
Understanding Misinformation on Mobile Instant Messengers (MIMs) in Developing Countries

May 2020

Authors

Irene V. Pasquetto,
Shorenstein Center, Harvard Kennedy
School

Eaman Jahani,
MIT Institute for Data, Systems and
Society

Alla Baranovsky,
Harvard Department of Government

Matthew A. Baum,
Shorenstein Center, Harvard Kennedy
School



HARVARD Kennedy School
SHORENSTEIN CENTER
on Media, Politics and Public Policy

The views expressed in Shorenstein Center Discussion Papers are those of the author(s) and do not necessarily reflect those of Harvard Kennedy School or of Harvard University.

Discussion Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important issues and challenges in media, politics and public policy. These papers are published under the Center's Open Access Policy. Papers may be downloaded and shared for personal use.

Note: this report was updated on June 3rd 2020 to correct an error in the images illustrating the types of misinformation shared on MIMs. A previous version of this report wrongfully characterized a piece of news from the website caravanmagazine.in as click-bait rumor.

Executive Summary

We conducted a mixed-methods research project in Nigeria, India, and Pakistan consisting of surveys, survey experiments, and semi-structured interviews, to better understand the spread and impact of misinformation, and in particular of misinformation on mobile messaging apps (MIMs).

One of our goals was to evaluate the relevance and prevalence of viral false claims in each country. We found evidence that popular “false claims” debunked by fact-checking companies are widely recognized, but, unsurprisingly, news from mainstream media are more widely recognized than the false claims. This is certainly a good sign, which confirms similar trends reported elsewhere (Allen et al., 2020; Guess et al., 2019).

However, we also found preliminary evidence that misinformation circulates wide on messaging apps. When asked whether and where they encountered researcher-selected false claims, participants reported to be exposed to more false claims than “mainstream claims” (i.e., true) on messaging apps, while this was not the case for traditional media, such as newspapers and TV News. While others have found false claims to be more prevalent on social media than on mainstream media (Stecula et al., 2020), this is the first systematic evidence that messaging apps might be the primary source for spread of misinformation.

Qualitative findings add nuance to our results. We asked our participants to forward us concrete examples of misinformation (*forwards*) that they had recently received on WhatsApp or other messaging apps. Only ⅓ of the forwards we collected (about 50) contained a link to a mainstream media, a website, or a social media page. Our analysis of the interview data and of these forwards strongly suggest that misinformation shared on messaging apps mostly consists of anonymous rumors that do not link to any external sources. Interview data also suggest that “advanced MIM users” are members of public groups and actively engage with content shared in these groups (journalists, politicians, religious personalities etc.). However, the “average MIM user” communicates and shares content within private, small-scale groups that are invisible to the rest of the world (no trace of these groups can be found online). This highlights one of the principal challenges researchers face in understanding, and ultimately mitigating, the patterns of misinformation distribution within the MIM ecosystem.

Finally, along on this last point, we sought to test and identify best practices for encouraging messaging apps’ users to correct other users when they share misinformation on messaging apps, leveraging what others have defined as “volunteer fact-checkers” (Kim & Walker, 2020). We empirically studied the conditions under which recipients are like-

Participants reported to be exposed to more false claims than “mainstream claims” (i.e., true) on messaging apps, while this was not the case for traditional media, such as newspapers and TV News.

ly to believe and re-share corrections (messages debunking claims that are determined as false by fact-checking agencies) shared via messaging apps. We found evidence that:

- Users tend to pay more attention to **voice-based** corrections, than to text or image-based corrections;
- Corrections received from a **family member** or a **close friend** are re-shared more than corrections received from an acquaintance;
- Similarly, corrections received from a **like-minded individual** are re-shared more than corrections received from a non-like-minded individual;
- Overall, sharing corrections on messaging apps is perceived as a highly stressful, unusual activity in which only **highly motivated individuals** are willing to engage.

Implications

Access to Content. Researchers and fact checkers need access to content shared in private groups to establish prevalence of misinformation on messaging apps. Without access to the content shared on MIMs' closed groups, it remains difficult, if not impossible, to measure the exact prevalence of misinformation on these platforms. In this study we used self-reported data to study prevalence. This is a methodology with obvious limitations, in particular it relies on the participants' ability to remember what they have seen and where they have seen it. Others have scraped and analyzed content shared on MIMs' public groups, which is a useful approach, but, like ours, also this approach might present limitations given that interview data indicate that only specific kinds of MIM users (i.e., advanced) are active in open groups. Promising forthcoming research further indicates that (perceived) private WhatsApp groups composed of ideologically heterogeneous individuals are primary vehicles of misinformation (Kuru et al., 2020).

Credibility is Key. It is important to enlist credible messengers as volunteer fact-checkers. This means finding people who are likely to be credible to recipients of the message (like-minded messengers) and making them the "source" of the debunk, or persuading them to endorse the debunk. Close personal connections and people with similar ideological or political preferences are more credible, and people are more likely to share debunks when they are received from such high-credibility sources. Not only are debunks received from credible sources more effective, but they are also more likely to proliferate. Also in this case, others are finding similar patterns (Kuru et al.,

2020). Any intervention in this direction, however, would present challenges. Most notably, in practice, any kind of divisive information, including misinformation, tends to be shared within networks of individuals who agree with one another (the infamous filter bubbles or echo-chambers). This seems to be especially the case for WhatsApp private groups. Consequently, on private chats as well as offline, volunteers fact-checkers for the most part tend to be exposed to information with which they already agree.

Audio vs. Video or Text. Further testing is needed to see whether the greater effect of voice-messages is cultural- or context-specific, or general. Our findings suggest that fact-checkers should concentrate on using audio-based debunks, at least in mobile messaging apps contexts, and at least in India and Pakistan (we were not able to test multiple formats in Nigeria). If it generalizes, then this finding would have clear implications for the modalities of fact-checking, going forward.

Different Platforms, Different Levels and Types of Misinformation. Most false claims on MIMs take the form of rumors, which are harder to debunk than stories with clearly identifiable sources. Participants reported encountering misinformation more frequently on MIMs than on traditional media (TV News and newspapers). In addition, qualitative findings suggest that misinformation for the most part takes a specific form and character on MIMs that is unique to these spaces. Misinformation on MIMs is more personal, it sounds and looks like a suggestion from a close friend, rather than a top-down piece of information created by a specific group to influence another. It is not clear, at this point, if this grass-root, platform-specific feature of MIMs' misinformation makes it actually more or less credible than misinformation found on traditional media and social media.

Misinformation Exploits Cognitive Biases. Further research is needed to determine the extent to which decisions about sharing false stories follow from assessments of form and style rather than, or in addition to, content. Our findings also suggest that false stories can be “remembered” even when they were never encountered (that is, stories that we fabricated) because they “appear” similar in form and style to stories that people have seen. Participants don't seem to pay great attention to full details of a false claim, but they do remember the general contours of the story. Recalling the generic theme of a similar claim can later render a new false claim as previously established, hence more plausible. This was a surprising but potentially important finding that requires further research. (We hope to further address this issue in a subsequent report focused on a fourth country included in the larger study.)

Introduction

An important emerging challenge in fact-checking is the spread of misleading and uncivil content on mobile instant messaging (MIM) applications. WhatsApp, Telegram, and Signal, among others, continue to grow in popularity worldwide. In January 2020, WhatsApp had the highest number of active users of all mobile apps worldwide, followed by Facebook messenger and WeChat (Kemp, 2020; Wardle, 2020).

