2 Effects after 2 years

In addition to the 2-week follow-up, we also 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 was a 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.[28] We cannot track and merge individuals across the two waves due to logistical reasons but can identify participants and those exposed to the information using self-reported measures. We find that out of a total of 469 individuals, 60 report being previous participants, 37 report having heard information about their community's views on talking about mental health and finances shared in the main experiment, and 24 were previous participants who report having heard this information. The remainder had neither participated nor heard about the previous experiment.

[27] We also present robustness checks controlling for these unbalanced variables.

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

For each measure above, we study the effect of being in that category on own beliefs, dialogue intensity, willingness to engage, and volatility of consumption in 2023. The above measures of capturing exposure to the treatment are imperfect in that (a) they combine individuals in treatment, control, and spillover conditions and (b) they may capture differences 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.

The results are reported in Tables C.1 - C.4. We find evidence that individuals who report being in the above categories 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 is 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.

4.7 Replication Exercise

Before proceeding, we briefly present the results of replication exercises conducted with around 800 individuals in 2 NGO centres in 2023. This exercise was conducted to ensure that (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 own engagement with the network, are not specific to the small sample size or to the timing when the experiment was conducted.

As previously discussed, these baseline patterns of low dialogue and inaccurate beliefs also hold in this larger sample in 2023. In addition to this, replication of a weaker treatment (i.e. where we provided information about others' willingness to engage in 2021) continues to have a significant causal impact on willingness to engage. The direction of the effect and underlying mechanisms are different and are discussed in detail in the appendix in Section M. We also 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.

5 Mechanisms

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

5.1.1 Network Prediction Experiment

We first present the result of an additional experiment run with around 800 waste-pickers and their family members across 2 NGO centres in New Delhi in 2023. 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 in their community, (2) whether A has contacts and knows people who do private jobs, and (3) whether A has attended a mental health sensitivity training organised by the NGO to talk sensitively about these issues. These three binary characteristics are varied to test the impact of the following costs respectively: (1) whether asking for support from A can impose a reputational cost if A is network central and can gossip, (2) Whether asking for support from A can impose a signalling cost if A can make negative inferences about B's type and be less likely to tell B about jobs or recommend them, and (3) Whether asking for support from A can impose an interaction cost in that A is not sensitive and might mock/mistreat B if B approaches them. For a randomly chosen vector of characteristics, the respondent was asked to predict whether B would approach A for mental health-related and financial support respectively. We varied the vector 3 times for each respondent thereby giving us about 2370 predictions for each randomly chosen vector of characteristics.

Crucially, whenever individuals rejected a link, they were also asked for the reason why they thought B would not approach A for support. This is to understand how they interpreted the presented characteristic. For example, if an individual reports that B would not seek financial support from A if A is very central in the network, it could be because they worry about the impact of A gossiping about them to several of their peers or it could be that they might think A would not have time for them given that they are so connected already. We ask individuals about these reasons to get a clearer idea about why they rejected a link. Finally, this experiment was conducted before any selected individuals were provided with any information about their community to prevent any contamination.

We use this experiment to causally identify the impact of various costs that can reduce network interactions in this setting. We find that 23% of links are rejected in the case of mental health-related support and 24% in the case of financial support. Given that this is a setting with very low dialogue and advice-taking around these issues, it is the hypothetical nature of the exercise that results in fewer rejections than expected. We find that a surveyed individual i is 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, as i becomes more and more pessimistic about their community's willingness to engage with them. In other words, individuals who think

A's characteristics such as their centrality and attendance of sensitivity training are likely to be the reason why B would not form a link with A are those who are themselves more pessimistic about the community's willingness to engage. This pattern holds for both mental health and finance-related support.[29]

Why are individuals who are more pessimistic about the social norm of engaging more likely to reject these links? Interestingly, and contrary to our expectations, we find that attendance at the sensitivity training acts as a negative signal about A in that they needed to be trained to act sensitively. 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, signalling for jobs is a dominant concern for many individuals in this setting. In particular, we find that 30% of individuals who reject 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.

The evidence from this hypothetical network formation experiment suggests that the reputational 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 costs are significantly higher for those who are more pessimistic about the social norm. While these actual costs cannot be reduced without alternative policies, the results from our randomised controlled trial show that economic outcomes can still be improved via a belief correction intervention as it targets precisely the subgroup for whom these costs are dominant.

As a result, we posit that the intervention works because it shifts individual beliefs about others' willingness to engage, which as we have seen, acts as a proxy for their concerns about 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".

We also use additional experiments and survey-based evidence to rule out alternative mechanisms. These are discussed in the next subsection.

5.2 Alternative Mechanisms

5.2.1 Social-desirability and Experimenter Demand

It could be the case that individuals are more likely to report a willingness to engage to the experimenter in the baseline and following the treatment to appear socially desirable or due to experimenter demand. We use survey evidence from the RCT run in 2021 and two additional experiments with around 800 individuals run in 2023 to show that social desirability and experimenter demand are not likely to be driving our results. We first present evidence from the two experiments that help us detect whether social desirability is a concern

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

at the level of the sample and the individual respectively.

Experiment 1: List Experiment

In this experiment, we randomly divided individuals into two treatments and one control group. The control group receives 3 statements and is asked to report how many of these opinions they agree with. The treatment groups are given the same 3 opinions as the control and in addition to this, either 1 or 2 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. 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. *(Only to Treatment Group 1 and 2) Individuals should take time to listen to the mental health concerns of their peers.*
5. *(Only to Treatment Group 2) Individuals should take time to listen to the employment or money-related concerns of their peers.*

If individuals only reported favourable views about engaging with peers around these topics due to experimenter demand or social desirability, the experiment provides them with a platform to mask their true opinions. In that case, there should be no difference between the number of statements that the control and treatment groups agree with. We find the opposite. The treated groups are significantly more likely to agree with more statements than the control group. In particular, 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. This experiment helps reassure us that on average, the high willingness to engage observed in the baseline sample must not be due to social desirability. We run another experiment to rule this out at the individual level.

Experiment 2: Increasing the Distance between the Enumerator and Respondent and Cross-Randomisation with the Information Treatment

In this experiment, we ask the additional sample of about 800 respondents whether they are willing to sign up for savings groups and listening services respectively. For a randomly chosen half of the participants, the enumerator asked the question and entered the response as usual. However, for the other half, we increased the distance between the enumerator and the respondent by informing respondents that they could answer these two questions on their screen and the enumerator was not able to see their responses. Moreover, this random revelation was cross-randomised with the replication treatment where randomly chosen participants were provided information about others. If the responses to these outcomes are due to experimenter demand, we should detect a difference between those who were asked to tell the enumerator their response and those who filled it privately. Table I.2 shows the results. We find that the willingness to participate in these services is not dependent on the distance between the enumerator and the respondent.

Moreover, this does not have a significant effect when interacted with the replication information treatment either. In other words, the reduction of distance between the enumerator and respondent does not affect outcomes on average and does not affect outcomes selectively for the replication treatment and control group.

We now present additional survey-based evidence to further show that social desirability is not likely driving the baseline patterns or the treatment effects that we observe.

Additional Survey Evidence against Social Desirability.

First, it is important to note that during the 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. Alternatively, in this setting, it can also be the case that a substantial proportion of individuals in both groups do not consider expressing support for mental health-related concerns as socially desirable in the first place.

Second, we measure willingness to engage along three distinct dimensions and find that individuals do not always agree to engage and further, they do not always agree to engage on all three dimensions – 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, the choice to engage is correlated with baseline beliefs about peers' willingness to engage. If individuals reported a willingness to engage solely due to experimenter demand, it is not clear why this would be correlated with beliefs about peers.

Fourth, balance along several variables, suggests that there should not have been differential selection into the treatment group based on the individual's propensity to appear socially desirable. However, the treatment group might wish to appear more socially desirable than the control group after receiving information about their community's willingness to engage. The above experiment rules this out as we find no effect of a reduction in the enumerator-respondent distance on willingness to engage for either the treatment or control groups. Moreover, recall that we find that individuals in the treatment group are willing to make significantly higher financial contributions to fund a listening service for their community. Positive effects on this incentivised outcome imply that the treatment effect cannot entirely be attributed to cost-less signals to appear socially likeable.

Finally, if treated individuals reported a willingness to engage only to appear socially desirable, then it is not clear why 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 – a choice that we have shown is correlated with their beliefs. This is shown in Table H.4, H.6, and H.7.

