Social Debunking of Misinformation on WhatsApp: The Case for Strong and In-group Ties

Matthew A. Baum, Harvard
Eaman Jahani, MIT
Shubham Atreja, Michigan
Irene Pasquetto, Michigan
Context: Growth of MIM Users Worldwide

Number of MIM Users in Billions

Source: Graham & Jones (2018)
MIM’s rival social media in popularity

![Graph showing the world's most-used social platforms in January 2022](image-url)
Context: Most popular MIM’s Worldwide

FB Messenger (~1B users)

WhatsApp (~2B users)

We Chat (~1.2B users)
The Problem

- Example: In US, only 3% of Americans claim to get Ukraine news from MIMs, but...

In the last 24 hours, did you get any news or information about Russia’s invasion of Ukraine from the following sources? (Please select all that apply).

<table>
<thead>
<tr>
<th>Source</th>
<th>Holds misperceptions</th>
<th>No misperceptions, has uncertainty</th>
<th>No misperceptions, no uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print newspapers</td>
<td>9%</td>
<td>22%</td>
<td>68%</td>
</tr>
<tr>
<td>National network news</td>
<td>10%</td>
<td>35%</td>
<td>54%</td>
</tr>
<tr>
<td>Cable news</td>
<td>16%</td>
<td>35%</td>
<td>50%</td>
</tr>
<tr>
<td>Online news sites</td>
<td>14%</td>
<td>37%</td>
<td>50%</td>
</tr>
<tr>
<td>Radio programs</td>
<td>16%</td>
<td>36%</td>
<td>48%</td>
</tr>
<tr>
<td>Local news</td>
<td>14%</td>
<td>42%</td>
<td>44%</td>
</tr>
<tr>
<td>Friends and family</td>
<td>14%</td>
<td>47%</td>
<td>40%</td>
</tr>
<tr>
<td>Celebrity influencers</td>
<td>17%</td>
<td>47%</td>
<td>36%</td>
</tr>
<tr>
<td>Social media</td>
<td>16%</td>
<td>49%</td>
<td>35%</td>
</tr>
<tr>
<td>Entertainment media</td>
<td>17%</td>
<td>50%</td>
<td>32%</td>
</tr>
<tr>
<td>Mobile instant messaging apps</td>
<td>20%</td>
<td>53%</td>
<td>27%</td>
</tr>
<tr>
<td>I did not get any news about it</td>
<td>12%</td>
<td>77%</td>
<td>11%</td>
</tr>
</tbody>
</table>
The Problem

- Example: In US, only 3% of Americans claim to get COVID-19 news from MIMs, but...
Two-wave survey-experiments to understand misinformation on MIMs in 4 developing countries:

1. Nigeria
2. India
3. Pakistan
4. Myanmar
5. Brazil
Millions of WhatsApp Users in 2020, per Country

- India: 390.1, Pop 1.38B
- Pakistan: 46.2, Pop 225M
- Nigeria: 90, Pop 211M
MIMs vs. Social Media

- MIM users primarily interact with strong ties (family members, close friends, etc.)
- Users form groups of up to 256 members
- Misinfo can be shared across groups

- Social media users more likely to interact with weak ties (acquaintances, co-workers, etc.)
- Misinfo shared with “friends” or “public” via newsfeed
Two Primary Research Question Clusters (RQ’s)

1. **Prevalence:** How widely are misinformation stories recognized? How widely believed?

2. **Correcting:** When, why, and how will MIM-delivered debunks work?
<table>
<thead>
<tr>
<th>Location</th>
<th>Mobile Internet Users</th>
<th>Wave 1 N</th>
<th>Wave 2 N</th>
<th>Male %</th>
<th>Female %</th>
<th>Age Distribution</th>
<th>Education</th>
<th>Employment Status</th>
<th>Religion</th>
<th>Urban %</th>
<th>Rural %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nigeria</td>
<td>mobile internet users</td>
<td>1894</td>
<td>951</td>
<td>68</td>
<td>32</td>
<td>64% &lt; 35</td>
<td>&gt; High School Education</td>
<td>72% employed or full-time students</td>
<td>25% Muslim; 75% Christian or other faiths</td>
<td>85%</td>
<td>15%</td>
</tr>
<tr>
<td>India</td>
<td>mobile internet users</td>
<td>1799</td>
<td>1457</td>
<td>71</td>
<td>29</td>
<td>55% &lt; 30</td>
<td>&gt; High School Education</td>
<td>93% employed or full-time students</td>
<td>76% Hindu, 11% Muslim; 13% Christian or other faiths</td>
<td>88%</td>
<td>12%</td>
</tr>
<tr>
<td>Pakistan</td>
<td>mobile internet users</td>
<td>2003</td>
<td>1348</td>
<td>72</td>
<td>28</td>
<td>64% &lt; 30</td>
<td>&gt; High School Education</td>
<td>73% Household Chief Earner &gt; High School Education</td>
<td>98% Muslim</td>
<td>89%</td>
<td>11%</td>
</tr>
</tbody>
</table>
Misinformation Claims & Debunks