Ample anecdotal evidence suggests that misleading rumors, false claims and uncivil talk spread easily on messaging apps, leading to serious real-life consequences worldwide. In 2017, a WhatsApp hoax about child-napping con artists led to the beating of two people in Brazil (Aragão, 2017). In early 2018, misinformation surrounding monkeypox in Nigeria was linked to a drop in routine immunizations (Oye-banji et al., 2019). Between the summer of 2018 and the spring of 2019, during Brazil and India's general elections, WhatsApp was flooded with hyper-partisan, xenophobic and homophobic messages (Avelar, 2019; Bengani, 2019). Meanwhile, in Myanmar, state military operatives turned to Messenger, Facebook's messaging app, to promote a misinformation campaign against the Muslim Rohingya minority group. The campaign led to murders, rapes and a large forced human migration (Mozur, 2018). In 2020, during the COVID-19 pandemic, hoaxes and rumors of many sorts have spread widely on messaging apps worldwide, including the US (Collins, 2020; Gold & O'Sullivan, 2020). In Lombardy — the Italian region most affected by COVID-19 — traffic on WhatsApp and Facebook Messenger has doubled since the beginning of the pandemic, and along with this increase the viral sharing of false information and conspiracy theories has also spiked (Facebook, Inc., 2020).

When information spreads on Facebook or Twitter, it is “public” enough that it can be identified and tracked. However, with few exceptions, all activity on closed messaging applications that exists outside of one's immediate network is completely invisible. As such, it can be much harder to identify, track, debunk and respond to misinformation. Given the challenges related to the identification of misinformation on messaging apps, there exists little rigorous research devoted to understanding the prevalence of misinformation on these platforms and the ways in which users share news-type content on them (Bengani, 2019; de Freitas Melo et al., 2020; Melo et al., 2019).

Our project employs surveys (on prevalence), survey experiments (on debunking), follow-up interviews, and content analysis (of an archive of WhatsApp “misinformation forwards”) to study misinformation on messaging apps in developing countries. This is the first of several reports on a series of empirical studies that we conducted on misinformation on messaging apps. In this first report, we present evidence

on how messaging apps' users in three countries (India, Pakistan, and Nigeria) engage with fact-checked content on the platforms, and on how, when and why they might be willing to re-share and believe such corrections.

Research Questions Addressed in this Report

1. **Prevalence:** How widely are misinformation stories recognized? How widely believed?
2. **Correcting:** When, why, and how will MIM-delivered debunks work.

Summary Statistics

We begin by reviewing the numbers and characteristics of participants in our survey experiments conducted in the three countries included in this report.

Quantitative Data

Participant Characteristics: Nigeria

Sample of mobile internet users

- N= 1894 in Wave 1 (951 passed attention checks)
- 68% male; 32% female.
- 64% 25-34.
- 35% married; 65% single (or divorced, widowed etc).
- 82% of the respondents hold a certificate or degree beyond high school.
- 72% employed or full time students.
- 25% Muslim; 75% Christian or other faiths (Bahai, Other etc).
- 85% urban residents; 15% rural.

Participant Characteristics: India

Sample of mobile internet users

- N= 1799 in Wave 1 (954 passed attention checks)
- 71% male; 29% female.
- Age: 55% <30, 32% 30-40, 11% >40
- 53% married; 47% single (or divorced, widowed etc).
- 80% of the respondents hold a certificate or degree beyond high school.
- 93% employed or full time students.

- 76% Hindu, 11% Muslim; 13% Christian or other faiths (Buddhism, etc)
- 88% urban residents; 12% rural.

Participant Characteristics: Pakistan

Sample of mobile internet users

- N= 2003 in Wave 1 (946 passed attention checks)
- 72% male; 28% female.
- Age: 64% <30, 27% 30-40, 9% > 40
- 73% of the respondents live in households where chief earner holds a certificate or degree beyond high school.
- 80% employed or full time students.
- 98% Muslim
- 89% urban residents; 11% rural

Qualitative Data - Nigeria and India

A total of 60 participants were interviewed for this study, 30 from Nigeria and 30 from India. Participants were interviewed in two waves (15 participants for each wave, per country). The first wave was conducted before the surveys were administered (January - April 2019), and aimed at informing their design. Participants for the first wave of interviews were recruited from an online list of journalists and fact-checkers from the regions. The second wave of interviews was conducted after the surveys and had the goal of helping with the interpretation of results from our quantitative data analysis. Participants for the second wave were recruited from a list of volunteers who completed our surveys.

All participants are residents of Nigeria or India. Interviews and discussions with the participants were conducted on MIMs (mostly WhatsApp) through chat, video calls, and voice messages. During interviews, we discussed the composition of WhatsApp groups and the nature of WhatsApp forwards. We asked participants to forward us relevant content shared in their WhatsApp groups. We also discussed the accuracy and potential harm of selected “forwards,” and helped debunk misinformation, whenever possible.

We asked participants to share with us their thoughts about why other users might forward false or harmful content, and how being exposed to this content on MIMs impacts their sense of safety, on MIMs as well as offline. We discussed participants’ own processes of sense making and discerning falsehood from truth.

Chats and video calls were conducted in English. Whenever partici-

pants shared content in other languages, this was translated to English by research assistants from the respective countries. Research assistants from the countries also helped to culturally and socially contextualize WhatsApp forwards shared by the participants.

We developed a protocol for semi-structured interviews which focuses on five areas of investigation: 1) daily use of MIMs (scope, duration, frequency, etc.), 2) nature of content shared on MIMs (topics, formats, etc.), 3) nature of contacts on MIMs (personal knowledge of the contacts, etc.), 4) technological affordances of MIMs (encryption, closed vs public groups, etc.), and 5) attitudes towards fact-checking and corrections.

Research Question #1. Prevalence of misinformation

We conducted a survey to measure the prevalence of false claims in each country. We showed each respondent 10 claims, including 5 false and debunked claims, 3 claims from mainstream media, and 2 claims that we fabricated as control (i.e., placebos). We randomly selected the 5 false claims from a list of 10 and the 3 mainstream claims from a list of 100 previously collected to match the false claims in terms of social media reach. We selected recent, false claims that had been debunked by local fact-checking websites, and used CrowdTangle to ensure that these claims were widely shared on social media. The final false claim that we used for our study had the highest number of shares among all qualifying claims from fact-checkers in each country. Mainstream claims had similar sharing prevalence and appeared the same week as debunked claims.

We asked respondents three main questions:

1. Have you seen the claim before?
2. Do you believe the claim to be true?
3. In which platform/medium have you seen the claim?

In relation to questions 1 and 2, a high percentage of respondents (23% in India, 28% in Nigeria, 18% in Pakistan) reported seeing the false claims, but, as expected, respondents were more likely (34% in India, 36% in Nigeria, 33% in Pakistan) to have seen mainstream stories than the false claims that we selected from fact-checking agencies (Figure 1), and this difference is statistically significant. Similarly, respondents in all three countries were more likely to believe mainstream stories than false claims (Figures 2). This finding is

A high percentage of respondents reported seeing the false claims, but, as expected, respondents were more likely to have seen mainstream stories than the false claims that we selected from fact-checking agencies.

Examples of Debunked, False Claims

"The United States' CIA issued a posthumous apology to Osama Bin Laden after new evidence cleared him of involvement in the 9/11 attacks."

"Hot coconut water kills cancer cells."

"Tomato paste and Coca-Cola is an emergency blood tonic that helps blood donors replenish their blood."

encouraging because it shows that the reach of false claims is narrower than that of mainstream news (assuming that mainstream news presents accurate information). This pattern holds even for the most successful and far-reaching false claims, and confirms similar patterns found by others (Allen et al., 2020; Guess et al., 2019).

In relation to question 3, we observed that participants reported to have seen more mainstream claims (26% in India, 27% in Nigeria, 30% in Pakistan) than false claims (15% in India, 14% in Nigeria, 20% in Pakistan) on traditional media, such as on newspapers and TV News (as expected). However, false claims are seen at a higher rate (37% in India, 48% in Nigeria, 46% in Pakistan) than mainstream claims (29% in India, 38.5% in Nigeria, 38% in Pakistan) on messaging apps. Table 1 illustrates these results for India. While other studies (e.g., Stecula et al., 2020) have found high correlations between social media usage and exposure to false information, our findings suggest that messaging apps might be the primary key source for spread of misinformation.

Looking at both of these findings in combination, the following scenario emerges: false claims are more transient, and less recalled than mainstream claims, but they dominate MIMs and that is where they mostly appear. Indeed, participants remember having seen more false claims than mainstream claims specifically on messaging apps, but, overall, they reported to have seen more mainstream claims than false claims.