5.2.2 Updating Beliefs about the Incidence of Stress.

It can be argued that the treatment leads individuals to positively update their beliefs about the incidence of mental health concerns in their community based on the information about willingness to engage that is provided to them. This can explain the improvement in the community engagement outcomes that we measure. However, we do not think that this is the case due to the following reasons.

First, we measured individual beliefs about stress in their community during the baseline and found that over 60% 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 in the first place as they would have otherwise engaged in dialogue to ameliorate this. It should not, therefore, be the case that the treatment informs individuals about the stress levels of their community and makes them realise that others require their assistance.

Additionally, as shown in Table I.3, 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. We would expect the interaction term to be significantly positive if that were to be the case.

On the other hand, it can also be argued, that the treatment makes individuals realise that others are not as stressed as they believed and can therefore make time to address their concerns. This is in line with the overestimation of stress that we detect during the baseline. Let us consider the extreme case in which when individuals are told that six out of any ten individuals are willing to engage with them, they believe that six out of any ten individuals are not stressed and can make time to engage with them. We re-define over-estimators of stress as individuals whose baseline beliefs about the proportion of stressed individuals in their community is above four. Under this interpretation, these individuals should benefit most from the treatment. We show in Table I.4 that the interaction term is in fact negative and insignificant. This allows us to rule this mechanism out as a potential explanation.

5.2.3 Altruism or Social Pressure.

It could also be the case that the treatment increases community engagement due to alternative mechanisms such as altruism or feeling pressured to be supportive towards the community after being informed about the community's willingness to engage. We do not expect this to be the main driver of our results. First, treated individuals make costly decisions in terms of time and money to engage with others which is unlikely in our setting purely out of social pressure, especially since they are told that their decisions will not be visible to others.

Second, we find in the baseline that there is low dialogue even though the vast majority of individuals overestimate levels of stress in their community. This provides evidence that pure altruism may not be the sole channel driving the results since it does not motivate individuals to engage with others in the status quo.

Third, if selfless reciprocation were the only channel through which engagement was improving i.e. individuals only engage for the benefit of others and not themselves, then it is not clear why treated individuals are significantly less likely than the control group to listen to helpline numbers that can benefit them. While 59% of the control group is willing to listen to helpline numbers, treatment reduces this by around 20 percentage points. Substitution between help from the community and help from outside the network is likely to explain this effect. In other words, treated individuals do not report a higher willingness to engage with the community without seeing a benefit in it for themselves. In that case, they would have demanded to listen to helpline numbers just as much as the control group.

Finally, we ask individuals if they would like their names to be included in the list of potential participants of an information session around mental health. Crucially, they are told that the list can be used to motivate others to participate. We do not find any significant differences between the treatment and control group in this outcome. This provides additional evidence that social pressure to conform or wanting to appear supportive in front of others may not be the relevant mechanism. If anything, the social visibility of an action may be preventing individuals from interacting as outlined earlier.

5.2.4 Updating Beliefs about Benefits of Interacting

Finally, it can also be the case that participants update their beliefs about how beneficial it is to interact with others after they receive information that the majority of their community is willing to engage, contrary to their priors. We do not think this is the main mechanism at work as the majority of participants reveal a high willingness to engage to the experimenter in the baseline. This is indicative that they are aware that there is a benefit to engaging with the network. That this willingness doesn't translate into 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 dialogue is low, such interactions are not uncommon and as we have outlined, those who are more connected are also correlated with having better consumption outcomes. It does not seem plausible then that those who do not interact are unaware of the potential benefits others in their community are receiving from interacting. To this effect, many respondents in our qualitative interviews also say that they are aware that it is important to interact with their community around these issues.

5.3 Participant Predictions

We briefly discuss the predictions that our participants made for the effect of the experiment and how they interpret our results. To this end, we asked the additional sample of around 800 waste-pickers and their families in 2023, how they think *others* would respond if they were provided information on average willingness to engage that was provided to the RCT sample. As discussed earlier, their predictions are in line with the treatment effects and suggest that the treatment was indeed an information shock. Around 38% of participants predict that there will be a large increase in willingness to participate in savings groups and listening services while approximately 26% think there will be a small increase. At the same time, around 36% predict that there will be no change (14%) or that it will decrease (22%).[30] While the majority of participants predict correctly, there exists a substantial number that underestimates the effect of the treatment. This might be, for example, because they underestimate the benefits that others receive from engaging.

Further, out of those who predict that there will be an increase in engagement, roughly 24% think that it will be because individuals feel comfortable discussing their concerns with others and wouldn't think that others would think of them poorly if they engage, 31% feel it will be because they will learn that others are not stressed and can make time, 29% feel it will happen because everyone will see a benefit in engaging as others also need help and 16% feel that individuals will engage due to social pressure. It is important to note that these explanations can be interrelated: for instance, individuals can feel comfortable engaging with others and feel that they won't be judged harshly for violating the norm if (a) they think others can make time and (b) if others also need help. Otherwise, there can be costs of violating the norm of low dialogue. So, when individuals are informed that the true willingness to engage is high, these perceived costs of violating the social norm are reduced.

Having said that, while these participant predictions are informative, they are likely plagued by the same misperceptions about others that lead individuals to have pessimistic beliefs in the first place. As a result, we do not rely on them to disentangle mechanisms. It is important to note, however, that the correction of misperceptions about social norms is one dimension via which we can increase engagement. There can exist misperceptions about other aspects such as the time that others have or how stressed others are that can also be corrected to increase engagement. These misperceptions might also get corrected naturally as individuals interact more. We do not delve deeper into this as it is outside the scope of the paper. However, we will now develop a framework to make relative comparisons between belief correction and alternative policies that can address some of the other systemic concerns that we outlined in this section.

6 Theoretical Framework

We now present a model to show why individuals might have inaccurate information about the social norm in the first place. The model will achieve three objectives. First, it allows us to understand why individuals can have inaccurate beliefs in the first place and the implications of these inaccuracies. Second, we will estimate the model to predict whether our treatment effects correspond to a persistent change in equilibrium engagement. Finally, we will run counterfactuals of alternative interventions that do not target beliefs. We will also be able to comment on the short-run and long-run impacts of our intervention and benchmark these estimates against counterfactual interventions. The model follows from Jackson & Yariv (2007) and Jackson (2019) and is adapted to our setting to study engagement with concerns around mental and financial well being.

[30] These percentages do not exclude those who were in the replication treatment or who had been exposed to information from the main experiment. Prediction patterns are very similar for these subsamples as well.

Let N be the set of individuals in a society connected to each other 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 that they have in this network where $d_i = \sum_j g_{ij}$. The distribution of degrees of all agents is denoted by $P(d)$. Individuals choose whether they wish to 'engage' which is modelled as a discrete action $e_i \in \{0, 1\}$. Engagement can be interpreted as an individual's choice to take a visible, discrete action such as the choice to participate in a savings group or to train and participate as a listening volunteer to hear the economic anxieties of other community members. Critically, we do not model the link-by-link discrete choice to interact with a specific peer —engagement is a decision taken at the level of the peer-neighbourhood and is assumed to be observed by all an individual's peers. We do not incorporate network formation as our focus is to explain how misperceptions about engagement can arise in an existing network. However, we extend the model and discuss the implications of network formation in the appendix in Section J.2.

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 i.e. $1 - E[e_{j:j\in N}]$ and their network degree d_i. We assume that c is weakly increasing in both its arguments, 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 $1 - H(c(d_i, 1 - E[e_{j:j\in N}]))$. Finally, we place restrictions on the information that is available to agents. We assume that agents only know their own degree d_i, benefits of engagement b_i, the distribution of benefits H, and the degree distribution in the network $P(d)$.

6.1 Empirical Evidence for Model Assumptions

Let us first consider how the costs of engagement depend on the expected disengagement in the network. These costs can include those associated with reputation, interaction, and signalling, as discussed in the previous section. Engagement can lead to gossip that can affect an individual's reputation. Individuals can also be met with insensitivity if they choose to engage. Finally, engagement can make others infer that individuals are a low type which can affect their ability to use network ties for benefits such as receiving information about jobs or receiving referrals. While we disentangle these costs empirically, we combine them here into a single term that depends on the proportion of others who are choosing to engage.

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 disengaging. Existing literature on social norms suggests 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).

Alternatively, costs can be rising in degree for other reasons. For example, more connected individuals may not have as much time to engage making it costlier for them to do so. We can be agnostic about which of these explanations is true. However, in line with the assumption, we find empirical evidence that more connected individuals are also those who may be less likely to engage. It is difficult to establish a causal link between degree centrality and engagement decisions due to the underlying endogeneity which is why these findings are only indicative.