- Each respondent shown 10 claims (5 debunked, 3 mainstream, 2 placebo)
  - randomly assigned 5 of 10 debunks and 3 of 100 mainstream claims
- Examples of 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.”
- False claims were most recent that exceeded median shares across all qualifying claims identified in 2019 from fact checkers in each country;
- Mainstream claims had similar sharing prevalence and appeared same week as debunked claims
RQ Cluster #1: Prevalence - How widely are misinformation stories recognized? How widely believed?
Pakistan

Have you encountered this claim before?

Claim Type
- Misinfo
- MSM
- Placebo

% of Respondents
- No: 29.8% Misinfo, 41.4% MSM, 33.7% Placebo
- Maybe: 37.4% Misinfo, 39.2% MSM, 17.8% Placebo
- Yes: 32.9% Misinfo, 19.4% MSM
India

Have you encountered this claim before?

% of Respondents

Claim Type
- Misinfo
- MSM
- Placebo

No: 52.6% Misinfo, 39% MSM, 27.7% Placebo
Maybe: 52.9% Misinfo, 19.6% MSM, 21.5% Placebo
Yes: 36.2% Misinfo, 24.7% MSM, 25.7% Placebo
Have you encountered this claim before?

<table>
<thead>
<tr>
<th>Claim Type</th>
<th>No</th>
<th>Maybe</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misinfo</td>
<td>45.3%</td>
<td>32.2%</td>
<td>22.5%</td>
</tr>
<tr>
<td>MSM</td>
<td>44.2%</td>
<td>38%</td>
<td>33.9%</td>
</tr>
<tr>
<td>Placebo</td>
<td>28.1%</td>
<td>31.6%</td>
<td>24.2%</td>
</tr>
</tbody>
</table>

% of Respondents
Have you seen the claim?

% of Respondents Who Saw Claim

<table>
<thead>
<tr>
<th>Country</th>
<th>False Claim</th>
<th>True Claim</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>37</td>
<td>29</td>
</tr>
<tr>
<td>Nigeria</td>
<td>48</td>
<td>39</td>
</tr>
<tr>
<td>Pakistan</td>
<td>46</td>
<td>38</td>
</tr>
</tbody>
</table>

- False Claim
- True Claim
Have you seen the claim on MIMs vs. Newspapers/TV News? (India only)

% of respondents who saw claim

<table>
<thead>
<tr>
<th></th>
<th>True Claims</th>
<th>False Claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspapers/TV News</td>
<td>26</td>
<td>15</td>
</tr>
<tr>
<td>MIMs</td>
<td>29</td>
<td>37</td>
</tr>
</tbody>
</table>

(India only)
RQ Cluster #2: Correcting - When, why, and how will MIM-delivered debunks work
Research Questions

1) **Correction Format.** What fact-check formats are most effective in changing beliefs: audio, video, or text?

2) **Relational Closeness of Source.** Are fact-checks shared via MIMs that come from strong ties (close friends, family) more persuasive than those emanating from weak ties (acquaintances)?

3) **Attitudinal Similarity of Source.** Are debunks to pro-attitudinal false claims shared by an in-group member (same party affiliation) more effective than same message by out-group member (opposition party affiliation).
Hypotheses

RQ1: Correction Format (vividness)

H1: Image- and voice-based corrections will generate greater message interest than text-based corrections.

H2: Image- and voice-based corrections will change beliefs more than text-based corrections.

H3: Image- and voice-based corrections will be more likely to be shared on social media and messaging apps than text-based corrections.

RQ2: Relational Closeness of Source (trust/credibility)

H4: Corrections shared by strong ties will have larger effects on re-sharing than corrections shared by weak ties.

RQ3: Attitudinal Similarity of Source (trust/shared interests)

H5: Corrections shared by an in-group member will have larger effects on re-sharing than corrections shared by an out-group member.
STRENGTH-TREATMENT: is very dear to you – like a close friend, family member or romantic partner

GROUP-TREATMENT: strongly agrees with your political beliefs

Please write the first name of a person that is STRENGTH-TREATMENT and GROUP-TREATMENT

STRENGTH-TREATMENT: you do not know well, but interact with occasionally (online or offline)

GROUP-TREATMENT: strongly disagrees with your political beliefs
How confident are you that the claim below is inaccurate or accurate?