INDIA	Newspapers and TV News	Messaging Apps
Mainstream claims	26%	29%
False claims	15%	37%

Table 1. Where have you seen the claim? Example results from India.

How can we make sense of this? It might be that information found on messaging apps is less credible hence users don't register the claims they see on messaging apps with the same rigor as they do with newspaper or TV claims. Foundational research in social science suggest that there is a connection between credibility and retention of infor-

mation: according to the Theory of Minimalism (Campbell et al., 1980), information that is consistent with our prior beliefs is more credible, all else equal, and we are more likely to engage in the effort to commit it to memory; while for information that we view as inconsistent with our prior beliefs -- and hence less credible, all else equal -- we are more likely to summarily discard it and so not retain in memory. In other words, attitudinally dissonant information is more quickly discarded than attitudinally consonant information. Then, it might be that, for some reason, users consider information on MIMs not very credible, and so they tend to not remember it. We could then hypothesize that while false claims circulate widely on MIMs, MIMs are at the same time perceived as less trustworthy than traditional media, and this is why false claims seen on MIMs are nevertheless remembered less. Further research is needed to understand if this is actually the case.

Finally, we also found that, overall, participants do not appear to distinguish between false claims and “placebos” (Figures 1 and 2), as they indicated that they had encountered debunked false claims and “placebos” (fabricated false claims) at very similar rates. Placebos are claims that resemble existing false claims in their grammatical structure, slant, and intent, but that have never circulated online. We created placebos to verify to what extent respondents remember false claims and details about them.

Example of false claim:

“The United States’ CIA issued a posthumous apology to Osama Bin Laden after new evidence cleared him of involvement in 9/11 attacks.”

Example of placebo claim:

“PM Imran Khan revealed that his views of U.S. President Donald Trump have changed; and voiced his support for Trump’s effort to build a wall between the U.S. and Mexico.”

Each respondent saw 5 false claims, 3 pieces of news from mainstream media, and 2 placebos. We found that respondents similarly remembered false claims that did actually circulate in social media and MIMs and fabricated false claims (placebos), suggesting that they conflate the two. This finding suggests that respondents rarely remember the details about a false claim, they only remember the general message and its “rhetorical style.” Others have found that readers can distinguish between credible and not credible sources of information only by reading the headline of an article (Dias et al., 2020), suggesting that, again, false claims have a recognizable style that readers can easily distinguish from mainstream claims, independently from the substantive content.

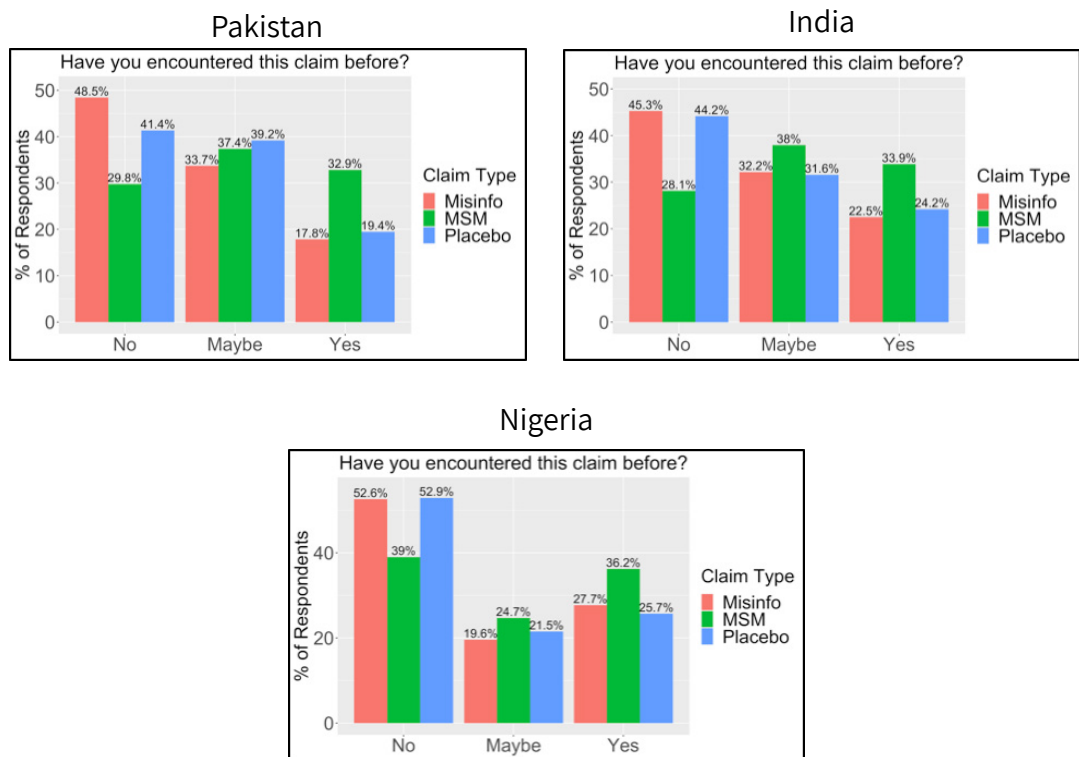


Figure 1. In all three countries, respondents were more likely to have seen stories from mainstream media (MSM) than false media (the difference is significant). Survey participants were equally likely to recall placebo and false claims, suggesting that they rarely remember the details of false claims.

During the interviews with our selected sample of survey participants in Nigeria and India we discussed concrete examples of misinformation as they have encountered them on MIMs, and we also asked participants to forward us all pieces of information that they were receiving on WhatsApp and that they found potentially false, misleading or offensive. Because most interviews were conducted between February and April 2020, most examples are related to the COVID-19 outbreak. This process resulted in an archive of over 50 examples of misinformation (i.e., “forwards”) shared on WhatsApp in Nigeria and India.

Significantly, most of these WhatsApp forwards were rumor-based and only $\frac{1}{3}$ contained a link to a mainstream media outlet, a website, or a social media page. This qualitative finding confirms survey results indicating that mainstream news stories are rarely seen on messaging apps.

Our analysis of the interview data and of the *forwards* that the participants sent us suggest that misinformation shared on messaging apps can be classified into at least five categories:

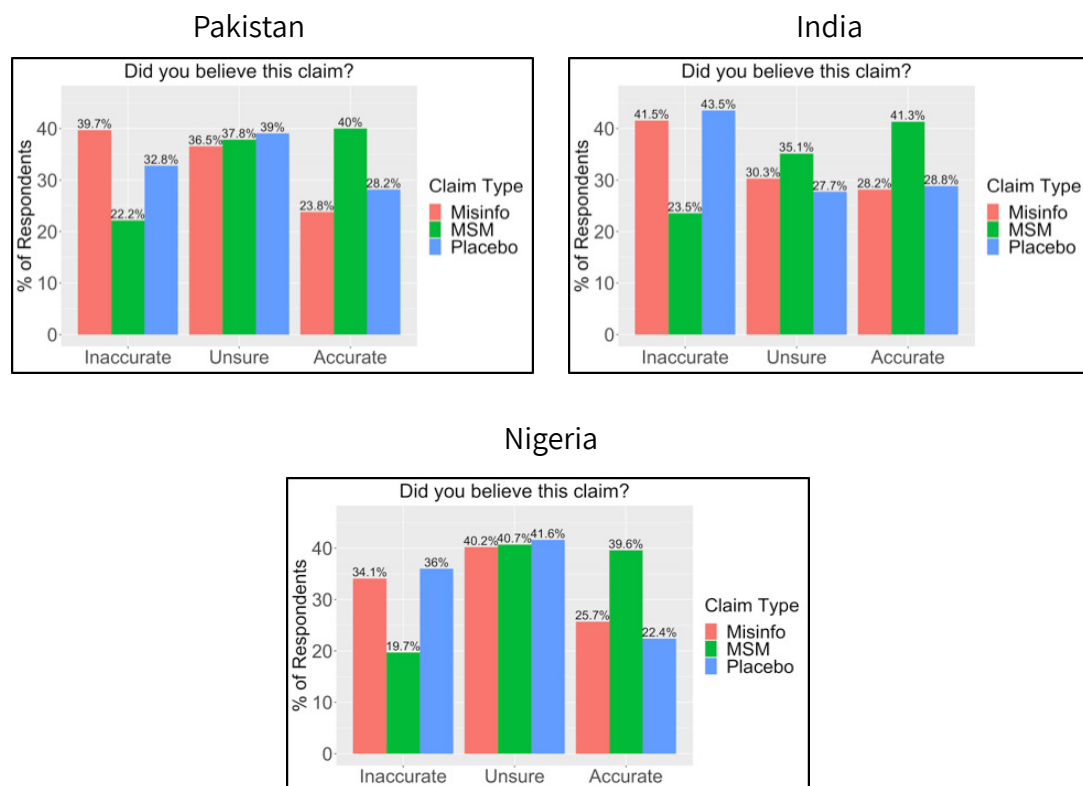


Figure 2. In all three countries, respondents were more likely to believe mainstream stories than false claims. Survey participants were equally likely to recall placebo and false claims, suggesting that they rarely remember the details of false claims.

- Anonymous rumors
- Influencer-driven rumors
- Impostor rumors
- Click-bait rumors
- Scams

Anonymous rumors consist of unverified rumors that do not reference or link to any source. These are often created in-platform and shared either by text or audio files. The texts might be quite long, sometimes exceeding 1000 words, and audio files can last up to ten minutes. In our archive, and also according to our participants, anonymous rumors seem to be the most common type of misinformation circulating on MIMs. No URL can be retrieved for these false claims.

Impostor rumors use logos and other designer tricks to forge documents or data that resemble legitimate documents or data from well-known institutions, companies, and organizations, in order to exploit

their credibility for spreading false and highly alarming claims. These rumors are often in image format and they do not link to any other source. No URL can be retrieved for these false claims.

Influencer-driven rumors are created and signed by specific individuals who are (or inspire to be) popular online (i.e., influencers and online celebrities). These individuals often offer some kinds of qualifications to enhance their credibility (a pastor, a doctor, a journalist, an entrepreneur, etc.), and provide links to their personal social media pages and accounts, such as Facebook and Instagram pages or YouTube channels.

Click-bait rumors contain links to for-profit websites specialized in the production of click-bait, sensational fake news. Such websites resemble legitimate news sources, but produce low-quality, highly sensational news articles aimed at attracting users to their platforms to collect revenues via online advertisement. In our sample, click-bait rumors were the least common forwards.

Scams aim at stealing money or credentials. Participants reported an insurgence of this kind of misinformation during the COVID-19 emergency. Multiple versions of the same COVID-19 scam were circulating in both Nigeria and India: scammers asked users to share their bank credentials in order to supposedly receive financial aid that the governments allocated in response to the COVID-19 crisis.

In addition, most participants reported being members of “closed groups,” while very few participants reported to have ever joined an “open group.” Close groups cannot be automatically joined by scraping algorithms, because no “invite links” exist for these groups on the open web. Users can join closed groups only when administrators decide to personally add them. If a user does not know the administrator of a group, or someone who knows the administrator, the user does not know that the group exists, and has no way of finding out. On the contrary, administrations of “open groups” share invite links on the open web, encouraging their audiences to join. In recent years, researchers of misinformation on messaging apps crawl the open web and social media looking for such invite links, they automatically join them, and then extract data from the open groups. However, these approaches might have limitations, because our findings suggest that only very specific individuals who operate in very specific sectors are members of these groups or regularly check the content shared on these groups (politicians, journalists, NGOs volunteers, etc.), while the “average MIM user” solely joins and knows groups created by close family members and friends.

Examples of Anonymous Rumors

Forwarded

We in Nasarawa State are save by God' grace. Read the secrete
COVID-19 REMEDIES THAT WORK.
Don't be deceived, Covid-19 has been defeated, conquered and subdued. I hereby present to you 6 proofing remedies that work. Choose which ever is convenient for you.

1. Dr Zev Zelenko treatments of New York treated above 700 positive cases with no hospitalization and no death. His prescription: Hydroxychloroquine 200mg twice daily for 5 days; Azithromycin 500mg once a day for 5 days, Zinc sulfate 220mg for 5 days.
2. UCH, CMD REMEDY. Professor Otegbayo , tested positive and within a week tested negative again.
His remedy is: Vitamin C-1000mg, Chloroquine, turmeric, lemon, unripe pineapple and steam inhalation, plus prayer.
3. H.E. Governor Seyi Makinde. He test... [Read more](#)


7:12 AM

Forwarded

Finally FBI arrested a professor from Boston university who was in connection with Chinese university and research lab in Wuhan, and was highly paid by China.....Now it is very close to clear that corona virus is a Bio Attack planned and conducted by China.
- A Chinese expert assures everybody that inhalation of a steam from hot water kills Corona virus 100 percent. Even if the virus entered the Nose, throat or Lungs. Corona Virus can not stand the hot water steam...
PLEASE CIRCULATE THIS INFORMATION TO ALL YOUR FAMILY MEMBERS AND FRIENDS.

11:48 AM

Examples of Imposter Rumors


Beverages Africa

people

DATE : 16th MARCH 2020
TO : ALL CCBA EMPLOYEES

Dear Colleagues,

SECURITY ALERT!

Please be informed there are reports of criminals taking advantage of the COVID-19 situation and masquerading as Health Service Staff and asking to spray houses as a preventive measure for Coronavirus.

Individuals/Staff members are urged to be vigilant and mindful of people you allow into your homes. Report any suspicious case to the police.

DO NOT accept offers for free masks or sanitizers from strangers as they could easily be laced with sedatives creating an opportunity for criminals to easily victimize you/perpetrate crime.

Regards,

HUMAN RESOURCES

Saint Luke's
Hospital
OF KANSAS CITY

03/07/2020

Topic: COVID-19

To whom it may concern,

After extensive research, our findings show that consuming alcoholic beverages may help reduce the risk of infection by the novel coronavirus; COVID-19.

Vodka is the most recommended for drinking, cleaning , and sanitizing.

Kansas City
The 7th of March 2020



Examples of Influencer-Driven Rumors

Forwarded

WUHAN VIRUS

By Ibn Taofeeq Abdul Azeez

The United States Of America(USA) now has more known cases of coronavirus disease than any other county. 155,705 cases and still counting. Mr Trump wants us to regard the virus as The Wuhan or Chinese Virus. The US claims the virus was invented by China because they want to control the world power. They want to gain global voltage and the COVID-19 is their master plan.

Fine, will you also agree the US is worse than China? They started the dirty game. Were they not the ones who created invisible divisions amongst other countries? They started wars everywhere especially The Middle East. They almost started war in China trying to instigate Hong Kong against Mainland. Nobody cautioned her! When they invaded Libya and the likes, it is still under world power.

When they started unnecessary trade war because of Chinese leading 5G technology, slowing the world advancement growth due to selfish interest, It is still all world power!

Forwarded



Instagram post by Bonjour ❤️ • Mar 31, 2020 at 5:58pm UTC
Bonjour ❤️ posted on Instagram • See 119 photos and www.instagram.com

Barack Obama is asking Africans not to accept the vaccines that will come from America and Europe...

Barack Obama: I'll be an accomplice if I don't denounce this evil act white people want to do to Africans, first of all I was born in America but I'm African blood, I'm not going to allow white people to ki...

[Read more](#)

12:30 PM

Forwarded

French Doctors proposes Coronavirus vaccine test on Africans, "Africans are not Lab Rats" Trends" - Video

<https://youtu.be/1yJdsonIJZE>

2:29 PM

Examples of Click-Bait Rumors

Forwarded



China officially announces victory over the Corona virus

The Chinese president at Wuhan Hospital without thespeaktoday.com

The Chinese president at Wuhan Hospital without muzzle ..
China officially announces victory over the Corona virus

Did the Chinese deceive the world with the Corona Virus? And they saved their economy ?!

