DISCUSSION PAPER SERIES

DP16492
(v. 2)

Social Proximity and Misinformation: Experimental Evidence from a Mobile Phone-Based Campaign in India

Alex Armand, Britta Augsburg, Antonella Bancalari and Kalyan Kumar Kameshwara

DEVELOPMENT ECONOMICS

PUBLIC ECONOMICS

ISSN 0265-8003

Social Proximity and Misinformation: Experimental Evidence from a Mobile Phone-Based Campaign in India

Alex Armand, Britta Augsburg, Antonella Bancalari and Kalyan Kumar Kameshwara

Discussion Paper DP16492
First Published 28 August 2021
This Revision 03 June 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Social Proximity and Misinformation: Experimental Evidence from a Mobile Phone-Based Campaign in India

Abstract

We study how social proximity between the sender and the receiver of information shapes the effectiveness of preventive health behaviour campaigns and the persistence of misinformation. We implement a field experiment among a representative sample of slum residents in two major Indian cities characterized by Hindu-Muslim tensions. We show that informative messages are effective at improving evidence-based behavior, but not non-evidence-based behavior. These findings do not differ by social proximity, signalled by religion. However, when sender and receiver share the same religion, the intervention significantly reduces misinformation carrying in-group salience, highlighting the role of social proximity in fighting misinformation.

JEL Classification: C93, D91, I12, I15, O12

Keywords: Misinformation, religion, identity, Social proximity, India, social media, COVID-19, Randomized field experiment

Alex Armand - alex.armand@novasbe.pt
Nova School of Business and Economics, NOVAFRICA, Institute for Fiscal Studies and CEPR

Britta Augsburg - britta_a@ifs.org.uk
Institute for Fiscal Studies

Antonella Bancalari - antonella.bancalari@st-andrews.ac.uk
School of Economics and Finance, University of St. Andrews, and Institute for Fiscal Studies

Kalyan Kumar Kameshwara - k.k.kameshwara@bath.ac.uk
University of Bath and Institute for Fiscal Studies

Acknowledgements
Declarations of interest: none. We would like to thank Oriana Bandiera, Maithreesh Ghatak and Michael Callen, as well as seminar participants at the LSE-STICERD Development Economics seminar series and the IFS for helpful comments. This project would have not been possible without the insightful support of our partners in the field, Bhartendu Trivedi, Yashashvi Singh, Tatheer Fatima and the field team of Morsel Research and Development. We are also indebted to study participants for their willingness to contribute to the study. We gratefully acknowledge financial support from the IGC (grant ref. IND-20077), the LSE Special Grant for Coronavirus Research (grant ref. 106404), and the ESRC's Centre for the Microeconomic Analysis of Public Policy at the IFS (grant ref. ES/M010147/1). Kameshwara acknowledges the support from the ESRC's SWDTP Placement Scheme. Ethics approval was secured from the London School of Economics and Political Science (REC ref. 1132). The pre-analysis plan was registered on the AEA RCT registry (Armand et al., 2020).

1 Introduction

Despite the importance of communication media for social mobilization (Enikolopov et al., 2020; Manacorda and Tesei, 2020), evidence about the mechanisms that facilitate the spread of (mis)information on these platforms remains limited and mainly targeted at higher-income countries (DellaVigna and Kaplan, 2007; Allcott and Gentzkow, 2017; Lazer et al., 2018; Bursztyn et al., 2020). In lower-income settings, where the steep increase in penetration rates of these channels is combined with lower literacy levels, the spread of both information and misinformation presents distinct opportunities, as well as challenges. Recent evidence shows how technology can be effective at raising awareness and promote factual information in these settings (see, e.g., Banerjee et al., 2020; Siddique et al., 2022). However, the simultaneous persistence of wrong or inaccurate beliefs remains incompletely understood (Guess et al., 2020; Zhuravskaya et al., 2020). In particular, despite the large evidence on the way social diversity influences decisions (Easterly and Levine, 1997; Alesina and Ferrara, 2005; Habyarimana et al., 2007), there is scarce evidence on its role in fighting misinformation.

In this paper, we study how social proximity between the sender and the receiver of information shapes the effectiveness of information campaigns and the persistence of misconceptions.[1] We implement a field experiment among slum residents in the Indian State of Uttar Pradesh (UP) in the context of a global outbreak of an infectious disease – the COVID-19 pandemic. The state is home to the largest Muslim population in India, representing 19.3% of the internal population, as compared to 79.7% of the Hindu population (Government of India, 2011). Conscious of the divisions between the two communities, we choose religious identity as a salient measure of social proximity. The Hindu-Muslim tensions in India goes back to the pre-partition era and flared up at regular frequency since (Jha, 2013; Mitra and Ray, 2014). In line with evidence showing that religion is more salient in the presence of unpredictable life events (Sinding Bentzen, 2019; Atkin et al., 2021), the onset of the pandemic saw a sudden rise in misinformation tied to religion: misleading claims about the role of Muslim citizens in the spread of the virus were the primary driver of fake news on social media, and spurred further violence (Yasir, 2020).[2]

[1] People assess their *social proximity* with others based on the extent to which they share common traits, characteristics, and identities (Bicchieri et al., 2022). The shaping of information to such individual characteristics of an information recipient is known in the literature as *microtargeting*. This approach has been used to influence public opinion (Bostrom et al., 2013; Goldberg et al., 2019), and the behavior of voters (Borgesius et al., 2018; Bradshaw et al., 2020; McKay and Tenove, 2021), consumers (Rosário et al., 2021), and medical patients (Yom-Tov et al., 2018; Alsan and Eichmeyer, 2021).

[2] UP was repeatedly in the news for religious incidents both before and during the pandemic (Arya, 2020; Al Jazeera, 2020). Discrimination and violence against stigmatized groups has been documented in other settings during the pandemic (Desai and Amarasingam, 2020).

In this context, we designed a mobile phone-based campaign in response to the urgent need to raise awareness among citizens about evidence-based practices mitigating the spread of disease. Using a unique sampling frame obtained from a census of slum residents in the two major cities of the state, we randomly allocated residents to either a *doctor message* group or a *control* group. In the *doctor message* group, participants received pre-recorded voice messages containing an introduction by a local citizen, the *sender*, followed by statements from doctors of locally-renowned hospitals debunking common misconceptions about the virus and reminding about evidence-based policy recommendations. In our setting, 95% of the targeted population report doctors as the most trusted source of COVID-19 information, highlighting a high degree of credibility of the message (see, e.g., O'Keefe, 2016; Vraga and Bode, 2017; Khan et al., 2021). In the *control* group, participants received instead a message with the same introduction, but followed by a Bollywood gossip. All households were targeted by two voice messages between October 2020 and January 2021.

To vary social proximity, we randomized in both the doctor message and the control message the initial greeting of the sender to be either a Muslim or a Hindu greeting. In contrast to previous literature, we do not vary the informative content of the message. This design creates exogenous variation in the *religion concordance* between the sender and the receiver of the message. Beyond information campaigns, inter- and intra-group contact have shown to generate different behavior in a variety of settings, from patient-doctor interaction (Greenwood et al., 2018; Alsan et al., 2019; Greenwood et al., 2020; Hill et al., 2020), to nation-building (Bazzi et al., 2019). In particular, inter-religious contact influences integration (Mousa, 2020; Lowe, 2021), and economic outcomes (Bhalotra et al., 2014; Fisman et al., 2017, 2020).

We gathered information about slum residents' preventive practices against COVID-19, compliance with evidence-based recommendations, trust in information, fact-checking and beliefs. Information was collected using phone-based interviews during a baseline survey in June–July 2020, which covered 3,991 respondents, followed by two panel data waves in October–November 2020, and December 2020–January 2021. We base our main analysis on intention to treat (ITT) effects of the doctor message and of religion concordance in both the doctor message and the control message. Using administrative data on the take-up of the interventions, we complement ITT estimates with treatment on the treated (ToT) estimates of the effect of interventions among compliers. Results are robust to alternative specifications, and to correcting inference for multiple hypothesis testing.

We show that mobile phone technology can be an effective way to inform citizens in a low-income setting and to promote the adoption of welfare-improving behavior. As compared to the control message, the doctor message increases significantly the awareness of evidence-based practices to prevent infection with COVID-19, as well as the compliance with them. These results complement recent evidence on the use of technology to raise health awareness in the US (Alsan et al., 2020; Breza et al., 2021; Torres et al., 2021), and in rural India and Bangladesh (Siddique et al., 2022). Religion concordance leads participants to listen to a larger share of the doctor message (3 percentage points more), but does not introduce any differential effect in awareness of, and compliance with, evidence-based practices.

The doctor messages have no significant effect on the degree to which respondents agree with non-evidence-based practices to prevent infection, despite being debunked in the message. The doctor messages, however, reduce fact-checking, an important determinant of factual knowledge (Barrera et al., 2020), while keeping the trust in the information shared by doctors or by other citizens unaffected. Again, religion concordance introduces no differential effect.

To find a central role for social proximity, we need to look directly into the level of misinformation among participants. Because social proximity to peers influences norm compliance (Bicchieri et al., 2022), individuals can avoid disagreement when group salience and inter-group tensions are higher (Akerlof and Kranton, 2000; Tajfel and Turner, 2004; Chen and Li, 2000). To address this issue, we developed a novel survey instrument that allows us to isolate the influence of group identity in belief formation. We elicit the level of misinformation by measuring agreement with a sequence of randomly-ordered statements. We distinguish between views that are not necessarily based on facts or knowledge (labeled as *opinions*), or incorrect views based on faulty knowledge or understanding (labeled as *misconceptions*). We ask respondents about agreement with these statements as stated by a third person living in Uttar Pradesh, which we label the *interlocutor*. For each statement, we randomize the name of the interlocutor to signal different religious identities, and thereby control for the group salience of each statement.

The doctor message has no impact on respondents' opinions, but decreases the level of agreement with misconceptions. The reduction is solely driven by a decrease in agreement in misconceptions stated by a person *outside* the religious group of the respondent (the 'out-group' interlocutor). This effect, which is a reduction of 3.2% over the control mean, remains highly significant after multiple hypothesis testing. The doctor message has no average effect on agreement with the same misconceptions but stated by an in-group interlocutor. In-group agreement with misconceptions is instead altered only in presence of religion concordance with the message's sender, in which case agreement is reduced by

4.6% over the mean for the religion-discordant doctor message. Religion concordance has no impact on other measures of agreement with opinions and misconceptions. These results highlight how informative messages can leverage social norms by potentially challenging the assumption that in- and out-groups agree with prevailing norms.

Our findings provide important insights on the design of information campaigns in low-income settings. We highlight the role of social proximity in the persistence of misinformation, reinforcing the importance of identity in the way beliefs and decisions are formed. We provide novel evidence on the role of religious identity in information campaigns, contributing to the field studying the role of religion among interacting citizens (see, e.g., Iyer, 2016), and complementing evidence on the role of identity for cooperation, political mobilization and violence (Philpott, 2007; Mitra and Ray, 2014; Laborde, 2021; Lowe, 2021).

2 Intervention and experimental design

The intervention targets the population of slum residents in the two largest urban agglomerations in UP, Lucknow and Kanpur. We draw the study population from a census of slum residents conducted in the second half of 2018 in both cities as part of a distinct study (Armand et al., 2021), which provided a unique sampling frame of more than 30,000 households living in the slums of the study area before the pandemic. From the sampling frame, we targeted a random sub-sample of 4,000 households. The resulting study population is comparable to the average slum resident in the state and in the rest of India (further details are provided in Armand et al., 2021; Solís Arce et al., 2021). Figure 1 summarizes the study timeline and compares it with COVID-19 regulations in UP in the corresponding period. Appendix B provides a description of the study area.

The intervention takes place at the onset of the COVID-19 pandemic. Similar to other states of India, UP was hit hard by the pandemic during the period of the study, with a rapid spread in the number of COVID-19 cases and a steep increase in the number deaths (Appendix Figure B3). Guidelines of social distancing and wearing of face masks remained in place throughout the study period.

The intervention consists of sending voice messages targeted at individual citizens using mobile phone technology. Each message has two components: the *introduction of a sender*, i.e. a local citizen, and the *informative content* of the message. In the *doctor message* treatment, the informative content is presented by doctors from locally renowned medical institutions debunking common misconceptions about ways to prevent COVID-19 and reminding about the confirmed ways to protect against infection. Qualified medical practitioners were chosen for the informative content because, at baseline, 95% of the respondents named doctors and health experts as their most trusted source of COVID-19 information. We sent two rounds of messages varying the misconception object of the message: first that eating a vegetarian diet protects against COVID-19 (sent in October–November 2020), and second that the immune system of Indians is resilient to COVID-19 (sent in December 2020–January 2021). At baseline, these two misconceptions were the two most prevalent ones in the targeted population: in an open-ended question, they represent one fifth and one-fourth of all reported misconceptions, respectively (Appendix Figure B5).

The informative content for the *doctor message* treatment was built by first asking several doctors from renowned local institutions to reply, unscripted, to the questions: "is it true that eating a vegetarian diet protects against COVID-19?" and "is it true that the immune system of Indians is resilient to COVID-19?". Responses were then collated ensuring that every message was composed by a first part debunking the misconception and a second part providing a reminder about the proven ways to protect against COVID-19. The scripts of the messages are provided in Appendix B.3.

A *control message* was designed with a similar structure to hold all features not varied experimentally constant: the message begins with the introduction of the local citizen (the same introduction performed in the doctor message) followed by the 'informative content', in this case an unsubstantiated gossip concerning Bollywood stars. Sending a control message, rather than no message, allows us to disentangle the effects of the intervention from receiving a message through mobile phone technologies. The final message duration was 95 and 123 seconds for the doctor messages and 39 seconds for the control message.[3]

To introduce variation in religious identity associated with the doctor and control messages, we exploit the religious diversity in UP. In the slum setting, the representation of religious groups is comparable to one in the whole state (Appendix B2), with 79% represented by Hindu citizens, and 21% by Muslim citizens. While Hindi is the official state language, in our setting, each group has a distinctive greeting.

[3]The intervention included a short video with the same content as voice messages sent through a WhatsApp chatbot. Only a negligible share of the sample saw the video: 38.5% received the message, and only 2.5% replied accurately with 'Hi' to the first message containing a 'Hi', a precondition imposed by the app to start the chat and visualize the content of the interventions. We therefore focus on the voice messages only. Including controls for the visualization of the video messages on Whatsapp or excluding these participants from the sample does not alter any of the results. Alternative approaches to debunk misconceptions include live phone calls (Sadish et al., 2021), communication via instant messaging platforms among subscribers (Bowles et al., 2020), or pedagogical interventions (Badrinathan, 2021).

The greeting "namaste" signals a Hindu identity of the speaker, while the greeting "salam alaikum" signals a Muslim identity. We exploit this characteristic by introducing two additional variations in the doctor and control messages by changing the initial greeting of the sender, which represents the first word of the message, while keeping all else constant. In our analysis, we refer to *religion concordance* of the message when the initial greeting of the sender is signalling the same religion of the receiver of the message, and *religion discordance* when it is signalling a different religion. Differently from previous studies (see, e.g., Alsan and Eichmeyer, 2021), this design allows varying only the initial signal, keeping the informative content of the message indistinguishable in terms of identity.

All messages were incentivized to increase attention to the message by giving participants the chance to enter a lottery if they replied correctly to a follow-up question about the message. During the introduction of the message, the sender announced the financial incentive to be paid out through mobile top-up.[4] Financial incentives are particularly relevant when faced with potentially extremely low uptake (see, e.g., Azrieli et al., 2018). For instance, in the context of video messages sent to Indian citizens by mobile phone and urging them to comply with COVID-19 policies, Banerjee et al. (2020) achieved a viewing rate of 1.1%, consistent with the low rates of click-through studies (Richardson et al., 2007; Kanich et al., 2008).

We perform household-level randomization by randomly allocate targeted households, independent of the number of mobile phones in each household, to receive doctor or control messages, and for each of these, to receive a message introduced by a Hindu or a Muslim greeting. We stratify by religion of the household head and by city. Randomization into the experimental arms was conducted at the household level because the intervention is directed one-to-one through mobile phones. Using households as the unit of randomization allows us to take advantage of greater variation in response to the intervention within slums. Concerns over spillover effects are mitigated because the voice message is in the form of an automatic call which cannot be forwarded or shared. There remains the possibility of word-of-mouth information sharing. To the extent that it did occur, our estimates would provide lower bounds to the true treatment effects.

[4] We introduced a lower-incentive lottery with a value of Rs. 2,500 (US$32) and a higher-incentive lottery with a value of Rs. 5,000 (US$64). We allocated one lottery award per city for each round of messages. The two modalities were cross-randomized in the process of randomization. We discuss the effects of higher incentives in Appendix D.3.

3 Data

We draw on two data sources: a panel survey of slum residents, and administrative data about the interventions. Appendix E details ethical considerations related to data collection activities.

Primary panel data. We collected primary data among slum residents on households' experiences during the COVID-19 pandemic, such as their knowledge on how to prevent the virus, compliance with policies, as well as information on sources of information and trust, and beliefs. We collected a baseline survey in June–July 2020 by reaching 3,991 households. Two waves of follow-up panel data were collected in October–November 2020, and December 2020–January 2021 (3.5 and 5.5 months after the baseline survey), reaching 3,816 households during the first follow-up and 3,906 during the second follow-up survey. To keep the time gap between the intervention and follow-up data collection similar across individuals, we split the sample in four batches due to the operational capacity of the field team. In each batch, we interviewed households two weeks after sending the voice messages by conducting phone conversations.[5] Combining both follow-up surveys, we re-interviewed 87% of residents at least once, with a low implied attrition rate (13%) compared to phone surveys conducted in similar settings. Response rates are typically around 50% in non-crisis contexts, while during crisis contexts this is expected to be lower. For instance, a study during the Ebola crisis was able to re-interview only 38% (Himelein et al., 2020). A likely driver of attrition in our setting is represented by people using multiple SIM cards to avail discounts offered by different providers (Silver and Huang, 2019). Attrition is orthogonal to treatment allocation (Panel B, Table C1), with being female and a dwelling owner reducing significantly attrition (Appendix Table C2).

Appendix Table C1 presents descriptive statistics of the sample. Twenty-one percent of respondents are Muslim, almost 80% of respondents are male, mostly represented by the household head, with an average age of 40 years. More than 80% live in a strong dwelling with four other members, and 38% have a ration card (i.e., an official document giving access to the subsidized purchase of essential commodities). At the time of the baseline survey, 12% of respondents report that at least one member was having COVID-19 symptoms, and respondents knew on average 1.6 COVID-19 symptoms. Further information about vaccine acceptance at baseline in this sample is provided in Solís Arce et al. (2021).

[5]The sampled households that were not reachable at the time of the survey were replaced with replacement households randomly selected from the sampling frame discussed in Section 2. To guarantee both a high quality of information and a concise interview, some modules were administered to a random subset of households only. This is the case for compliance with COVID-19 regulations: some households were asked about social distancing, while other households about hand-washing.

To measure knowledge about ways to prevent infection from COVID-19, we asked respondents about the level of agreement with a number of evidence-based preventive practices (i.e., present in policy recommendations) and non-evidence-based preventive practices (i.e., not present in policy recommendations), all of which were discussed in the doctor message (Section 2). Evidence-based practices included wearing a face mask, hand-washing, and keeping physical distance; non-evidence-based practices included instead the two most-common views collected at baseline on how to protect from the virus: to rely on the Indian immune system or on vegetarianism. Panel A of Figure 2 provides descriptive statistics on respondents' levels of agreement with evidence- and non-evidence-based practices over the course of study, separated by the religion of the respondent.[6]

A key part of the surveys was to measure the presence of misinformation among participants. In line with the literature (see, e.g., Scheufele and Krause, 2019), we define misinformation as the holding of inaccurate views or being uninformed about scientific facts and processes. In the literature, the reporting of misinformation has shown to be closely associated with motivated thinking (Kunda, 1990; Taber and Lodge, 2006), i.e., the set of emotional biases leading individuals to agree with views based on desirability rather than evidence. This is particularly important in our setting for two reasons. First, biases can lead individuals to not be able to distinguish between views that are not necessarily based on facts or knowledge (labeled as *opinions*), and incorrect views based on faulty knowledge or understanding (labeled as *misconceptions*). Opinions have shown to be less responsive to fact-checking as compared to misconceptions (Walter and Salovich, 2021). Second, in the presence of group identity and inter-group tensions, it is hard to disentangle agreement with misinformation from group identity (Tankard and Paluck, 2016; Nyhan, 2021). For instance, bias in the reporting of misinformation might arise in the presence of social desirability or cognitive dissonance associated to misconceptions that are common in Hindu or Muslim communities.