First, we find that degree centrality is negatively correlated with wanting to participate in savings groups and signing up as a listening volunteer among our additional sample of 800 individuals. These correlations are shown in Figures J.1 and J.2 in the appendix. The correlation is weakly negative in the second case, but we would have expected it to be strongly positive if the standard strategic complementarities channel was in place i.e. if more connected individuals enjoyed greater benefits of engagement.

Second, in our hypothetical network formation experiment, we find that 42% of individuals do not think that advice-taking links will be formed when the advisor is more connected in the network because the advisor will not have time. This suggests, again, that more central agents might be more time-constrained and find it more costly to engage.

Third, as discussed earlier, we also find in the experiment that more connected agents respond more to the treatment. 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 disengagement around mental or financial well being, the coefficient on the interaction term is almost always negative.[31]

6.2 Equilibrium

Analogous to Jackson & Yariv (2007) where agent types are determined by their costs and degrees, agent types here are determined by their benefits and degrees. 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. $\hat{e}$)

[31] The coefficient is weakly positive (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 three cases when 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.

must be equal to the probability that a person with degree d would engage, summed across all degrees, weighted by the degree distribution. Now, as in Jackson (2019), 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 neighbourhood $N(i)$ who do not engage 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 may form estimates about it using what they observe among their peers. The distribution of the degree of an individual's neighbours is equal to $\widetilde{P}(d) = \frac{d}{E[d]} P(d)$ as per the friendship paradox (Feld 1991, Jackson 2019). Crucially, We know from Feld (1991), 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:

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

where $\tilde{e}$ denotes the equilibrium engagement when individuals use $\widetilde{P}(d)$ instead of $P(d)$. 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, we rewrite the equilibrium as follows.

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

Let us denote the probability that a randomly chosen peer disengages 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 disengages 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 disengagement 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 J.4.

The following proposition compares equilibria with and without the friendship paradox in the simple example where there exist three equilibria of disengagement out of which there is one stable equilibrium at zero and two stable and unstable 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 disengagement and a_h and a_l be the non-zero symmetric Bayes Nash equilibrium disengagement that solves equation 2 such that $a_h > a_l$. Let $E[a_{j:j\in N(i)}]$ be the expected disengagement 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 J.5 in the Appendix. Figure 4 illustrates this result. The proposition implies that in a society where there are three possible equilibrium values of disengagement out of which two are non-zero, one stable, and one unstable, the highest stable equilibrium disengagement 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 disengagement. 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.

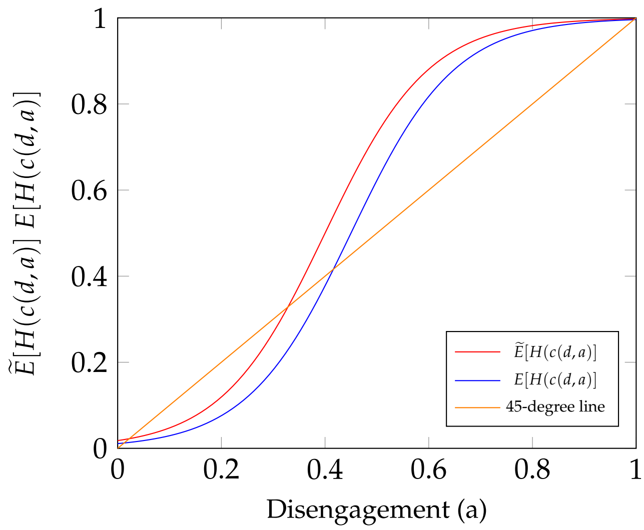


Figure 4: Illustration for Proposition 1 under standard assumptions about H.

6.3 Dynamics

Consider the dynamic version of Equation 3:[32]

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

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

What will happen when we provide individuals with information about the proportion of others in the community who are willing to engage? Provision of society-level willingness to engage will be useful in the case where individuals overestimate disengagement 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 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.

We now proceed to make a few assumptions to provide more structure to the model and prepare for estimation using the RCT data.

7 Structural Estimation

Now, we proceed with estimating the model using data from the randomised controlled trial. The purpose of this exercise is two-fold. First, we estimate the best-response curve $\sum_d \widetilde{P}(d)(H(c(d,a)))$ and compute the type and magnitude of equilibria that exist in the data. This will allow us to evaluate the scope for the success of belief correction in our setting and compute the equilibrium that we anticipate these communities are at. Moreover, it will allow us to comment on the long-run effects of our intervention by checking whether our credibly executed belief correction is large enough to lead to a long-run change in equilibrium beliefs. The shape of the estimated best response curve will allow us to infer this. Second, the exercise will allow us to compare the effects of our belief-shifting intervention with other alternative policy instruments that do not target beliefs. These include interventions that increase the perceived benefits of engagement or reduce the costs of engagement. For example, we will be able to compute the magnitude by which the individual benefits b_i would have to be increased to lead to the same short-run or long-run effects as those achieved by shifting beliefs about peers. We can use this to assess whether increasing an individual's perceived benefits of engaging by setting up savings groups or reducing their costs of engaging by setting up formal job referral services can be a more effective alternative to increase network engagement than implementing a belief-shifting intervention.

7.1 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. This directly follows from the assumptions of the model. The benefit of engagement b is drawn from a log-logistic distribution H. Recall that the probability that a randomly chosen peer of degree d disengages under equilibrium disengagement a is given by $H(c(d,a))$. We assume that $H(c(d,a)) = (1 + (\theta da/\alpha)^{-\kappa})^{-1}$ where we set $\alpha = 1$. We show below that this functional form allows for complete flexibility in the number and type of possible equilibria.[33] Figure K.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 will use the estimation exercise to recover estimates of our two parameters: θ i.e. the extent of complementarities in disengagement and κ i.e. the parameter which affects the shape of the distribution H of the benefit of engagement. As we have seen, both θ and κ affect every agent's predicted choice and also affect the magnitude, type, and number of equilibria.

7.2 Estimation Strategy

The estimation strategy proceeds 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.[34] 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 expected degree $E(d)$. Note that we do not compute degree distributions for each NGO centre due to small samples in each centre. There is no reason to believe that the networks should be systematically different across centres. As a result, we pool the sample to compute the degree distribution.

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:

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

[33] 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.

[34] 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.

All solutions to this equation are candidate equilibria that the individuals in the control group could use to decide whether to engage or not. In order 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), 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 $a_t^* = 0.4$, 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 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. 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 & Dennis Jr 2002). It 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. We perform this exercise by using the willingness to participate in the listening service as the measure of engagement. The reasons behind choosing this as the measure of engagement are discussed in the next subsection.

7.2.1 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

K.1. In short, 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.

7.3 Estimation Results

The first row of Table 1 shows the minimised 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 untargetted moments such as the standard deviation and skewness of engagement well in that the squared gap between the predicted and actual value in both cases is very low.

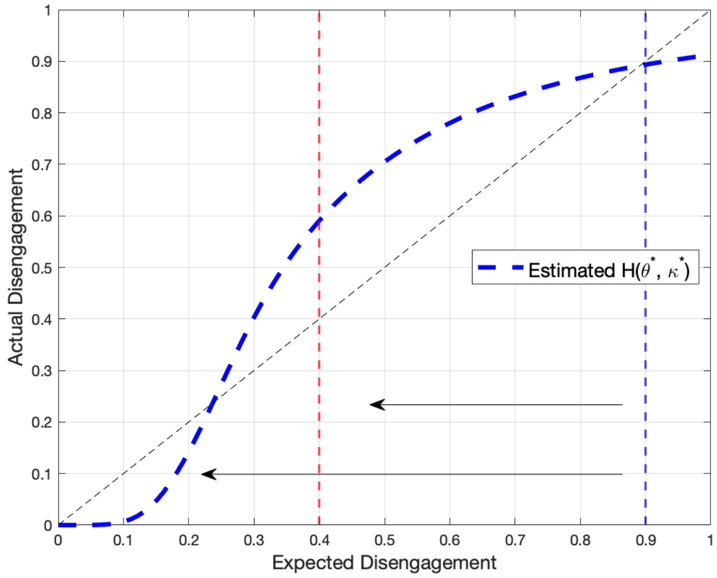
Table 1: Structural Estimation Model Fit

	Targetted Moments				Untargetted Moments	
Objective Function	Predicted Mean (Treated)	Actual Mean (Treated)	Predicted Mean (Control)	Actual Mean (Control)	Standard Deviation (Squared Gap)	Skeweness (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.

