

the
shades of truth
study series

WhatsApp Research Awards for Social Science
and Misinformation

Preface

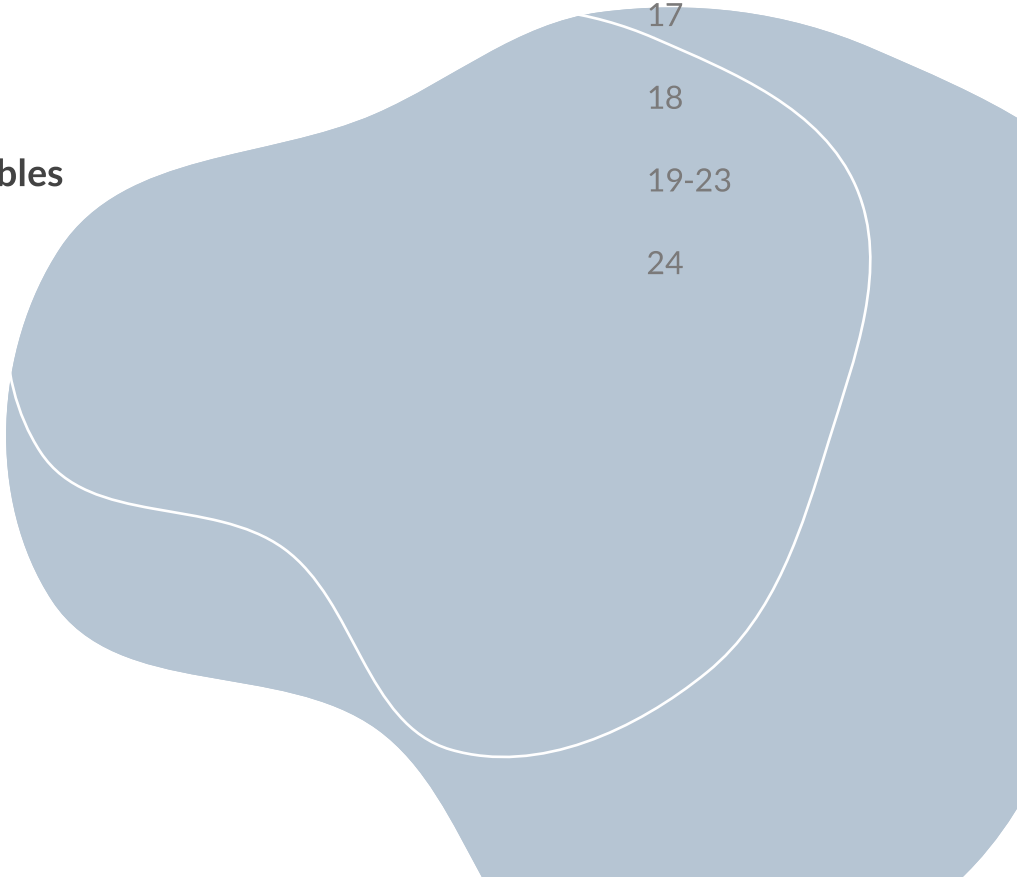
Online misinformation has become a central issue in the fight against COVID-19. This study series focuses on how we, social media users, interact with information that arrives on our screens in varying hues of truth during a pandemic. We investigate how age and scientifically accurate information affect beliefs around misinformation. And discover factors that shape behaviours around users correcting each other on WhatsApp.

We hope that the findings of this report and our research will strengthen the efforts of health agencies and WhatsApp in designing interventions to manage the infodemic.

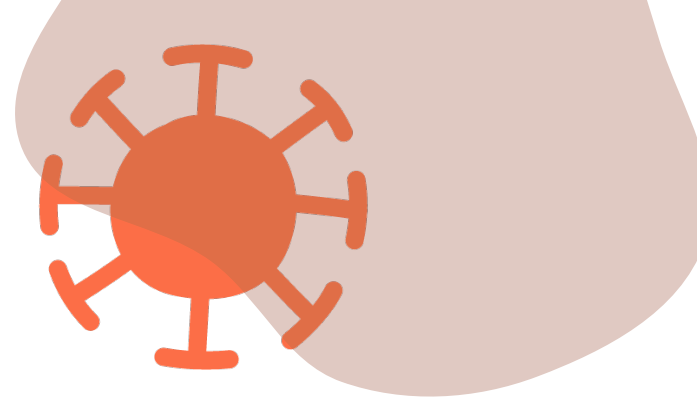


Table of contents

Problem Context	4-5
Participant Profile	6
Study 1 : Age and Shades of Truth	7-8
Study 2 : Information Vetting and Sharing	9-10
Study 3 : Social Correction Behaviour	11-12
Study 4 : Memory and Misinformation	13-14
Implications for Public Health Communication	15
Recommendations for UX Design	16
Key Takeaways	17
Future Research	18
Glossary of Variables	19-23
Team	24



Problem Context



The infodemic begins

December 2019

China identifies novel coronavirus after a cluster of pneumonia cases from Wuhan in Hubei province

Unknown virus creates information lacuna that triggers a misinformation outbreak on global social media platforms including WhatsApp

March 2020

World Health Organization warns against "Infodemic" before declaring COVID-19 a pandemic