To address these potential issues, we introduce a novel survey instrument to elicit the level of misinformation among participants. We focus on agreement with a sequence of statements related to COVID-19, presented in a random order during the interview to avoid question order bias. We distinguish between statements that we classify as defined above as *opinions*) and as *misconceptions*. We selected statements to capture religious salience, and to cover topics characterized by a high degree of misinformation in

[6]For ease of presentation, we average together the agreement with individual protection mechanisms (wearing of a face mask and practice hand-washing), and the agreement with non-evidence-based practices. Appendix Figure B6 provides descriptive statistics for the separate variables. We do not have baseline information for these variables because the baseline question elicited practices through an open-ended question, rather than in levels of agreement.

the media (see, e.g., Al-Zaman, 2021; World Health Organisation, 2020). We included in the questionnaire 5 statements, 3 capturing opinions and 2 capturing misconceptions. The first opinion, ”religious gatherings should be allowed”, carries general religious salience and is particularly relevant in the study context due to the early outbreak linked to the Islamic missionary movement *Tablighi Jamaat*, which led to Islamophobic reactions across media (Menon, 2020). The second opinion, ”unity and brotherhood will help us fight the coronavirus”, is typically connected with Islam, but in the context of India, it is also associated with the Hindu nationalist ruling party Bharatiya Janata Party (BJP). Slogans such as *Vasudhaiva Kutumbakam* (brotherhood of mankind) and *unity among religions* were evoked multiple times by party leadership before and during the pandemic as one of the ideological foundations of Hinduism (Kulkarni, 2017; Choudhury, 2020; Hindustan Times, 2020). The third opinion, ”the virus was created in a laboratory”, is related to conspiracy theories about the creation of the virus, which often lead to conspiracy theories targeting Muslim in India (The Guardian, 2020). The first misconception, ”if you are vegetarian, you do not need to worry about the coronavirus”, carries specific religious salience since, in the context of India, vegetarianism is widely associated with the dominant ideology of Hinduism. The second misconception, ”if you are a good person, you do not need to worry about the coronavirus”, carries general religious salience, with the idea that religion helps a person to become a good person.

To further limit social desirability bias in the reporting of agreement, we ask respondents about agreement with these statements as stated by a third person living in UP, which we label the *interlocutor*. To remove the link between group identity and agreement, we randomize the name of the interlocutor to signal different religious identities using 5 options: 1 male Muslim name, 1 female Muslim name, 1 male Hindu name, 1 female Hindu name, or the generic ”people”. The names were selected using information on the most common names by religion from the census of slum residents (Section 2).[7] Because the list of statements is constant in the survey, but names vary in each interview, we can build measures of *in-group* and *out-group* agreement for both opinions and misconceptions depending on whether the respondent shares the same religion with the interlocutor. In absence of any bias associated with religious identity, the agreement should be independent from being in-group or out-group.

Levels of agreements with opinions and misconceptions are computed by averaging agreement with individual statements. Panel B of Figure 2 provides descriptive statistics on respondents’ levels of agreement

[7] The exact script of the question reads as follows: ”We have surveyed a few people from UP and we would like to hear if you agree with their opinion. Note that responses to the statements are a matter of opinion. There is no scientific evidence about their truthfulness. On a scale of 1 to 5, where 1 means you strongly disagree and 5 you strongly agree, how much do you agree or disagree with the following statements. [Interlocutor] says that [statement].”

with opinions and misconceptions independently from the interlocutor assigned to the statement, separated by the religion of the respondent. Appendix Figure B7 shows these statistics for each statement separately.

Comparing Panel A and Panel B in Figure 2 provides descriptive insights in the relationship between knowledge and the persistence of misinformation. First, the agreement with non-evidence-based ways to protect from the virus is significantly lower than agreement with evidence-based practices, a relationship that is similar when comparing misconceptions with opinions. Second, differences by religion are specific to beliefs about non-evidence-based ways and agreement with misconceptions. On average, Hindu respondents are significantly more likely to agree with non-evidence-based ways and misconceptions than Muslim respondents, a difference that is mainly driven by beliefs about vegetarianism, the predominant diet among the Hindu population. Third, while opinions tend to change over time, agreement with misconceptions tend to be relative constant over time.

Administrative data. We gather information about the delivery of voice messages, and about the duration and the share of the voice message that each user played. The voice messages were sent to the whole sample: 37% picked-up the phone when receiving the call, and once picked-up, the average listening time was 42 seconds or 55% of the message.

Listening times vary between the experimental arms because the duration of the control message is constant, while the doctor message is different in the two rounds (see Section 2). Conditional on having picked up the call, Panel A in Figure 3 shows the distribution of the share of each message that is listened by study participants in the control group versus the doctor message group. In the control group, the average listening time was 33 seconds or 67% of the message. Respondents in the doctor message group listen to a smaller share of the message as compared to the control group (on average 25 percentage points less, Appendix Figure B9). Because the control message is shorter in length, we also show that on average respondents in the doctor message group listen to the message for a larger number of seconds (18 seconds more).

Panel B in Figure 3 shows instead the distribution of the share of each message that is listened by study participants in presence of religion-concordant or -discordant message for both the control group (left figure) and the doctor message group (right figure). Religion concordance leads to a higher share of the message that is listened throughout the distribution, suggesting that religion concordance is an important driver of listening time. The difference is particularly relevant for the doctor message.

4 Empirical approach

To assess treatment impacts we rely on post-baseline data, in line with the trial registry (Armand et al., 2020), and justified by having successfully created observationally-equivalent groups. Appendix Table C1 shows mean differences at baseline between the different treatment arms for respondent's characteristics, highlighting the balance across groups in terms of observable characteristics. To estimate the impact of the doctor message, we estimate the following specification:

$$Y_{ijt} = \beta_D \, doctor_i + \alpha \, \mathbf{X}_{ij} + \delta_t + \epsilon_{ijt} \tag{1}$$

where Y_{it} are outcomes of interest of respondent i in slum j at time t. The variable $doctor_i$ is an indicator variable equal to 1 if the receiver i is in the doctor message treatment group, and 0 otherwise. $\mathbf{X}_{ij}$ is a set of indicator variables for randomization strata, and δ_t are period-of-survey indicator variables. The error term ϵ_{it} is assumed to be clustered at the slum level.[8]

To estimate the effect of religion concordance with the sender of the message, we restrict the sample to either the doctor message group or the control group, and estimate the following specification:

$$Y_{ijt} = \beta_C \, concordance_i + \alpha \, \mathbf{X}_{ij} + \delta_t + \epsilon_{ijt} \tag{2}$$

where $concordance_i$ is an indicator variable equal to 1 if the receiver i received a message in which the sender and the receiver shares the same religion, and 0 otherwise. When the sample is restricted to the doctor message group, the parameter β_C captures the differential effect of receiving a religion-concordant doctor message as compared to a religion-discordant doctor message. Similarly, when the sample is restricted to the control group, the parameter β_C captures the differential effect of receiving a religion-concordant Bollywood message as compared to a religion-discordant Bollywood message.[9]

Following McKenzie (2012), to estimate both equation 1 and equation 2, we pool data from the two follow-up surveys together and we estimate the average impact in the follow-up period (assuming β_D and β_C are constant over time). Especially in presence of outcomes variables measured in high temporal proximity, this approach allows averaging out the noise in the outcome variables and increasing power.

[8] Results are robust to alternative assumptions about standard errors, such as assuming heteroskedasticity or clustering at the individual level. Using weights to increase representativeness of the sample with the original census does not affect results.

[9] Appendix D.2 provides instead estimates of the effect of a Hindu versus a Muslim greeting, independently from the content of the message. We observe no effect for these comparisons.

Results are presented in Tables 1–3. In each table, panel A provides estimates of treatment effects using equation 1, while panel B presents estimates using equation 2 restricting the sample to the doctor message group to estimate the differential impact of religion concordance in the doctor message. Panel C presents estimates using equation 2 restricting the sample to the doctor message group, serving as placebo test for the effect of religion concordance because the control message does not carry any informative content.

In each table, we supplement standard inference with multiple hypothesis testing adjusting p-values for the significance of each individual coefficient in the table using the List et al. (2019) bootstrap-based procedure. In Appendix Tables D10 and D11, we confirm robustness of our results to using an ANCOVA specifications, controlling for the baseline value of the dependent variable. Appendix D.5 reports estimates of heterogeneous impacts by estimating equation (1) and and equation 2 in sub-samples defined by pre-specified variables (religion of the respondent, and % of Muslim living in the slum), and by other relevant dimensions (caste, strength of religious identity, trust in the government, and social desirability).

To support the interpretation of β_D and β_C, Figure 4 shows estimates of the effect of the doctor message (Panel A) and of a religion-concordant doctor message (Panel B) on the probability to have picked up the call with the doctor message (left panels), and on the share of the doctor message that is listened (right panel). Both variables are computed from administrative data. Panel A shows that just 41% of the respondents in the doctor message group picked up the call containing the doctor message, and, conditional on picking up the call, they listened to on average 34% of the message. These statistics are similar across Hindu and Muslim participants.

Restricting the analysis to the doctor message group, Panel B of Figure 4 indicates that the likelihood of picking up the call with the doctor message is not different in presence of religion concordance or discordance. This is an expected result because the identity of the sender is disclosed only upon picking up the call. However, religious concordance leads to longer listening time by 3 percentage points. The increase is 5 percentage points among Hindu respondents, and negative and not significant among Muslim respondents.

Because compliance to treatment is therefore not perfect, as standard in mass information campaigns, the parameters β_D and β_C identify the intention to treat (ITT) impact of the treatment. In Appendix D.1, using administrative data about the endogenous take-up of the intervention, measured by the share of each message that is effectively listened on the phone of the participant, we supplement ITT impacts with estimates of the treatment on the treated (TOT) effects. The endogenous share of the doctor message and share of the religion-concordant doctor message that are listened are instrumented with the treatment indicators $doctor_i$ and $concordance_i$, respectively. Following Imbens and Angrist (1994), in light of the likely heterogeneity in the (potential) impacts of the intervention, we interpret these estimates as the local average treatment effects for participants that comply with the intervention.

5 Results

5.1 Preventive practices

Table 1 presents estimates of treatment effects on behavioral outcomes. We first focus on agreement with preventive practices against COVID-19 by studying, in columns (1)–(3), the effect of interventions on the agreement with evidence-based practices, and with non-evidence-based practices. Column (4) is an indicator variable equal to 1 if the respondent reports to comply with a randomly-asked World Health Organization's recommendation to protect from infection (either leaving the slum and receiving visitors, or washing hands after all events), and 0 otherwise.

The doctor message treatment leads to changes in evidence-based practices (panel A). We find a significant increase in agreement with using face masks and practicing hand-washing to protect against the virus by 0.6 percentage points (0.75% over the control mean), while agreement with social distancing is also increased but not significantly, a result potentially driven by the high population density of slums. These effects are driven by increases among respondents with a higher strength of religious identity, and by respondents living in slums with a smaller share of Muslim households (Appendix D.5). While agreement with official recommendations increases in response to the doctor message, we observe no effect for non-evidence-based practices that were debunked in the message. This highlights the persistence of misconceptions even in presence of increased agreement with evidence-based preventive practices, suggesting that evidence- and non-evidence-based practices are not perfect substitutes in the preventive health behavior.

The doctor message increases significantly also the compliance with policy recommendations by 4 percentage points (7.09% over the control mean). Inference for the effects on both evidence-based preventive practices and on compliant behavior is robust to multiple hypothesis testing. Estimates increase significantly in magnitude when considering ToT effects (Appendix Table D1): listening to the full doctor message leads to an increase of 4.4 percentage points in the level of agreement with using face masks and practicing hand-washing, and by 28.2 percentage points in compliant behavior. These effects are particularly encouraging in light of the previous literature underlining the challenges to adhere to COVID-19 guidelines in presence of limited access to clean water and overcrowding (Patel, 2020; Wasdani and Prasad, 2020). Both characteristics are crucial features of slum settings.

Turning our attention to the role of religion concordance of the messages, we do not find any differential effect for either the doctor or the control message (panel B and C of Table 1). While panel C serves as a placebo test because the message has no content related to preventive practices, panel B highlights how the effect of the doctor message is statistically homogeneous with respect to religion concordance of the message, suggesting that preventive behavior is affected by the informative content of the message rather than social proximity with the sender.

5.2 Fact-checking and trust in information

This section focuses on impacts related to information sharing. Table 2 presents the effects on evidence-based behavior (or *fact-checking*), and on trust in information shared by alternative sources. In column (1), fact-checking is measured as an indicator variable equal to 1 of the respondent frequently checks the truthfulness of the information he/she shares or discusses, and zero otherwise. Columns (2)–(3) measure the effect on the average level of trust in the information shared by doctors and health experts, and by other citizens, respectively. Trust is measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust.

Beyond improving evidence-based preventive behavior (Section 5.1), the doctor message also affects behavior associated with misinformation by crowding-out fact-checking. The need of respondents to verify information discussed with family and friends decreases significantly by 2.3 percentage points (3.54% over the control mean of 0.64), and remains significant after adjusting p-values for multiple hypothesis testing. This effect translates in a ToT estimate of 14.7 percentage points when the respondent listens to the full doctor message (Appendix Table D1). At the same time, we do not find any changes in reported levels of trust.

Reductions in fact checking following the doctor message are higher when respondents were incentivized with the higher lottery amount (Appendix D.3), potentially driven by participants paying closer attention to the message. This result is accompanied by a reduction in trust in information shared by other citizens. These effects are however only significant at the 11 and 10% significance level after correcting for multiple hypothesis testing. No other outcomes are affected by lottery amounts, in line with Porter and Whitcomb (2003).

Similarly to results presented in Table 1, religion concordance does not introduce any significant differential effect for both the doctor and the control message, highlighting again that behavior related to preventive and misinformation-related behavior is more-closely associated with the informative content of the message. However, among respondents with a lower strength of religious identity, we find a significant reduction in the alignment with non-evidence-based practices when the doctor message is religion-concordant (Appendix Figure D4). This finding highlights how the persistence of misconceptions is more prevalent among individuals with a stronger sense of identity.

5.3 Misinformation

Because non-evidence-based behavior tend to be unaffected (Section 5.1) or negatively (Section 5.2) by the doctor message, with homogeneous effects relative to religion concordance, we turn our attention to direct measures of misinformation. Table 3 presents estimates of treatment effects on the level of agreement with public views related to COVID-19. The outcome variables and their construction is described in Section 3. Columns (1)–(2) focus on agreement with opinions and misconceptions, independently from the interlocutor being in-group or out-group. Overall, respondents tend to be more aligned with views not necessarily based on facts or knowledge (opinions) as compared to misconceptions, with an average agreement in the control group of 0.62 and 0.49, respectively. Nevertheless, these statistics highlight the widespread nature of misinformation among the study participants.

The doctor message has no impact on agreement with opinions, but it reduces agreement with misconceptions by 0.8 percentage points, significant at the 10% level, but not robust to multiple hypothesis testing. This effect corresponds to a decrease of 5.4 percentage points when the doctor message is listened fully (Appendix Table D3). In line with the results on non-evidence-based preventive practices (Section 5.1), the effect of the doctor message on misconceptions is larger for respondents with a lower strength of religious identity (Appendix Figure D4). Adding religion concordance to the doctor message does not introduce any differential effect in neither the doctor message group nor the control group. However, religion concordance in the doctor message amplifies the effect of the doctor message for respondents with lower strength of religious identity, lower trust in the government, and lower social desirability (Appendix Figure D6).

To disentangle the mechanisms behind these effects from social norms associated with group identity, in columns (3)–(6), we focus on agreement distinguishing by whether the interlocutor is in-group or out-group. Note that, thanks to the design of the survey instrument, the content of the statements used to measure agreement with opinions and misconceptions is orthogonal to in/out-group, i.e., when comparing in- and out-group agreement, opinions and misconceptions are constant, while the interlocutor varies exogenously.

Similar to column (1), we find no effect of the doctor message on opinions disaggregated by in-group and out-group agreement. Instead, for misconceptions, we highlight that the doctor message leads to a significant reduction in agreement with out-group misconceptions by 1.6 percentage points (an effect of 3.24% over the control mean), while in-group misconceptions are not affected by the doctor message. This effect corresponds to a ToT estimate of 10.5 percentage points reduction in agreement with out-group misconceptions after listening to the doctor message fully (Appendix Table D1). These results highlight how the ability to effectively measure impacts of misinformation requires considerations about social norms and group identity.

Turning our attention to religion concordance in the doctor message (panel B), we find no differential impact on agreement with opinions and out-group agreement with misconceptions. However, religion concordance in the doctor message is crucial in influencing agreement with in-group misconceptions. When the doctor message is introduced by a religion-concordant greeting, agreement with in-group misconceptions is reduced by 2.3 percentage points. This effect is highly significant and robust to multiple hypothesis testing. The magnitude of the ToT estimate is a reduction of 15.0 percentage points in agreement with in-group misconceptions after listening to a religion-concordant doctor message fully (Appendix Table D1). These effects are specific to the combination of religion concordance with a doctor message, as receiving the religion-concordant greeting with the control message does not affect agreement with any of the variables in Table 3 (panel C).

Overall, these results suggest that the persistence of misinformation is mainly due to non-factual views, which are never affected by either the doctor message or its combination with religion concordance. Evidence-based views are instead more flexible with respect to information shared by trusted sources.

However, while information campaigns can shift misinformation not carrying in-group salience, when in-group salience is stronger, misinformation can be affected only in presence of social proximity.

6 Conclusions

The internet and social media have become popular sources for news. At the same time, they have become spreaders of inaccurate and misleading information. Ensuring that misconceptions resulting from misinformation and fake news do not drown-out scientific evidence and affect behavior in consequential ways is of global importance.

We demonstrate that low-cost interventions relying on mobile phone technology can be effectively deployed to improve evidence-based preventive behavior in low-income settings. However, the campaign has limited effect on non-evidence-based preventive practices, often the result of persistent misconceptions. We highlight that persistence is closely related to social norms, especially in presence of tensions across social group. Informative content can reduce agreement with misconceptions when in-group identity is not signalled, but has no effect when in-group identity is signalled. Including social proximity in the information campaign, by signaling religion concordance between the sender and the receiver of information, can instead reduce the agreement with misconceptions when in-group identity is signalled.

This study provides important insights into how agencies can rely on mobile phone messages to counter misinformation, and improve adherence to policy guidelines. While results highlight the importance of shaping messages to the characteristics of individual citizens, further research is needed to understand how to target communication against misconceptions in a more effective way.

References

Akerlof, G. A. and R. E. Kranton (2000). Economics and Identity. *Quarterly Journal of Economics 115*(3), 715–753.

Al Jazeera (2020). ‘Do not buy from Muslims’: BJP leader in India calls for boycott. https://www.aljazeera.com/news/2020/4/29/do-not-buy-from-muslims-bjp-leader-in-india-calls-for-boycott.

Al-Zaman, S. (2021). COVID-19-Related Social Media Fake News in India. *Journalism and Media 2*, 100–114.

Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature 43*(3), 762–800.

Allcott, H. and M. Gentzkow (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives 31*(2), 211–236.

Alsan, M. and S. Eichmeyer (2021). Experimental Evidence on the Effectiveness of Non-Experts for Improving Vaccine Demand. National Bureau of Economic Research (No. w28593).

Alsan, M., O. Garrick, and G. Graziani (2019). Does diversity matter for health? Experimental evidence from Oakland. *American Economic Review 109*(12), 4071–4111.

Alsan, M., S. Stantcheva, D. Yang, and D. Cutler (2020). Disparities in coronavirus 2019 reported incidence, knowledge, and behavior among us adults. *JAMA Network Open 3*(6), e2012403–e2012403.

Armand, A., B. Augsburg, and A. Bancalari (2020). COVID-19, fake news and religious tensions: experimental evidence from India. AEA RCT Registry, October 08. https://doi.org/10.1257/rct.6564-1.0.

Armand, A., B. Augsburg, and A. Bancalari (2021). Public Service Delivery and Free Riding: Experimental Evidence from India. CEPR discussion paper DP16284, Centre for Economic Policy and Research.

Arya, S. (2020). Muslims face social boycott in Buldhana, business too hit. The Times of India, https://timesofindia.indiatimes.com/city/nagpur/muslims-face-social-boycott-in-buldhana-business-too-hit/articleshow/7.

Atkin, D., E. Colson-Sihra, and M. Shayo (2021). How do we choose our identity? A revealed preference approach using food consumption. *Journal of Political Economy 129*(4), 1193–1251.

Awasthi, P. (2020). UP seals off hotspots in 15 districts, including Lucknow, Noida to contain spread. *The Week*.

Azrieli, Y., C. P. Chambers, and P. J. Healy (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy 126*(4), 1472–1503.

Badrinathan, S. (2021). Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India. *American Political Science Review 115*(4), 1325–41.