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 K.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 maximises the likelihood is the high, stable level of disengagement $a^* = 0.89$. This is expected in our setting as we measured low levels of dialogue around these issues in the baseline survey. These two 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. It is easy to 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 $a^* = 0.23$. However, the model predicts large short-run effects that we observe in our RCT effects – this is by design since the model is estimated by matching the predicted mean engagement in the treatment and control group with that in the data. For example, we have shown in our reduced form results that the willingness to participate in the listening service depends on beliefs about others. However, the structural estimation results emphasise not only that our sample is stuck at a stable equilibrium of high expected disengagement, but that shifting this equilibrium is possible via a belief-correction intervention only if the credibly delivered belief about the proportion of others willing to engage is greater than equal to 80%. This is a large increase and it may not always be possible to credibly implement such a change. We will now compare this to a counterfactual intervention that instead increases the benefits of engagement or reduces the individual's concerns about violating the social norm. Crucially, both of these counterfactuals do not target beliefs.

Before proceeding with the counterfactual analysis, it is important to note that we have used willingness to participate in the listening service as the main measure of engagement. This is because the intervention delivered concrete 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, say in the savings group. Since the estimation relies on the precise information that was delivered, it seems sensible to choose a measure of engagement for which the perceived participation in the next period is likely to shift by the same amount as the information delivered i.e. it is likely that treated individuals would believe that 60% of others are likely to participate in the listening service in the next period. As a result, we proceed by treating participation in the listening service as the main measure.

Robustness to other measures and distributional assumptions: We also 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 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 K.1 shows the parameter estimates and value of the objective function for the various outcomes.

Finally, before proceeding with the counterfactuals, we 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 K.4 in the appendix. The shape of the curve and the resultant equilibria are remarkably similar to that estimated under the log-logistic. We now proceed with the counterfactuals.

7.3.1 Counterfactual 1: Shifting the perceived benefits of interactions.

The top panel of Figure E.1 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 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 $b > 0$.

This shift in benefits can be achieved via a cash transfer that incentivises individuals to interact. More realistically and feasibly, this can be achieved, for example, via several information or awareness sessions informing individuals about the benefits of network interactions. 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 the 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 sudden reduction in disengagement to zero. In the absence of such a big push, it is worth noting that the belief-shifting intervention is a cheaper alternative that can even be used to finance such a big push as our empirical results have shown. 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 2 years later.

7.3.2 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 E.1. 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.

8 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. We find that generating permanent effects would require more costly interventions that increase the benefits of interactions or reduce how much individuals care about the social norm. 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. We learn that, in addition to generating large, positive short-run effects on network engagement, belief-correction interventions can also be used to increase individual demand for network interactions 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 A.1: Summary Statistics

	Mean	SD
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	0.668	(0.472)
Degree (Advice)	2.787	(3.378)
Degree (Overall)	3.954	(3.945)
Stress (Index; Scale 1-5)	3.076	(0.864)
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)
Observations	352	

Notes: The above table shows the summary statistics (mean and standard deviation) for various demographic characteristics of interest for the baseline sample in 2020. The binary variable "Difficulties" is equal to 1 if the individual reports having felt that "difficulties were piling up so high that they could not overcome them" often or very often and 0 otherwise.

Table A.2: Summary Statistics

	Mean	SD
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)
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)
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 additional summary statistics (mean and standard deviation) for the baseline sample in 2020 such as willingness to engage, dialogue intensity, stigma, and beliefs about peers. All variables are binary except beliefs and dialogue intensity. 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

B.1 Endline Results

Table B.1: Treatment Effect on Community Engagement

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) 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

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 centre 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.

Table B.2: Treatment Effect on Own Health Outcomes

VARIABLES	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

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 centre 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.

Click to go back.

Table B.3: Treatment Effect on Other Outcomes

VARIABLES	(1) Memory (Numbers Remembered)	(2) 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

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 centre as the cluster unit.

Click to go back.

C Long Run Effects

Table C.1: Long Run Effects on Beliefs about Peers

VARIABLES	(1) Beliefs (MH)	(2) Beliefs (MH)	(3) Beliefs (MH)	(4) Beliefs (FC)	(5) Beliefs (FC)	(6) 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

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

Note: 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 C.2: Long Run Effects on Willingness to Engage

VARIABLES	(1) Willingness to Engage (MH)	(2) Willingness to Engage (MH)	(3) Willingness to Engage (MH)	(4) Willingness to Engage (FC)	(5) Willingness to Engage (FC)	(6) 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

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

Note: 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 C.3: Long Run Effects on Dialogue

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

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

Note: 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.

Table C.4: Long Run Effects on Consumption Outcomes

VARIABLES	(1) Consumption Crisis (Yes/No)	(2) Consumption Crisis (Yes/No)	(3) Consumption Crisis (Yes/No)	(4) Consumption varies a lot	(5) Consumption varies a lot	(6) 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

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

Note: 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. "Consumption crisis" is a binary variable equal to 1 if the respondent reports that they have faced a consumption crisis in the last 6 months. "Consumption varies a lot" is a binary variable equal to 1 if the respondent reports that their consumption fluctuates a lot as opposed to little or not at all.

Click to go back.

D Mechanisms

Table D.1: 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

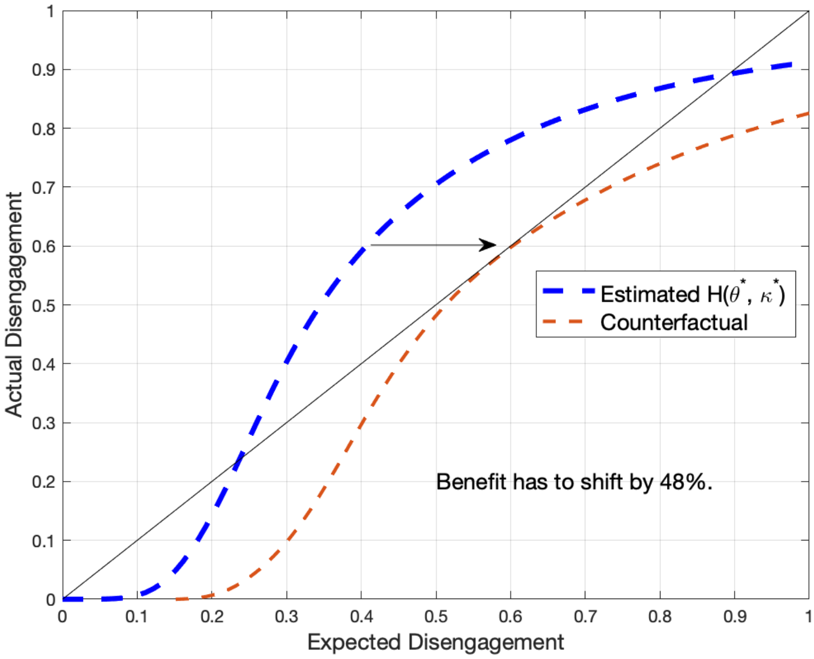
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: 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 (standardised) 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.

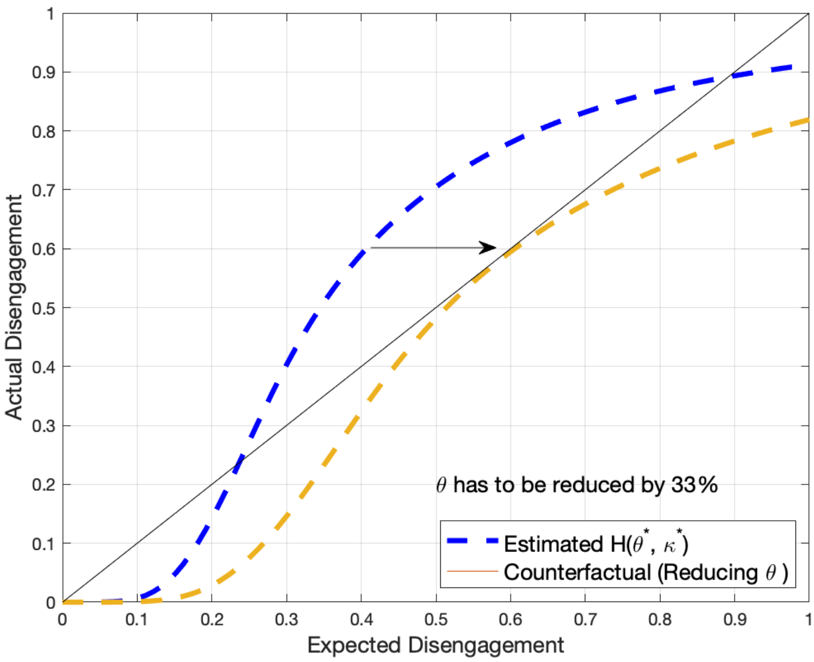
E Structural Estimation: Counterfactual Results

Figure E.1: 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. & 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. & Rasul, I. (2018), 'Consumption and investment in resource pooling family networks', *The Economic Journal* **128**(615), 2613–2651.