Definition

(health misinformation is) any health-related claim of fact that is false based on current scientific consensus¹

Spectrum of misinformation



Impact of misinformation

CONFUSION

Public unsure about what information to believe or ignore

VIOLENCE

Conspiracy theories lead to attacks on 5G towers

STIGMA

Discriminatory attacks on minority communities because of reference to COVID-19 as "Chinese Virus"



WHO initiates ACTIONS



EPI-WIN Platform



Collaborations with social media companies to block misinformation



Global infodemic management consultation



WhatsApp Tiplines

1. Chou, W.-Y. S., Gaysynsky, A., & Cappella, J. N. (2020). Where We Go From Here: Health Misinformation on Social Media. *American Journal of Public Health, 110*(S3), S273-S275. doi:10.2105/ajph.2020.305905

Factors affecting psychological responses to misinformation

Shades of truth

Previous research has shown that users' and audiences' response to messages could vary based on the level of detail, source and factual content.

But we understand little about the psychological response of WhatsApp users to misinformation that arrives in various hues or shades of truth.



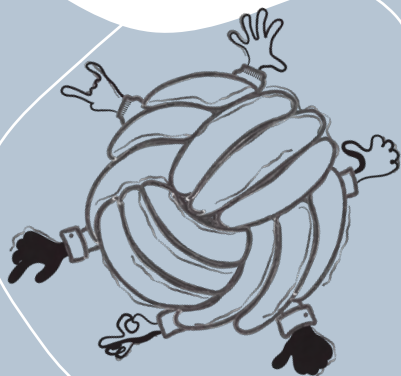
Corrective information

- Interventions**
- Fact-checking
 - Appeals to credibility
 - Source of corrective information
 - Warnings of possible presence of misinformation
 - Alternative explanation
 - Credibility of the misinformation itself
 - Corrections with factual elaboration
 - Social discrediting of misinformation source

Several interventions to combat online health misinformation have been tested and launched by academics, data literacy entrepreneurs and social media companies.





However, little is known about the effectiveness of corrective information interventions by public health agencies like the World Health Organization.

Linkages between individual characteristics and health misinformation susceptibility



Misinformation pollutes our memory of original events with details that do not exist, effectively making our memory malleable

Older adults have previously been shown to

-  possess a greater confidence in false memories
-  demonstrate a tendency to forget the source of the original misinformation
-  possess reduced cognitive functioning and abstract reasoning
-  exhibit limited capabilities to differentiate between various 'shades of truth'

Study Locations



N = 725
UK



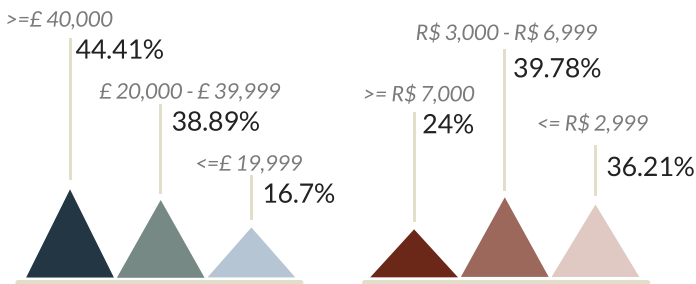
N = 729
Brazil

UK and Brazil have reported the highest number of lost to COVID-19 in Europe and South America respectively.

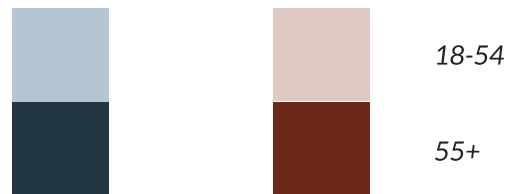
WhatsApp is among the most popular social media apps in both these countries; however it is also a proven vector of misinformation.