Banerjee, A., M. Alsan, E. Breza, A. G. Chandrasekhar, A. Chowdhury, E. Duflo, P. Goldsmith-Pinkham, and B. A. Olken (2020, July). Messages on COVID-19 Prevention in India Increased Symptoms Reporting and Adherence to Preventive Behaviors Among 25 Million Recipients with Similar Effects on Non-recipient Members of Their Communities. NBER Working Paper 27496, National Bureau of Economic Research.

Banerjee, A., M. Faye, A. Krueger, P. Niehaus, and T. Suri (2020). Effects of a Universal Basic Income during the pandemic. Working Paper, UC San Diego.

Barrera, O., S. Guriev, E. Henry, and E. Zhuravskaya (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics 182*, 104123.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2019). Unity in diversity? How intergroup contact can foster nation building. *American Economic Review 109*(11), 3978–4025.

Bhalotra, S., I. Clots-Figueras, G. Cassan, and L. Iyer (2014). Religion, politician identity and development outcomes: Evidence from india. *Journal of Economic Behavior & Organization 104*, 4–17.

Bicchieri, C., E. Dimant, S. Gaechter, and D. Nosenzo (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior 132*, 59–72.

Borgesius, F., J. Möller, S. Kruikemeier, R. Fathaigh, K. Irion, T. Dobber, B. Bodo, and C. De Vreese (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review 14*(1), 82–96.

Bostrom, A., G. Böhm, and R. E. O'Connor (2013). Targeting and tailoring climate change communications. *Wiley Interdisciplinary Reviews: Climate Change 4*(5), 447–455.

Bowles, J., H. Larreguy, and S. Liu (2020). Countering misinformation via WhatsApp: Preliminary evidence from the COVID-19 pandemic in Zimbabwe. *PloS one 15*(10), e0240005.

Bradshaw, S., H. Bailey, and P. N. Howard (2020). Industrialized Disinformation – 2020 Global Inventory of Organized Social Media Manipulation. Technical report, Oxford Internet Institute, Computational Propaganda Research Project.

Breza, E., F. C. Stanford, M. Alsan, B. Alsan, A. Banerjee, A. G. Chandrasekhar, S. Eichmeyer, T. Glushko, P. Goldsmith-Pinkham, K. Holland, and Others (2021). Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial. *Nature Medicine*, 1–7.

Bursztyn, L., A. Rao, C. Roth, and D. Yanagizawa-Drott (2020). Misinformation During a Pandemic. NBER Working Paper Series.

Chen, Y. and S. X. Li (2000). Group Identity and Social Preferences. *American Economic review 99*(1), 431–457.

Choudhury, C. (2020). Under the BJP's Big Tent. *The Caravan. A Journal of Politics and Culture*.

DellaVigna, S. and E. Kaplan (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics 122*(3), 1187–1234.

Desai, S. and A. Amarasingam (2020). CoronaJihad: Covid-19, misinformation, and anti-Muslim violence in India. ISD Global, Report.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: policies and ethnic divisions. *The Quarterly Journal of Economics*, 1203–1250.

Enikolopov, R., A. Makarin, and M. Petrova (2020). Social Media and Protest Participation: Evidence from Russia. *Econometrica 88*(4), 1479–1514.

Facebook's Crowd Tangle Team (2020). CrowdTangle.

Fisman, R., D. Paravisini, and V. Vig (2017). Cultural proximity and loan outcomes. *American Economic Review 107*(2), 457–92.

Fisman, R., A. Sarkar, J. Skrastins, and V. Vig (2020). Experience of communal conflicts and intergroup lending. *Journal of Political Economy 128*(9), 3346–3375.

Goldberg, M. H., A. Gustafson, M. T. Ballew, S. A. Rosenthal, and A. Leiserowitz (2019). A social identity approach to engaging Christians in the issue of climate change. *Science communication 41*(4), 442–463.

Government of India (2011). Census of India 2011.

Government of India (2021). Orders prescribing lockdown for containment of COVID-19 Epidemic in the country and guidelines for phased re-opening (unlocks). https://www.mha.gov.in/sites/default/files/PRNationalLockdown26032020.pdf.

Greenwood, B. N., S. Carnahan, and L. Huang (2018). Patient–physician gender concordance and increased mortality among female heart attack patients. *Proceedings of the National Academy of Sciences of the United States of America 115*(34), 8569–8574.

Greenwood, B. N., R. R. Hardeman, L. Huang, and A. Sojourner (2020). Physician-patient racial concordance and disparities in birthing mortality for newborns. *Proceedings of the National Academy of Sciences of the United States of America 117*(35), 21194–21200.

Guess, A. M., M. Lerner, B. Lyons, J. M. Montgomery, B. Nyhan, J. Reifler, and N. Sircar (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences 117*(27), 15536–15545.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review 101*(4), 709–725.

Hill, A., D. Jones, and L. Woodworth (2020). Physician-Patient Race-Match and Patient Outcomes. SSRN Working Paper.

Himelein, K., S. Eckman, and D. McKenzie (2020). Mobile Phone Surveys for Understanding COVID-19 Impacts: Part II Response, Quality, and Questions. Development Impact Blog.

Hindustan Times (2020). Modi calls for unity, says virus doesn't see religion. https://www.hindustantimes.com/india-news/modi-calls-for-unity-saysvirus-doesn-t-see-religion/story-odKsutBpjYQ6hdv5lVWj, New Delhi.

Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–475.

Iyer, S. (2016). The new economics of religion. *Journal of Economic Literature 54*(2), 395–411.

Jha, S. (2013). Trade, institutions, and ethnic tolerance: Evidence from south asia. *American political Science review 107*(4), 806–832.

Kanich, C., C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxsonf, and S. Savage (2008). Spamalytics: An empirical analysis of spam marketing conversion. *Proceedings of the ACM Conference on Computer and Communications Security*, 3–14.

Khan, A., S. Nasim, M. I. Shaukat, and A. Stegmann (2021). Building Trust in the State with Information: Evidence from Urban Punjab. *Journal of Public Economics*.

Kulkarni, S. (2017). *RSS-BJP Symbiosis: On the Cusp of Culture and Politics*. New Delhi: Prabhat Prakashan.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin 108*(3), 480.

Laborde, C. (2021). Minimal Secularism: Lessons for, and from, India. *American Political Science Review 115*(1), 1–13.

Lazer, D. M., M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, and J. L. Zittrain (2018). The science of fake news: Addressing fake news requires a multidisciplinary effort. *Science 359*(6380), 1094–1096.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics 22*(4), 773–793.

Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review 111*(6), 1807–1844.

Manacorda, M. and A. Tesei (2020). Liberation Technology: Mobile Phones and Political Mobilization in Africa. *Econometrica 88*(2), 533–567.

McKay, S. and C. Tenove (2021). Disinformation as a threat to deliberative democracy. *Political Research Quarterly 74*(3), 703–717.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics 99*(2), 210–221.

Menon, S. (2020). Coronavirus: The human cost of fake news in India. BBC News, BBC Reality Check https://www.bbc.co.uk/news/world-asia-india-53165436.

Mitra, A. and D. Ray (2014). Implications of an economic theory of conflict: Hindu-Muslim violence in India. *Journal of Political Economy 122*(4), 719–765.

Mousa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science 369*(6505), 866–870.

Nyhan, B. (2021). Why the backfire effect does not explain the durability of political misperceptions. *Proceedings of the National Academy of Sciences 118*(15), e1912440117.

O'Keefe, D. J. (2016). *Persuasion: Theory and Research* (3 ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Patel, A. (2020). Preventing COVID-19 Amid Public Health and Urban Planning Failures in Slums of Indian Cities. *World Medical and Health Policy 12*(3), 266–273.

Philpott, D. (2007). Explaining the Political Ambivalence of Religion. *American Political Science Review 101*(3), 505–525.

Porter, S. R. and M. E. Whitcomb (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education 44*(4), 389–407.

Reserve Bank of India (2019). State-wise Poverty Rate. Technical report, National Sample Survey Organisation, NITI Aayog, Delhi, India.

Richardson, M., E. Dominowska, and R. Ragno (2007). Predicting clicks: Estimating the click-through rate for new ads. *16th International World Wide Web Conference, WWW2007*, 521–530.

Rosário, A., L. B. Moniz, and R. Cruz (2021). Data science applied to marketing. *Journal of Information Science and Engineering 37*(5), 1067–1081.

Sadish, D., A. Adhvaryu, and A. Nyshadham (2021). (Mis)information and anxiety: Evidence from a randomized Covid-19 information campaign. *Journal of Development Economics 152*(102699).

Scheufele, D. A. and N. M. Krause (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences of the United States of America 116*(16), 7662–7669.

Siddique, A., T. Rahman, D. Pakrashi, A. Islam, and F. Ahmed (2022, 03). Raising Health Awareness in Rural Communities: A Randomized Experiment in Bangladesh and India. *The Review of Economics and Statistics*, 1–45.

Silver, L. and C. Huang (2019). Smartphone, Social Media Users Have Broader Social Networks In Emerging Economies. Pew Research Center, Report.

Sinding Bentzen, J. (2019). Acts of God? Religiosity and natural disasters across subnational world districts. *The Economic Journal 129*(622), 2295–2321.

Solís Arce, J. S., S. S. Warren, N. F. Meriggi, A. Scacco, N. McMurry, M. Voors, G. Syunyaev, A. A. Malik, S. Aboutajdine, O. Adeojo, et al. (2021). COVID-19 vaccine acceptance and hesitancy in low-and middle-income countries. *Nature medicine 27*(8), 1385–1394.

Taber, C. S. and M. Lodge (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science 50*(3), 755–769.

Tajfel, H. and J. C. Turner (2004). The social identity theory of intergroup behavior. *Political psychology: Key readings*, 276–293.

Tankard, M. E. and E. L. Paluck (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review 10*(1), 181–211.

The Guardian (2020). Coronavirus conspiracy theories targeting Muslims spread in India. The Guardian.

Torres, C., L. Ogbu-Nwobodo, M. Alsan, F. C. Stanford, A. Banerjee, E. Breza, A. G. Chandrasekhar, S. Eichmeyer, M. Karnani, T. Loisel, P. Goldsmith-Pinkham, B. A. Olken, P. L. Vautrey, E. Warner, and E. Duflo (2021). Effect of Physician-Delivered COVID-19 Public Health Messages and Messages Acknowledging Racial Inequity on Black and White Adults' Knowledge, Beliefs, and Practices Related to COVID-19: A Randomized Clinical Trial. *JAMA Network Open 4*(7), e2117115–e2117115.

United Nations (2019). World Urbanization Prospects: The 2018 Revision. United Nations, Department of Economic and Social Affairs, Population Division, New York, US.

Vraga, E. K. and L. Bode (2017). Using expert sources to correct health misinformation in social media. *Science Communication 39*(5), 621–645.

Walter, N. and N. A. Salovich (2021). Unchecked vs. uncheckable: How opinion-based claims can impede corrections of misinformation. *Mass Communication and Society 24*(4), 500–526.

Wasdani, K. P. and A. Prasad (2020). The impossibility of social distancing among the urban poor: the case of an Indian slum in the times of COVID-19. *Local Environment 25*(5), 414–418.

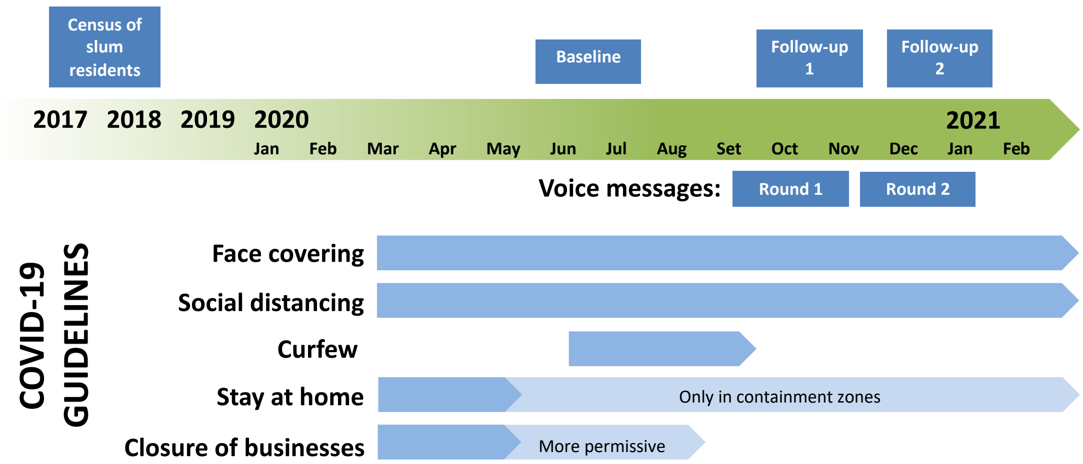
World Health Organisation (2020). COVID-19 Response. Seventy-third world health assembly (a73.conf./1 rev. 1), World Health Organization.

Yasir, S. (2020). India is scapegoating Muslims for the spread of the Coronavirus. *Foreign Policy*.

Yom-Tov, E., J. Shembekar, S. Barclay, and P. Muennig (2018). The effectiveness of public health advertisements to promote health: a randomized-controlled trial on 794,000 participants. *NPJ Digital Medicine* (24).

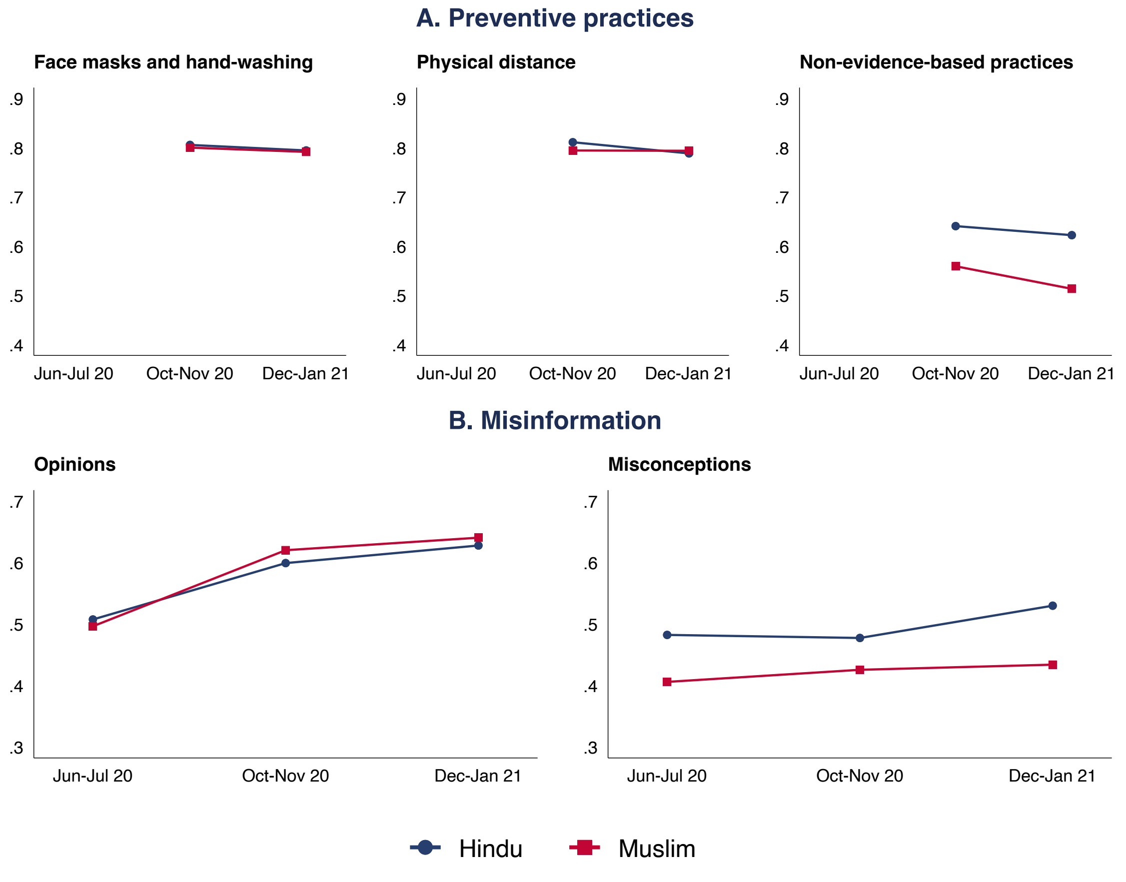
Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual Review of Economics 12*, 415–438.

Figure 1: Study timeline and comparison with COVID-19 guidelines in UP



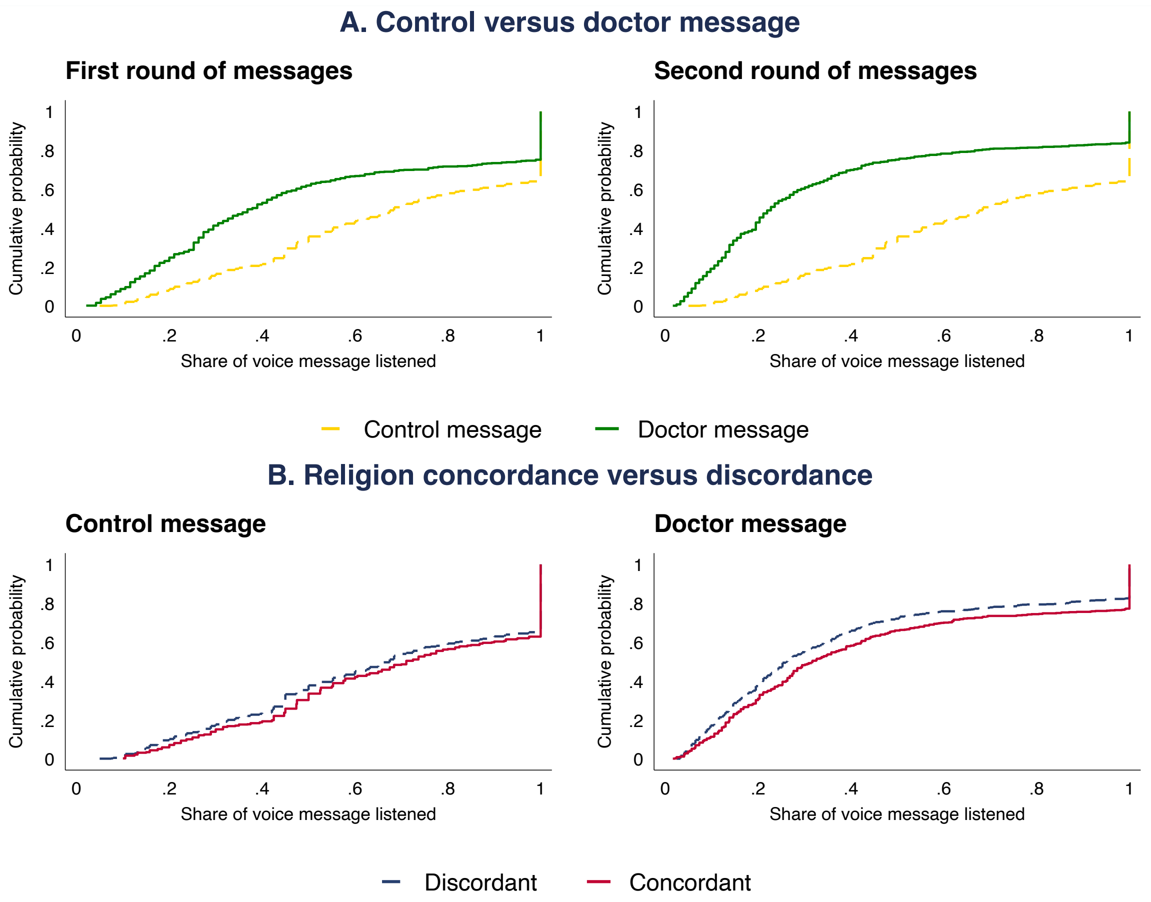
Notes. Guidelines are compiled from official sources (Government of India, 2021; Awasthi, 2020). Lucknow and Kanpur were included in the red zone in May 2020. Red zones are the areas with high coronavirus cases and high doubling rate in the previous 21 days. The first phase of the closure of businesses included all businesses apart from essential shops and services, while the second more permissive phase allowed the re-opening of the following activities: shopping malls, religious places, hotels and restaurants in June 2020 (unlock phases 1 and 2); gyms and yoga centers in August 2020 (unlock phase 3); entertainment, sport, political, academic and social functions and gatherings with a limited number of participants in September 2020 (unlock phases 4, 5 and 6). Curfews were first characterized by night curfews from 9pm to 5am in June and July 2020, and then to weekend curfews until September 2020. Local authorities had the power to impose curfews based on local conditions.