This is what the Americans and Europeans think, after they sold their shares in high-value-added t... [Read more](#)

3:50 PM

**Report updated on June 3rd 2020. A previous version of this report wrongfully characterized a piece of news from the website caravanmagazine.in as click-bait rumor.*

Example of Scams

Forwarded



FG Lockdown Funds
bit.ly

Free N5,000 Lockdown Funds to sustain During this lockdown



FG has started giving out free N5,000 to sustain and support citizens during the lockdown period

I just claimed mine now and you too can claim yours

Note : You are only eligible to be credited once

Click below to claim yours intantly

<https://bit.ly/free-N5000>

Its available for just few people so claim yours now!!

5:00 AM

Research Question #2. Correcting misinformation on messaging apps

As observed by many commentators, it is crucial to understand what constitutes effective fact-checking before designing policies aimed at combating misinformation (Feldman, 2020; Pennycook & Rand, 2020). MIMs present specific challenges to fact-checking. Most MIMs, such as WhatsApp, limit the number of groups or individuals with whom one can share a piece of content at a given time. Then, the shareability of MIM-based fact-checks become paramount (Singh, 2020). Therefore, it is through repeated sharing within and across small groups that a fact-check message (also called “debunk”) can achieve an extensive reach and become effective in terms of changing beliefs. In this second section of the report we present evidence regarding what factors can make debunks more likely to be shared on MIMs, and eventually become widespread.

Research questions

Based on an analysis of recent scholarship on what might constitute a successful fact-checking strategy, we decided to conduct a randomized experiment that tested the efficacy of three factors: 1) the **format** of the debunk (text, audio, or image), 2) the nature of the **personal relationship** between those who send the fact-checking messages and those who receive it (close relationship or acquaintance), 3) and the **level of agreement** on certain political topics between those who send the fact-checking messages and those who receive it (consonant or dissonant political preferences).

Research Design

We implemented a *factorial research design* to study how the three factors identified above interact with one another in determining the success of a fact-checking effort. We used an interactive online survey to test how our three different factors influence how people evaluate corrections of misinformation. For example, we wanted to know if any of the formats that we identified (a text message, a voice message, or an image) fares better in attracting the curiosity of our participants, or can be more successful than other formats in correcting people’s inaccurate beliefs. Similarly, we wanted to see whether participants who receive a correction from a close relation (friend or family member) are more likely than participants who receive a correction from an acquaintance to believe the correction and re-share it with other friends. Finally, we wanted to test whether participants tend to weigh more heavily corrections that come from people they agree with politically, as opposed to people with whom they do not agree politically. In order to do so, we randomly assigned participants to different “con-

Factorial Design

Condition	Tie Strength (Correction Sharer)	Political Agreement (Correction Sharer)	Correction Format
1	Strong Tie	In-Group	Image
2			Voice Memo
3			Text
4		Out-Group	Image
5			Voice Memo
6			Text
7	Weak Tie	In-Group	Image
8			Voice Memo
9			Text
10		Out-Group	Image
11			Voice Memo
12			Text

Table 2. The debunk experiment design outlining the 12 treatment conditions. Each survey participant is assigned to one treatment condition.

ditions” (or possible combinations of factors). We tested 12 different conditions, meaning 12 different possible ways in which our three factors interact with one another. The table above presents our factorial design: 2 (tie strength) x 2 (political agreement) x 3 (correction format).

In order to assign participants to different conditions, we asked them to provide us with the following information:

- Rate their feelings towards various political parties (in order to measure political beliefs);
- Provide (only) the first name of a close friend or family member who strongly agrees with their political beliefs (a “strong in-group tie”), or a close friend or family member who strongly disagrees with their political beliefs, (a “strong out-group tie”) or a casual acquaintance who strongly agrees with their political beliefs (a “weak in-group tie”) or a casual acquaintance who strongly disagrees with their political beliefs (a “weak out-group tie”);

We then showed participants a pro-attitudinal example of misinformation (that is, a piece of news that, based on their political party identification, they would tend to agree with - see examples below). Subsequently, we presented the debunk message in one of three formats (image, video, text) correcting the misinformation they just saw and asked them to imagine they had received the debunk message from the tie they named. This allowed us to compare the effect of the debunk message originating from various social “sources”.



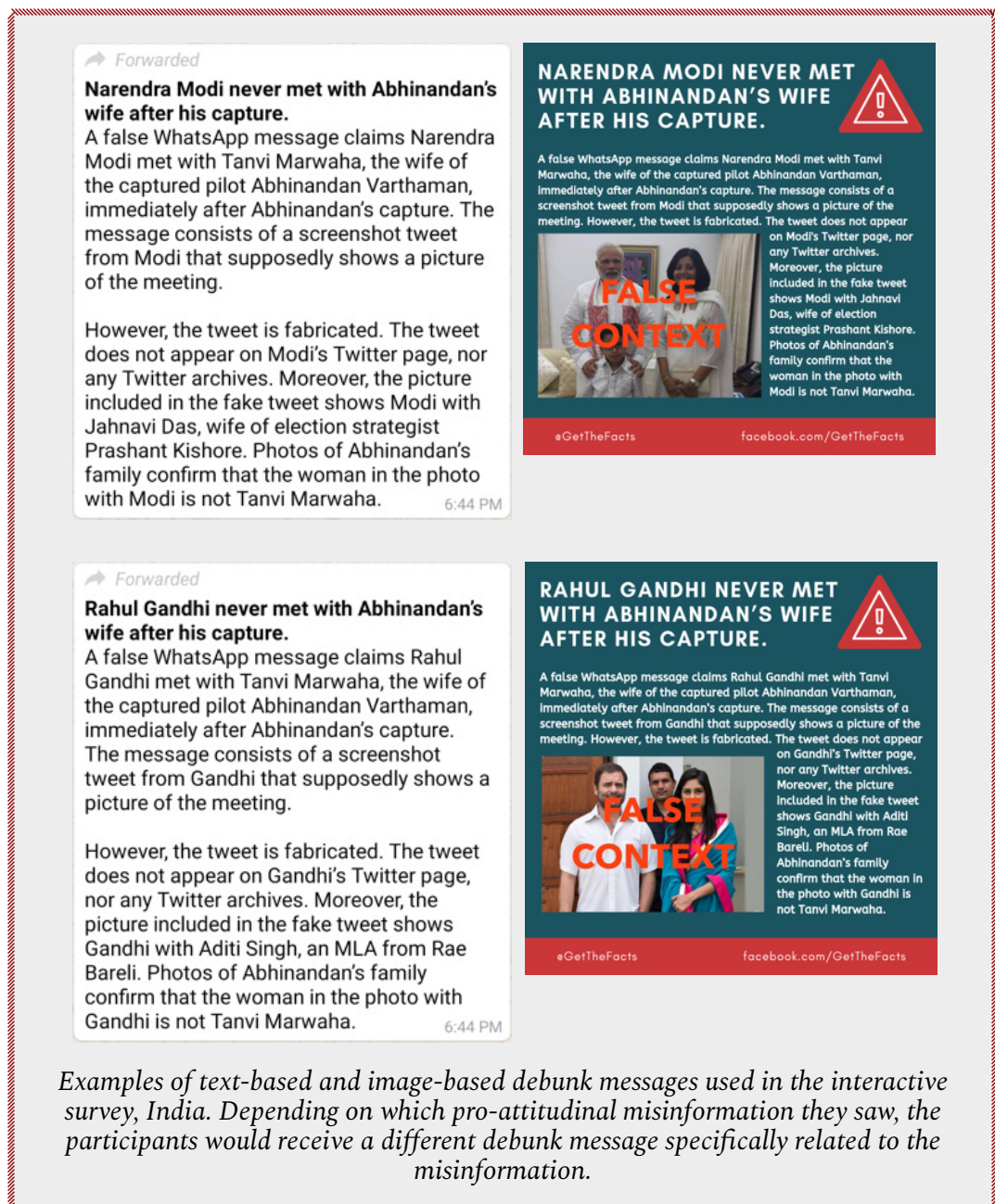
Results

Finding 1. Voice-based corrections generate more interest than text or image-based corrections

Each survey participant was randomly assigned to one correction format (either audio, text, or image, see examples below). After the participants saw the correction, they answered three questions designed to measure their interest in the debunk message: :

- How much attention did you pay to the message?
- How much did you think about the message?
- How interesting was the message?