Participant profile

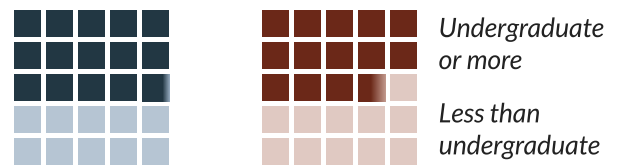
Annual Income



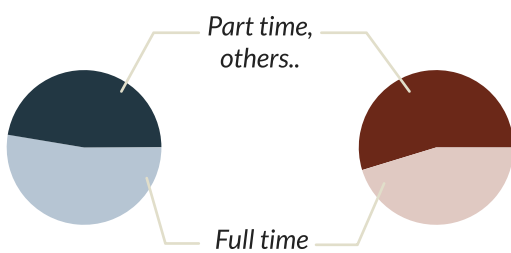
Age



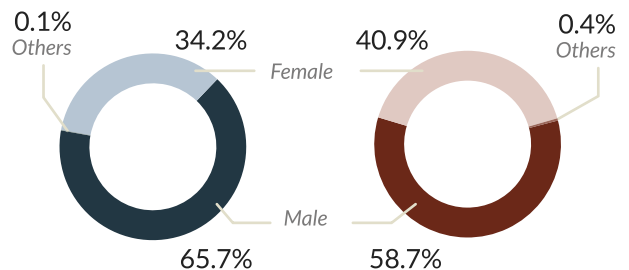
Education



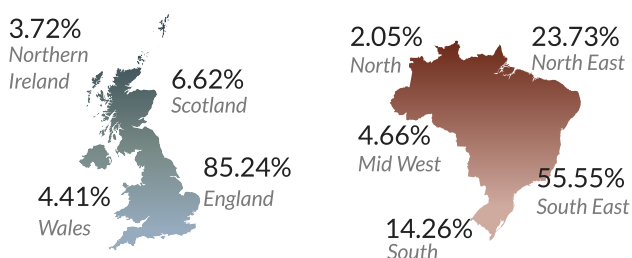
Employment



Gender



Geography

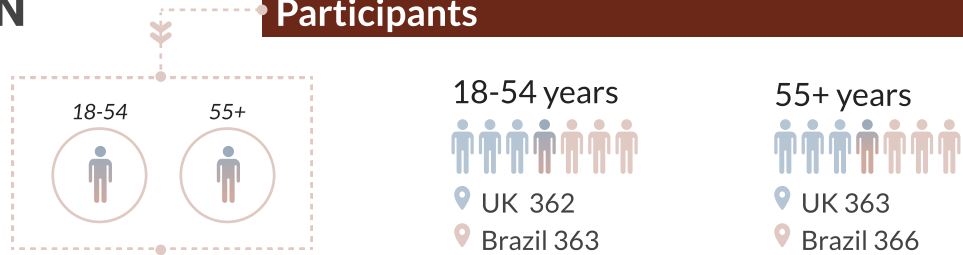


Participants in both UK and Brazil were recruited through Qualtrics' panel of survey respondents

How is the effectiveness of corrective information affected by age and different types of misinformation that people have been exposed to?

STUDY DESIGN

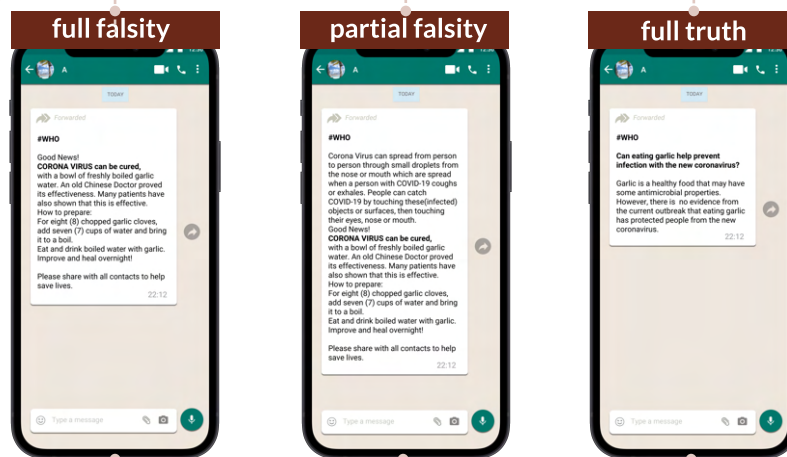
Participants



Participants were then randomly assigned one of three messages based on 'shades of truth'

Exposure 1

Stimuli: Misinformation type or 'shades of truth'



The stimuli were created by extracting specific elements from two kinds of existing WhatsApp messages already in circulation in Brazil and the UK

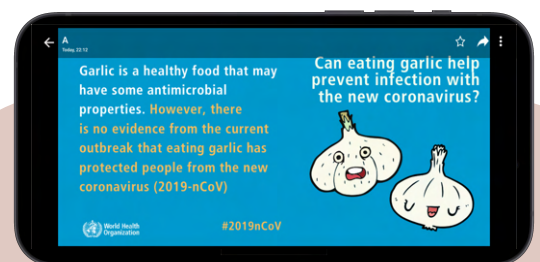
Misinformation features embedded in all three stimuli :

- false attribution to a health authority (WHO)
- garlic as a claimed COVID-19 cure

Exposure 2

Stimulus: Corrective information

All participants were then presented with a corrective message in the form of a COVID-19 mythbuster infographic from WHO's EPI-WIN online resource website



Main Outcomes (measured after exposure to each stimuli)

Misinformation Belief

Message Credibility

Intention-to-Share

2. Vijaykumar, S., Jin, Y., Rogerson, D., Lu, X., Sharma, S., Maughan, A., Fadel, B., Costa, M.S.O., Pagliari, C., Morris, D., (In Press). How shades of truth and age affect responses to COVID-19 (mis)information: Randomized survey experiment among WhatsApp users in UK and Brazil. *Humanities and Social Sciences Communications (Palgrave Communicatons)*, doi: 10.1057/s41599-021-00752-7.