Figure 2: Preventive practices and misinformation about COVID-19, by religion



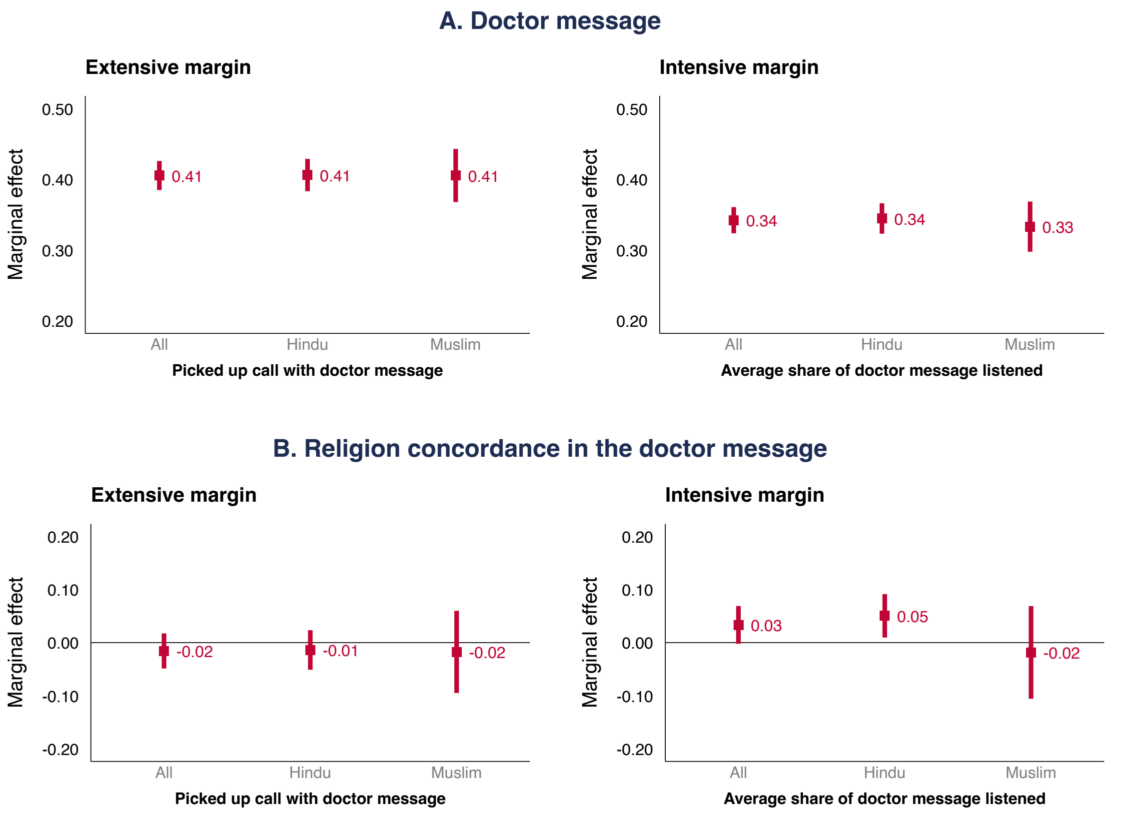
Notes. *Hindu* and *Muslim* refer to the religion of the respondent. Each figure shows the average level of agreement of the respondent with preventive practices against COVID-19 (Panel A) or with misinformation about COVID-19 (Panel B). *Opinions* are views that are not necessarily based on facts or knowledge, while *misconceptions* are incorrect views based on faulty knowledge or understanding. For the rationale behind such distinction in the measurement of misinformation, refer to Section 3 The sample is restricted to the control group. Statements are aggregated by averaging responses using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. In panel B, statements are averaged independently from the interlocutor linked to the statement. Statistics separated for each statement are presented in Appendix Figure B7. Agreement with opinions and misconceptions distinguishing by whether the interlocutor linked to the statement is in- or out-group is presented in Appendix Figure B8. Statements and categorization are described in Section 3.

Figure 3: Share of voice messages listened by study participants



Note. The figures show the share of the messages listened by study participants, conditional on having picked up the call. Information is based on administrative data from the intervention. Panel A includes the full sample separated by round of intervention, panel B restricts the sample to the control message group in the left figure and to the doctor message group in the right figure. Differences by treatment arm for the doctor message are reported in Figure 4, for any message in Appendix Figure B9.

Figure 4: The effect of the doctor message and of religion concordance on compliance with treatments



Note. Estimates based on OLS regressions using equation (1) and aggregating information in both rounds of data collection. Standard errors are clustered at the slum level, confidence intervals reported at 95% level. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). *Extensive margin* is an indicator variable equal to 1 if the respondent picked up the call containing the doctor message in any of the two rounds of interventions, and 0 otherwise. *Intensive margin* is the average share of the doctor message that is listened, conditional on having picked up the call containing any message at least once. Both variables are set to zero for the control group because it was not targeted by a doctor message call. Panel A reports the effect of being assigned to the doctor message treatment group, while Panel B reports the effect of religion concordance (conditional on being assigned to the doctor message treatment group). The distributions of the share of each message that is listened by study participants (conditional on having picked up the call) is presented in Appendix Figure 3.

Table 1: The effect on preventive practices

	Preventive practices against COVID-19			**Compliance with evidence-based practices**
	Face masks and hand-washing	**Physical distancing**	**Non-evidence-based practices**	
	(1)	(2)	(3)	(4)
A. Full sample				
Doctor message	0.006	0.005	-0.003	0.041
	(0.003)	(0.004)	(0.004)	(0.015)
	[0.013 ; 0.040]	[0.195 ; 0.359]	[0.485 ; 0.502]	[0.006 ; 0.026]
Mean (Control message)	0.799	0.799	0.612	0.578
Observations	7700	7698	7699	5079
Slums	142	142	142	142
Observation rounds	2	2	2	2
B. Sample restricted to doctor message group				
Religion concordance	-0.004	-0.006	-0.009	0.026
	(0.004)	(0.004)	(0.006)	(0.022)
	[0.336 ; 0.331]	[0.202 ; 0.460]	[0.131 ; 0.418]	[0.241 ; 0.449]
Mean (religion discordant)	0.807	0.806	0.613	0.604
Observations	3851	3849	3851	2519
Slums	142	142	142	142
Observation rounds	2	2	2	2
C. Sample restricted to control group				
Religion concordance	-0.002	-0.007	-0.006	-0.017
	(0.004)	(0.004)	(0.005)	(0.021)
	[0.622 ; 0.626]	[0.126 ; 0.427]	[0.260 ; 0.590]	[0.437 ; 0.684]
Mean (religion discordant)	0.800	0.802	0.615	0.586
Observations	3849	3849	3848	2560
Slums	142	142	142	141
Observation rounds	2	2	2	2

Notes. Estimates based on OLS regressions using equation (1). Panel B restricts the sample to participants allocated to the doctor message, Panel C restricts the sample to participants allocated to the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table, separate for panels A and B. Dependent variables in columns (1)–(3) indicate respondent's level of agreement in the way they protect themselves against COVID-19, measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. In particular, column (1) *Face masks and hand-washing* concerns the average agreement with wearing face masks in crowded places and washing hands with soap more frequently and for longer; column (2) *Social distancing* concerns keeping physical distance with other people; column (3) *Non-evidence-based practices* concerns the average agreement with relying on a stronger immune system and on eating a vegetarian diet. Column (4) *Compliance with evidence-based practices* is an indicator variable equal to 1 if the respondent complied with a randomly-asked World Health Organization's recommendation to protect from infection (either leaving the slum and receiving visitors or washing hands), and 0 otherwise. This outcome was collected only for a random sub-set of study participants. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table 2: The effect on fact-checking and trust in information

	Fact-checking	**Trust in information shared by...**	
		Doctors and health experts	**Other citizens**
	(1)	(2)	(3)
A. Full sample			
Doctor message	-0.023	0.002	-0.002
	(0.010)	(0.003)	(0.004)
	[0.029 ; 0.088]	[0.449 ; 0.703]	[0.552 ; 0.561]
Mean (Control message)	0.648	0.801	0.685
Observations	7700	7700	7700
Slums	142	142	142
Observation rounds	2	2	2
B. Sample restricted to doctor message group			
Religion concordance	0.006	-0.001	-0.002
	(0.015)	(0.006)	(0.005)
	[0.687 ; 0.974]	[0.813 ; 0.823]	[0.740 ; 0.934]
Mean (religion discordant)	0.674	0.804	0.683
Observations	3851	3851	3851
Slums	142	142	142
Observation rounds	2	2	2
C. Sample restricted to control group			
Religion concordance	-0.017	-0.005	-0.006
	(0.017)	(0.005)	(0.005)
	[0.321 ; 0.558]	[0.334 ; 0.344]	[0.243 ; 0.544]
Mean (religion discordant)	0.655	0.803	0.688
Observations	3849	3849	3849
Slums	142	142	142
Observation rounds	2	2	2

Notes. Estimates based on OLS regressions using equation (1). Panel B restricts the sample to participants allocated to the doctor message, Panel C restricts the sample to participants allocated to the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table, separate for panels A and B. Dependent variables in column (1) *Fact-checking* is an indicator variable equal to 1 if the respondent always or very frequently check the truthfulness of information shared or discussed with family and friends, and 0 otherwise; column (2) *Trust in doctors and health experts* is the level of trust in the information shared by doctors and health experts; column (3) *Trust in other citizens* is the average level of trust in the information shared by people from UP and by people from other religions. Variables in columns (2)–(3) are measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table 3: The effect on misinformation

	Opinions	**Misconceptions**	**Opinions**		**Misconceptions**	
			In-group	**Out-group**	**In-group**	**Out-group**
	(1)	(2)	(3)	(4)	(5)	(6)
A. Full sample						
Doctor message	-0.000	-0.008	0.007	-0.005	0.004	-0.016
	(0.003)	(0.005)	(0.005)	(0.005)	(0.007)	(0.006)
	[0.882 ; 0.892]	[0.074 ; 0.260]	[0.180 ; 0.490]	[0.277 ; 0.575]	[0.557 ; 0.799]	[0.005 ; 0.017]
Mean (Control message)	0.617	0.489	0.617	0.622	0.483	0.494
Observations	7700	7700	6704	7700	5182	6704
Slums	142	142	142	142	142	142
Observation rounds	2	2	2	2	2	2
B. Sample restricted to doctor message group						
Religion concordance	0.003	-0.004	-0.000	0.003	-0.023	0.006
	(0.005)	(0.007)	(0.008)	(0.007)	(0.009)	(0.008)
	[0.507 ; 0.905]	[0.537 ; 0.895]	[0.958 ; 0.958]	[0.644 ; 0.879]	[0.011 ; 0.050]	[0.464 ; 0.913]
Mean (religion discordant)	0.615	0.483	0.625	0.615	0.497	0.476
Observations	3851	3851	3340	3851	2587	3340
Slums	142	142	142	142	142	142
Observation rounds	2	2	2	2	2	2
C. Sample restricted to control group						
Religion concordance	-0.004	0.003	-0.007	0.002	0.001	0.002
	(0.004)	(0.007)	(0.009)	(0.006)	(0.012)	(0.008)
	[0.398; 0.899]	[0.680; 0.974]	[0.484 ; 0.935]	[0.745 ; 0.984]	[0.952 ; 0.948]	[0.848 ; 0.978]
Mean (religion discordant)	0.619	0.487	0.620	0.620	0.483	0.493
Observations	3849	3849	3364	3849	2595	3364
Slums	142	142	140	142	141	140
Observation rounds	2	2	2	2	2	2

Notes. Estimates based on OLS regressions using equation (1). Panel B restricts the sample to participants allocated to the doctor message, Panel C restricts the sample to participants allocated to the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table, separate for panels A and B. Dependent variables by column indicate respondent's level of agreement with statements reporting public views concerning opinions and misconceptions. *Opinions* are views that are not necessarily based on facts or knowledge, while *misconceptions* are incorrect views based on faulty knowledge or understanding. For the rationale behind such distinction in the measurement of misinformation, refer to Section 3. Statements are aggregated by averaging responses using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. In columns (1)–(2), the outcome variables include all statements, independently from the person with whom the statements is associated. In column (3) and (5), the outcome variables include only statements from an interlocutor with the same religion of the respondent. In column (4) and (6), the outcome variables include only statements from an interlocutor with a religion different from the one of the respondent or from the generic term "people". In columns (3)–(6), identity of the interlocutor is randomly allocated at respondent level. Individual statements and categorization are described in Appendix B.2. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). Differences in observations across columns are due to unit non-response.

ONLINE APPENDIX

Social Proximity and Misinformation: Experimental Evidence from a Mobile Phone-Based Campaign in India

Alex Armand, Britta Augsburg, Antonella Bancalari and Kalyan Kumar Kameshwara

A Variable definition

Variable	Description
Respondent's characteristics	
Gender	Indicator variable equal to 1 for male respondents, and 0 for female respondents.
Head is male	Indicator variable equal to 1 if household head is male, and 0 otherwise.
Muslim	Indicator variable equal to 1 if respondent is Muslim, and 0 otherwise.
Incomplete primary education	Indicator variable equal to 1 if household head did not complete primary education, and 0 otherwise.
Caste: general	Indicator variable equal to 1 if respondent belongs to General caste, and 0 if the respondent is of other backward caste, scheduled caste, or scheduled tribe.
Share females	Number of women in the household.
No children	Indicator variable equal to 1 if household has no children (less than five years old), and 0 if household has children.
BPL ration card	Indicator variable equal to 1 if household possess a below poverty line ration card, and 0 if it does not.
Own dwelling	Indicator variable equal to 1 if respondent owns the dwelling, and 0 otherwise.
Own latrine	Indicator variable equal to 1 if household has its own latrine, and 0 otherwise.
Member with COVID-19	Indicator variable equal to 1 if any household member has tested positive with COVID-19, and 0 otherwise.
COVID-19 symptoms known	Indicator variable equal to 1 if any household member has COVID-19 symptoms, and 0 otherwise.
Trust in government	Indicator variable equal to 1 if respondent trusts or strongly trusts information shared by government officials (measured at the baseline), and 0 otherwise. Trust is measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust.
Trust in religious leaders	Indicator variable equal to 1 if respondent trusts or strongly trusts information shared by religious leaders (measured at the baseline), and 0 otherwise. Trust is measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust.
Traditional sources of information	Indicator variable equal to 1 if respondent uses traditional sources to gain information about coronavirus pandemic (tv/radio, printed media, government letters), and 0 otherwise. Measured at baseline only.

(continued on next page)

Variable	Description
Technology as sources of information	Indicator variable equal to 1 if respondent uses technology as source of information on coronavirus pandemic (online news sites, social media, Whatsapp, messenger, emails, voice/text messages), and 0 otherwise. Measured at baseline only.
Intervention (% listened)	
Doctor message	Proportion of the audio message containing the doctor message that is listened by the respondent. It is coded as 0 for the respondents in the control group. Information is obtained from administrative data.
Religion-concordant doctor message	Proportion of the audio message containing the doctor message and being introduced by a messenger of the same religion of the respondent that is listened by the respondent. It is coded as 0 for the respondents in the control group. Information is obtained from administrative data.
Religion-concordant control message	Proportion of the audio message containing the control message and being introduced by a messenger of the same religion of the respondent that is listened by the respondent. It is coded as 0 for the respondents in the treatment group. Information is obtained from administrative data.
Extensive margin (doctor message)	Indicator variable equal to 1 if the respondent picked up the call containing the doctor message in any of the two rounds of interventions, and 0 otherwise. The variable is set to zero for the control group because it was not targeted by a doctor message call.
Intensive margin (doctor message)	Average share of the doctor message that is listened, conditional on having picked up the call containing any message at least once. The variable is set to zero for the control group because it was not targeted by a doctor message call.
Outcomes	
Compliance with policy guidelines	Indicator variable equal to 1 if the respondent complied with a randomly asked COVID-19 World Health Organization (consisting of either leaving the slum and receiving visitors or washing hands), and 0 otherwise. The question was asked to a random sub-set of the sample to improve data quality by allowing for a shorter phone interview.
Face masks and hand-washing	Respondent's level of agreement with wearing face masks in crowded places and washing hands with soap more frequently and for longer to protect themselves against COVID-19. Agreement is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree.
Fact-checking	Indicator variable equal to 1 if the respondent always or very frequently check the truthfulness of information shared or discussed with family and friends, 0 otherwise.
Opinions	Respondent's average level of agreement with statements reporting public views concerning opinions. Statements are aggregated by averaging responses using a re-scaled likert scale in which 0 refers to strongly disagree and 1 refers to strongly agree. Individual statements and categorization are described in Appendix B.2.
Opinions (in-group)	Respondent's average level of agreement with statements reporting public views concerning opinions. Statements are aggregated by averaging responses using a re-scaled likert scale in which 0 refers to strongly disagree and 1 refers to strongly agree, and including only statements from a person with the same religion of the respondent.. Individual statements and categorization are described in Appendix B.2.

(continued on next page)

Variable	Description
Opinions (out-group)	Respondent's average level of agreement with statements reporting public views concerning opinions. Statements are aggregated by averaging responses using a re-scaled likert scale, where 0 refers to strongly disagree and 1 to strongly agree, and including only statements from a person with a religion different from the one of the respondent or from the generic term "people". Individual statements and categorization are described in Appendix B.2.
Misconceptions	Respondent's average level of agreement with statements reporting public views concerning misconceptions. Statements are aggregated by averaging responses using a re-scaled likert scale in which 0 refers to strongly disagree and 1 refers to strongly agree. Individual statements and categorization are described in Appendix B.2.
Misconceptions (in-group)	Respondent's average level of agreement with statements reporting public views concerning opinions. Statements are aggregated by averaging responses using a re-scaled likert scale in which 0 refers to strongly disagree and 1 refers to strongly agree, and including only statements from a person with the same religion of the respondent.. Individual statements and categorization are described in Appendix B.2.
Misconceptions (out-group)	Respondent's average level of agreement with statements reporting public views concerning opinions. Statements are aggregated by averaging responses using a re-scaled likert scale in which 0 refers to strongly disagree and 1 refers to strongly agree, and including only statements from a person with a religion different from the one of the respondent or from the generic term "people". Individual statements and categorization are described in Appendix B.2.
Social distancing	Respondent's level of agreement with keeping physical distance with other people to protect themselves against COVID-19. Agreement is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree.
Trust in doctors and health experts	Respondent's level of trust in the information shared by doctors and health experts. Trust is measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust.
Trust in other citizens	Respondent's average level of trust in the information shared by people from UP and by people from other religions. Trust is measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust.
Unconfirmed practices	Respondent's level of agreement with relying on a stronger immune system and on eating a vegetarian diet as preventive practices against COVID-19. Agreement is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree.
Heterogeneity dimensions	
Muslim households in the slum	Share of Muslim households in the slum. High % of Muslim households in the slum is an indicator variable equal to 1 if the share of Muslim households in the slum is below the median of the sample distribution, and 0 otherwise.

(continued on next page)

Variable	Description
Strength of religious identity	Strength of religious identity is measured as the average of the following three variables: 1) respondent's level of trust in the information shared by religious leaders, measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust; 2) respondent's agreement with the statement ”when I take important decisions, my religious faith/philosophy of life plays a considerable role”, measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree; 3) respondent's disagreement with the statement ”I have friends that are of a different religion than me”, measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. High strength of religious identity is an indicator variable equal to 1 if the strength of religious identity is below the median of the sample distribution, and 0 otherwise.
Trust in government	High trust is an indicator variable equal to 1 if the respondent trusts or strongly trust information from government officials, and 0 otherwise.
Social desirability	Social desirability is measured using the short version of the Marlowe–Crowne Social Desirability Scale (MC–SDS). High social desirability is an indicator equal to 1 if social desirability is below the median of the sample distribution, and 0 otherwise.

B Study setting

The Indian state of Uttar Pradesh (UP) provides an ideal setting to study these issues. Out of 29 states, it is the largest (home to 200 million people), the 4th most-densely populated, and the 6th in terms of share of population living in slums (corresponding to more than 6 million people) (Government of India, 2011). While UP presents a higher poverty rate as compared to the average for India (29.43% versus 21.92%, Reserve Bank of India, 2019), its slum population is highly comparable to the average slum population in the country. The share of adult males (0.53 in UP versus 0.52 in India), of adult females (0.47 versus 0.48), and of children (0.14 versus 0.12), as well as the sex ratio (1.12 versus 1.08) and the share belonging to Scheduled Castes (0.22 versus 0.20) are indicative of close similarities between these two populations. In terms of literacy rates, the average slum in UP outperforms the one of the whole India (0.69 versus 0.78).