In order to measure the efficacy of debunk messages in correcting beliefs, we asked the participants to rate their belief about the misinformation twice, once immediately after seeing the misinformation but before seeing the debunk message, and second after seeing the debunk message. The difference between these two measures the effectiveness of debunk messages in correcting the beliefs about the misinformation.



We found that, overall, audio formats create more interest than image or text (differences are statistically significant¹), but we didn't find significant differences between text and image formats. Similarly, we found that audio corrections are more effective than text- or image-based corrections in changing beliefs (differences are statistically significant²), but, again, we did not find a significant difference between text and image formats (Figure 3).

1 p = 0.005 in India, p = 0 in Pakistan
2 p = 0.007 in India, p = 0.058 in Pakistan

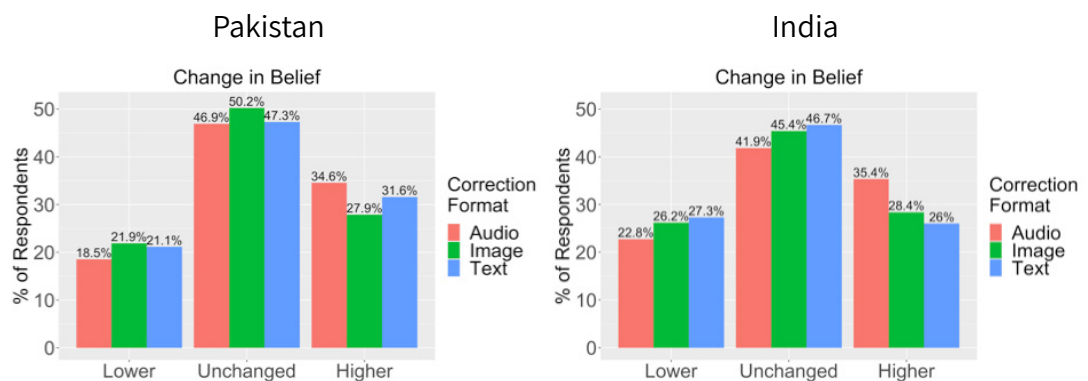


Figure 3. Audio treatment appears more effective in reducing belief in misinfo than text. (“Higher” means reduced belief in misinformation after exposure to debunk).

Intuitively, we might anticipate that richer, audio and image-based formats would be more interesting, and hence engaging, than text-based formats, and that this should result in greater belief change in cases when the content of the message was persuasive. The fact that participants showed greater interest in audio than image was thus unexpected.

We found that, overall, audio formats create more interest than image or text... Intuitively, we might anticipate that richer, audio and image-based formats would be more interesting, and hence engaging, than text-based formats... The fact that participants showed greater interest in audio than image was thus unexpected.

Our qualitative findings might help explain why this is the case. Interview data suggest that participants use MIMs to send frequent voice messages, often multiple times per day. While rare on other platforms, recording and sending voice messages is a very common way of communicating on MIMs. Participants perceive audio as the fastest and easiest way of sharing information: it does not require any typing, and users can record and listen to voice messages while doing other activities. It is possible to lock the recording button until the recording is over, freeing hands for other activities. Often users send “chains” of voice messages, which can be auto-played in sequence, without interruption. The phone screen can be locked, and the phone placed in a pocket, while listening to a voice message or to a chain of voice messages. That is, participants perceive audio messages as an effortless way of consuming information, one they are familiar with outside of the experimental setting. This might explain the ease and familiarity of voice based corrections. Higher interest in audio format versus images might also be driven by a more direct human interaction with audio than text or image. It might also be that voice-based corrections give a more personal experience than text or image based corrections that do not directly involve another individual. More research is nevertheless needed to clarify this finding, and also to understand why audio messages seem to be more persuasive than text or images in changing peoples’ beliefs.

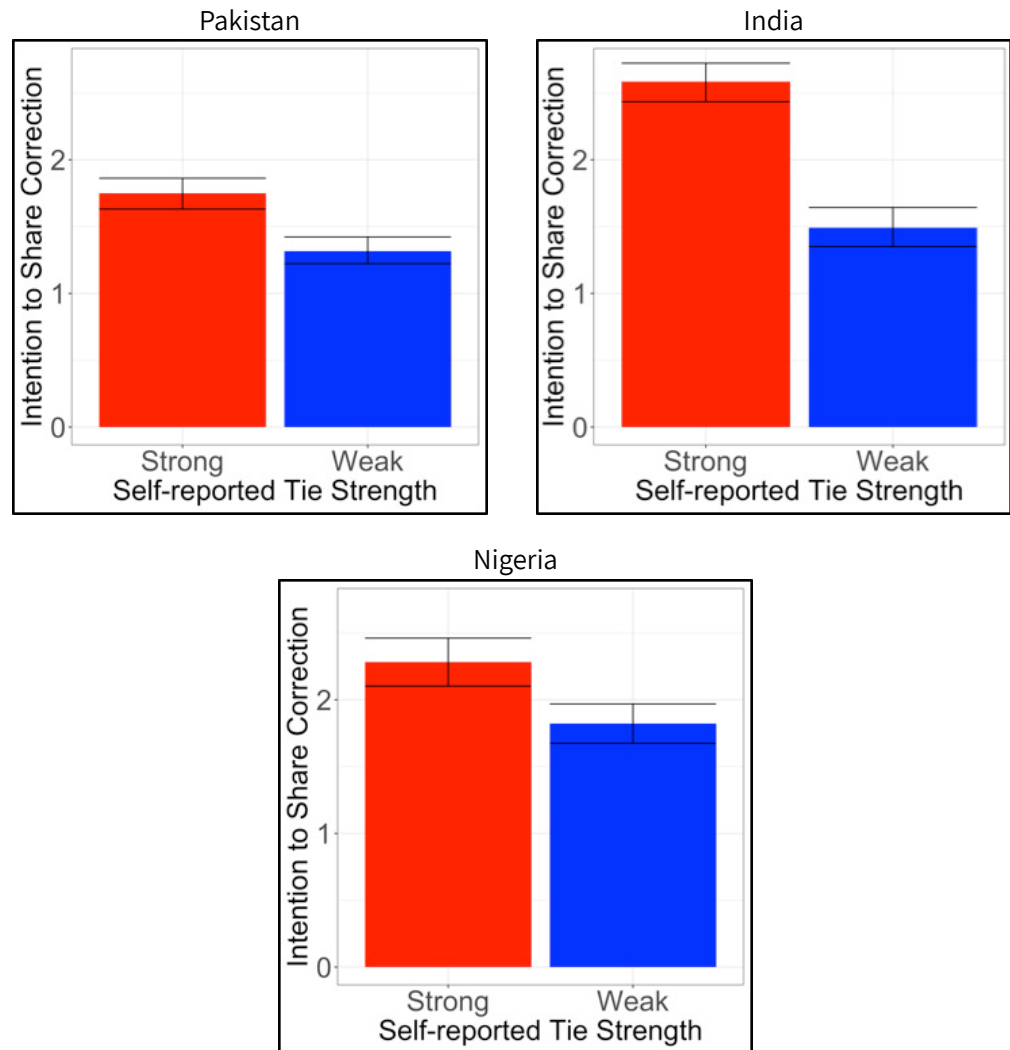


Figure 4. Respondents assigned to strong tie treatment are significantly more likely to share correction. Bars correspond to 95% confidence intervals. NOTE: y-axes represent the average intention to share the debunk message where intention ranges from 1 [not at all likely] to 5 [very likely].