Age and Shades of Truth

INSIGHTS

Misinformation Belief

United Kingdom

2.95 1.42

18-54 years

55 years and over

Age

Younger adults were significantly more likely to believe misinformation than older adults and also more likely to share it

Older adults are able to deploy their more extensive general knowledge to critically evaluate new information

Belief in fully or partially false messages increased after older participants were exposed to corrective information

Misinformation type

Messages which are partially true can be particularly dangerous in entrenching misinformation belief and may trigger further sharing based on individuals' false impressions that such information appears to be correct

Users correctly identified accurate information irrespective of its stated source, which may have implications for so-called imposter messages

Participants rated the full truth messages as more credible and were intent on sharing them more than the full-falsity or partial-falsity messages, demonstrating critical message evaluation abilities

Perceived Credibility of Misinformation

2.53

UK

1.95

Brazil

Perceived Credibility of Corrective Information

3.26

UK

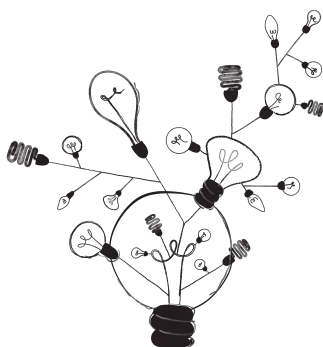
2.85

Brazil

Corrective information

Corrective information from the WHO reduced misinformation beliefs in most sub-groups, and significantly increased perceived credibility and intention-to-share across all sub-groups

These findings suggest that corrective information from public health authorities is critical in



sustaining previously held accurate information



debunking misinformation



intervening against harmful misinformation-sharing

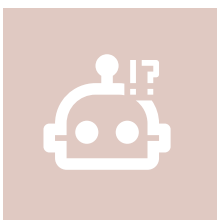
In addition to prior exposure, how do emotional and cognitive factors affect users' digital behaviours in relation to misinformation?

Predictors



Cognitive Responses

- Misinformation belief
- Misinformation familiarity



Emotional Response

Feeling of uncertainty due to exposure to varied shades of truth might evoke different emotions among individuals

- Confusion
- Anxiety
- Hope



Information Vetting

A two-stage psychological process users might take to determine whether the information is accurate or not:

Primary Vetting

User makes a judgment on whether the information one received is accurate in terms of multiple characteristics of the information itself

Secondary Vetting

User makes a further judgment on whether the conclusion one made in the primary vetting stage is indeed valid

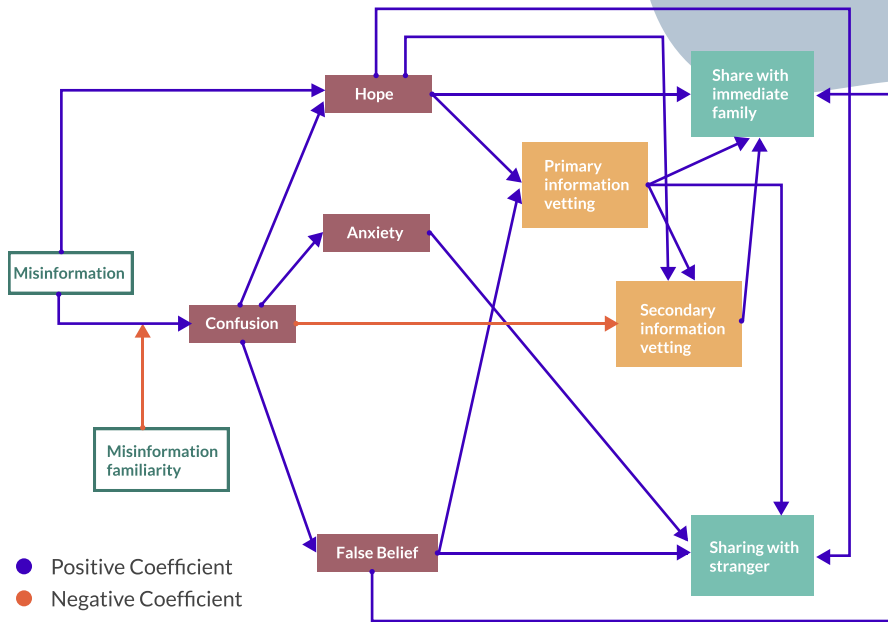
Main Outcomes (measured after exposure to each stimuli)

Intention to share misinformation

Information vetting

Information Vetting and Sharing

INSIGHTS



Conceptual model describing the effects of emotions on information vetting and sharing behaviours

Individuals' hope, anxiety, misinformation belief might be aroused by their confusion.

Hope and misinformation belief predict users' intention to engage in primary information vetting.

Hope, confusion and primary information vetting might influence users' intention to engage in secondary information vetting.

Linkage between emotions and sharing behavior

Individuals with higher misinformation beliefs and hope are more likely to share the message with others

Anxiety functions as a trigger for people to share the information with strangers via social media

Confusion inhibits people from sharing the (mis)information with their immediate family members via WhatsApp

Linkage between misinformation veracity and sharing behavior

Confusion may evoke hope and anxiety, which in turn, could influence user behaviours around information vetting and sharing

Individuals may be motivated to take a two-stage vetting process, first by assessing the information itself (primary vetting), followed by judging their own assessments (secondary vetting)

Secondary vetting is a supplementary vetting process to make individuals more confident, which trigger their intentions to share (mis)information with immediate family

What types of social correction behaviours do WhatsApp users engage in? What are the health, technological and demographic factors that affect these behaviours?

Cross sectional study

Focuses on the role of and extent to which WhatsApp users demonstrate the tendency to correct their social peers who might deliberately or unwittingly share COVID-19 misinformation on this popular messaging platform



Independent variables



Demographic factors

- Age
- Gender
- Household income
- Education



Health beliefs

- Perceived severity to COVID-19
- Perceived susceptibility to COVID-19



Misinformation factors

- Misinformation exposure
- Message credibility



Technological factors

- Information seeking norms
- Critical message evaluation
- Time spent discussing COVID-19

Main Outcomes (measured after exposure to each stimuli)

Active feedback to group

Active feedback to sender

Passive or no feedback

Social Correction Behaviour

INSIGHTS



We found greater passivity among younger participants to engage in peer correction.