This study focuses on the two largest urban agglomerations of UP, Lucknow and Kanpur. Figure B1 shows their geographic location. Similar to many expanding cities in South Asia and Sub-Saharan Africa, Lucknow and Kanpur are characterized by a relatively large prevalence of informal settlements, and a prospect of rapid population growth. In 2015, among all urban agglomerations with more than 300,000 inhabitants, Lucknow and Kanpur were respectively the 129th and 141st worldwide (United Nations, 2019). In the period 2015–2035, Lucknow is expected to grow from 3.2 to 5.2 million (+59%), and Kanpur from 3 to 4.1 million inhabitants (+37%). Across agglomerations of similar size (1.5–4 million inhabitants), this growth prospect is similar to cities such as Accra (Ghana), Amman (Jordan), Jaipur (India), or Hyderabad (Pakistan). Among largest cities, this is similar to Karachi (Pakistan), Cairo (Egypt) or Manila (The Philippines). In terms of proportion of slum households, the share of households living in slums is 12.95% in Lucknow and 14.5% in Kanpur. This is comparable to other major cities in India, such as Delhi, where 14.66% of households live in slums (Government of India, 2011). Figure B2 shows the distribution of the share of the Muslim population at slum level in the study area.

B.1 The COVID-19 pandemic and infodemic in the study area

Figure B3 reports the time series of the number of COVID-19 cases and deaths in UP from the beginning of 2020 until April 2021. Figure 1 shows the timeline of study activities, and a comparison with the restrictions that were in place in our study location. Figure B4 displays trends in social media interactions

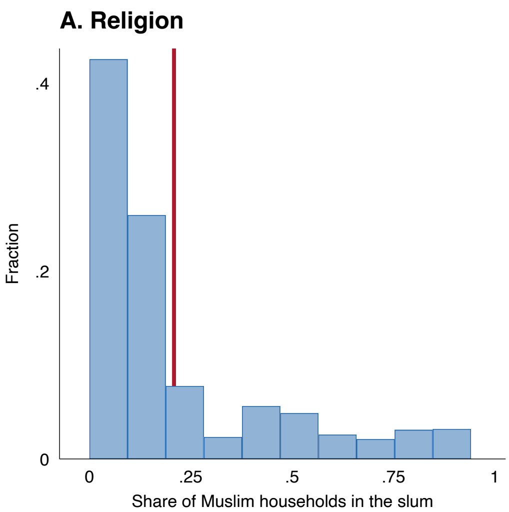
Figure B1: Study location

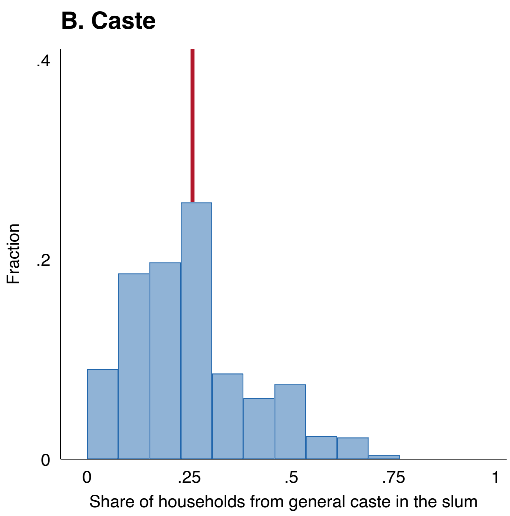
Panel A. State

Panel B. Cities

Notes. Panel A shows the location of the state of UP, while Panel B show the location of Lucknow and Kanpur in the state. Basemap source: Esri.

Figure B2: Distribution of Muslim and general caste population at slum level

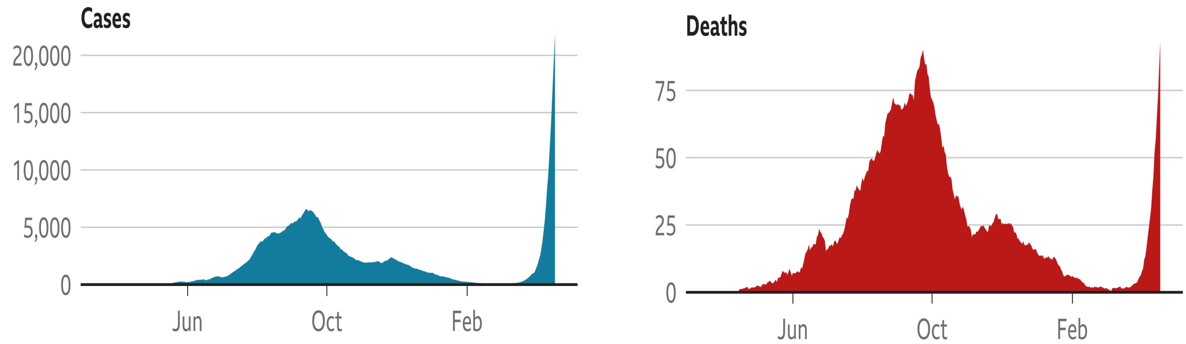




Notes. The Muslim and the general caste population is computed at slum level. The vertical lines indicate the sample mean.

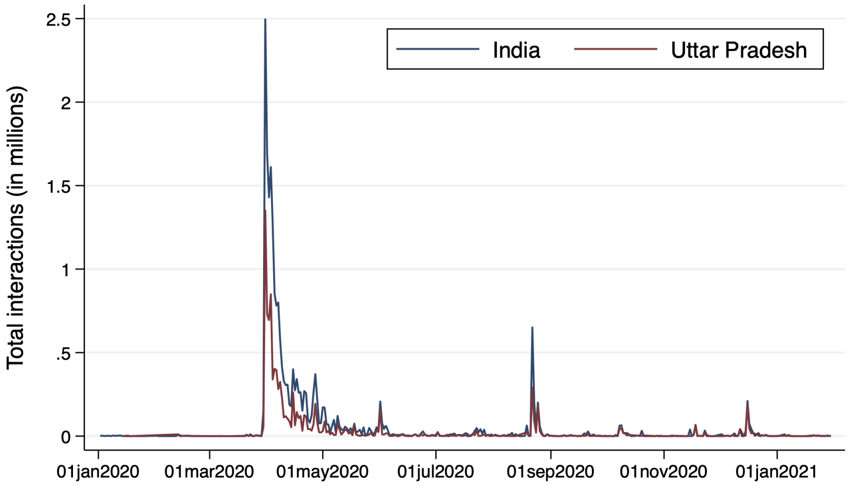
(Facebook and Facebook-related media) targeting and blaming the Muslim population for the spread of the virus. It highlights the sharp increase early on in the pandemic in India and UP in particular.

Figure B3: Daily COVID-19 cases and deaths in Uttar Pradesh



Notes. Reported number of cases and deaths from the beginning of 2020 until April 2021. The source of data is the Indian Ministry of Health and Family Welfare. Graphic elaboration produced by BBC (https://www.bbc.com/news/world-asia-india-56799303).

Figure B4: Muslim-related social media posts about COVID-19 spread

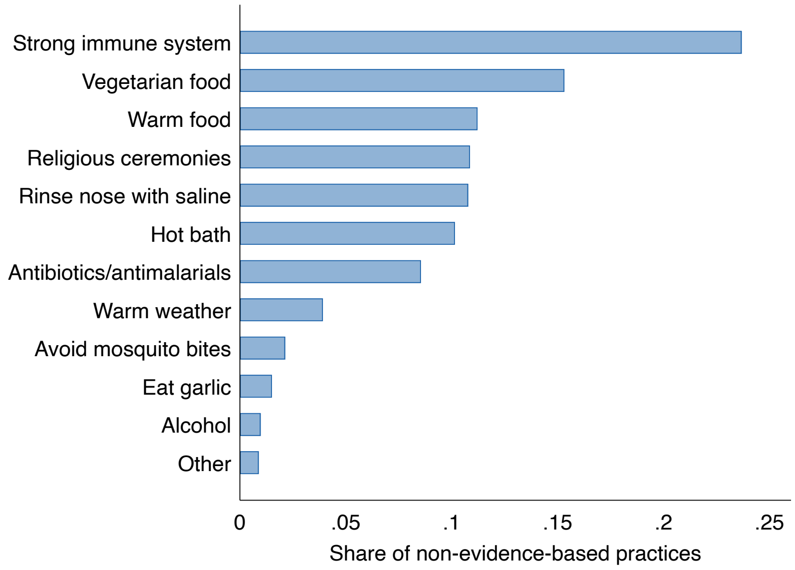


Notes. The data shows the trends between January 2020 and February 2021. The vertical axis depicts the total number of times a Facebook post created on a given date is liked, shared or commented upon. The blue and red lines show the trends across India and UP, respectively. Trends are constructed using data from Facebook's Crowd Tangle Team (2020). To select posts on misinformation related to the Muslim population we select the following keywords (both in Hindi and in latin transliteration): Corona_Jihad, CoronaJihad, Corona Jihad, Tablighi, Tablighi jamat, Tablighijamat, Tablighi_jamat, jihadivirus, Muslim virus, Nizamuddin Markaz. *CoronaJihad*, *Jihadivirus* and *Muslim virus* were popular phrases used on social media accusing the Muslim population of using coronavirus for purposes of jihad. *Tablighi jamat* is an international Islamic missionary movement which organized a mass religious congregation in Delhi in March 2020, and became a target word when a group of attendees tested positive with COVID-19. *Nizamuddin Markaz* refers to the mosque where the religious congregation took place and was popular on social media.

B.2 Preventive practices and misinformation: measurement

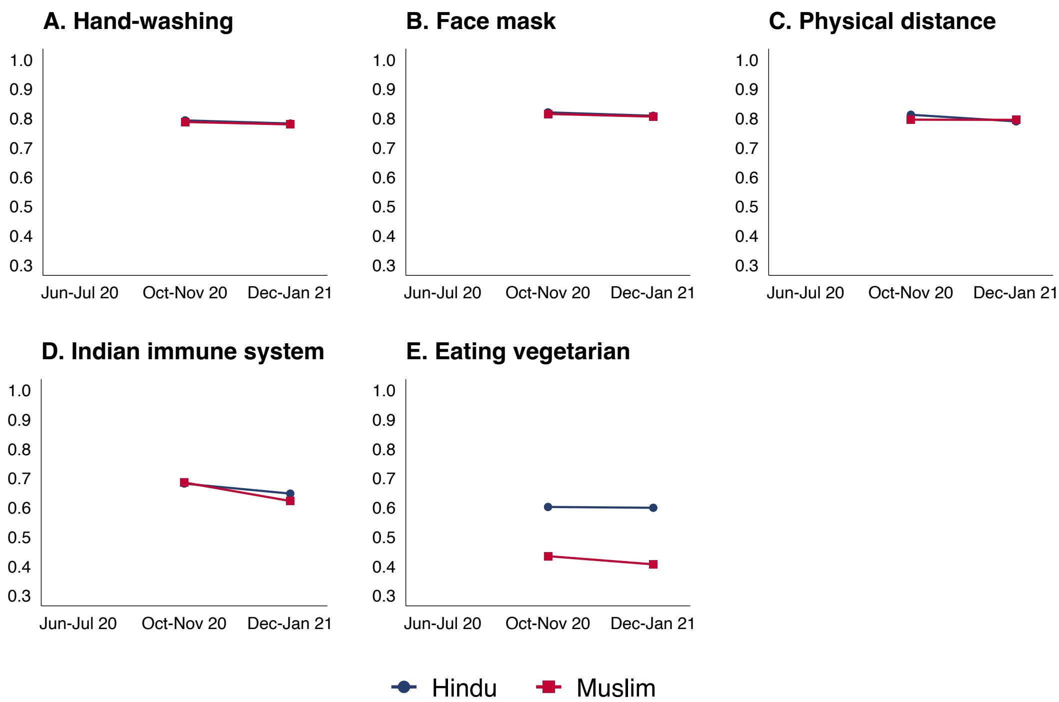
To understand misinformation related to preventive practices at baseline, we collected misconceptions from sample respondents using an open-ended question about preventive practices. Figure B5 reports the main reported misconceptions on the way to protect against COVID-19. Table B1 estimates the likelihood of identifying misinformation from the baseline sample as a function individual characteristics. Figure B6 shows descriptive statistics over time for agreement with evidence-based and non-evidence-based preventive practices, as reported by the control group.

Figure B5: Non-evidence-based preventive practices, at baseline



Notes. Respondents were asked about what, according to their opinion, would help in protecting them, or their family, from getting coronavirus. The questions were open-ended and responses were categorized into evidence-based and non-evidence-based preventive practices. We present the share of each non-evidence-based practice out of all non-evidence-based practices reported by the respondent. The sample is restricted to baseline observations and to respondents that reported at least one non-evidence-based practice.

Figure B6: Preventive practices against COVID-19, by respondent's religion



Notes. Hindu and *Muslim* refer to the religion of the respondent. Each figure shows the average level of agreement with evidence-based and non-evidence-based preventive practices. Each outcome is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. The outcomes are missing at baseline because preventive practices were collected using an open-ended question in the baseline survey and with individual questions in the follow-up surveys.

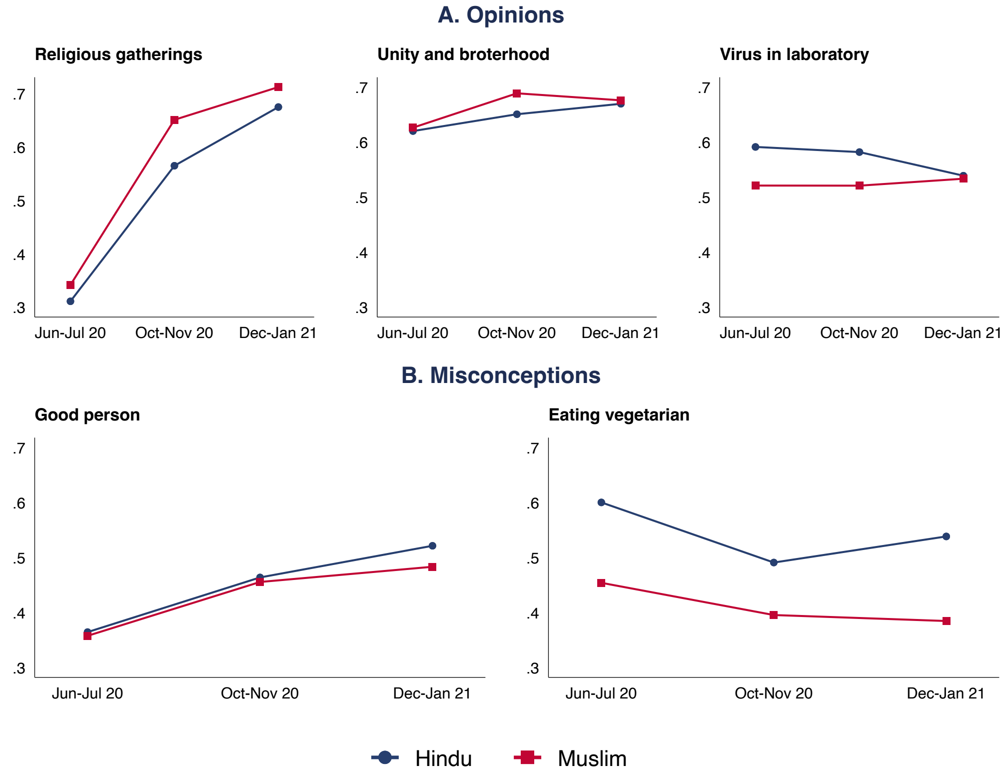
Table B1: Baseline correlates of non-evidence-based preventive practices

	Number of reported preventive practices		
	Evidence-based	**Non-evidence-based**	**At least 1 non-evidence-based practice**
	(1)	(2)	(3)
Respondent is male	-0.03	-0.03	-0.01
	(0.08)	(0.03)	(0.02)
Head is male	0.11	0.05	0.02
	(0.07)	(0.04)	(0.02)
Respondent is Muslim	-0.08	-0.04	-0.01
	(0.10)	(0.03)	(0.02)
Caste: General	0.24***	0.05	0.02
	(0.08)	(0.04)	(0.02)
Age	-0.01**	-0.00	-0.00
	(0.00)	(0.00)	(0.00)
Household members	0.01	0.01	0.00
	(0.01)	(0.01)	(0.00)
Share females	-0.03	0.06	0.03
	(0.09)	(0.06)	(0.03)
No children	0.03	-0.01	-0.01
	(0.06)	(0.03)	(0.02)
Own dwelling	0.13	0.01	0.01
	(0.10)	(0.03)	(0.02)
BPL ration card	-0.06	-0.01	-0.00
	(0.07)	(0.02)	(0.01)
Member with COVID-19	0.23	0.13**	0.11***
	(0.18)	(0.05)	(0.03)
COVID-19 symptoms known	-0.03	-0.07*	-0.05***
	(0.09)	(0.04)	(0.02)
Slums	142	142	142
Households	3,975	3,975	3,975

Notes. The dependent variables in column (1) *Number of evidence-based preventive practices* is the number of practices reported by the respondent that are evidence-based; (2) *Number of non-evidence-based preventive practices* is the number of practices reported by the respondent that are non-evidence-based; column (3) *At least 1 non-evidence-based practice* is an indicator equal to one if the respondent reported at least 1 non-evidence-based preventive practice, and 0 otherwise. All specifications include strata (city and managed by main provider) variables as controls. Standard errors clustered at the slum level are presented in parenthesis.

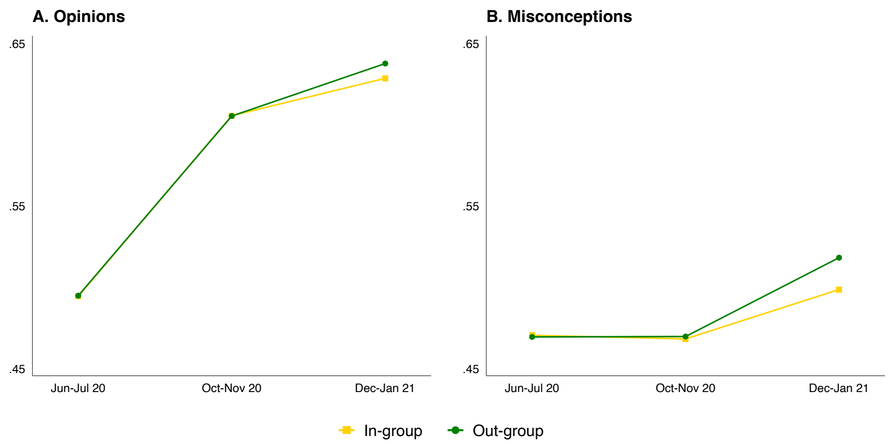
Figure B7 shows average levels of agreement with public opinions and misconception statements for the control group, aggregating statements reported by all options for [Person]. We compute agreement with opinions and with misconceptions by averaging within each category all non-missing observations for agreement with each individual statements. To understand further the role of agreement of the respondent with statements reported by someone of the same religion, we average separately statements reported by a person of the same religion or statements reported by someone of a different religion or by the generic ”people”. Figure B8 shows the evolution over time of the average agreement with opinions and misconceptions, distinguishing along these dimensions.

Figure B7: Views about COVID-19, by respondent's religion



Notes. Hindu and *Muslim* refer to the religion of the respondent. Each figure shows the average level of agreement with public views over time in the control group, distinguishing by the religion of the respondent. Panel A focuses on opinions, while Panel B focuses on misconceptions. *Opinions* are views that are not necessarily based on facts or knowledge, while *misconceptions* are incorrect views based on faulty knowledge or understanding. For the rationale behind such distinction in the measurement of misinformation, refer to Section 3. Each outcome is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. Details about the survey instrument is described in Section 3. For each individual, we use the agreement with the statement independently from the interlocutor reporting the statement.

Figure B8: Views about COVID-19, by interlocutor assigned to statements



Notes. Opinions are views that are not necessarily based on facts or knowledge, while *misconceptions* are incorrect views based on faulty knowledge or understanding. For the rationale behind such distinction in the measurement of misinformation, refer to Section 3. *In-group* averages only statements in which the respondent and the interlocutor assigned to the statement share the same religion. *Out-group* averages only statements in which the respondent and the interlocutor assigned with the statement do not share the same religion or the statement is associated with the generic "people". Panel A shows the average level of agreement with opinions, while Panel B shows the average level of agreement with misconceptions. Each outcome ranges from 0 to 1, with 0 referring to strongly disagree and 1 referring to strongly agree. Details about the survey instrument is described in Section 3.

B.3 Intervention content

The **control message** was 39 seconds long and the informative content included a gossip about popular actresses of Bollywood. The **doctor message** treatment included two messages sent as voice messages to the study participants twice during the study. The first message was 95 seconds long and the second message was 123 seconds long. Both treatment messages had a very similar structure, but addressed a different topic.

The script of the messages reads as follows. This Introduction is included in both control and doctor message.