Finding 2. Corrections received by a close friend or family member are re-shared more than corrections received by a casual acquaintance

We found clear evidence that participants would rather re-share corrections when they receive them from a person close to them, such as a family member or a close friend (i.e., strong tie), than from a casual acquaintance (i.e., weak tie) (statistically significant difference³, Figure 4). However, our results suggest that this greater tendency to share the correction with a close tie only emerges when the respondent believes the message to be accurate. In other words, when the

³ $p < 10^{-6}$ in India, $p < 10^{-7}$ in Pakistan, $p = 0.0001$ in Nigeria

Interview data suggest that for MIM users fact-checking is an unusual, socially awkward practice.

respondents do not believe the correction is credible, they are neither more nor less likely to re-share it regardless of whether it came from a strong or weak tie. The plots above are based on self-reported strength of ties, which are not necessarily causal. Nevertheless, we did find causal evidence of the same finding, but only among those respondents who passed our attention checks.

Finding 3. Corrections received by a like-minded individual are re-shared more than corrections received by an unsympathetic individual

In India and in Pakistan, we found clear evidence that participants would rather re-share corrections when they receive them from a like-minded individual, such as a person they agree with politically (i.e., in-group), rather than from someone who has opposite views, such as someone who disagrees with them politically (i.e., out-group) (differences are statistically significant, or nearly so in the case of Nigeria⁴, Figure 5).

Finding 4: Fact-checking in the comfort zone

Data analysis also shows that participants who receive a correction from a close friend who also agrees with them politically (both conditions are met), are much more likely to re-share such corrections than those who receive the correction from an acquaintance with whom they agree politically or a close friend with whom they do not agree politically (that is, when only one condition is met) (difference is statistically significant⁵). These results suggest that closeness and agreement are two components of trust and have additive effects on the intention to re-share the correction. In other words, the strongest effects appear when the contact is both a close and an in-group tie.

Interview data suggest that for MIM users fact-checking is an unusual, socially awkward practice. While nearly all interview participants reported being willing to verify the accuracy of potentially misleading forwards before re-forwarding them, when we asked how often they do so, and to provide a recent example of such a practice, only a few respondents were able to comment further. The implication is that most participants, while they are aware that some content shared on MIMs might be misleading, rarely engage in fact-checking practices as part of their daily information and media consumption routine. Most participants also shared that – while they are willing to fact-check MIMs forwards for their own sake – they are willing to correct other MIM users only if they are close friends with the person who sent the misinformation. These users believe that otherwise, such practice would make them feel uncomfortable, and that it would not be “their place” to correct others.

⁴ $p < 10^{-6}$ in India, $p < 10^{-7}$ in Pakistan, $p = 0.005$ in Nigeria.

⁵ $p < 10^{-8}$ in India, $p = 0.0003$ in Pakistan

In India and in Pakistan, we found clear evidence that participants would rather re-share corrections when they receive them from a like-minded individual, such as a person they agree with politically.

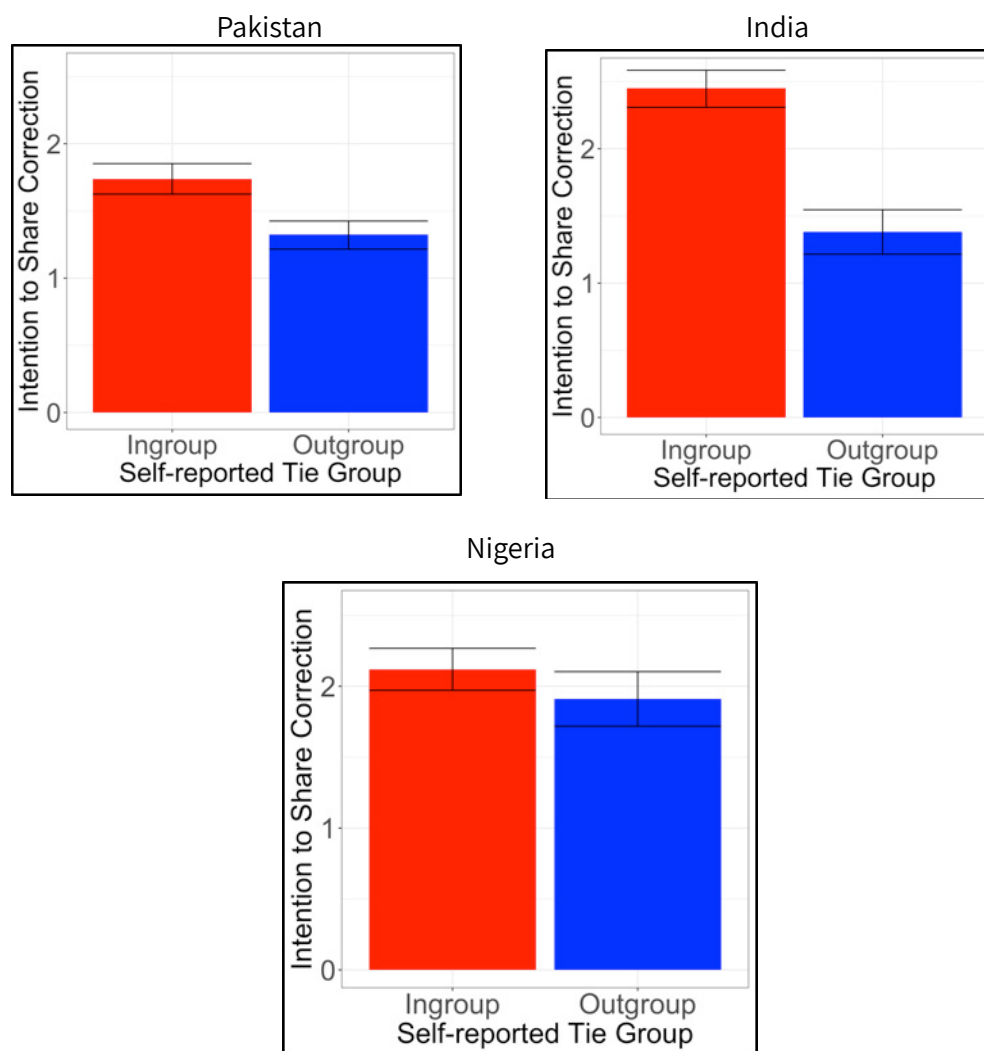


Figure 5. In India and Pakistan, respondents assigned to an in-group treatment are significantly more likely to share corrections. Bars correspond to 95% confidence intervals.

NOTES: (1) For ease of visualization, we have collapsed the 5-point group-of-ties scales into two categories (above and below the median). When collapsed in this way, the differences for Nigeria are not quite statistically significant. However, when we employ the full scale, the differences in intent to share between in-group and out-group ties are statistically significant; (2) y-axes represent the average intention to share the debunk message where intention ranges from 1 [not at all likely] to 5 [very likely].)

Overall, participants expressed several reservations about the intention of sharing corrections. Some other participants told us that they believe that people should be able to decide and verify by themselves what is true or not. Others told us that they believe that there is no point in trying to correct MIM users because they either simply do not

care about accuracy as a value worth pursuing, or they are politically motivated. Still other respondents indicated that they might encounter social sanctioning if they engage in correcting others, for example respondents from Nigeria indicated that it is against their cultural practices to correct seniors or people with a higher social status.

“I do not want to argue, I laugh and smile, and I correct only people that I personally know, otherwise I do not. People are lazy, they just forward whatever they think they agree on, they don't care if it is real or not. [...] I think it is WhatsApp's responsibility to clean up these messages, not mine. Otherwise their [WhatsApp] credibility will go down, people are getting tired of WhatsApp. There is no mechanism for fact check, it is not right.” (study participant #5, II wave of interviews, February 2020)

“To be frank with you, why should I try to comment on these fake news? People who share it have an agenda, they are trying to help or attach others. They are sharing the news with that intention and they all know about it, I would simply look naive if I would try to correct it.” (study participant #18, II wave of interviews, April 2020)

Obviously, there are exceptions. We identified four individuals who routinely correct others on WhatsApp. Two were professional journalists, one a trained librarian working at an academic library, and one was a retired professor. With the exception of these “crusade users,” most participants perceived sharing corrections as an extraordinary action that requires them to make an extra effort outside of their comfort zone, something that they would not normally perform. Participants seem to be willing to engage in such a stressful, unusual action only if they are confident that they are likely to be taken seriously. Then, participants seem to be willing to share corrections solely with like-minded individuals who they personally know and are trusted by, reducing the likelihood that their intervention would be dismissed out of hand by the recipient. These findings suggest that when WhatsApp users voluntarily engage in sharing debunks, corrections are shared (and, consequently, welcomed) mostly between like-minded individuals who are close to each other.