Higher educational attainment was associated with providing private feedback to the original sender.

Passive/No Feedback	2.33 18-54 years	2.12 55 years and over
Active Private Feedback	3.82 lesser educated	3.97 more educated

Perceived severity affects, information seeking norms

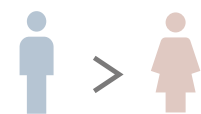
Critical message evaluation significantly affected all three peer feedback behaviours

- Active Private Feedback
- Active Group Feedback
- Passive/No Feedback

Impact on peer feedback behaviour



Younger participants had a greater preference than older participants for engaging in passive or no feedback



Male participants indicated a higher preference than female participants for engaging in active group feedback



Participants with an undergraduate level degree or higher showed a preference to engage more in active private feedback than those without an undergraduate degree



No significant difference of income was found for any of the dependent variables

How effective are corrective information interventions when they are followed by misinformation?

STUDY DESIGN

online experiment through Qualtrics' panel of survey respondents

A group of participants from UK were randomly selected from the age groups

18-34 years
N = 100

55+ years
N = 100

Baseline Phase

Measurement of existing views

10 statements containing misinformation about common COVID-19 myths were presented to each participant

Phase 1

Exposure to corrective information

10 infographics from WHO were presented about common COVID-19 myths

Phase 2

Exposure to corrective information vs misinformation

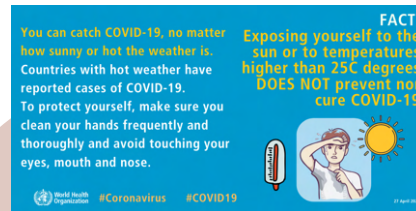
10 WhatsApp-like messages - 5 true, 5 false - about COVID-19 were shown to participants

Sample statement

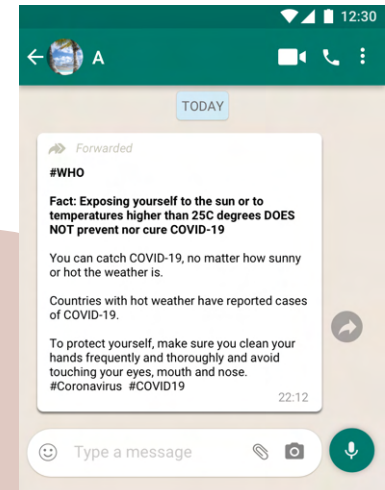
Please rate the truthfulness of the statement

“Exposing yourself to high temperatures can prevent or cure Coronavirus (COVID-19)”

Sample infographic from WHO



Sample WhatsApp like message



Main Outcomes (measured after exposure to stimuli in each phase)

Truthfulness of the statement

Perceived credibility of the statement

Willingness to share misinformation in the statement

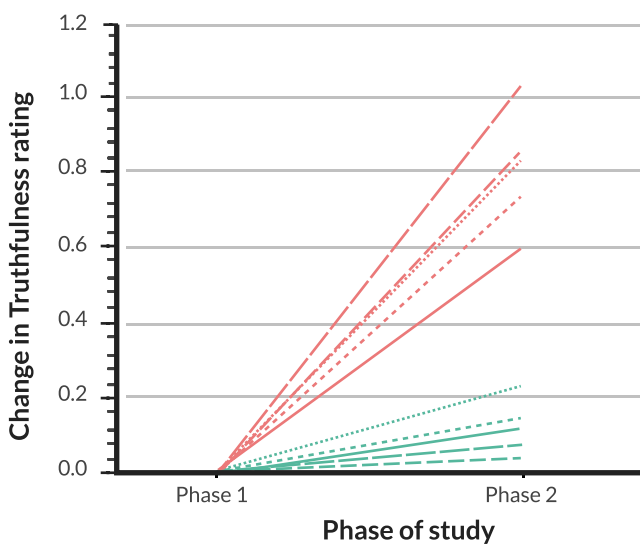
Memory and Misinformation

INSIGHTS



Experimental manipulation of information experienced shortly after retrieval influences subsequent memory accuracy

Change in truthfulness ratings after exposure to false (red) and truthful (green) information



Individual memories pertaining to the truthfulness of common COVID-19 myths can be