Audet, C. & 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. & Golub, B. (2018), When less is more: Experimental evidence on information delivery during india's demonetization, Technical report, National Bureau of Economic Research.

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

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

Belloni, A., Chernozhukov, V. & 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. & Yekutieli, D. (2006), 'Adaptive linear step-up procedures that control the false discovery rate', *Biometrika* **93**(3), 491–507.

Bisin, A., Moro, A. & 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. & Parvathaneni, A. (2019), 'Networks in economic development', *Oxford Review of Economic Policy* **35**(4), 678–721.

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

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

Cameron, A. C., Gelbach, J. B. & 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. & Schilbach, F. (2022), Blue spoons: Sparking communication about appropriate technology use, Technical report, National Bureau of Economic Research.

Chandrasekhar, A. G., Golub, B. & 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. & Khan, I. (2018), *Wastepickers: Delhi's Forgotten Environmentalists?*, Chintan Environmental Research and Action Group.

Chen, X., Christensen, T. M. & 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. & 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. & 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. & Gubert, F. (2007), 'The formation of risk sharing networks', *Journal of development Economics* **83**(2), 326–350.

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

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

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

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

Ivaschenko, O., Rodriguez Alas, C. P., Novikova, M., Romero Robayo, C., Bowen, T. V. & 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. & Tan, X. (2012), 'Social capital and social quilts: Network patterns of favor exchange', *The American Economic Review* **102**(5), 1857–1897.

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

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

Jackson, M. & 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. & Szeidl, A. (2009), 'Trust and social collateral', *The Quarterly Journal of Economics* **124**(3), 1307–1361.

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

Marx, B., Stoker, T. & 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. & 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. (2014), 'Community networks and the process of development', *Journal of Economic Perspectives* **28**(4), 49–76.

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

Munshi, K. & 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. & 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. & 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.

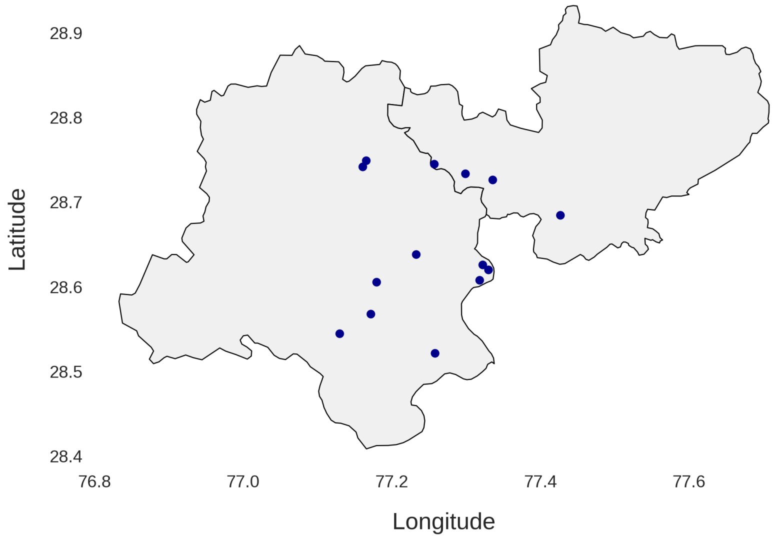
Van de Rijt, A. & Buskens, V. (2005), 'A nash equilibrium refinement for myerson's network formation game', *ISCORE paper* **230**.

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

Appendix

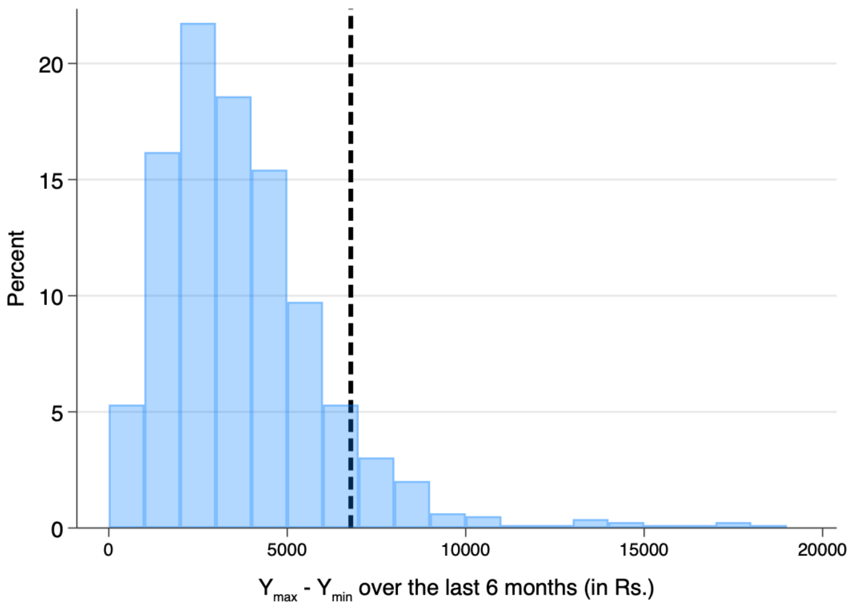
F Baseline

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



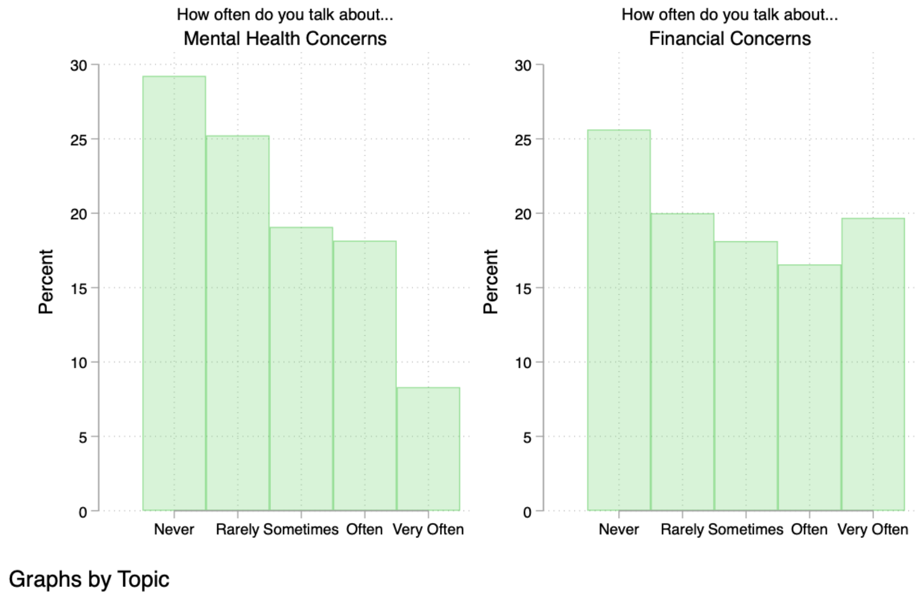
Notes: The figure plots the locations of various centres where we conducted the surveys (in blue) on the map of the national capital region.

Figure F.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.