Conclusions

Overall, our findings paint a scenario in which messaging apps are widely used to exchange rumors that are created on the platform itself and rarely link to outside sources. More research is needed to understand the extent to which participants pay attention to these rumors, believe and remember them. However, MIM rumors seem to be shared and re-shared mostly on private chats, which are encrypted and inaccessible to the outside world. This means that, currently, there is not

Our findings paint a scenario in which messaging apps are widely used to exchange rumors that are created on the platform itself and rarely link to outside sources. More research is needed to understand the extent to which participants pay attention to these rumors, believe and remember them.

a systematic way to access and study MIM rumors comprehensively.

Anecdotal evidence suggests that MIMs can play an important role in disseminating misinformation, and that such information can sometimes prove quite dangerous. Consequently, it is important that researchers gain access to these data so that they can begin to assist the MIM platforms in developing effective tools to combat misinformation on their platforms. Absent such capacity, for the foreseeable future there will likely remain a paradoxical situation in which MIMs are an increasingly problematic source of misinformation, yet researchers and fact-checkers face significant challenges in identifying and implementing effective counter-measures.

One important step forward would be to **secure cooperation from the platforms** who could elect to make their data available to researchers while implementing steps to insure privacy. A model along the lines of **Social Science One** might be helpful in this regard.

Finally, we found that fact-checking activities are performed and welcomed primarily among individuals who are like-minded and feel close to each other. This reinforces the centrality of credible messengers to the chain of influence that starts with encountering and then sharing a rumor via MIMs. If credibility is the key to its spread of misinformation on MIMs, it also appears to be the key to mitigating that spread.

Acknowledgements

This research is funded by the Bill & Melinda Gates foundation and Omidyar Network. The authors would like to thank Claire Wardle and Nico Mele, who brought attention to the need for empirical research on messaging apps and misinformation and collected the necessary funding for this project. We would also like to thank all our research assistants at the Shorenstein Center for Media, Politics and Public Policy for their help in collecting and analyzing contextual and background data that informed our surveys, in particular: Pedro Armelin, Syed Zubair Noqvi, Chrystel Oloukoï, Tout Lin, and Mahnoor Umair. Finally, we would like to thank colleagues and collaborators from the Kennedy School and elsewhere for comments on earlier drafts of this report, and in particular Taha Siddiqui, Nic Dias, Julie Ricard, Yasmin Padamsee, David Ajikob, Sangeeta Mahapatra, Esther Htusan and Sai Htet Aung.

Bibliography

- Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6(14), eaay3539. <https://doi.org/10.1126/sciadv.aay3539>
- Aragão, A. (2017, May 31). WhatsApp Has A Viral Rumor Problem With Real Consequences. *BuzzFeed News*. <https://www.buzzfeednews.com/article/alexandre-aragao/whatsapp-rumors-have-already-provoked-lynch-mobs-a>
- Avelar, D. (2019, October 30). WhatsApp fake news during Brazil election 'favoured Bolsonaro.' *The Guardian*. <https://www.theguardian.com/world/2019/oct/30/whatsapp-fake-news-brazil-election-favoured-jair-bolsonaro-analysis-suggests>
- Bengani, P. (2019). India had its first 'WhatsApp election.' We have a million messages from it. *Columbia Journalism Review*. https://www.cjr.org/tow_center/india-whatsapp-analysis-election-security.php
- Campbell, A., Center, U. of M. S. R., Converse, P. E., Miller, W. E., & Stokes, D. E. (1980). *The American Voter*. University of Chicago Press.
- Collins, B. (2020, March 16). As platforms crack down, coronavirus misinformation finds a new avenue: Digital word of mouth. *NBC News*. <https://www.nbcnews.com/tech/tech-news/false-coronavirus-rumors-surge-hidden-viral-text-messages-n1160936>
- de Freitas Melo, P., Vieira, C. C., Garimella, K., de Melo, P. O. S. V., & Benevenuto, F. (2020). Can WhatsApp Counter Misinformation by Limiting Message Forwarding? In H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, & L. M. Rocha (Eds.), *Complex Networks and Their Applications VIII* (pp. 372–384). Springer International Publishing. https://doi.org/10.1007/978-3-030-36687-2_31
- Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School Misinformation Review*, 1(1). <https://doi.org/10.37016/mr-2020-001>
- Feldman, B. (2020, March 9). How Facebook Fact-Checking Can Backfire. *Intelligencer*. <https://nymag.com/intelligencer/2020/03/study-shows-possible-downside-of-fact-checking-on-facebook.html>
- Gold, H., & O'Sullivan, D. (2020, March 19). Facebook has a coronavirus problem. It's WhatsApp. *CNN*. <https://www.cnn.com/2020/03/18/tech/whatsapp-coronavirus-misinformation/index.html>
- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), eaau4586. <https://doi.org/10.1126/sciadv.aau4586>

Kemp, S. (2020). *Digital 2020: 3.8 billion people use social media*. <https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media>

Kim, H., & Walker, D. (2020). Leveraging volunteer fact checking to identify misinformation about COVID-19 in social media. *The Harvard Kennedy School (HKS) Misinformation Review*, 1(Special Issue on COVID-19 and Misinformation). <https://doi.org/10.37016/mr-2020-021>

Kuru, O., Campbell, S., Bayer, J., Baruh, L., & Cemalcilar, Z. (2020, Forthcoming). Understanding Information and News Processing in WhatsApp Groups: A Comparative Survey of User Perceptions and Practices in Turkey, Singapore, and the USA. *Comparative Approaches to Disinformation*, Harvard University, Cambridge MA.

Melo, P., Messias, J., Resende, G., Garimella, K., Almeida, J., & Benevenuto, F. (2019). WhatsApp Monitor: A Fact-Checking System for WhatsApp. *Proceedings of the International AAAI Conference on Web and Social Media*, 13, 676–677.

Mozur, P. (2018, October 15). A Genocide Incited on Facebook, With Posts From Myanmar's Military. *The New York Times*. <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>

Oyebanji, O., Ofonagoro, U., Akande, O., Nsofor, I., Ukenedo, C., Mohammed, T. B., Anueyiagu, C., Agyenyi, J., Yinka-Ogunleye, A., & Ihekweazu, C. (2019). Lay media reporting of monkeypox in Nigeria. *BMJ Global Health*, 4(6). <https://doi.org/10.1136/bmjgh-2019-002019>

Pennycook, G., & Rand, D. (2020, March 24). Opinion | The Right Way to Fight Fake News. *The New York Times*. <https://www.nytimes.com/2020/03/24/opinion/fake-news-social-media.html>

Singh, M. (2020, April). WhatsApp introduces new limit on message forwards to fight spread of misinformation. *TechCrunch*. <https://social.techcrunch.com/2020/04/07/whatsapp-rolls-out-new-limit-on-message-forwards/>

Stecula, D. A., Kuru, O., & Jamieson, K. H. (2020). How trust in experts and media use affect acceptance of common anti-vaccination claims. *Harvard Kennedy School Misinformation Review*, 1(1). <https://doi.org/10.37016/mr-2020-007>

Wardle, C. (2020). 7. Monitoring and Reporting Inside Closed Groups and Messaging Apps. In *Verification Handbook For Disinformation And Media Manipulation*. <https://datajournalism.com/read/handbook/verification-3/investigating-platforms/7-monitoring-and-reporting-inside-closed-groups-and-messaging-apps>