improved and maintained through repeated exposure to correct information

distorted through a single (30-second) exposure to false information

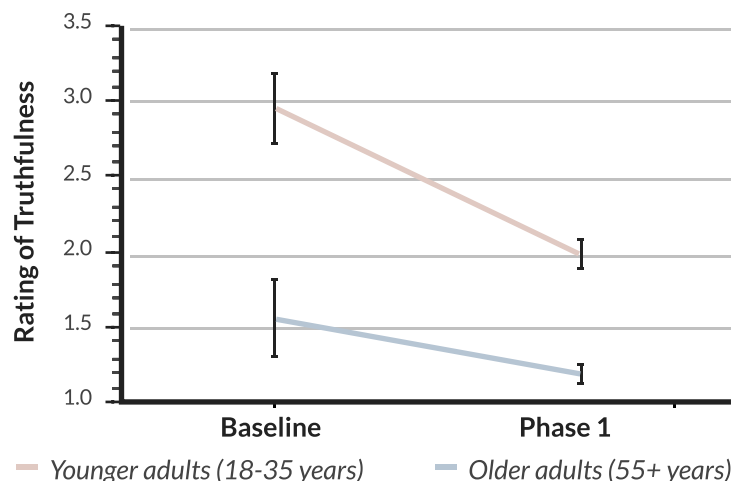
- Topic 1: Antibiotics
- Topic 2: High Temperatures
- Topic 3: Hot Food
- Topic 4: Hand dryers
- Topic 5: 5G
- Topic 6: Pneumonia Vaccines
- Topic 7: Cold Weather
- Topic 8: Garlic
- Topic 9: Rinsing with saline
- Topic 10: Houseflies

Memory accuracy was inversely related to the likelihood to share misinformation about COVID-19

These effects presented in younger (18-35 years old) and older adults (55+ years old), are in keeping with behavioural and neuroscientific evidence demonstrating that

memories are not fixed once encoded and re-enter a malleable state following retrieval

Improvement in truthfulness ratings in response to corrective information



Implications

for Public Health Communication



Actively engage with younger adults

Our study highlights the need to actively engage with younger adults given their increased vulnerability to misinformation belief, lower levels of knowledge and increasing susceptibility to COVID-19.



Incorporate audience segmentation activities

Our findings also clearly speak to the need for public health organisations to incorporate audience segmentation activities as part of their infodemic management initiatives.

Ongoing efforts (e.g. EPI-WIN) have made a promising start by creating targeted communication materials for different audience groups. These efforts need to be bolstered by identifying demographic segments through audience segmentation so that risk communication can be better tailored to suit specific informational and belief sensitivities.



Pre-test messages

This important step will help health agencies evaluate the effectiveness of these messages before making them available for public consumption. This inturn would help to avoid unintended backfire effects of myth-busting activities.



Evaluate the impact of corrective interventions

Consistent with our approach, we recommend greater emphasis on evaluating corrective interventions using scientifically robust study designs.

Such evaluations will help create an evidence base of strategies to combat misinformation which could prove valuable in containing and managing future infodemics.



Calibrate communications for older adults

Communications reaching older adults will need to be calibrated in terms of frequency and content in order to avoid reinforcement of beliefs in misinformation that they receive on social media platforms like WhatsApp.



Identify specific cognitive factors that trigger information sharing

Because the viral spread of misinformation is driven by sharing behaviors, we suggest that future research identify cognitive factors – e.g. misinformation beliefs and perceived credibility – activated by misinformation exposure that tend to influence sharing behaviours.

Recommendations

for UX Design

Envelop messages with intent and emotion

Capture the intentions and emotions of the sender before a message is shared. This may inculcate habits of responsible sharing behaviour by adding a moment of thoughtful pause and including a context to the message.



Provide indicator for forwards beyond the platform

Help users recognize (rather than recall) the source/sender of the multimedia content, whether they are using WhatsApp or not. This will help users to identify the source and take informed decision if they choose to share the content.

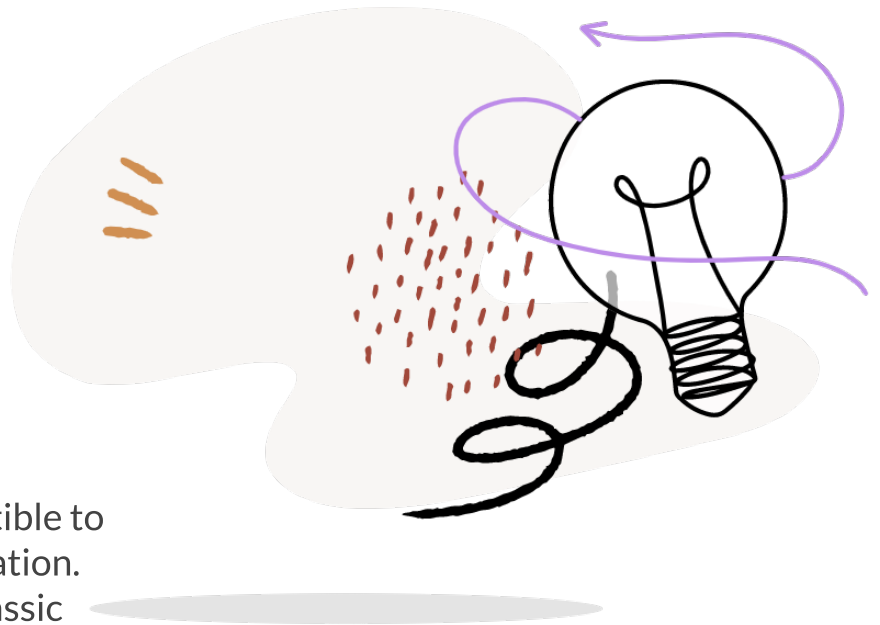
Reduce notability of sharing function

Limit the notability, weightage and convenience of the sharing functionality. Its ease-of-use may be retained but with reduced access, which may help users who feel social anxiety and peer pressure to share.