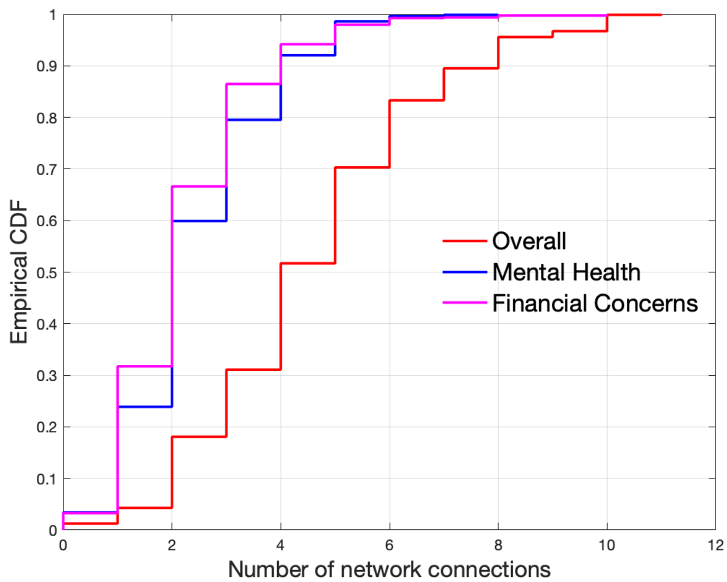
Figure F.3: Intensity of Dialogue in the last two weeks



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

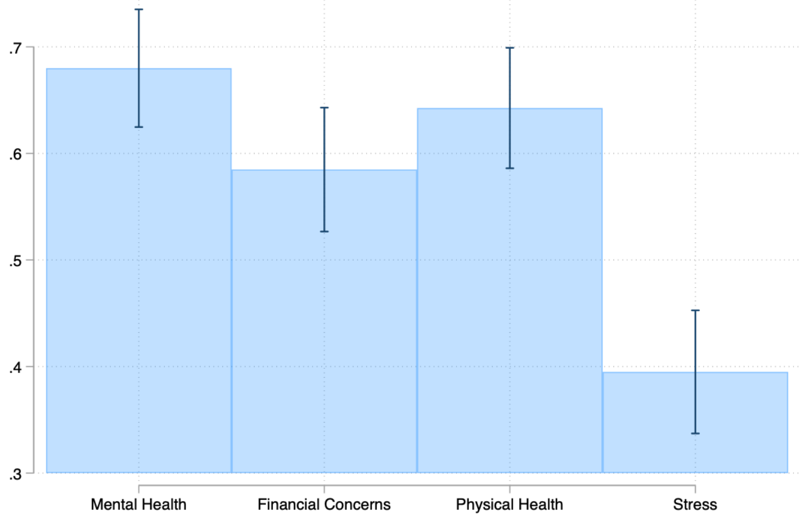
Click to go back.

Figure F.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 F.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 centre).

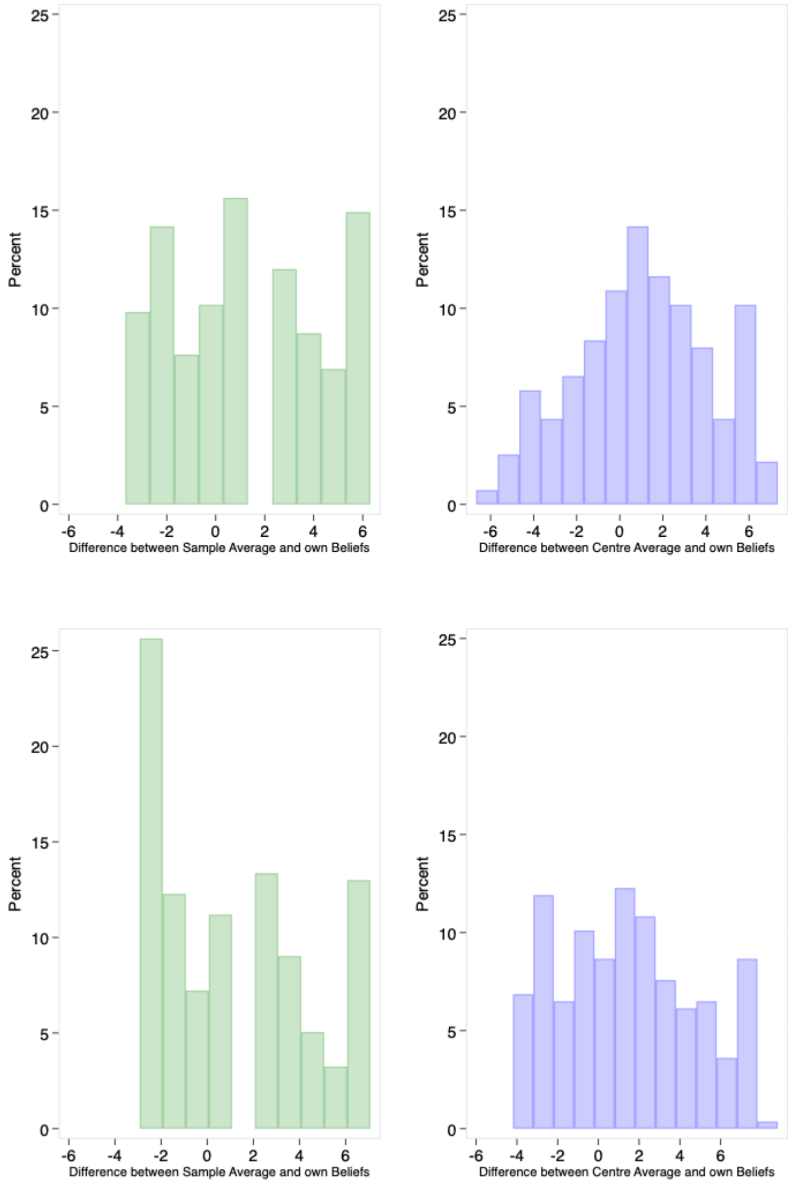
Table F.1: Correlations between Consumption Volatility, Demographic Characteristics, Network Characteristics, and Beliefs about Peers in 2023.

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 regression 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 dummy 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 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). "MH" refers to mental health 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.

Figure F.6: Magnitude of Misperceptions around Willingness to Engage with Mental Health



Notes: The figure at the top plots the difference between own beliefs around community's willingness to engage around mental health and the true sample level willingness to engage (across all communities) on the left and true willingness to engage in their own community (on the right). The figure at the bottom shows the same for financial concerns.

Click to go back.

Table F.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 F.3: Correlations of Network Gaps and Dialogue Intensity with Beliefs

VARIABLES	(1) Degree Gap	(2) Dialogue (MH)	(3) Dialogue (PH)	(4) 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

Note: This table reports regression results where we regress degree gaps and dialogue intensity on various measures of beliefs. 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 centre level.

Table F.4: Correlations of Network Gaps and Dialogue Intensity with being an Underestimator

VARIABLES	(1) Degree Gap	(2) Degree Gap	(3) Dialogue (MH)	(4) 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 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. Standard errors are robust and clustered at the centre level.

Table F.5: 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 centre.

Table F.6: 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).

G Design

G.1 Endline Balance

Table G.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 G.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.

H Endline Results

H.1 Additional Endline Results

Table H.1: Demand for Information

	(1) Information Session	(2) Listening to Good Practices (Immediate)	(3) 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

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 centre 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.

Table H.2: 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

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 centre 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.

Table H.3: 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 centre 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.

Click to go back.

H.2 Heterogeneity

Heterogeneity by Baseline Willingness to Talk about Mental Health

Table H.4: Treatment Effects by Baseline Willingness to Talk about Mental Health -1

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) Contribution (>0)	(6) 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

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

Note: This table shows heterogeneous effects on measures of community engagement 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 centre as the cluster unit.

Table H.5: Treatment Effects by Baseline Willingness to Talk about Mental Health -2

VARIABLES	(1) Information Session	(2) Listening to Good Practices (Immediate)	(3) 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 centre as the cluster unit.

Click to go back.

Heterogeneity by Baseline Dialogue Frequency

Table H.6: Treatment Effects by Baseline Dialogue around Mental Health

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) Contribution (>0)	(6) 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

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 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 centre as the cluster unit.

Table H.7: Treatment Effects by Baseline Dialogue around Financial Concerns

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) Contribution (>0)	(6) 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 financial concerns 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 centre as the cluster unit.

Click to go back.

Heterogeneity by Baseline Network Degree

Table H.8: 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 centre as the cluster unit.

Heterogeneity by Baseline Stigma towards Mental Health

Table H.9: 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 centre as the cluster unit.

Click to go back.

Heterogeneity by Underestimation of other's Willingness to Engage

Table H.10: Treatment Effects by Underestimators (Mental Health)

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) Contribution (>0)	(6) 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

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 sample willingness to engage around mental health i.e. whether their belief about their community's willingness to engage is less than the delivered information. 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 centre as the cluster unit.

Table H.11: Treatment Effects by Underestimators (Financial Concerns)

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) Contribution (>0)	(6) 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 financial concerns (FC). 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 centre as the cluster unit.

Click to go back.