> **Introduction**
>
> *Messenger*: Greeting! ['Namaste' or 'Salam Walekum', according to randomization] I am a resident of UP and like me, you might also be confused about information shared on social media. If this is the case, then the following messages might be helpful for you. After watching this video, if you answer the question correctly, then you can get a chance to win the lottery of up to Rs. [high or low amount, according to randomization] in the form of mobile recharge.
>
> **First round of the doctor message**
>
> So, let's listen to what the renowned doctors have to say about this question: Is it correct that being a vegetarian or eating only a vegetarian diet fully protects from contracting the virus?
>
> *Doctor 1*: No, this misconception is spread inside the society, there is no such thing. You can see that people all over the world are non-vegetarians or vegetarians and everyone is getting infected.
>
> *Doctor 2*: Yes, it is true that vegetarian food is good food and healthy food. It also increases some immunity. But it is a misconception that if we take vegetarian food then there is no need to do other measures and we will not be infected from Corona.
>
> *Doctor 3*: The most important thing to avoid coronavirus is to use masks, social distance, wash hands frequently with soap, use of sanitizer.
>
> **Second round of the doctor message**
>
> So, let's listen to what the renowned doctors have to say about this question: Is it correct that we Indians need not worry about the coronavirus because our immune system is quite strong?
>
> *Doctor 1*: This is a myth. It can lead to false beliefs among people that they we will not get the disease. Please do not live with this false belief. In fact, the Indian population has contracted many diseases in the past. Please look at how many people are contracting the virus: the number of people getting the disease is increasing in the country and the world.

Doctor 2: Coronavirus is a threat to the entire human civilization today. Do not stay under the misconception that we are immune to the virus. We need to be careful, protect ourselves from the virus, and follow the guidelines set by the government.

Doctor 3: Maintain physical distance, use face mask and sanitizer and take nutritious diet. All these things are being emphasized, so keep doing all these. Avoid fake news and the confusion that is being spread, and follow all these things.

All rounds of the doctor message

Messenger: We thank the doctors. Now, things are clear for me and hopefully for you too. If you have understood the message, please spread it to others. If each of us makes this contribution, we can save a lot of lives together. To enter the lottery, you would have to answer the following question correctly:

(First round of Doctor message) "Can we Indians be carefree and not worry about coronavirus because our immune system is very strong?"

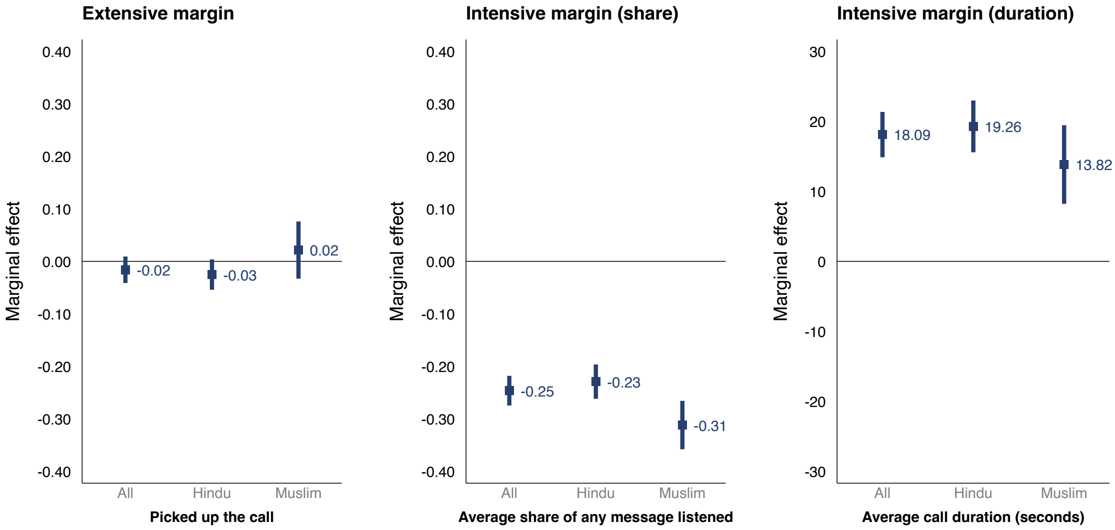
(Second round of Doctor message) "When eating pure vegetarian, you cannot get coronavirus."

Press 1 for true or 2 for false.

For the **religion concordance** treatment, we randomly varied the greetings used by the messenger in the very beginning of the message. For a random subset of participants, the messenger's message was introduced with *Namaste*, signaling its Hindu identity, while the rest was introduced by *Salam Walekum*, signaling the Muslim identity of the messenger. This variation does not affect the duration of messages.

Figure B9 shows estimates of the effect of the doctor message treatment and of the religion concordance treatment (within the doctor message treatment group) on the probability to pick up the call containing any voice message (left panels), and, conditional on having picked up the call, on the average share of the message that is listened (middle panels), and on the average duration of the call (right panels). On average, respondents in the doctor message treatment group listen to a lower share of the message (25 percentage points less), but are exposed to a call duration that is longer by 18 seconds.

Figure B9: Effect on exposure to mobile phone calls from the intervention



Note. Estimates based on OLS regressions using equation (1). Standard errors clustered at the slum level, confidence intervals reported at 90% level. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). *Extensive margin* is an indicator variable equal to 1 if the respondent picked up the call containing any message from the intervention in any of the two rounds of interventions, and 0 otherwise. *Intensive margin (share)* is the average share of any message that is listened, conditional on having picked up the call at least once. *Intensive margin (duration)* is the average duration of the call (reported in seconds), conditional on having picked up the call at least once. Panel A reports the effect of being assigned to the doctor message treatment group, while Panel B reports the effect of religion concordance (conditional on being assigned to the doctor message treatment group).

C Study population, balance and attrition

Tables C1 reports descriptive statistics for observable characteristics of the respondent and the household and of outcome variables. Column (1) reports the mean and standard deviation of the each variable for the control group in the doctor message treatment, while column (2) shows the difference to this mean of those who were sent the doctor messages. Column (3) reports the joint sample size. Columns (4)–(6) report the same information comparing those that were sent the message with a Muslim messenger to those that were sent a message with a Hindu messenger, hence focusing on the Doctor sample only. Table C2 reports correlates of attrition.

Table C1: Respondent characteristics and attrition

	Full sample			**Sample restricted to doctor message**		
	Control mean	Difference with doctor message treatment	N	Muslim messenger (mean)	Difference with Hindu messenger	N
	(1)	(2)	(3)	(4)	(5)	(6)
A. Respondent characteristics						
Respondent is male	0.79	-0.00	3983	0.78	0.02	1996
	[0.41]	(0.01)		[0.41]	(0.02)	
Head is male	0.82	-0.00	3983	0.81	-0.00	1996
	[0.39]	(0.01)		[0.39]	(0.02)	
Respondent is Muslim	0.21	-0.00	3983	0.23	0.01	1996
	[0.41]	(0.00)		[0.42]	(0.01)	
Caste: General	0.16	0.01	3983	0.16	0.02	1996
	[0.36]	(0.01)		[0.36]	(0.02)	
Age	39.77	-0.50	3983	39.34	-0.15	1996
	[11.41]	(0.38)		[11.59]	(0.47)	
Household members	5.17	0.05	3983	5.32	-0.18*	1996
	[2.20]	(0.07)		[2.37]	(0.10)	
Share females	0.35	-0.01	3983	0.34	0.01	1996
	[0.16]	(0.01)		[0.16]	(0.01)	
No children	0.72	-0.02	3983	0.71	-0.02	1996
	[0.45]	(0.01)		[0.45]	(0.02)	
Own dwelling	0.73	-0.00	3983	0.73	0.00	1996
	[0.44]	(0.01)		[0.44]	(0.02)	
Own latrine	0.61	0.00	3977	0.61	0.02	1995
	[0.49]	(0.02)		[0.49]	(0.02)	
BPL ration card	0.38	-0.01	3983	0.38	-0.01	1996
	[0.49]	(0.02)		[0.49]	(0.02)	
Member with COVID-19	0.12	0.01	3983	0.14	-0.01	1996
	[0.32]	(0.01)		[0.34]	(0.01)	
COVID-19 symptoms known	1.60	-0.03	3975	1.58	-0.02	1991
	[0.66]	(0.02)		[0.66]	(0.02)	
Trust: Doctors and health experts	0.95	0.01	1586	0.96	-0.01	753
	[0.23]	(0.01)		[0.20]	(0.01)	
Trust: Government official	0.84	-0.01	1586	0.83	-0.01	753
	[0.37]	(0.02)		[0.38]	(0.03)	
B. Attrition						
Attrition BL-any FU	0.13	-0.01	3983	0.14	-0.01	1996
	[0.34]	(0.01)		[0.34]	(0.01)	
Attrition BL-FU1	0.28	0.00	3983	0.29	-0.00	1996
	[0.45]	(0.01)		[0.45]	(0.02)	
Attrition BL-FU2	0.24	-0.00	3983	0.23	-0.00	1996
	[0.42]	(0.01)		[0.42]	(0.02)	

Notes. Column (1) reports the mean and standard deviation of the each variable for the control group in the doctor message treatment, while column (2) shows the difference to this mean of those who were sent the doctor messages. Column (3) reports the joint sample size. Columns (4)–(6) report the same information comparing those that were sent the message with a Muslim messenger to those that were sent a message with a Hindu messenger, hence restricting the sample to the doctor message treatment group.

Table C2: Correlates of attrition

	Attrition			
	Full sample		Sample restricted to doctor message group	
	(1)	(2)	(3)	(4)
Doctor message	-0.01	-0.01		
	(0.01)	(0.01)		
Doctor message x Muslim		0.01		
		(0.02)		
Religion concordance intro			-0.01	-0.02
			(0.01)	(0.02)
Religion concordance intro x Muslim				0.04
				(0.04)
Respondent is male	0.03**	0.03**	0.04*	0.04*
	(0.01)	(0.01)	(0.02)	(0.02)
Head is male	-0.04**	-0.04**	-0.04*	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)
Respondent is Muslim	0.04**	0.03	0.04*	0.02
	(0.02)	(0.02)	(0.03)	(0.03)
Caste: General	-0.02*	-0.02*	-0.03	-0.03
	(0.01)	(0.01)	(0.02)	(0.02)
Age	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Household members	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share females	0.00	0.00	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
No children	-0.01	-0.01	0.00	0.00
	(0.01)	(0.01)	(0.02)	(0.02)
Own dwelling	-0.04**	-0.03**	-0.04*	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)
BPL ration card	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.02)	(0.02)
Member with COVID-19	-0.02	-0.02	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
COVID-19 symptoms known	0.02**	0.02**	0.02	0.02*
	(0.01)	(0.01)	(0.01)	(0.01)
Attrition Rate	0.13	0.13	0.13	0.13
Slums	142	142	142	142
Households	3,975	3,975	1,991	1,991

Notes. The dependent variable is an indicator variable equal to 1 if a households was neither re-interviewed in follow-up 1 or follow-up 2, and 0 otherwise. Standard errors are clustered at the slum level and presented in parenthesis. Columns (1) and (2) are for the full sample, columns (3) and (4) restrict the sample to the doctor message treatment group.

D Additional analysis

D.1 Estimates of treatment effects using 2SLS

Appendix Tables D1–D3 show 2SLS estimates of the treatment on the treated effect. In Panel A, the endogenous intensity of the treatment is represented by the share of the doctor message that is effectively listened by the respondent (set to zero for the control group). The variable is instrumented by the doctor message treatment indicator. In panels B and C, the endogenous intensity of the treatment is represented by the share of the religion-concordant message that is effectively listened by the respondent (set to zero for the group in which the message is religion discordant). Panel B restricts the sample to the doctor message group, while panel C restricts the sample to the control group. In both panels B and C, the endogenous take-up is instrumented by the religion concordance treatment indicator.

Table D1: The effect on preventive practices, 2SLS

	Preventive practices against COVID-19			**Compliance with**
	Face masks and hand-washing	**Physical distancing**	**Non-evidence-based practices**	**evidence-based practices**
	(1)	(2)	(3)	(4)
A. Full sample				
% listened of doctor message	0.044	0.033	-0.019	0.282
	(0.018)	(0.025)	(0.027)	(0.101)
	[0.012]	[0.193]	[0.482]	[0.005]
Mean (Not listened)	0.802	0.800	0.612	0.594
Observations	7700	7698	7699	5079
Slums	142	142	142	142
Observation rounds	2	2	2	2
B. Sample restricted to doctor message group				
% listened of religion-concordant message	-0.024	-0.038	-0.060	0.175
	(0.025)	(0.030)	(0.040)	(0.148)
	[0.335]	[0.201]	[0.132]	[0.239]
Mean (religion discordant)	0.807	0.806	0.613	0.604
Observations	3851	3849	3851	2519
Slums	142	142	142	142
Observation rounds	2	2	2	2
C. Sample restricted to control group				
% listened of religion-concordant message	-0.008	-0.026	-0.024	-0.059
	(0.016)	(0.017)	(0.021)	(0.076)
	[0.620]	[0.123]	[0.256]	[0.433]
Mean (religion discordance)	0.797	0.797	0.614	0.577
Observations	3849	3849	3848	2560
Slums	142	142	142	141
Observation rounds	2	2	2	2

Notes. Estimates based on 2SLS regressions. Panel B restricts the sample to participants allocated to the doctor message, Panel C restricts the sample to participants allocated to the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables in columns (1)–(3) indicate respondent's level of agreement with the way they protect themselves against COVID-19, measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. In particular, column (1) *Face masks and hand-washing* concerns the average agreement with wearing face masks in crowded places and washing hands with soap more frequently and for longer; column (2) *Social distancing* concerns keeping physical distance with other people; column (3) *Unconfirmed* concerns the average agreement with relying on a stronger immune system and on eating a vegetarian diet. Column (4) *Compliance with evidence-based practices* is an indicator variable equal to 1 if the respondent complied with a randomly-asked World Health Organization's recommendation to protect from infection (either leaving the slum and receiving visitors or washing hands), and 0 otherwise. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). The outcome in column (4) was collected only for a random sub-set of study participants.

Table D2: The effect on fact-checking and trust in information, 2SLS

	Fact-checking	Trust in information shared by...	
		Doctors and health experts	**Other citizens**
	(1)	(2)	(3)
A. Full sample			
% listened of doctor message	0.147	0.016	-0.015
	(0.066)	(0.021)	(0.025)
	[0.027]	[0.446]	[0.550]
Mean (Control message)	0.648	0.801	0.685
Observations	7700	7700	7700
Slums	142	142	142
Observation rounds	2	2	2
B. Sample restricted to doctor message group			
% listened of religion-concordant message	-0.037	-0.009	-0.010
	(0.091)	(0.037)	(0.029)
	[0.685]	[0.812]	[0.739]
Mean (religion discordant)	0.674	0.804	0.683
Observations	3851	3851	3851
Slums	142	142	142
Observation rounds	2	2	2
C. Sample restricted to control group			
% listened of religion-concordant message	0.064	-0.018	-0.023
	(0.064)	(0.019)	(0.019)
	[0.313]	[0.332]	[0.239]
Mean (religion discordance)	0.345	0.800	0.686
Observations	3849	3849	3849
Slums	142	142	142
Observation rounds	2	2	2

Notes. Estimates based on 2SLS regressions. Panel B restricts the sample to participants allocated to the doctor message, Panel C restricts the sample to participants allocated to the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables in column (1) *Confidence* is an indicator variable equal to 1 if the respondent does not always or very frequently check the truthfulness of information shared or discussed with family and friends, and 0 otherwise; column (2) *Trust in doctors and health experts* is the level of trust in the information shared by doctors and health experts; column (3) *Trust in other citizens* is the average level of trust in the information shared by people from UP and by people from other religions. Variables in columns (2)–(3) are measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D3: The effect on misinformation, 2SLS

	Opinions	Misconceptions	Opinions		Misconceptions	
			In-group	Out-group	In-group	Out-group
	(1)	(2)	(3)	(4)	(5)	(6)
A. Full sample						
% listened of doctor message	-0.003	-0.054	0.045	-0.034	0.027	-0.105
	(0.021)	(0.030)	(0.033)	(0.031)	(0.046)	(0.037)
	[0.882]	[0.073]	[0.177]	[0.272]	[0.554]	[0.005]
Mean (Control message)	0.617	0.489	0.617	0.622	0.483	0.494
Observations	7700	7700	6704	7700	5182	6704
Slums	142	142	142	142	142	142
Observation rounds	2	2	2	2	2	2
B. Sample restricted to doctor message group						
% listened of religion-concordant message	0.020	-0.027	-0.003	0.021	-0.150	0.037
	(0.030)	(0.044)	(0.050)	(0.045)	(0.059)	(0.051)
	[0.504]	[0.534]	[0.958]	[0.642]	[0.011]	[0.462]
Mean (religion discordance)	0.615	0.483	0.625	0.615	0.497	0.476
Observations	3851	3851	3340	3851	2587	3340
Slums	142	142	142	142	142	142
Observation rounds	2	2	2	2	2	2
C. Sample restricted to control group						
% listened of religion-concordant message	-0.014	0.011	-0.024	0.008	0.003	0.006
	(0.017)	(0.028)	(0.035)	(0.024)	(0.043)	(0.030)
	[0.395]	[0.678]	[0.482]	[0.743]	[0.951]	[0.847]
Mean (religion discordance)	0.619	0.490	0.618	0.622	0.485	0.494
Observations	3849	3849	3364	3849	2595	3364
Slums	142	142	140	142	141	140
Observation rounds	2	2	2	2	2	2

Notes. Estimates based on 2SLS regressions. Panel B restricts the sample to participants allocated to the doctor message, Panel C restricts the sample to participants allocated to the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables by column indicate respondent's level of agreement with statements reporting public views concerning opinions and misconceptions. *Opinions* are views that are not necessarily based on facts or knowledge, while *misconceptions* are incorrect views based on faulty knowledge or understanding. For the rationale behind such distinction in the measurement of misinformation, refer to Section 3. Statements are aggregated by averaging responses using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. In columns (1)–(2), the outcome variables include all statements, independently from the person with whom the statements is associated. In column (3) and (5), the outcome variables include only statements from a person with the same religion of the respondent. In column (4) and (6), the outcome variables include only statements from a person with a religion different from the one of the respondent or from the generic term "people". In columns (3)–(6), the person reporting the statement is randomly allocated at respondent level. Individual statements and categorization are described in Appendix B.2. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). Differences in observations across columns are due to unit non-response.

D.2 Effect of Hindu introduction

Tables D4–D6 show estimates of the effect of the Hindu greeting at the beginning of the message, independently from whether the message is a doctor message or a control message. Estimates using the full sample are based on equation (1), in which T_i is an indicator variable for whether the greeting in the message is Hindu, and zero otherwise.

Table D4: The effect of the Hindu greeting on preventive practices

	Preventive practices against COVID-19			**Compliance with evidence-based practices**
	Face masks and hand-washing	**Physical distancing**	**Non-evidence-based practices**	
	(1)	(2)	(3)	(4)
Hindu greeting	0.001	-0.004	0.001	0.008
	(0.003)	(0.003)	(0.004)	(0.016)
	[0.685]	[0.210]	[0.763]	[0.617]
Mean (Muslim greeting)	0.802	0.803	0.610	0.594
Observations	7700	7698	7699	5079
Slums	142	142	142	142
Observation rounds	2	2	2	2

Notes. Estimates based on OLS regressions using the full sample are based on equation (1), in which T_i is an indicator variable for whether the greeting in the message is Hindu, and zero otherwise. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables in columns (1)–(3) indicate respondent's level of agreement with the way they protect themselves against COVID-19, measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. In particular, column (1) *Face masks and hand-washing* concerns the average agreement with wearing face masks in crowded places and washing hands with soap more frequently and for longer; column (2) *Social distancing* concerns keeping physical distance with other people; column (3) *Unconfirmed* concerns the average agreement with relying on a stronger immune system and on eating a vegetarian diet. Column (4) *Compliance with evidence-based practices* is an indicator variable equal to 1 if the respondent complied with a randomly-asked World Health Organization's recommendation to protect from infection (either leaving the slum and receiving visitors or washing hands), and 0 otherwise. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). The outcome in column (4) was collected only for a random sub-set of study participants.