Provide ethical affordance of features

Suggest ethically permissible usage of its features from time to time for WhatsApp users who are new to technology and have limited digital literacy. This may be done by making help and usage tips accessible.

Key Takeaways



Younger adults are more susceptible to believing and sharing misinformation. This key finding conflicts with classic psychological research around the misinformation effect but is consistent with emerging evidence in the area of health misinformation research.

Our findings add useful nuances around the effects of partial truths, around younger or older users' fact-checking behaviours and about the role of misinformation in their social worlds.

WhatsApp users vet information using cognitive processes. However, this process of information vetting may be influenced by emotions triggered by misinformation.

WhatsApp users may be actively correcting social peers who share misinformation on the platform. Younger, and lesser literate age groups might be more resistant to offering such corrections.

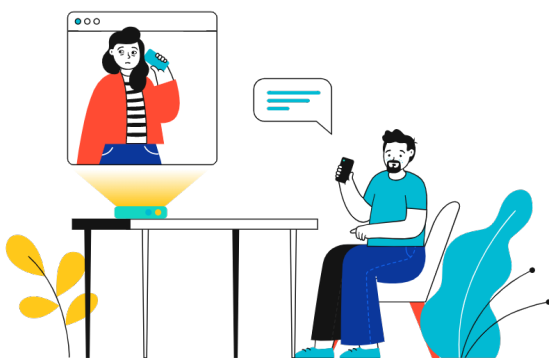
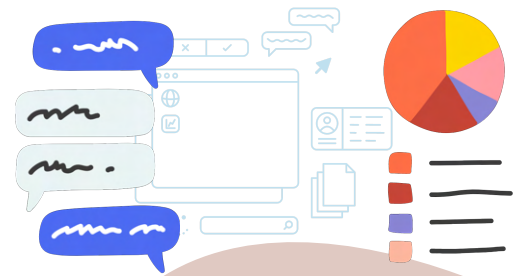
The WHO myth-buster infographics can be effective in reducing misinformation beliefs. However, it will be important to repeatedly expose audiences to such corrective messages for the effects to sustain.

Future Research



Our study designs may be replicated in the context of other types of misinformation, for example anti-vaccine mythology where beliefs may be more entrenched.

Further research is needed to understand the social affordances/value of misinformation sharing, including the role of in-group/out-group pressure and aspects of identity, social comparison and self-censorship.



There is a need to examine how different types of actors derive perceived or actual value from forwarding different types of health misinformation, and how this intersects with value systems, feelings, cultures and relationships in online spaces.

Glossary of Variables

Active feedback to group

If you find that there is incorrect or fake Covid19 misinformation in a WhatsApp forward you have just received you will...

1. Inform the whole group that the forward had inaccurate information
 2. Address the sender individually but send the message to the entire group
 3. Supply the accurate information to the whole group
 4. Address the sender individually but supply the accurate information to the entire group
-

Active feedback to sender

If you find that there is incorrect or fake Covid19 misinformation in a WhatsApp forward you have just received you will...

1. Inform the sender immediately
 2. Inform the sender privately/separately that the forward had inaccurate information
 3. Supply the accurate information to the sender privately/separately
-

COVID-19 knowledge

1. The most common symptoms of Coronavirus (COVID-19) are fever, tiredness, and dry cough
2. The time between catching the Coronavirus (COVID-19) and beginning to have symptoms of the disease ranges from 1-14 days
3. Coronavirus (COVID-19) is mainly transmitted through contact with respiratory droplets than through the air
4. Antibiotics are not effective in preventing or treating Coronavirus (COVID-19)
5. Wearing multiple masks is not effective against Coronavirus (COVID-19)

Score / Scale

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly agree

Score / Scale

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly agree

Score / Scale

True / False

Critical message evaluation

When I view social media messages posted by my friends, peers, or people like me, I think about...

1. The purpose behind a message/post
2. Who created a message I see on social media
3. What people who made a media message want me to believe
4. The things that advertisers do to get my attention
5. Whether the things that message senders want me to do are good for me

Score / Scale

- 1 = Never
- 2 = Rarely
- 3 = Sometimes
- 4 = Often
- 5 = Always

Emotional responses

Please indicate your level of agreement with the following statement.

"Immediately after reading the WhatsApp forward, I felt..."

1. Hope
2. Anxiety
3. Confusion

Score / Scale

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly agree

Info seeking

Please indicate your level of agreement with the following statement.

1. I will consider others' Coronavirus (COVID-19) experience on WhatsApp before I make decisions regarding COVID-19
2. I intend to seek Coronavirus (COVID-19) related information on WhatsApp frequently
3. I will ask others on WhatsApp to provide me with their suggestions before I make decisions regarding Coronavirus (COVID19)

Score / Scale

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly agree

Info vetting motivation

I would further verify the information in this message because...

1. It is important to me personally.
2. The information affects me.
3. The information interests me.

Score / Scale

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly agree

Intention-to-share

How likely would you be to share this WhatsApp forward with...