H.3 Endline Robustness Checks

Table H.12: Community Engagement (Robust)

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) Community Engagement
Treatment	0.147**	0.194**	0.115*	6.333**	0.138***
	(0.0711)	(0.0814)	(0.0649)	(3.184)	(0.0516)
Constant	0.667***	0.483***	0.733***	21.44***	0.662***
	(0.0548)	(0.0940)	(0.0811)	(3.653)	(0.0393)
Observations	150	147	144	138	150
R-squared	0.028	0.046	0.023	0.033	0.046

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

Note: We additionally control for any baseline variables that are not balanced between the treatment and control for the subsample for which the outcome variable is non-missing. We report robust standard errors.

Click to go back.

Table H.13: Demand for Information (Robust)

	(1) Information Session	(2) Listening to Good Practices (Immediate)	(3) Listening to Good Practices
Treatment	0.101	-0.0507	0.0263
	(0.0731)	(0.0846)	(0.0636)
Constant	0.753***	0.551***	0.821***
	(0.0848)	(0.0603)	(0.0472)
Observations	152	141	139
R-squared	0.017	0.003	0.001

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

Note: We additionally control for any baseline variables that are not balanced between the treatment and control for the subsample for which the outcome variable is non-missing. We report robust standard errors.

Click to go back.

Table H.14: Own Health Outcomes (Robust)

VARIABLES	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)
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

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

Note: We additionally control for any baseline variables that are not balanced between the treatment and control for the subsample for which the outcome variable is non-missing. We report robust standard errors.

Click to go back.

Table H.15: Self Efficacy (Robust)

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)
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

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

Note: We additionally control for any baseline variables that are not balanced between the treatment and control for the subsample for which the outcome variable is non-missing. We report robust standard errors.

Click to go back.

Table H.16: Stigma (Robust)

VARIABLES	Stigma (Information Session)	List Experiment	Depression Score Revelation
Treatment	-0.0105	-0.0763	-0.0917
	(0.0336)	(0.211)	(0.0881)
Constant	0.956***	3.507***	0.925***
	(0.0339)	(0.157)	(0.0423)
Observations	119	154	64
R-squared	0.002	0.001	0.020

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

Note: We additionally control for any baseline variables that are not balanced between the treatment and control for the subsample for which the outcome variable is non-missing. We report robust standard errors.

Click to go back.

Table H.17: Other Outcomes (Robust)

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

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

Note: We additionally control for any baseline variables that are not balanced between the treatment and control for the subsample for which the outcome variable is non-missing. We report robust standard errors.

Click to go back.

I Mechanisms

Table I.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 3 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 I.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.

Heterogeneity by Stress

Table I.3: Treatment Effects by Baseline Under-estimation about Stress in the Community

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) Contribution (>0)	(6) 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

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 underestimates the level of stress in their community i.e. their belief is less than the true community-level stress. 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 centre as the cluster unit.

Click to go back.

Table I.4: Treatment Effects by Baseline Beliefs about Proportion who are Stressed

	(1) Savings Group	(2) Listening Volunteer	(3) Listening Contribution	(4) Contribution (in Rupees)	(5) Contribution (>0)	(6) 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

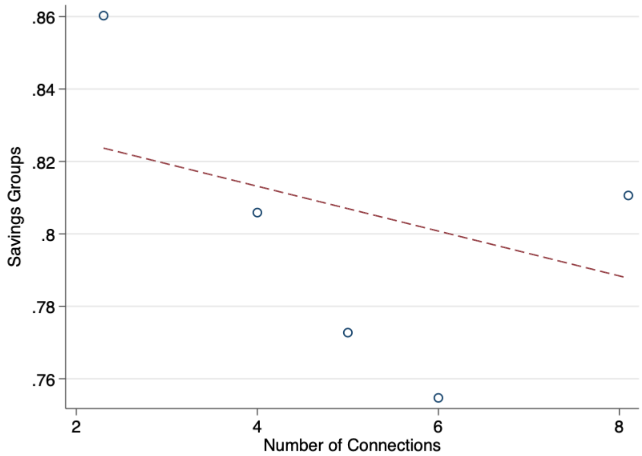
Note: This table shows heterogeneous effects on various measures of community engagement by whether the individual's belief about the level of stress in their community is above the proportion who are actually not willing to engage (i.e. 40%). This is to check if those who think more than 40% are stressed 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 centre as the cluster unit.

J Theory

J.1 Evidence for Model Assumptions

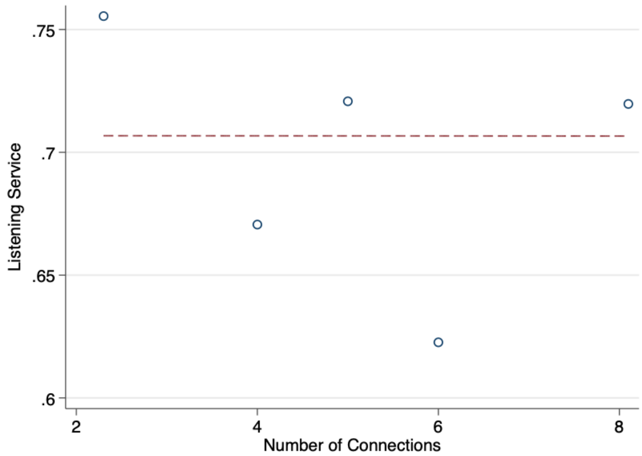
The following figures show correlations between degree centrality and survey measures of engagement with the network.

Figure J.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.

Figure J.2: Correlation between Degree and Willingness to Participate in Listening Service



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

J.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.

J.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.

J.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-

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

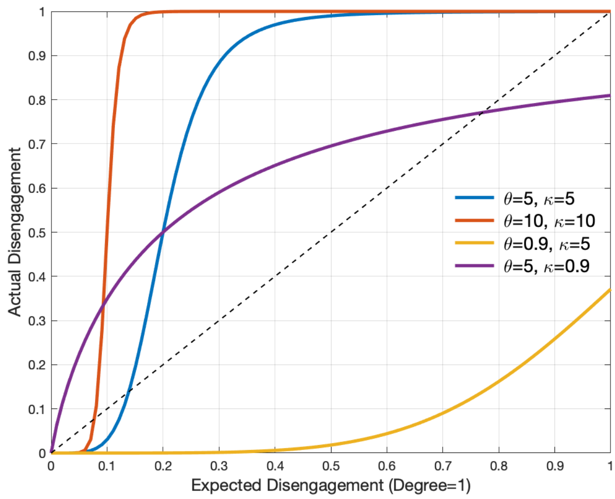
.

J.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.

K Structural Estimation

Figure K.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 K.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.

K.1 Constructing Confidence Sets using Quasi MCMC Methods

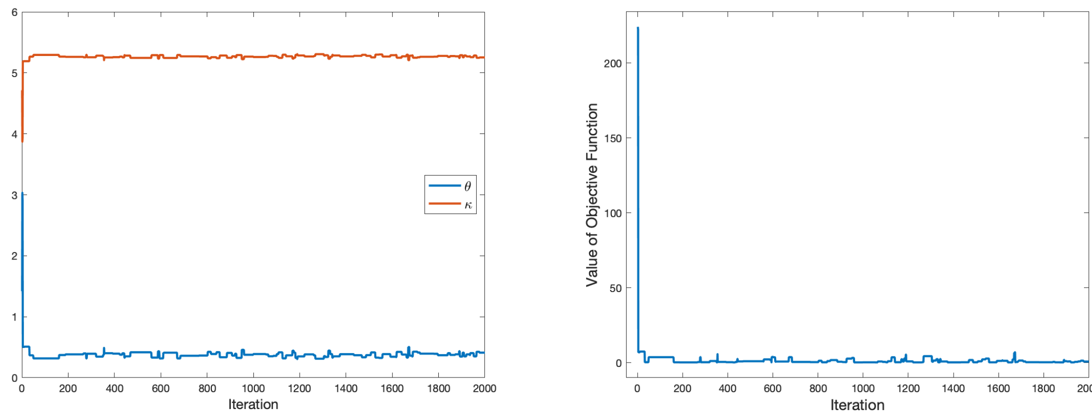
We estimate the confidence intervals for θ an κ as follows. We first compute the covariance matrix for the moment conditions (i.e. the mean of the treatment group and the control group) using the variance-covariance matrix returned via the OLS regression on treatment and control indicators and no constant. 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 1e-3 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($\hat{cov}$) * 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:

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

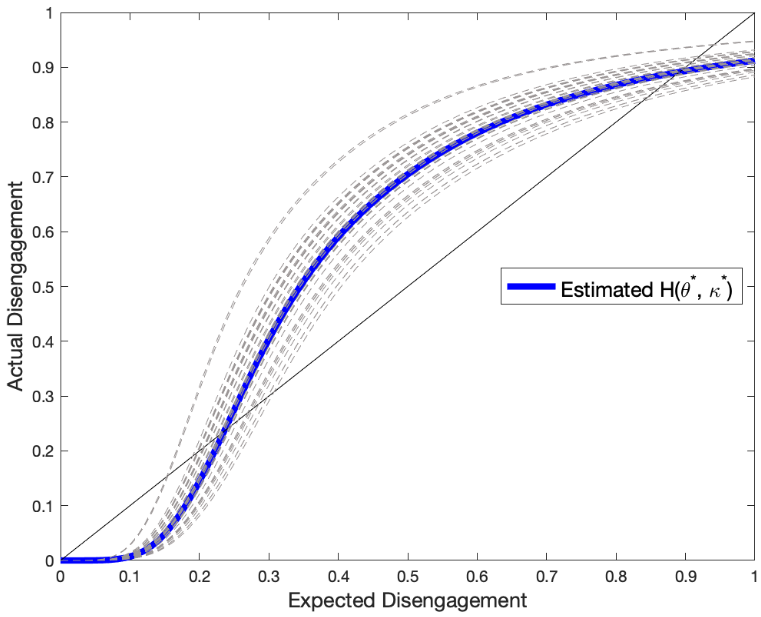
Crucially, α_t is close to zero if $Loss_t$ is very high compared to $Loss_{t-1}$ but is positive if the difference is not very large. If θ_1 and κ_1 are selected, then θ_2 and κ_2 are constructed as before by adding a shock to θ_1 and κ_1. If not, θ_2 and κ_2 are constructed by adding a shock to θ_0 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 K.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 K.3

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



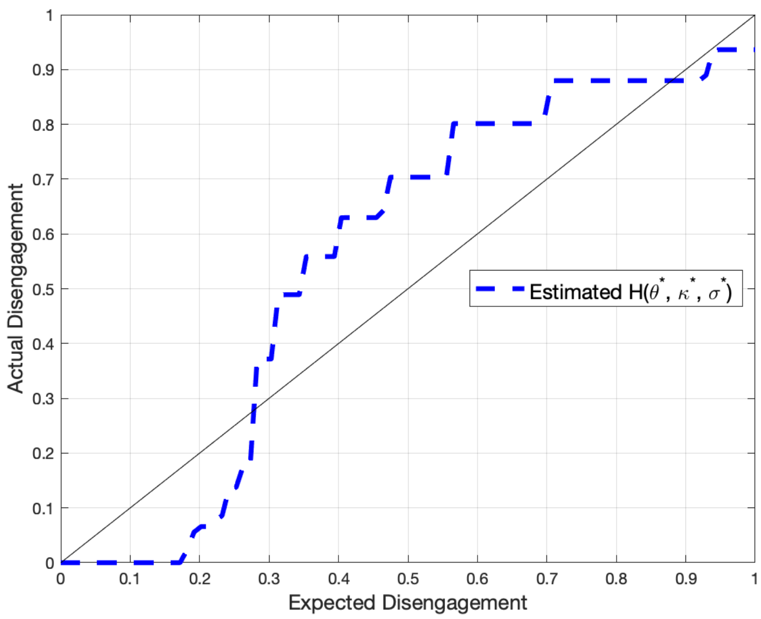
Notes: The figure on the left plots θ and κ over various iterations. In each iteration t, a candidate value of θ and κ is chosen with probability 1 if $Loss_t$ is less than $Loss_{t-1}$ and with a probability $e^{-0.5(Loss_t - Loss_{t-1})}$ if it is higher than before. The figure on the right plots the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

Figure K.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 K.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 $H(c(d,a)) = 1/(1+e^((-\theta * d * a) + \kappa)/\sigma)$.

Click to go back.

L Follow up Survey

L.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

L.2 Balance and Attrition

We test balance on 44 baseline variables as before and find that the sample is unbalanced for 2 variables. We also show that attrition from the endline survey is not correlated with treatment status and not correlated with various baseline characteristics.

First, we find in Table L.1 individuals who received the treatment are $11 - 19\%$ more likely to have reported that they made a call to their peers to discuss their own or their peers' mental health or employment related concerns respectively. We also document positive effects at the intensive margin since we also asked individuals how many peers they spoke to. The treated group has spoken to significantly more number of peers about both mental health and employment related concerns. The results are robust to the inclusion of unbalanced controls. It is worth noting that the effect on dialogue is stronger, and in the case of mental health, only present, when individuals in the treated sample are making the calls rather than receiving them. Increase in this one-sided interaction could suggest that the treatment is at work since it only impacts the participants and not their peers.

Next, we find in Table L.3 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 a 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 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 L.2 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.

L.3 Follow up Survey: Main Results

Table L.1: Dialogue

VARIABLES	(1) Employment (Made Call)	(2) Employment Calls (Frequency)	(3) Employment (Got Call)	(4) Mental Health (Made Call)	(5) Mental Health Calls (Frequency)	(6) Mental Health (Got Call)
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

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.

Table L.2: Other Outcomes

VARIABLES	(1) Masks (How often)	(2) Safety Gear (How often)	(3) Maintain Distance (How often)	(4) 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

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.

Table L.3: Well Being

VARIABLES	(1) Happiness	(2) Life Satisfaction	(3) Stress	(4) 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.

Click to go back.

M Replication Exercise

We conducted a demographic survey with a sample of $\sim$ 800 individuals in 2 NGO centres 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 M.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 replication of the 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.[35] Further, unlike the original treatment where individuals were told that the information we are providing is from their own community, in the replication 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 centre was not a part of our original sample and even if it were, the centre composition would have changed due to COVID-induced migration.

Table M.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

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

et al. 2014) to estimate the treatment effects. The effect on willingness to participate in savings groups and listening services is shown in Tables M.4 and M.6. We continue to find evidence that beliefs about community willingness to engage significantly affects 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 M.5 and M.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. We will not delve further into the mechanisms behind this treatment effect. 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 M.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.[36] 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.

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

M.1 Replication Tables

Table M.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 M.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 6 i.e. the information provided to the treatment group. Degree refers to the number of connections

Table M.3: Balance Tests for Centre 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 centre in the replication sample. We present these balance tests separately as the mental health fund outcome was only accurately measured for this centre. 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 6 i.e. the information provided to the treatment group. Degree refers to the number of connections

Table M.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 (standardised) 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 M.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 (standardised) 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 M.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 (standardised) 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 M.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 (standardised) 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 M.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 1 centre 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 (standardised) 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.

References

Ambrus, A. & 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. & Rasul, I. (2018), 'Consumption and investment in resource pooling family networks', *The Economic Journal* **128**(615), 2613–2651.

Audet, C. & 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. & Golub, B. (2018), When less is more: Experimental evidence on information delivery during india's demonetization, Technical report, National Bureau of Economic Research.

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

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

Belloni, A., Chernozhukov, V. & 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. & Yekutieli, D. (2006), 'Adaptive linear step-up procedures that control the false discovery rate', *Biometrika* **93**(3), 491–507.

Bisin, A., Moro, A. & 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. & Parvathaneni, A. (2019), 'Networks in economic development', *Oxford Review of Economic Policy* **35**(4), 678–721.

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

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

Cameron, A. C., Gelbach, J. B. & 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. & Schilbach, F. (2022), Blue spoons: Sparking communication about appropriate technology use, Technical report, National Bureau of Economic Research.

Chandrasekhar, A. G., Golub, B. & 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. & Khan, I. (2018), *Wastepickers: Delhi's Forgotten Environmentalists?*, Chintan Environmental Research and Action Group.

Chen, X., Christensen, T. M. & 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. & 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. & 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. & Gubert, F. (2007), 'The formation of risk sharing networks', *Journal of development Economics* **83**(2), 326–350.

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

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

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

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

Ivaschenko, O., Rodriguez Alas, C. P., Novikova, M., Romero Robayo, C., Bowen, T. V. & 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. & Tan, X. (2012), 'Social capital and social quilts: Network patterns of favor exchange', *The American Economic Review* **102**(5), 1857–1897.

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

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

Jackson, M. & 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. & Szeidl, A. (2009), 'Trust and social collateral', *The Quarterly Journal of Economics* **124**(3), 1307–1361.

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

Marx, B., Stoker, T. & 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. & 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. (2014), 'Community networks and the process of development', *Journal of Economic Perspectives* **28**(4), 49–76.

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

Munshi, K. & 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. & 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. & 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. & 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.