Table D5: The effect of the Hindu greeting on fact-checking and trust in information

	Fact-checking	**Trust in information shared by...**	
		Doctors and health experts	**Other citizens**
	(1)	(2)	(3)
Hindu greeting	0.006	-0.001	-0.001
	(0.010)	(0.004)	(0.003)
	[0.584]	[0.685]	[0.778]
Mean (religion discordance)	0.339	0.803	0.684
Observations	7700	7700	7700
Slums	142	142	142
Observation rounds	2	2	2

Notes. Estimates based on OLS regressions using the full sample are based on equation (1), in which T_i is an indicator variable for whether the greeting in the message is Hindu, and zero otherwise. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables in column (1) *Fact-checking* is an indicator variable equal to 1 if the respondent does not always or very frequently check the truthfulness of information shared or discussed with family and friends, and 0 otherwise; column (2) *Trust in doctors and health experts* is the level of trust in the information shared by doctors and health experts; column (3) *Trust in other citizens* is the average level of trust in the information shared by people from UP and by people from other religions. Variables in columns (2)–(3) are measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D6: The effect of the Hindu greeting on misinformation

	Opinions	Misconceptions	Opinions		Misconceptions	
			In-group	Out-group	In-group	Out-group
	(1)	(2)	(3)	(4)	(5)	(6)
Hindu greeting	0.002	0.003	-0.004	0.006	-0.001	0.004
	(0.003)	(0.005)	(0.006)	(0.005)	(0.007)	(0.006)
	[0.558]	[0.567]	[0.472]	[0.224]	[0.901]	[0.525]
Mean (Muslim greeting)	0.616	0.483	0.623	0.616	0.485	0.485
Observations	7700	7700	6704	7700	5182	6704
Slums	142	142	142	142	142	142
Observation rounds	2	2	2	2	2	2

Notes. Estimates based on OLS regressions using the full sample are based on equation (1), in which T_i is an indicator variable for whether the greeting in the message is Hindu, and zero otherwise. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables by column indicate respondent's level of agreement with statements reporting public views concerning opinions and misconceptions. *Opinions* are views that are not necessarily based on facts or knowledge, while *misconceptions* are incorrect views based on faulty knowledge or understanding. For the rationale behind such distinction in the measurement of misinformation, refer to Section 3. Statements are aggregated by averaging responses using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. In columns (1)–(2), the outcome variables include all statements, independently from the person with whom the statements is associated. In column (3) and (5), the outcome variables include only statements from a person with the same religion of the respondent. In column (4) and (6), the outcome variables include only statements from a person with a religion different from the one of the respondent or from the generic term "people". In columns (3)–(6), the person reporting the statement is randomly allocated at respondent level. Individual statements and categorization are described in Appendix B.2. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). Differences in observations across columns are due to unit non-response.

D.3 Effect of higher versus lower financial incentives

We introduced two levels of incentives, cross-randomized across treatment arms: a lower-incentive lottery with a value of Rs. 2,500 (US$32) and a higher-incentive lottery with a value of Rs. 5,000 (US$64). Appendix Tables D7–D9 show estimates of the effect of offering a higher financial incentive estimated using the following specification restricted to either the doctor message group (Panel A) or the control group (panel B):

$$Y_{ijt} = \beta_D \, Higher_i + \alpha \, \mathbf{X}_{ij} + \delta_t + \epsilon_{ijt} \quad (3)$$

where Y_{it} are outcomes of interest of respondent i in slum j at time t. The variable $higher_i$ is an indicator variable equal to 1 if the receiver i is offered a higher financial incentive, and 0 otherwise. $\mathbf{X}_{ij}$ is a set of indicator variables for randomization strata, and δ_t are period-of-survey indicator variables. The error term ϵ_{it} is assumed to be clustered at the slum level.

Table D7: The effect of higher incentives on preventive practices

	Preventive practices against COVID-19			Compliance with evidence-based practices
	Face masks and hand-washing	**Physical distancing**	**Non-evidence-based practices**	
	(1)	(2)	(3)	(4)
A. Sample restricted to doctor message group				
Higher incentive	-0.000	-0.002	-0.005	-0.010
	(0.003)	(0.004)	(0.006)	(0.024)
	[0.985 ; 0.984]	[0.606 ; 0.935]	[0.331 ; 0.767]	[0.678 ; 0.892]
Mean (lower incentive)	0.806	0.805	0.612	0.624
Observations	3851	3849	3851	2519
Slums	142	142	142	142
Observation rounds	2	2	2	2
B. Sample restricted to control group				
Higher incentive	0.003	-0.000	0.000	-0.001
	(0.004)	(0.004)	(0.006)	(0.023)
	[0.427 ; 0.875]	[0.963 ; 1.000]	[0.978 ; 0.978]	[0.968 ; 1.000]
Mean (lower incentive)	0.798	0.798	0.612	0.580
Observations	3849	3849	3848	2560
Slums	142	142	142	141
Observation rounds	2	2	2	2

Notes. Estimates based on equation (3) restricted to the sample to participants allocated to the doctor message (Panel A), or the sample to participants allocated to the control group (Panel B). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table, separate for panels A and B. Dependent variables in columns (1)–(3) indicate respondent's level of agreement with the way they protect themselves against COVID-19, measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. In particular, column (1) *Face masks and hand-washing* concerns the average agreement with wearing face masks in crowded places and washing hands with soap more frequently and for longer; column (2) *Social distancing* concerns keeping physical distance with other people; column (3) *Unconfirmed* concerns the average agreement with relying on a stronger immune system and on eating a vegetarian diet. Column (4) *Compliance with evidence-based practices* is an indicator variable equal to 1 if the respondent complied with a randomly-asked World Health Organization's recommendation to protect from infection (either leaving the slum and receiving visitors or washing hands), and 0 otherwise. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). The outcome in column (4) was collected only for a random sub-set of study participants.

Table D8: The effect of higher incentives on fact-checking and trust in information

	Fact-checking	**Trust in information shared by...**	
		Doctors and health experts	**Other citizens**
	(1)	(2)	(3)
A. Sample restricted to doctor message group			
Higher incentive	-0.030	0.003	-0.010
	(0.016)	(0.005)	(0.005)
	[0.057 ; 0.118]	[0.618 ; 0.625]	[0.038 ; 0.101]
Mean (lower incentive)	0.344	0.801	0.687
Observations	3851	3851	3851
Slums	142	142	142
Observation rounds	2	2	2
B. Sample restricted to control group			
Higher incentive	-0.019	0.004	-0.001
	(0.016)	(0.006)	(0.006)
	[0.228 ; 0.560]	[0.481 ; 0.732]	[0.850 ; 0.852]
Mean (lower incentive)	0.361	0.798	0.685
Observations	3849	3849	3849
Slums	142	142	142
Observation rounds	2	2	2

Notes. Estimates based on equation (3) restricted to the sample to participants allocated to the doctor message (Panel A), or the sample to participants allocated to the control group (Panel B). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table, separate for panels A and B. Dependent variables in column (1) *Confidence* is an indicator variable equal to 1 if the respondent does not always or very frequently check the truthfulness of information shared or discussed with family and friends, and 0 otherwise; column (2) *Trust in doctors and health experts* is the level of trust in the information shared by doctors and health experts; column (3) *Trust in other citizens* is the average level of trust in the information shared by people from UP and by people from other religions. Variables in columns (2)–(3) are measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D9: The effect of higher incentives on misinformation

	Opinions	Misconceptions	Opinions		Misconceptions	
			In-group	Out-group	In-group	Out-group
	(1)	(2)	(3)	(4)	(5)	(6)
A. Sample restricted to doctor message group						
Higher incentive	-0.001	-0.006	-0.003	-0.003	-0.006	-0.008
	(0.005)	(0.007)	(0.009)	(0.007)	(0.011)	(0.009)
	[0.828 ; 0.834]	[0.443 ; 0.897]	[0.743 ; 0.908]	[0.700 ; 0.953]	[0.614 ; 0.958]	[0.383 ; 0.883]
Mean (lower incentive)	0.617	0.483	0.626	0.618	0.488	0.482
Observations	3851	3851	3340	3851	2587	3340
Slums	142	142	142	142	142	142
Observation rounds	2	2	2	2	2	2
B. Sample restricted to control message group						
Higher incentive	0.004	0.001	0.005	0.002	0.013	-0.003
	(0.004)	(0.007)	(0.008)	(0.006)	(0.012)	(0.008)
	[0.356 ; 0.825]	[0.892 ; 0.889]	[0.508 ; 0.915]	[0.760 ; 0.939]	[0.285 ; 0.770]	[0.735 ; 0.969]
Mean (lower incentive)	0.615	0.488	0.614	0.621	0.477	0.496
Observations	3849	3849	3364	3849	2595	3364
Slums	142	142	140	142	141	140
Observation rounds	2	2	2	2	2	2

Notes. Estimates based on equation (3) restricted to the sample to participants allocated to the doctor message (Panel A), or the sample to participants allocated to the control group (Panel B). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table, separate for panels A and B. Dependent variables by column indicate respondent's level of agreement with statements reporting public views concerning opinions and misconceptions. *Opinions* are views that are not necessarily based on facts or knowledge, while *misconceptions* are incorrect views based on faulty knowledge or understanding. For the rationale behind such distinction in the measurement of misinformation, refer to Section 3. Statements are aggregated by averaging responses using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. In columns (1)–(2), the outcome variables include all statements, independently from the person with whom the statements is associated. In column (3) and (5), the outcome variables include only statements from a person with the same religion of the respondent. In column (4) and (6), the outcome variables include only statements from a person with a religion different from the one of the respondent or from the generic term "people". In columns (3)–(6), the person reporting the statement is randomly allocated at respondent level. Individual statements and categorization are described in Appendix B.2. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). Differences in observations across columns are due to unit non-response.

D.4 Treatment effects with ANCOVA specification

Tables D10 and D11 present estimates of treatment effects of the doctor message and of the religion concordance treatments using ANCOVA specifications. Estimates based on OLS regressions using equation (1) and controlling for the baseline value of the dependent variable.

Table D10: Compliance, fact-checking and trust, ANCOVA

	Compliance with evidence-based practices	**Fact-checking**	**Trust in doctors and health experts**	**Trust in other citizens**
	(1)	(2)	(3)	
Panel A				
Doctor message	0.040	-0.023	0.003	-0.002
	(0.014)	(0.010)	(0.003)	(0.004)
	[0.007]	[0.028]	[0.443]	[0.553]
Mean (Control message)	0.578	0.352	0.801	0.685
Observations	5079	7700	7700	7700
Slums	142	142	142	142
Observation rounds	2	2	2	2
Panel B				
Religion concordance	0.026	0.006	-0.001	-0.002
	(0.022)	(0.015)	(0.006)	(0.005)
	[0.238]	[0.678]	[0.803]	[0.740]
Mean (Other religion)	0.604	0.326	0.804	0.683
Observations	2519	3851	3851	3851
Slums	142	142	142	142
Observation rounds	2	2	2	2

Notes. Estimates based on OLS regressions using equation (1) and controlling for the baseline value of the dependent variable. When the dependent variable is missing at baseline, we impute it with the slum-level average value of the dependent variable at baseline. Panel B restricts the sample to participants allocated to the doctor message. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table, separate for Panels A and B. Dependent variables in column (1) *Compliance with evidence-based practices* is an indicator variable equal to 1 if the respondent complied with a randomly-asked World Health Organization's recommendation to protect from infection (either leaving the slum and receiving visitors or washing hands), and 0 otherwise; column (2) *Confidence* is an indicator variable equal to 1 if the respondent does not always or very frequently check the truthfulness of information shared or discussed with family and friends, and 0 otherwise; column (3) *Trust in doctors and health experts* is the level of trust in the information shared by doctors and health experts. *Trust in other citizens* is the average level of trust in the information shared by people from UP and by people from other religions. Because this variable was not collected at baseline, we use the average of all trust measures collected at baseline to impute the missing value. Trust is measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D11: The effect on misinformation, ANCOVA

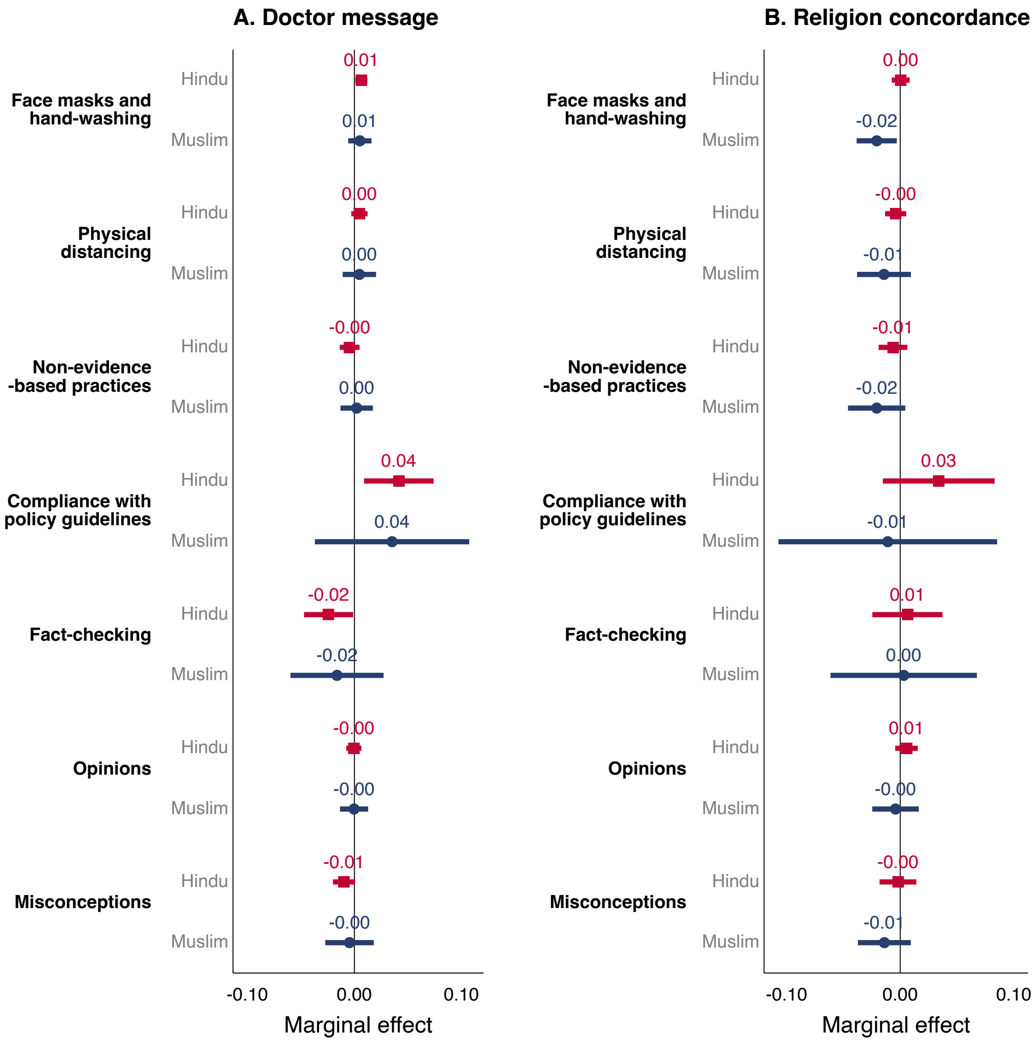
	Opinions	**Misconceptions**	**Opinions**		**Misconceptions**	
			In-group	**Out-group**	**In-group**	**Out-group**
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Doctor message	-0.000	-0.008	0.007	-0.005	0.004	-0.016
	(0.003)	(0.005)	(0.005)	(0.005)	(0.007)	(0.006)
	[0.882]	[0.087]	[0.179]	[0.268]	[0.548]	[0.006]
Mean (Control message)	0.617	0.489	0.617	0.622	0.483	0.494
Observations	7700	7700	6704	7700	5182	6704
Slums	142	142	142	142	142	142
Observation rounds	2	2	2	2	2	2
Panel B						
Religion concordance	0.003	-0.004	-0.001	0.004	-0.023	0.006
	(0.005)	(0.007)	(0.008)	(0.007)	(0.009)	(0.008)
	[0.507]	[0.534]	[0.939]	[0.622]	[0.010]	[0.442]
Mean (Other religion)	0.615	0.483	0.625	0.615	0.497	0.476
Observations	3851	3851	3340	3851	2587	3340
Slums	142	142	142	142	142	142
Observation rounds	2	2	2	2	2	2

Notes. Estimates based on OLS regressions using equation (1) and controlling for the baseline value of the dependent variable. When the dependent variable is missing at baseline, we impute it with the slum-level average value of the dependent variable at baseline. Panel B restricts the sample to participants allocated to the doctor message. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table, separate for Panels A and B. Dependent variables by column indicate respondent's level of agreement with statements reporting public views concerning opinions and misconceptions. *Opinions* are views that are not necessarily based on facts or knowledge, while *misconceptions* are incorrect views based on faulty knowledge or understanding. For the rationale behind such distinction in the measurement of misinformation, refer to Section 3. Statements are aggregated by averaging responses using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. In columns (1)–(2), the outcome variables include all statements, independently from the person with whom the statements is associated. In column (3) and (5), the outcome variables include only statements from a person with the same religion of the respondent. In column (4) and (6), the outcome variables include only statements from a person with a religion different from the one of the respondent or from the generic term "people". In columns (3)–(6), the person reporting the statement is randomly allocated at respondent level. Individual statements and categorization are described in Appendix B.2. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). Differences in observations across columns are due to unit non-response.

D.5 Heterogeneous analysis

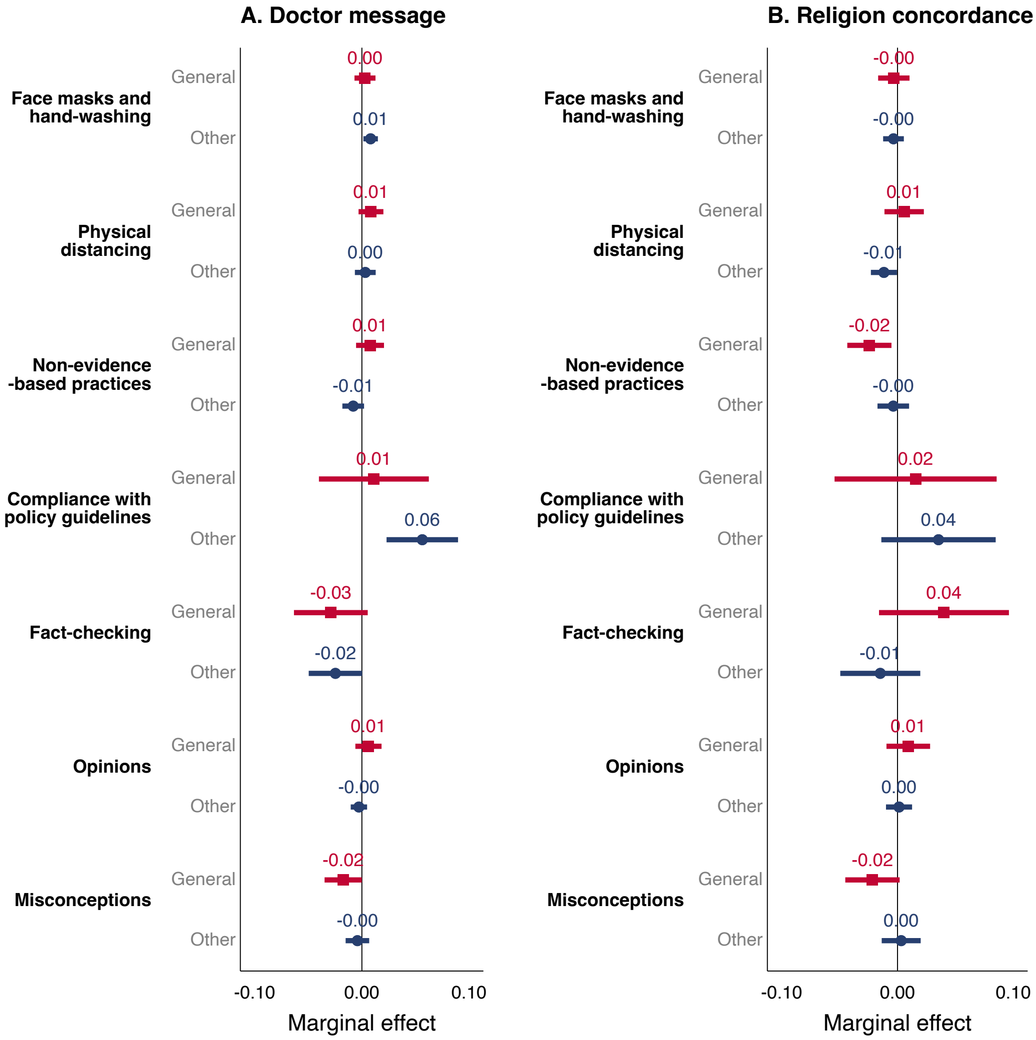
Figures D2–D6 reports estimates of heterogeneous treatment effects for the effect of the doctor message (Panel A), and for the effect of the doctor message being delivered by a religion-concordant messenger (Panel B). Estimates of the effects are obtained by estimating equation (1) separately by restricting the sample to the categories reported in column.