1. Friends
2. Family (Immediate)
3. Family (Extended)
4. Colleagues
5. Strangers
6. Nobody

Misinformation belief

Garlic can cure me of the coronavirus (COVID-19)

Message credibility

Please indicate how well the following words describe the WhatsApp forward you have just read:

1. Accurate
2. Believable
3. Authentic

Misinformation exposure

Please indicate if you have come across these kinds of message on WhatsApp:

1. Drinking water more frequently can prevent Coronavirus (Covid19)
2. Natural remedies can protect you from Coronavirus (Covid19), such as using colloidal silver, essential oils, garlic, MMS (chlorine dioxide) or vitamin C
3. Eating warm food/drink and avoiding cold food/drink can protect from Coronavirus (Covid19)
4. Salt-water gargling can protect you from Coronavirus (Covid19)
5. Putting clothes in the sun to disinfect them can prevent Coronavirus (Covid)

Score / Scale

(Studies 1 to 3)

1 = Highly unlikely to

5 = Very likely

(Study 4)

1 = Highly unlikely and

9 = Highly likely

Score / Scale

1 = Completely inaccurate

2 = Inaccurate

3 = Neither inaccurate nor accurate

4 = Accurate

5 = Completely accurate

Score / Scale

1 = Very poorly

2 = Poorly

3 = Fair

4 = Well

5 = Very well

Score / Scale

Yes/No

Misinformation familiarity

Have you seen a similar message before?

Score / Scale

Yes / No

Passive or no feedback

If you find that there is incorrect or fake Covid19 misinformation in a WhatsApp forward you have just received you will...

1. Inform the sender after waiting for a while
2. Not inform the sender at all
3. Take no action at all

Score / Scale

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly agree

Perceived severity of COVID-19

Please indicate your level of agreement to the following statements:

1. I believe Coronavirus (Covid19) is severe
2. I believe Coronavirus (Covid19) has serious negative consequences
3. I believe Coronavirus (Covid19) is not serious as media says

Score / Scale

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly agree

Perceived susceptibility to COVID-19

Please indicate your level of agreement to the following statements:

1. It is likely that I will get Coronavirus (Covid19).
2. I am at risk of getting Coronavirus (Covid19).
3. It is possible that I will get Coronavirus (Covid19)

Score / Scale

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly agree

Perceived truthfulness

Please rate the truthfulness of the statement:

"Garlic can cure me of the Coronavirus (COVID-19)"

Score / Scale

- 1 = Not at all
- 9 = Very

Primary info vetting

I think the information in this message is accurate because the information...

1. Provides me with a certain solution
2. Is from the source I trust
3. Is from a credible source
4. Is from the source I am familiar with
5. Is from an accountable source

Score / Scale

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly agree

Secondary info vetting

The message is accurate because...

1. It's consistent with my common sense
2. It looks conclusive to me
3. It was intended to deceive me
4. It satisfies my need for information
5. The information makes me feel confident

Score / Scale

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly agree

Time spent discussing COVID-19

How much time do you spend looking at or discussing COVID-19 information on WhatsApp each day?

Score / Scale

- 1 = No time spent at all
- 2 = Less than one hour
- 3 = Between one and three hours
- 4 = Between three and five hours
- 5 = More than five hours

Disclaimer: This is a visual report for the purposes of scientific dissemination to lay audiences, and hence the references section has been omitted. The only two instances where papers have been cited as footnotes are 1) where the verbatim definition of the original authors was used, and 2) where a research paper from this project was published in an academic journal.

References to any of the above scales can be obtained by contacting Dr. Santosh Vijaykumar (santosh.vijaykumar@northumbria.ac.uk). We are thankful to all authors whose work has informed and shaped our research.



Uncovering shades of truth

Created by Dr. Santosh Vijaykumar

9 Contributors



Dr. Santosh Vijaykumar

santosh.vijaykumar@northumbria.ac.uk
University of Northumbria

Principal Investigator



Dr. Yan Jin

yanjin@uga.edu
University of Georgia

Co-Principal Investigator



Dr. Claudia Pagliari

Claudia.Pagliari@ed.ac.uk
University of Edinburgh

Co-Principal Investigator



Ms. Swati Sharma

arcturus23.swati@gmail.com
Independent Consultant

Designer



Mr. Daniel Rogerson

daniel.t.rogerson@northumbria.ac.uk
University of Northumbria

Research Assistant



Ms. Xuerong Lu

Xuerong.Lu@uga.edu
University of Georgia

Research Assistant



Ms. Anna Maughan

anna3.maughan@northumbria.ac.uk
University of Northumbria

Research Assistant



Ms. Mariella Silva de Oliveira Costa

mariella.costa@fiocruz.br
Fiocruz Brasília - Brasil

Collaborator



Mr. Michael Craig

michael2.craig@northumbria.ac.uk
University of Northumbria

Collaborator



Ms. Bianca Fadel

bianca.fadel@northumbria.ac.uk
University of Northumbria

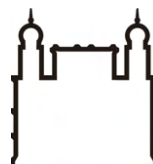
Collaborator

We would like to thank WhatsApp for their steadfast support to this project and all the participants in the studies we have conducted.

Copyright © 2021 Dr. Santosh Vijaykumar
All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means without prior written permission by the publisher. For permission requests, please write to or contact Dr. Santosh Vijaykumar at:

Email: santosh.vijaykumar@northumbria.ac.uk

We thank you for your continued support



**UNIVERSITY OF
GEORGIA**