Misinformation Belief

Misinformation tweet

Debunk: Image

Debunk: Voice

Debunk: Text

Misinformation Belief

Debunk Interest

Debunk Re-sharing

Please imagine that the message on the next page was forwarded to you by TIE.
Misinformation Stimuli (in-group vs. out-group)

India

Upon hearing about our brave soldier Abhinandan, I could not wait and immediately flew down to Chennai to see his wife and family. Have assured them my utmost support in this time of despair.

Nigeria

JUST IN: PDP representative spent ₦5 billion on faulty wooden bridge

Mpigi Propen, a PDP member of the House of Representatives in Rivers State, commissioned a small wooden bridge be constructed over a local stream. The project reportedly cost ₦5 billion and fell apart in a week, injuring four people.

JUST IN: APC representative spent ₦5 billion on faulty wooden bridge

Mpigi Propen, an APC member of the House of Representatives in Rivers State, commissioned a small wooden bridge be constructed over a local stream. The project reportedly cost ₦5 billion and fell apart in a week, injuring four people.
Correction Format and Belief Change (India).

Narendra Modi never met with Abhinandan’s wife after his capture.
A false WhatsApp message suggests Modi met with Tanvi Marwaha, the wife of the captured pilot Abhinandan Vartaman, immediately after Abhinandan’s capture. The message consists of a screenshot tweet from Modi that supposedly shows him meeting with Tanvi Marwaha.

However, the tweet is fabricated. The tweet does not appear on Modi’s Twitter page, nor any Twitter archives. Moreover, the picture included in the fake tweet shows Gandhi with Aditi Singh, an MLA from Rae Bareli. Photos of Abhinandan’s family confirm that the woman in the photo with Gandhi is not Tanvi Marwaha.

Rahul Gandhi never met with Abhinandan’s wife after his capture.
A false WhatsApp message claims Rahul Gandhi met with Tanvi Marwaha, the wife of the captured pilot Abhinandan Vartaman, immediately after Abhinandan’s capture. The message consists of a screenshot tweet from Gandhi that supposedly shows him meeting with Tanvi Marwaha.

However, the tweet is fabricated. The tweet does not appear on Gandhi’s Twitter page, nor any Twitter archives. Moreover, the picture included in the fake tweet shows Gandhi with Aditi Singh, an MLA from Rae Bareli. Photos of Abhinandan’s family confirm that the woman in the photo with Gandhi is not Tanvi Marwaha.
Effects of Debunk Format (H1, H2) (India)

(“Higher” means reduced belief in misinformation after exposure to debunk).
Effects of Debunk Format (H1, H2) (Pakistan)

("Higher" means reduced belief in misinformation after exposure to debunk).
H4: Respondents in **strong tie** group more likely to share debunk.
H5: Respondents more likely to share debunk from in-group
INDIA: Self-Reported **Strength x Group** of Tie & Intent to Share

ANOVA p-val: 0

- **Numeric Correction Sharing**
  - Strong-Ingroup
  - Strong-Outgroup
  - Weak-Ingroup
  - Weak-Outgroup

- **Self-reported Tie Strength-Group**
  - Strong-Ingroup
  - Strong-Outgroup
  - Weak-Ingroup
  - Weak-Outgroup
Summary of Findings

- Users encounter more misinformation than mainstream news on MIMs; opposite on traditional news outlets
- But more likely to encounter mainstream news overall
- Users tend to pay more attention to (H1) to and believe (H2) voice-or image-based corrections, than text-based corrections (especially voice-based); image type does not significantly affect re-sharing (H3).
- Corrections received from a family member or a close friend (H4) are re-shared more than corrections received from an acquaintance;
- Corrections received from a like-minded individual (H5) are re-shared more than corrections received from a non-like-minded individual.
Limitation of Findings

- Qualitative interviews (15 per country) revealed that while respondents are more likely to believe fact checks received from close ties and like-minded individuals, they are hesitant to re-share fact checks because they view it is impolite, or an imposition, especially when potential recipient is a close tie.
- So, for arguably the most effective fact-checking context for MIMs – strong ties with fellow partisans -- respondents are cross-pressured in deciding whether or not to re-share debunks
Clarity for friends, confusion for foes: Russian vaccine propaganda in Ukraine and Serbia

BY KATRINA KEEGAN

This paper examines how Russia tailors its vaccine propaganda to hostile and