Figure D1: Heterogeneous effects by respondent's religion



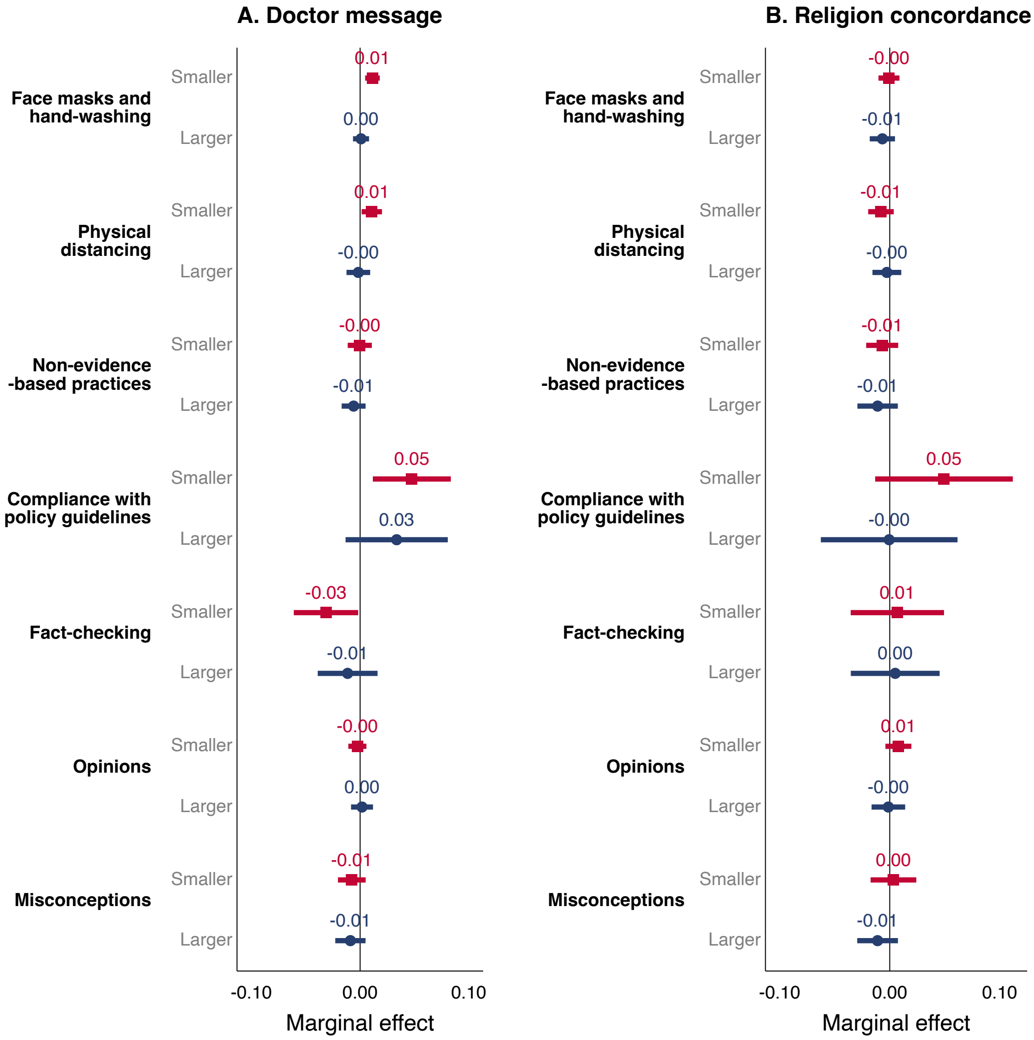
Notes. Heterogeneity is based on an indicator variable equal to 1 if the respondent is of Muslim religion, and 0 otherwise. Estimates based on OLS regressions using equation (1) and restricting the sample to the categories reported in column. Panel A presents estimates of the effect of the doctor message and includes all participants in the sample. Panel B presents estimates of the effect of religion concordance and restricts the sample to participants allocated to the doctor message. Standard errors are clustered at the slum level. Confidence intervals at 95% level. Variables are defined in Appendix A.

Figure D2: Heterogeneous effects by respondent's caste



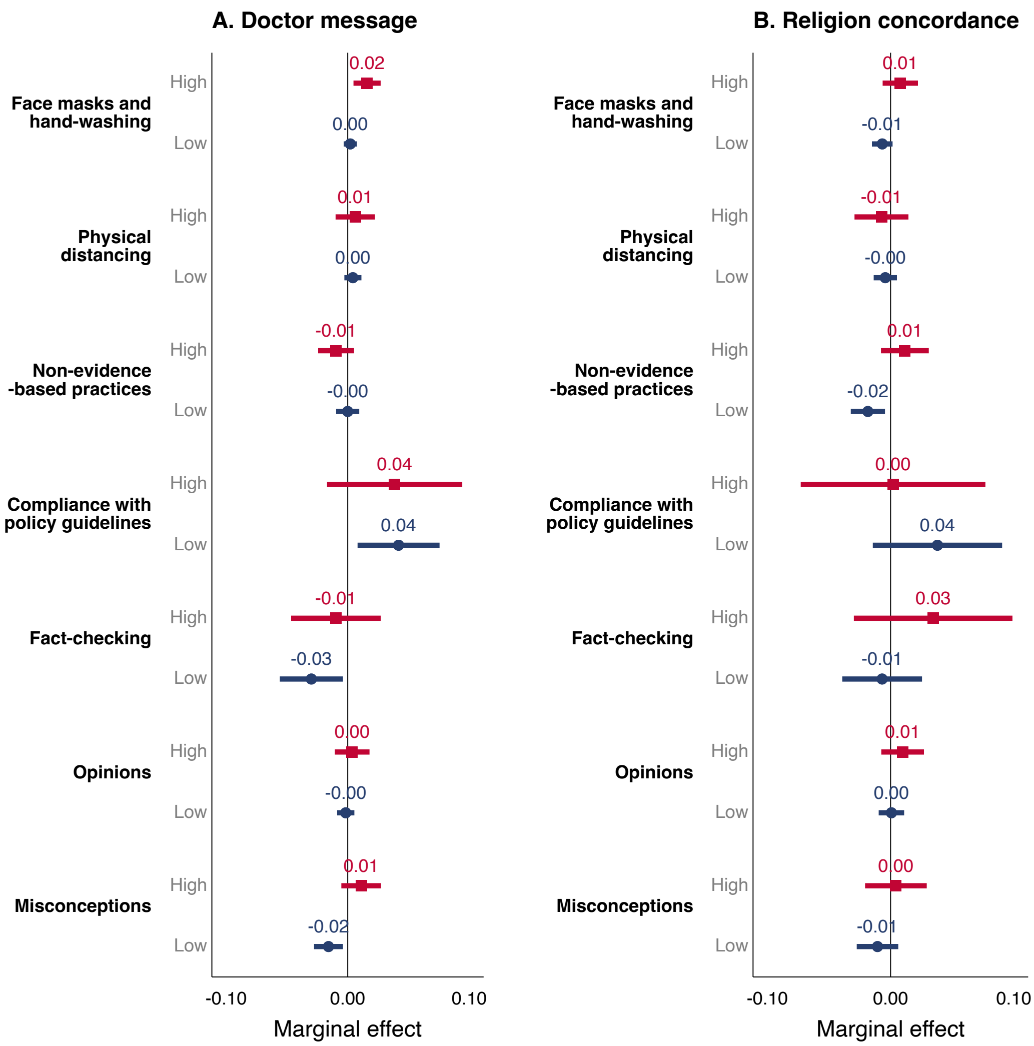
Notes. Heterogeneity is based on an indicator variable equal to 1 if the respondent is of a general caste, and 0 if the respondent is of other backward caste, scheduled caste, or scheduled tribe. Estimates based on OLS regressions using equation (1) and restricting the sample to the categories reported in column. Panel A presents estimates of the effect of the doctor message and includes all participants in the sample. Panel B presents estimates of the effect of religion concordance and restricts the sample to participants allocated to the doctor message. Standard errors are clustered at the slum level. Confidence intervals at 95% level. Variables are defined in Appendix A.

Figure D3: Heterogeneous effects by the Muslim share of the slum



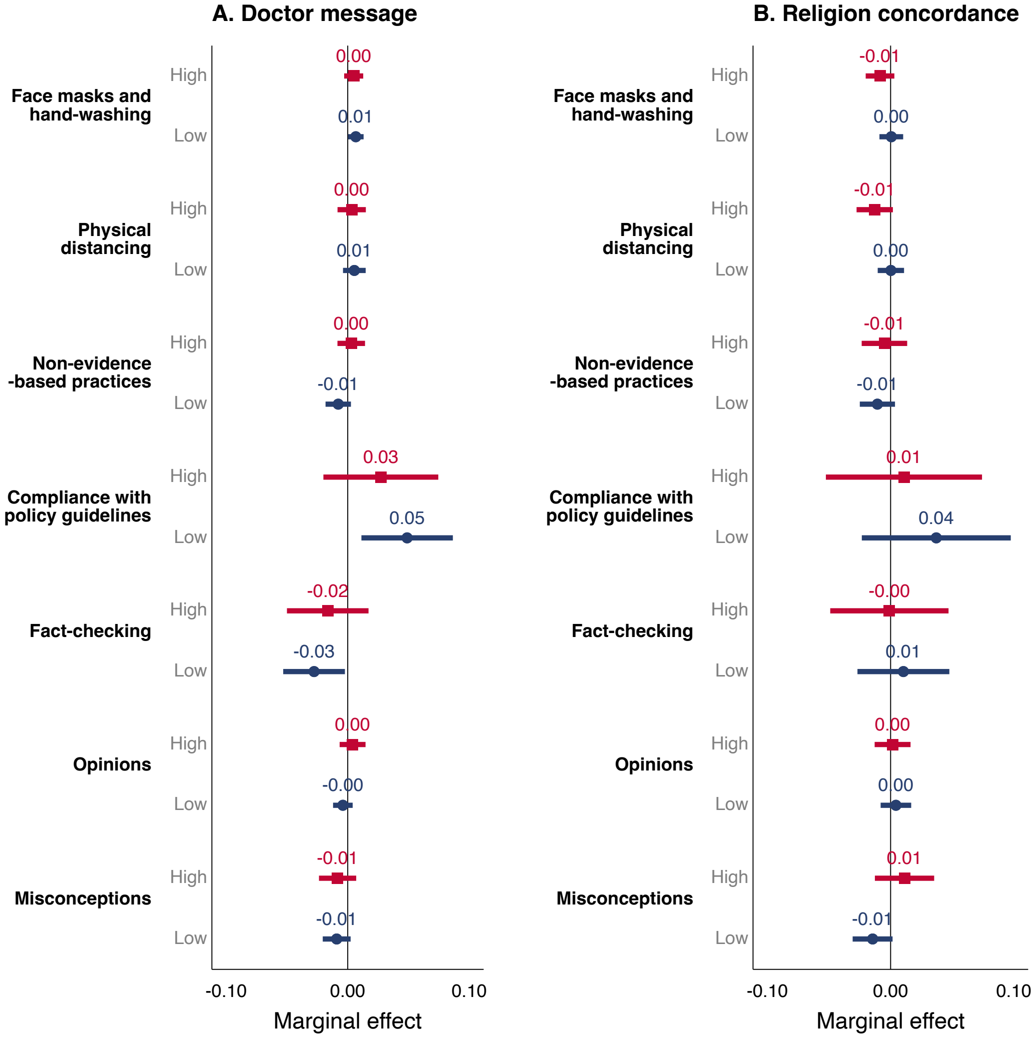
Notes. Heterogeneity is based on an indicator variable equal to 1 if the share of Muslim households in the slum is below the median of the sample distribution, and 0 otherwise. Estimates based on OLS regressions using equation (1) and restricting the sample to the categories reported in column. Panel A presents estimates of the effect of the doctor message and includes all participants in the sample. Panel B presents estimates of the effect of religion concordance and restricts the sample to participants allocated to the doctor message. Standard errors are clustered at the slum level. Confidence intervals at 95% level. Variables are defined in Appendix A.

Figure D4: Heterogeneous effects by strength of religious identity



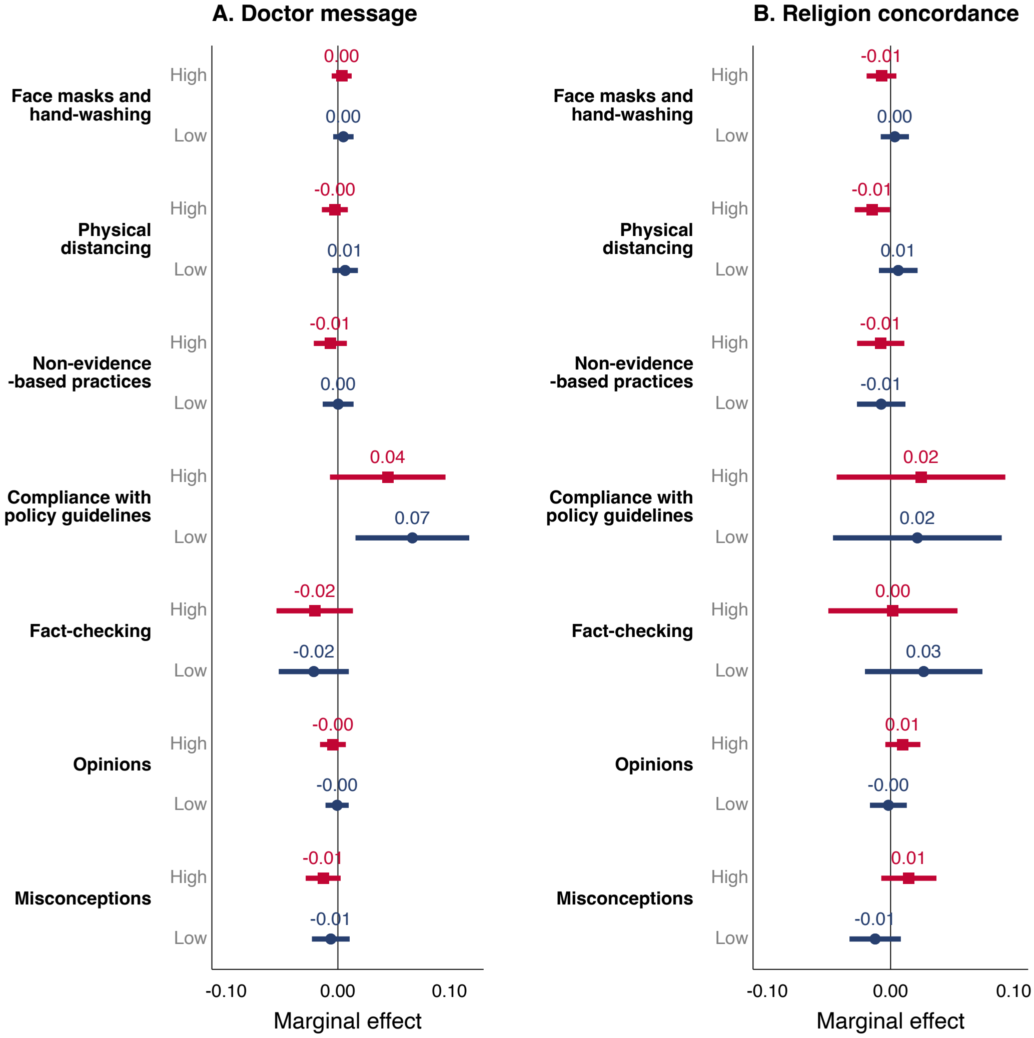
Notes. Heterogeneity is based on an indicator variable equal to 1 if the respondent's strength of religious identity is below the median of the sample distribution, and 0 otherwise. Estimates based on OLS regressions using equation (1) and restricting the sample to the categories reported in column. Panel A presents estimates of the effect of the doctor message and includes all participants in the sample. Panel B presents estimates of the effect of religion concordance and restricts the sample to participants allocated to the doctor message. Standard errors are clustered at the slum level. Confidence intervals at 95% level. Variables are defined in Appendix A.

Figure D5: Heterogeneous effects by trust in the government in the slum



Notes. Heterogeneity is based on an indicator variable equal to 1 if the average trust in the government in the slum is below the median of the sample distribution, and 0 otherwise. Trust is measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust. Estimates based on OLS regressions using equation (1) and restricting the sample to the categories reported in column. Panel A presents estimates of the effect of the doctor message and includes all participants in the sample. Panel B presents estimates of the effect of religion concordance and restricts the sample to participants allocated to the doctor message. Standard errors are clustered at the slum level. Confidence intervals at 95% level. Variables are defined in Appendix A.

Figure D6: Heterogeneous effects by social desirability



Notes. Heterogeneity is based on an indicator variable equal to 1 if social desirability is below the median of the sample distribution, and 0 otherwise. Social desirability is measured using the short version of the Marlowe–Crowne Social Desirability Scale (MC–SDS). Estimates based on OLS regressions using equation (1) and restricting the sample to the categories reported in column. Panel A presents estimates of the effect of the doctor message and includes all participants in the sample. Panel B presents estimates of the effect of religion concordance and restricts the sample to participants allocated to the doctor message. Standard errors are clustered at the slum level. Confidence intervals at 95% level. Variables are defined in Appendix A.

E Research methods and ethical concerns

Participants of the study were identified through previously-collected data. Participants were interviewed as part of a separate field experiment, completed at the end January 2020 and whose ethics approval was secured from the University College London Research Ethics Committee (ref. 2168/012). Informed consent was obtained for the participation in the study and for being contacted in the future for further survey rounds and related studies. All participants were above 18 years old and provided written consent. The pool of participants is representative of slum residents in the selected area of study, the sample size selected is therefore representative of the different groups composing the targeted population, and thus diverse by definition. Because the study was designed to evaluate a intervention in informal settlements, the targeted population is represented by a vulnerable group. For the study subject of this paper, the population of interest remains the same, especially considering the high risk of misinformation linked to low levels of literacy. The ethics approval was secured from the London School of Economics Research Ethics Review (ref. 1132). We discuss in this section potential concerns arising from ethical considerations.

Data were collected via mobile phone interviews due to constraints imposed by the COVID-19 pandemic. To respect the autonomy and the well-being of the participants, voluntary and informed consent from the participants was obtained orally at the beginning of each interview. Participation was voluntary and participants were not paid for responding the survey. The script of the consents form was as follows (translated to Hindi):

> Hello. My name is [NAME] and I work with Morsel Research and Development on a research project called "COVID-19 Spread in Informal Settlements" and funded by the London School of Economics (LSE). Researchers at the LSE, Institute for Fiscal Studies in the United Kingdom and the Nova School of Business and Economics in Portugal are interested in collecting information to assess slum dwellers' response to the COVID-19 pandemic. We are not affiliated to the government. Results from this research will be shared with policymakers and academics. However, they will not get any information about each participant, including names.
>
> We would like to interview you for approximately ten minutes. All the information you provide remains confidential and can be accessed only by selected members of the research team. You have the right to decline your participation or withdraw from the study at any time without the need to explain yourself and your decision will carry no consequences.
>
> Should you take part in the study, you agree that we can contact you again in the future to collect more informa-

> tion related to this study, at which point you will again be able to choose whether to participate or not.
>
> Please let me know if you have any questions at this point.
>
> We just send you a text message with contact information, should you have any queries about this study and your interview going forward.
>
> Please confirm that you have understood the information just provided and that you were given the opportunity to clarify any doubts or questions. [If respondent says ‘Yes’ proceed with the survey].

Respondents were informed that they could have asked questions regarding the study at any point before, during and after the interview. To this purpose, following the informed consent, a text message was sent to participants providing a contact number for questions related to the study. To support the well-being of participants, the text message provided also information on how to contact the COVID-19 helpline for issues and questions related to the pandemic.

To respect the confidentiality of responses, participants were informed that only anonymous data would have been analyzed. To this purpose, arrangements were made at three levels. First, interviewers used headphones to avoid that responses could be overheard by anyone. Second, interviewers were trained not to view or share any information about respondents other than for what was strictly required for the purposes of data collection. To guarantee this condition, the project used Computer-Assisted Personal Interviewing (CAPI) to collect data, a well-established system designed specifically with the needs of confidentiality and data-security in mind, including, for example, single log-in and access to data available only during the interview. Third, collected data were encrypted and stored on a secured server. Data on the servers are backed up to an off-site machine stored securely with a third-party company, and network access is restricted to technical support staff.

Concerning the content of the questionnaire, we do not highlight any concern relative to induced stress among participants, including delving into some deeply personal experience, exercising coercion or domination, or dealing with topics that are sacred to those being studied that they do not wish profaned. To verify these concerns, given that our target population is represented by a vulnerable group, we paid particular attention in the framing of questions and conducted a pilot of the survey instruments, during which respondents could provide feedback on types of questions, wordings, and the length of the interview. We did not encounter any issue reported by pilot participants.

Concerning the interventions (detailed in Appendix B.3), we did not intervene in any political process and provided informative content to participants in line with scientific evidence. In particular, the project did not engaged in deception of respondents. Notice that exposure to different religions is common in our population of interest, supported by the religious diversity encountered in our sample. In addition, our approach does not communicate discriminatory messages related to specific religions.